The Politics of Anti-Corruption: Bureaucratic Sabotage and Collusion

Anna Denisenko

DRAFT: Comments Welcome

Abstract

This paper investigates why anti-corruption reforms focused on strengthening political enforcement can paradoxically lead to more corruption. I develop a dynamic model of a strategic, rent-seeking bureaucrat operating under an elected politician. The bureaucrat covertly chooses the level of corruption, which influences the probability of the incumbent's re-election. The model demonstrates that the bureaucrat's incentives are shaped by the incumbent's commitment to anti-corruption. A lenient incumbent encourages "strategic collusion," where the bureaucrat reduces graft to help the incumbent win re-election. Conversely, a tough incumbent incentivizes "strategic sabotage," where the bureaucrat inflates corruption to induce the incumbent's removal. These opposing motives generate a non-monotonic relationship between enforcement and corruption: starting from low enforcement, marginally tougher anti-corruption stance erodes collusive incentives and *raises* corruption; beyond a certain threshold, higher personal costs dominate and corruption *falls* with enforcement. This paper demonstrates how the actions of unelected officials can subvert the mechanisms of democratic control, leading well-intentioned anti-corruption policies to backfire.

Introduction

The global effort to combat corruption often centers on a seemingly straightforward prescription: strengthen political will and enhance state enforcement. International organizations and domestic reformers alike have invested vast resources in policies designed to increase the costs of malfeasance, from establishing independent anti-corruption agencies to empowering voters to remove compromised incumbents. Yet, the results of this global push are decidedly mixed. In many countries plagued by systemic corruption, even the election of committed, reform-minded leaders fails to produce a sustained reduction in graft; in some cases, their tenures are marked by an increase in scandals that ultimately undermine their political project (Johnston, 2017; Cheeseman and Peiffer, 2022). This presents a critical puzzle: why do earnest anti-corruption campaigns, led by politicians with demonstrable political will, so often falter or even backfire?

This paper argues that the success or failure of anti-corruption efforts depends crucially on the strategic behavior of unelected, rent-seeking bureaucrats who can actively manipulate electoral outcomes to protect their interests. Standard principal-agent models, which conceptualize corruption as a problem of controlling self-interested agents (Rose-Ackerman, 2013; Groenendijk, 1997), fail to account for situations where the agent's behavior is not just about maximizing immediate rents but about actively selecting their future principal. I develop a dynamic game-theoretic model that moves beyond the traditional focus on bureaucratic shirking to analyze a more detrimental threat: electoral manipulations. The model demonstrates that a bureaucrat's response to an anti-corruption agenda in the presence of information asymmetry between the bureaucrat and a voter is conditional on the incumbent politician's type. When faced with a lenient incumbent — one who is ineffective against graft – the bureaucrat has an incentive to engage in strategic collusion, moderating corrupt activities to reduce the probability of a public scandal and thereby aid the incumbent's reelection.

Conversely, when faced with a tough, reform-minded incumbent, the bureaucrat's incentives flip to strategic sabotage. By inflating the level of corruption, the bureaucrat can increase the likelihood of a public scandal, damaging the incumbent's reputation and paving the way for their replacement. This dynamic generates the paper's central theoretical result: a non-monotonic relationship between enforcement and corruption. For incumbents with low levels of enforcement, a marginal increase in an incumbent's anti-corruption stance weakens the bureaucrat's incentive for strategic collusion, leading corruption to paradoxically rise with enforcement. However, beyond a certain threshold, the personal cost of engaging in graft becomes prohibitively high for the bureaucrat, and so corruption begins to fall as enforcement continues to increase. The model thus explains how a leader's sincere commitment to fighting corruption can, under certain conditions, initially generate more of the very behavior it is designed to prevent.

This theoretical framework offers a lens through which to interpret the otherwise puzzling political dynamics observed in several recent anti-corruption drives, where the arrival of a reform-minded leader paradoxically coincides with an eruption of public scandals. Following the presidency of Jacob Zuma in South Africa, an era widely seen as the apex of "state capture", Cyril Ramaphosa ascended to power in 2018 on an explicit promise to dismantle these entrenched patronage networks and restore integrity to the state (PARI, 2022; South African Institute of International Affairs, 2024). Yet, rather than a swift cleanup, his administration has been mired in debilitating intra-party warfare. Factions within the ruling African National Congress (ANC) have engaged in a conspicuous "fight back," and Ramaphosa's tenure has been plagued by a steady stream of corruption revelations that serve to undermine his authority and portray his reform agenda as ineffective. From the perspective of the model, this is not a simple failure of political will but a predictable strategic response: entrenched bureaucratic and political networks, facing a hostile principal, expose malfeasance to sabotage his political project and create the conditions for his electoral removal.

A similar dynamic is evident in Nigeria and Ukraine. Muhammadu Buhari's landmark election victory in Nigeria in 2015 was predicated almost entirely on his personal reputation as an incorruptible leader who would finally tame the country's notoriously rapacious political class (Hoffmann and Patel, 2017; Reuters, 2018). His presidency, however, has been marked by a series of high-profile corruption scandals within the civil service and state-owned enterprises that have left many citizens questioning his efficacy (U4 Anti-Corruption Resource Centre, 2023). Likewise, Ukraine's post-2014 "Revolution of Dignity" created immense popular and international pressure for a fundamental break with the country's oligarchic past, leading to the creation of a new architecture of anti-corruption institutions. The response from the old guard was not to acquiesce, but to launch a sustained campaign of sabotage. This resistance has been most visible in the actions of the judiciary and the prosecutor general's office, which have consistently worked to block investigations, dismiss cases against powerful figures, and even dismantle core elements of the anti-corruption legal framework (Reuters, 2025; Transparency International Ukraine, 2025).

The model's logic of strategic collusion, conversely, helps explain the stability of corrupt systems under lenient or complicit leadership. This is not the absence of corruption, but its careful management to avoid destabilizing public outrage. In modern Russia, the state operates as a highly centralized system of rent extraction and distribution, what some scholars have termed a "kleptocracy" (Lanskoy and Myles-Primakoff, 2018; Dawisha, 2015). The incentive for the vast majority of actors within the bureaucracy is to keep their rent-seeking within the informal, yet clearly understood, bounds of the system to ensure its preservation. Similarly, in Serbia under the rule of the Serbian Progressive Party, the state has been captured through the systematic politicization of public administration and the media, a system of partokratija (party-state) that ensures loyalty and manages the flow of information (Bertelsmann Stiftung, 2024; Freedom House, 2024). For a bureaucrat, whose position is dependent on the ruling party's continued dominance, the "retention value" of the incumbent leadership is exceptionally high. This creates a powerful collective incentive to suppress any information about malfeasance that could threaten the party's hold on power, demonstrating the logic of strategic collusion on a national scale.

The rest of the paper is organized as follows: Section **??** puts the argument within research on corruption, political agency, and sabotage. Section **??** presents the model, including comparative statics and the nonmonotonic enforcement–corruption relationship. Section **??** evaluates robustness of the predictions. Section **??** discusses implications of the model. Section **??** concludes.

Related Literature

This paper connects two large literatures: theories of corruption and political agency, and recent work on strategic policy sabotage.

Foundational work in economics models corruption as a principal–agent problem in which officeholders or bureaucrats trade-off illicit rents against the probability and costs of detection. See Shleifer and Vishny (1993) for the seminal formalization of how organizational structure and market competition shape bribe extraction and enforcement; see also Rose-Ackerman (2013); Groenendijk (1997). In these models, tougher monitoring or sanctions typically reduces graft monotonically. Political economy models of electoral accountability likewise emphasize how voters use noisy performance signals to select competent or honest incumbents (e.g., Barro (1973); Fearon (1999); Williams (2021)). A growing strand examines the role of state capacity and institutional design in enabling enforcement, often treating bureaucratic behavior as a static response to incentive changes (Bersch, Praça and Taylor, 2017). By contrast, my model treats the bureaucrat as a forward-looking political actor who can *select* his future principal — by manipulating the probability that the incumbent is retained — thereby endogenizing how enforcement affects both rent extraction *and* electoral selection.

This paper is most directly in conversation with recent theoretical work on policy sabotage, which studies how actors strategically worsen outcomes to damage political opponents. Gieczewski and Li (2022) and Hirsch and Kastellec (2022) analyze sabotage by an opposition party and show that the timing and intensity of sabotage depend on policy/incumbent popularity and electoral incentives. Related formal work highlights how organized actors can degrade policy implementation or withhold effort to influence electoral or legislative payoffs (e.g., Patty (2016); Fong and Krehbiel (2018); Delgado-Vega and Shaver (2024)). In particular, Heo and Wirsching (2024) develop a theory of bureaucratic sabotage where anti-reform bureaucrats strategically worsen service quality to influence voters.

The key departure from the existing frameworks is substantively driven: the policy of interest (anti-corruption) directly cuts into bureaucrats' rents. This creates environments in which bureaucrats have incentives not only to *sabotage* a tough principal but also to *collude* with a lenient one, sharply complicating the voter's inference problem. Technologically, I model a forward-looking bureaucrat who manipulates the incumbent's retention probability via the equilibrium level of (un)observed corruption — rather than an agent degrading a publicly observed policy stream — so the signal voters see is endogenously distorted. The result is a selection mechanism that generates a non-monotonic enforcement–corruption relationship and distinct conditions under which scandals rise under reformers.

Model

I analyze a dynamic game played over two periods, $t \in \{1, 2\}$. The strategic actors are a rent-seeking Bureaucrat (B) and a representative Voter (V). The political environment is populated by elected officials (Politicians) who are non-strategic actors: an Incumbent (I) who presides in period 1, and a Challenger (C) who may replace I in period 2 via an election.

The game begins with Nature determining the type of each potential politician, $P \in \{I, C\}$. The distribution of types is defined by a single parameter, $\overline{e} \in \mathbb{R}^+$, which represents the state's capacity for anti-corruption. Specifically, a politician's type, which equals their enforcement level e^P, is drawn independently and identically from a uniform distribution bounded by the state capacity: $e^P \sim U[0, \overline{e}]$. This enforcement level, e^P, captures a politician's intrinsic commitment to anti-corruption as constrained by the state's ability to implement policy. Crucially, elected politicians are non-strategic actors whose type, once realized, exogenously determines their choice of anti-corruption enforcement.

After observing the incumbent's enforcement level, e^I, the Bureaucrat chooses the level of corruption for the first period, $k_1 \in \mathbb{R}_+$. The Bureaucrat's payoff in period t is given by $u_t^B(k_t, e^P) = \pi(k_t) - c(k_t, e^P)$, where $P \in \{I, C\}$ is the politician in office. The function $\pi(\cdot)$ represents the benefits from corruption, and the function $c(\cdot, \cdot)$ captures the costs of corruption. I begin the analysis by specifying functional forms for the costs and benefits of corruption. This specific case provides a concrete foundation for the general model presented in the subsequent section.

Let the benefit and cost functions be given by

$$\pi(k) = k; \quad c(k, e) = \alpha \cdot k^2 \cdot e; \tag{1}$$

where $\alpha > 0$ measures the inherent difficulty of graft, independent of the political leadership's active enforcement efforts.

Following the Bureaucrat's choice of k_1, the Voter observes a public signal of corruption, $s_1 \in \{0, 1\}$. The probability that corruption is detected ($s_1 = 1$) is given by the function $p(k_1)$. Let the probability functions be given by

$$p(k) = \beta \cdot k - \gamma \cap [0, 1]. \tag{2}$$

This formulation captures two key features of real-world accountability. First, the parameter $\gamma \geq 0$ establishes an enforcement threshold. It implies that corruption below a certain level ($k_1 \leq \gamma/\beta$) goes undetected, creating a safe zone for low-level malfeasance. This reflects the practical limits of state audits or the media's attention capacity. Second, for corruption above this threshold, the parameter $\beta > 0$ measures the sensitivity of the signal, representing the quality of monitoring institutions. A higher β means that any given increase in graft is more likely to be exposed to the public.

At the end of period 1, an election is held. Based on the signal s_1, the Voter decides whether to retain the Incumbent ($r = 0$) or replace him with the Challenger ($r = 1$). The Voter's objective is to maximize their total expected utility. Her utility in each period $u_t^V(k_t, e^P) = B \cdot \mathbb{I}_{\{s_t=0\}}$, where B is the utility derived from the absence of detected corruption. The Voter's decision is thus a forward-looking choice to select the politician who is expected to induce a lower level of corruption in the second period.

In period 2, the Bureaucrat observes the enforcement level, e_2, of the elected politician (either e^I or e^C). The Bureaucrat then chooses the second-period corruption level, $k_2 \in \mathbb{R}^+$. Subsequently, the game ends and payoffs are realized.

The Bureaucrat and the Voter share the discount factor δ. I solve for PBE, which consists of a strategy for the Bureaucrat (k_1^*, k_2^*), a strategy for the Voter (r^*), and a belief system for the Voter about the incumbent's type.

Myopic Corruption Benchmark

The analysis begins by characterizing the Bureaucrat's behavior in a static setting. Denote by $k^M(e)$ the myopic level of corruption – the level of graft a Bureaucrat would choose if he did not consider the future electoral consequences of their actions. Formally, for any given enforcement level e, $k^M(e)$ is defined by

$$k^M(e) := \arg\max_{k \geq 0} \{\pi(k) - c(k, e)\}. \tag{3}$$

The optimal myopic level of corruption is thus implicitly defined by $\pi'(k^M(e)) = c_k(k^M(e), e)$. The specified functional forms allow to derive closed-form solutions for the Bureaucrat's problem.

Lemma 1. *The myopic level of corruption and the associated single-period payoff for the Bureaucrat are given by*

$$k^M(e) = \frac{1}{2\alpha e} \quad \textit{and} \quad u^{B,M}(e) \equiv u^B(k^M(e), e) = \frac{1}{4\alpha e}$$

Proof. See Appendix. □

This explicit solution shows that myopic corruption is inversely proportional to both the cost parameter α and the enforcement level e. The resulting payoff for the Bureaucrat is likewise a strictly decreasing and convex function of enforcement. This establishes an intuitive baseline: absent electoral concerns, greater anti-corruption enforcement leads to less corruption.

The following proposition formalizes these intuitive comparative statics.

Proposition 1. *The myopic level of corruption, $k^M(e)$, is*

1. *strictly decreasing in the enforcement level, e, and inherent difficulty of graft, α;*
2. *weakly decreasing in state capacity, $\bar{e}$.*

Proof. See Appendix. □

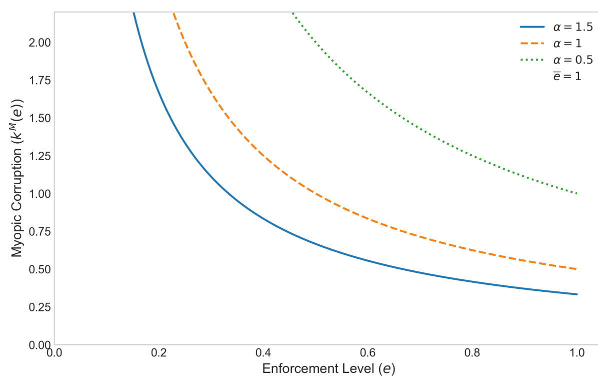


Figure 1: Myopic corruption as a function of the enforcement level.

Figure 1 depicts the optimal myopic corruption level ($k^M(e)$) as a function of the enforcement level, e, and the inherent difficulty of graft, α. As e and α increase, $k^M(e)$ decreases. Having established the myopic benchmark, I now turn to the analysis of the Bureaucrat's behavior.

Myopic Voter

To build intuition, I first analyze the case of a myopic Voter. A myopic Voter employs a simple, retrospective voting rule: she retains the Incumbent if and only if she does not observe evidence of corruption. Formally, the myopic Voter's retention rule, $r(s_1)$, such that

$$r(s_1) = \begin{cases} 1, & s_1 = 1; \\ 0, & s_1 = 0. \end{cases} \tag{4}$$

This assumption removes the Voter's signal-extraction problem and transforms the game into a decision problem for the Bureaucrat. The Bureaucrat knows that by choosing a corruption level k_1 they induce a probability of electoral replacement equal to $p(k_1)$.

In the second period, the election is over, and the Bureaucrat's objective is simply to maximize his static, single-period payoff given the enforcement level, e_2, of the politician in office. The Bureaucrat, thus, chooses the myopic level of corruption in the second period in every equilibrium $k_2^* = k^M(e_2)$.

The Bureaucrat's objective in the first period is to choose k_1 to maximize his expected payoffs, given the Voter's retention rule. The Bureaucrat, thus, selects k_1^* such that

$$k_1^* := \arg\max_{k_1 \geq 0} \left\{ u_B(k_1, e^I) + \delta \left[(1 - p(k_1)) u_B(k^M(e^I), e^I) + p(k_1) \mathbb{E}[u_B(k^M(e^C), e^C)] \right] \right\}$$

where $\delta \in (0, 1]$ is the Bureaucrat's discount factor, $u_B(k^M(e^I), e^I)$ is the continuation payoff if the Incumbent is retained, and $\mathbb{E}[u_B(k^M(e^C), e^C)]$ is the expected continuation payoff if the Incumbent is replaced by a Challenger of unknown type. The core trade-off the Bureaucrat faces depends on the *retention value* of the Incumbent – the difference between the Bureaucrat's expected continuation payoff under the Incumbent, $u_B(k^M(e^I), e^I)$, and under a randomly drawn Challenger, $V_B^C := \mathbb{E}[u_B(k^M(e^C), e^C)]$.

If the Incumbent is lenient, the retention value of the Incumbent is positive; the Bureaucrat prefers to keep the Incumbent in office. This creates an incentive to moderate corruption ($k_1^* < k^M(e^I)$) to reduce the probability of replacement, what I call an act of *strategic collusion*. Conversely, if the Incumbent is sufficiently tough, the retention value is negative; the Bureaucrat prefers a new draw from the pool of politicians. This creates an incentive to inflate corruption ($k_1^* > k^M(e^I)$) to increase the probability of replacement — an act of *strategic sabotage*.

First, I isolate the Bureaucrat's pure electoral incentives, deriving the unconstrained strategic level of corruption, denoted $k_1^U(e^I)$. This is the notional level of graft the Bureaucrat would choose if the probability of detection, $p(k)$, were not bounded. This benchmark reveals the Bureaucrat's latent strategic preference, stripped of the institutional constraints of the signaling technology.

Proposition 2. *The unconstrained strategic corruption is* $k_1^U(e^I) = \frac{4\alpha e^I \overline{e} + \beta\delta(2e^I - \overline{e})}{8\alpha^2 (e^I)^2 \overline{e}}$*, denote* $e^* := \overline{e}/2$*, then*

1. *when* $e^I \leq e^*$*, the Bureaucrat engages in strategic collusion* $k_1^U(e^I) \leq k^M(e^I)$*;*
2. *when* $e^I > e^*$*, the Bureaucrat engages in strategic sabotage* $k_1^U(e^I) > k^M(e^I)$*.*

Proof. See Appendix. □

The intuition for this result builds on the Bureaucrat's comparison of the Incumbent's known enforcement level, e^I, with the expected enforcement of a potential replacement. Since a Challenger's type is a random draw from the population, their expected enforcement level is the population mean, $\mathbb{E}[e^C] = \overline{e}/2$. When the Incumbent is more lenient than this expected Challenger ($e^I \leq \overline{e}/2$), the retention value of the incumbent is positive. This creates an incentive to moderate corruption to reduce the probability of electoral replacement — an act of strategic collusion. Conversely, when the Incumbent is tougher than average ($e^I > \overline{e}/2$), the retention value is negative, as the Bureaucrat expects a more favorable outcome from a new draw. This incentivizes the Bureaucrat to inflate corruption above the myopic level, engaging in strategic sabotage.

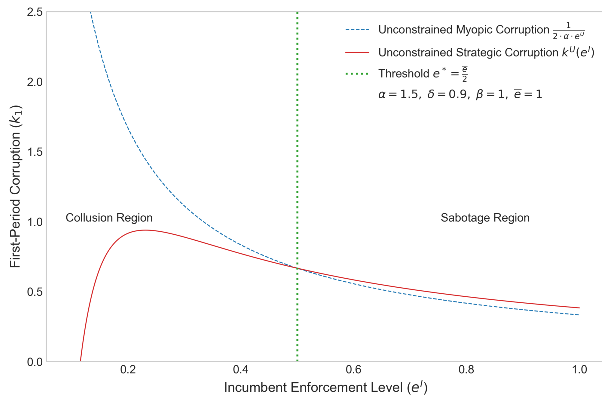


Figure 2: Unconstrained corruption as a function of enforcement

The threshold $e^* := \bar{e}/2$ endogenously arises as the Bureaucrat's point of indifference, where the incumbent is exactly as tough as the expected replacement, neutralizing any electoral incentive.

The next proposition describes how the level of unconstrained corruption, k_1^U, responds to marginal changes in the Incumbent's enforcement, e^I. An increase in e^I triggers two opposing forces. First, there is a *Direct Deterrence Effect*: a higher e^I increases the marginal cost of corruption, creating a standard disciplining incentive to reduce graft. Second, there is an *Indirect Political Effect*: a higher e^I lowers the Bureaucrat's continuation payoff under the incumbent, making that official a more attractive target for electoral removal.

Proposition 3. *The unconstrained strategic level of corruption, $k_1^U(e^I)$, is a non-monotonic function of the Incumbent's enforcement level, e^I. There exists a unique threshold, $e^{**} = \frac{\beta\cdot\delta\bar{e}}{\beta\cdot\delta+2\cdot\alpha\cdot\bar{e}}$, such that:*

1. *if $e^I < e^{**}$, the Political Effect dominates, and corruption is strictly increasing in enforcement ($\frac{dk_1^U}{de^I} > 0$).*
2. *if $e^I > e^{**}$, the Deterrence Effect dominates, and corruption is strictly decreasing in enforcement ($\frac{dk_1^U}{de^I} < 0$).*

Proof. See Appendix. □

Proposition 3 formalizes the net result of these competing forces. The threshold e^{**} represents the enforcement level at which the Direct Deterrence Effect balances the Indirect Political Effect. When incumbent enforcement is sufficiently low, $e^I < e^{**}$, the direct cost of graft is not yet prohibitive. In this region, a marginal increase in e^I primarily serves to make the Incumbent a more threatening political adversary, strengthening the Bureaucrat's incentive to sabotage them. The Political Effect dominates, and corruption rises with enforcement. However, as enforcement crosses the threshold e^{**}, the marginal cost of engaging in graft outweighs the political incentive to oust the Incumbent, and the level of corruption begins to fall.

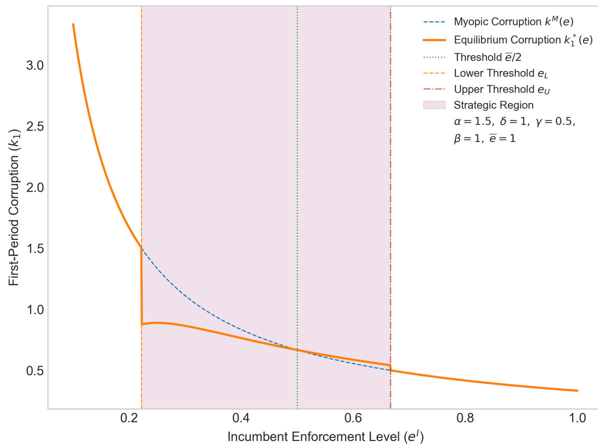


Figure 3: The equilibrium level of corruption as a function of the Incumbent's corruption.

While the analysis above outlines the Bureaucrat's strategic incentives, his ability to act on them is constrained by the mechanics of the signaling technology. The next proposition characterizes optiomal first-period level of corruption, $k_1^*(e^I)$.

Proposition 4. *The equilibrium level of first-period corruption, $k_1^*(e^I)$, is such that*

$$k_1^*(e^I) = \begin{cases} k_1^U(e^I) & \textit{if } e^I \in [e_L, e_U] \quad \textit{(Strategic Region)} \\ k^M(e^I) & \textit{if } e^I \notin (e_L, e_U) \quad \textit{(Myopic Region)}, \end{cases}$$

where $e_L := \frac{\beta}{2\alpha(1+\gamma)}$ and $e_U := \frac{\beta}{2\alpha\gamma}$.

Proof. See Appendix. □

The Bureaucrat's lever for influencing the election is the probability of detection, $p(k_1)$, which is naturally bounded, $p(k_1) \in [0, 1]$. When the Incumbent is lenient, the myopic level of corruption, k^M, saturates the public signal, driving the probability of detection to its maximum. Since electoral replacement is a certainty, the marginal electoral cost of further increasing graft is zero. With no political consequences to consider at the margin, the Bureaucrat's problem collapses to static rent maximization, and the myopic choice, k^M, becomes optimal.

In contrast, with a very strong incumbent, $e^I \geq e_U$, the high level of enforcement suppresses the myopic level of corruption so effectively that it becomes invisible to the public, $p(k^M) = 0$. The marginal electoral cost of corruption is again zero, as the default behavior already guarantees the incumbent's retention. It is only in the intermediate range, $e^I \in [e_L, e_U]$, that the Incumbent is strong enough to avoid automatic saturation, yet not so strong as to make corruption politically invisible.

Figure 3 plots the equilibrium level of first-period corruption, $k_1^*(e^I)$, against the myopic benchmark, $k^M(e^I)$. The two institutional thresholds e_L and e_U separate strategic region (shaded) and myopic region (not shaded). Figure 3 also suggests a non-monotonic relationship between enforcement and corruption. The following proposition characterizes conditions under which the strategic incentive for sabotage becomes powerful enough to overwhelm the direct effect of deterrence, causing corruption to rise with enforcement.

The following proposition provides comparative statics on the scope of the Bureaucrat's strategic behavior, showing how features of the political and institutional environment expand or contract the set of incumbent types that trigger a collusive versus a sabotaging response.

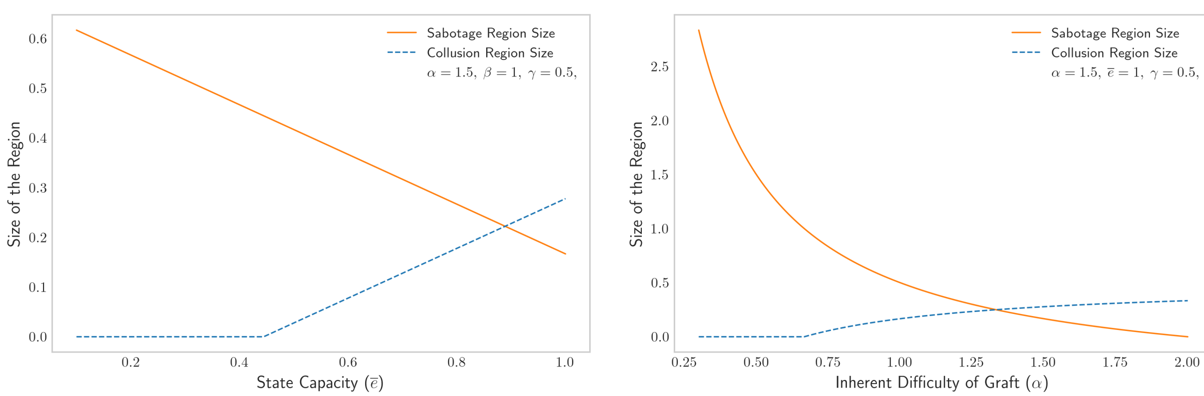
Proposition 5. *The set of incumbent types that induces*

1. *strategic collusion expands with higher state capacity, $\overline{e}$, greater inherent difficulty of graft, α, and a higher enforcement threshold, γ, but shrinks with better monitoring, β;*

2. *strategic sabotage shrinks with higher $\overline{e}$, α, and γ, but expands with β.*

Proof. See Appendix. □

Higher state capacity, $\overline{e}$, is disciplining sabotage while encouraging collusion. By increasing the expected enforcement level of a potential challenger, it makes the prospect of replacing the current incumbent more daunting. This encourages the Bureaucrat to expand the set of incumbents with whom they would rather collude than risk a new draw from a tougher pool of politicians. It implies that high-capacity states are less prone to scandal-driven turnover under reformers but more exposed to subtle, harder-to-detect collusion under permissive leaders. In contrast, better monitoring institutions, β, make the public signal of corruption more sensitive. This heightened transparency makes subtle collusion more difficult and risky, forcing the Bureaucrat's hand. Consequently, the incentive to engage in the more aggressive strategy of sabotage applies to a wider range of tough incumbents, shrinking the zone of cooperation.



(a) Effect of State Capacity ($\overline{e}$) (b) Effect of Inherent Difficulty of Graft (α)

Figure 4: The Size of Strategic Collusion and Sabotage Regions

Figure 4 plots the size of the strategic regions to visually demonstrate these comparative statics. The left panel shows that as state capacity, $\overline{e}$, increases, the range of incumbent types who induce sabotage shrinks, while the range that induces collusion expands. The right panel shows the effect of the inherent difficulty of graft, α. When corruption is easy (low α), the potential rewards are high, incentivizing the Bureaucrat to engage in sabotage against a very wide range of incumbents to secure a more favorable future principal. As corruption becomes intrinsically more costly (a higher α), the value of this political gamble falls sharply. Consequently, the Bureaucrat becomes less willing to engage in sabotage, and the scope for collusive behavior expands.

Proposition 6. *The equilibrium level of first-period corruption, $k_1^*(e^I)$, is non-monotonic in the incumbent's enforcement level, e^I, if and only if*

1. $\bar{e} > \frac{\beta\delta}{2\cdot\alpha} \cdot \frac{1}{\delta\cdot(1+\gamma)-1}$;
2. $\delta \cdot (1+\gamma) - 1 > 0$.

Proof. See Appendix. □

The first condition states that the state capacity ($\bar{e}$) must be sufficiently high. The intuition for this is not that high costs deter corruption. Instead, a high $\bar{e}$ ensures that the Bureaucrat is not automatically in the "myopic region" of signal saturation. It creates the political space for the Bureaucrat to operate within the strategic corridor, where his actions have an impact on the electoral outcome.

The second condition, $\delta(1+\gamma) > 1$, ensures, in turn, that the incentive for strategic sabotage is present. A higher δ ensures the Bureaucrat is patient enough to value the future prize of ousting a tough incumbent; a higher γ shrinks the strategic corridor from the right, generating incentives for sabotage. Jointly, these two conditions, result in the paradoxical relationship this paper seeks to explain: an anti-corruption policy centered on elevating enforcement can, after a point, become counterproductive, actively generating the very behavior it is designed to prevent.

Voter with Agency

Finally, I endogenize the Voter's strategy, assuming she is a rational, forward-looking agent. The Voter's objective is to have a politician in office who will induce the lowest possible level of corruption in the second period. Let $\mathbb{E}[e^I|s_1]$ be the Voter's posterior expectation of the Incumbent's enforcement level after observing signal s_1 and let $\mu(e^I|s_1)$ denote the Voter's posterior belief density over the Incumbent's type given the signal. The optimal retention rule is

$$r^*(s_1) = 0 \quad \iff \quad \int_0^{\bar{e}} p(k^M(e^I))\mu(e^I|s_1)de_I \geq \mathbb{E}[p(k^M(e^C))].$$

Since $k_2^* = k^M(e_2)$ is a strictly decreasing function of enforcement, this is equivalent to selecting a Politician with the highest expected enforcement lecel. I assume that when indifferent the Voter chooses to retain politician. The Voter's decision rule, then, is to retain the Incumbent if and only if the posterior expectation of their enforcement level is at least as high as the Challenger's. The following proposition demonstrates that the viability of the simple retrospective rule depends on the level of state capacity, $\bar{e}$.

Proposition 7. *When $\bar{e} > \frac{\beta}{\alpha\gamma}$ the simple retrospective voting rule (retain on $s_1 = 0$, replace on $s_1 = 1$) constitutes a Perfect Bayesian Equilibrium.*

The intuition behind this result is that state capacity acts as an anchor for the political system. A higher state capacity, $\bar{e}$, has a disciplining effect that operates independently of the Incumbent's type. By increasing the expected toughness of any potential challenger, it systematically lowers the Bureaucrat's incentive for strategic sabotage. This disciplining effect clarifies the meaning of the public signal. As state capacity increases, the pool of incumbents who generate a corruption signal ($s_1 = 1$) becomes increasingly composed of the genuinely leneint Politicians, rather than the strategically sabotaged. A "bad" signal becomes more reliable as "bad news."

Robustness Check

This section examines how sensitive the paper's core results are to modeling choices and parametric assumptions. I ask whether the two central findings — (i) that bureaucrats switch between *strategic collusion* and *strategic sabotage* depending on the incumbent's toughness, and (ii) that enforcement can have a *non-monotonic* effect on corruption — survive when I relax the model's structure.

As before, Nature acts first. It draws the type of each potential politician $P \in \{I, C\}$, denoted $e^P \sim f(\cdot)$ with support $[0, \bar{e}]$ and pdf f.

After observing e^I, the Bureaucrat chooses first-period corruption $k_1 \in \mathbb{R}_+$. Per-period utility is

$$u_t^B(k_t, e^P) = \pi(k_t) - c(k_t, e^P), \qquad P \in \{I, C\},$$

with benefits π strictly increasing and concave ($\pi' > 0$, $\pi'' \leq 0$) and costs c twice differentiable, increasing, and convex in each argument ($c_1, c_2 > 0$, $c_{11}, c_{22} \geq 0$). I maintain $c_{12} \geq 0$ (enforcement raises the marginal cost of graft) and normalize $u_t^B(0, e^P) = 0$.

The analysis begins by characterizing the Bureaucrat's behavior in a static setting. Let $k^M(e)$ denote the *myopic* graft choice at enforcement level e, defined by

$$\max_{k \geq 0} \{\pi(k) - c(k, e)\} \tag{5}$$

The optimal myopic level of corruption is thus implicitly defined by $\pi'(k^M(e)) = c_k(k^M(e), e)$. The following proposition establishes the comparative statics of this myopic level of corruption.

Proposition 8. *The myopic level of corruption ($k^M(e)$)*

1. *is a strictly decreasing function of the enforcement level (e^P);*

2. *is a weakly decreasing function of the state capacity ($\bar{e}$).*

Proof. See Appendix. □

The comparative statics in Proposition 8 is driven by the assumption of $c_{ke} \geq 0$: enforcement raises the marginal cost of graft, making e and k strategic substitutes for the Bureaucrat. Consequently, in the static benchmark — absent electoral considerations — stronger state capacity or a more committed politician reduces optimal corruption.

Consider now a voter who follows a purely retrospective rule: retain the incumbent iff no corruption is observed. Let $r(s_1) \in \{0, 1\}$ indicate retention ($r = 1$) vs. replacement ($r = 0$). The rule is

$$r(s_1) = \begin{cases} 1, & s_1 = 1; \\ 0, & s_1 = 0. \end{cases} \tag{6}$$

This assumption removes the Voter's signal-extraction problem and transforms the game into a decision problem for the Bureaucrat. The Bureaucrat knows that by choosing a corruption level k_1 they induce a probability of electoral replacement equal to $p(k_1)$.

Proposition 9. *Let $V_B^C := \mathbb{E}[u_B(k^M(e^C), e^C)]$ be the Bureaucrat's expected continuation payoff from electing the Challenger. There exists a unique enforcement threshold $e^* \in (0, \bar{e}]$ defined by $u_B(k^M(e^*), e^*) = V_B^C$.*

1. *If $e^I \leq e^*$, the Bureaucrat engages in strategic collusion: $k_1^*(e^I) \leq k^M(e^I)$.*
2. *If $e^I > e^*$, the Bureaucrat engages in strategic sabotage: $k_1^*(e^I) > k^M(e^I)$.*

Proof. See Appendix. □

Proposition 9 shows that electoral pressure generates opposing forces. Increasing e^I raises the marginal cost of graft in period 1 (*direct deterrence effect* pushes k_1^* down) but lowers the Bureaucrat's continuation value under the incumbent, reducing the retention value of I (*indirect political effect* pushes k_1^* up). The sign of $\partial k_1^* / \partial e^I$ is therefore ambiguous.

Proposition 10. 1. *The higher level of the incumbent's enforcement level, e_I,*

   (a) *increases the equilibrium level of corruption, k_1^*, when*

   $$-c_{ke}(k_1^*, e^I) + \delta \cdot p'(k_1^*) \cdot c_e(k^M(e^I), e^I) \geq 0;$$

   (b) *and decreases the equilibrium level of corruption, k_1^*, when*

   $$-c_{ke}(k_1^*, e^I) + \delta \cdot p'(k_1^*) \cdot c_e(k^M(e^I), e^I) \leq 0;$$

2. *The higher level of the state capacity, $\bar{e}$, decreases the equilibrium level of corruption, k_1^*.*

Proof. See Appendix. □

Intuitively, higher e^I cuts against graft contemporaneously (via c_{ke}) but also strengthens incentives to unseat a tough incumbent (via $p'(k_1^*)\, c_e$); either effect can dominate locally. By contrast, higher state capacity $\bar{e}$ raises expected challenger toughness and unambiguously lowers k_1^*.

I now endogenize the Voter's strategy by assuming she is a rational, forward-looking agent. Let $\mathbb{E}[e^I | s_1]$ be the Voter's posterior expectation of the Incumbent's enforcement level after observing signal s_1 and let $\mu(e^I|s_1)$ denote the Voter's posterior belief density over the Incumbent's type given the signal. The optimal retention rule is

$$r^*(s_1) = 0 \quad \iff \quad \int_0^{\bar{e}} p(k^M(e^I))\mu(e^I|s_1)de_I \geq \mathbb{E}[p(k^M(e^C))]$$

Because $k_1^*(e^I)$ maps types into endogenous detection probabilities, the signal is distorted by both collusion and sabotage. The next proposition characterizes conditions under which the Voter retains the Incumbent following $s_1 = 1$ and replaces the Incumbent following signal $s_0 = 0$.

Proposition 11. *The Voter replaces the Incumbent following* $s_1 = 1$ *and retains the Incumbent following* $s_1 = 0$ *if and only if* $Cov(e^I, p(k_1^*(e^I))) < 0$.

Proof. See Appendix. □

A negative covariance, $\mathrm{Cov}(e^I, p(k_1^*(e^I)))$, indicates that on average, higher enforcement is associated with a lower probability of a corruption signal. Voter, then, should follow the simple retrospective rule in equilibrium replacing Incumbents if and only if corruption is observed.

Discussion

Substantively, the model speaks to corporate governance, compliance, and organizational design by showing how oversight can backfire when agents internalize selection consequences. Methodologically, it brings a tractable dynamic selection mechanism to settings typically analyzed with static agency models, offering guidance for monitoring robust to endogenous signal manipulation. The presented theoretical argument highlights that strengthening monitoring is not uniformly disciplining. Near the lower end of enforcement, marginal tightening can *raise* observed failures if agents use incidents to unseat strict principals; the same tightening later *reduces* failures once deterrence dominates. Second, because signals are endogenously distorted, naïve retrospective evaluation can be systematically wrong.

Finally, although motivated by anti-corruption, the mechanism this paper proposes applies to organizations and firms when (i) middle agents privately control outcomes that affect the perceived performance of their principal, and (ii) the intensity of oversight both disciplines misconduct and changes which principal the agent expects to face in the future. Classic organizational economics emphasizes how measurement, multitasking, and authority allocation shape effort and gaming. My model adds a selection margin: agents manipulate observable outcomes to *choose* the principal they will face next (e.g., a strict new division head versus a more permissive incumbent CEO).

Appendix

A Lemma 1

The Bureaucrat chooses $k^M(e)$ s.t.

$$k^M(e) = \arg\max_{k \geq 0} \{k - \alpha k^2 e\}$$

Because $-2\alpha e < 0$, the objective function is concave in k and FOC is

$$1 - 2\alpha k e = 0$$

Solving for k produces $k^M(e) = \frac{1}{2\alpha e}$.

To find the Bureaucrat's utility, substitute $k^M(e)$ back into the utility function

$$u^{B,M}(e) = \left(\frac{1}{2\alpha e}\right) - \alpha \left(\frac{1}{2\alpha e}\right)^2 e = \frac{1}{2\alpha e} - \alpha \left(\frac{1}{4\alpha^2 e^2}\right) e = \frac{1}{2\alpha e} - \frac{1}{4\alpha e} = \frac{1}{4\alpha e}$$

B Proposition 1

Because

$$\frac{dk^M}{de} = \frac{d}{de}\left(\frac{1}{2\alpha e}\right) = -\frac{1}{2\alpha e^2} < 0$$

and

$$\frac{dk^M}{d\alpha} = \frac{d}{d\alpha}\left(\frac{1}{2\alpha e}\right) = -\frac{1}{2\alpha^2 e} < 0$$

the function is strictly decreasing in e and α. The enforcement level is $e = \min\{\theta^P, \bar{e}\}$. For any politician type θ^P, an increase in $\bar{e}$ either leaves e unchanged (if $\theta^P < \bar{e}$) or increases it (if $\theta^P \geq \bar{e}$). Since $k^M(e)$ is a decreasing function of e, an increase in $\bar{e}$ must, therefore, weakly decrease the myopic level of corruption for any politician type.

C Proposition 2

The Bureaucrat's unconstrained objective function is

$$\max_{k_1 \geq 0} \left\{ (k_1 - \alpha k_1^2 e^I) + \delta \left[(1 - \beta k_1) \frac{1}{4\alpha e^I} + \beta k_1 \frac{1}{2\alpha \bar{e}} \right] \right\}.$$

The first-order condition is

$$(1 - 2\alpha k_1 e^I) + \delta \left[-\frac{1}{4\alpha e^I} + \frac{1}{2\alpha \bar{e}} \right] = 0.$$

Solving for k_1^U

$$k_1^U = \frac{4\alpha e^I \bar{e} + \beta \cdot \delta \cdot (2 \cdot e^I - \bar{e})}{8 \cdot \alpha^2 \cdot (e^I)^2 \cdot \bar{e}}.$$

The myopic level of corruption is $k^M(e^I) = 1/(2\alpha e^I)$. Thus, $k_1^* > k^M(e^I)$ if and only if $e^I > \bar{e}/2$.

D Proposition 3

Given

$$\frac{\partial k_1^U}{\partial e^I} = \frac{-2\alpha e^I \overline{e} + \beta\delta(\overline{e} - e^I)}{4\alpha^2 (e^I)^2 \overline{e}},$$

unconstrained level of corruption, k_1^U, increases in e^I if and only if

$$\frac{-2\alpha e^I \overline{e} + \beta\delta(\overline{e} - e^I)}{4\alpha^2 (e^I)^2 \overline{e}} > 0 \quad \Leftrightarrow \quad e^I < \frac{\beta\delta\overline{e}}{\beta\delta + 2\alpha\overline{e}}.$$

E Proposition 4

The Bureaucrat chooses k_1 to maximize their total expected utility

$$V(k_1) = \underbrace{\pi(k_1) - c(k_1, e^I)}_{\text{Period 1 Payoff}} + \delta \underbrace{\left[(1 - p(k_1))u^{B,M}(e^I) + p(k_1)V_B^C\right]}_{\text{Expected Period 2 Payoff}}$$

where $u^{B,M}(e^I) = u^B(k^M(e^I), e^I) = \frac{1}{4\alpha e^I}$ is the continuation payoff if the Incumbent is retained, and $V_B^C := \mathbb{E}[u^{B,M}(e^C)] = \frac{1}{2\alpha\overline{e}}$ is the expected continuation payoff if the Incumbent is replaced by a Challenger.

The probability of detection function, $p(k_1)$, is piecewise

$$p(k_1) = \begin{cases} 1 & \text{if } k_1 \geq (1+\gamma)/\beta \quad \text{(Saturation Region);} \\ \beta k_1 - \gamma & \text{if } \gamma/\beta < k_1 < (1+\gamma)/\beta \quad \text{(Strategic Region);} \\ 0 & \text{if } k_1 \leq \gamma/\beta \quad \text{(Safe Zone).} \end{cases}$$

Let's analyze the Bureaucrat's optimal choice, k_1^*, in each of these three regions.

If the Bureaucrat chooses k_1 high enough such that detection is certain ($k_1 \geq (1+\gamma)/\beta$), the probability of replacement is 1. The objective function becomes

$$V(k_1) = (k_1 - \alpha(k_1)^2 e^I) + \delta V_B^C.$$

Since the second term is constant with respect to k_1, the Bureaucrat's problem reduces to maximizing their first-period payoff. The first-order condition is $\frac{\partial V}{\partial k_1} = 1 - 2\alpha k_1 e^I = 0$, which implies the myopic corruption level, $k_1 = \frac{1}{2\alpha e^I} = k^M(e^I)$.

This choice is valid if it lies within the saturation region, i.e., if $k^M(e^I) \geq (1+\gamma)/\beta$.

$$\frac{1}{2\alpha e^I} \geq \frac{1+\gamma}{\beta} \iff e^I \leq \frac{\beta}{2\alpha(1+\gamma)} := e_L.$$

For any Incumbent with enforcement $e^I \in [0, e_L]$, the Bureaucrat, thus, chooses the myopic level of corruption, $k_1^*(e^I) = k^M(e^I)$, which guarantees electoral replacement.

If the Bureaucrat chooses k_1 low enough such that detection is impossible ($k_1 \leq \gamma/\beta$), the probability of replacement is 0. The objective function becomes:

$$V(k_1) = (k_1 - \alpha(k_1)^2 e^I) + \delta u^{B,M}(e^I).$$

Again, the second term is constant in k_1, and the Bureaucrat's optimal choice is the myopic level, $k_1 = k^M(e^I)$.

This choice is valid if it lies within the safe zone, i.e., if $k^M(e^I) \leq \gamma/\beta$. This condition holds when:

$$\frac{1}{2\alpha e^I} \leq \frac{\gamma}{\beta} \iff e^I \geq \frac{\beta}{2\alpha\gamma} := e_U.$$

Thus, for any Incumbent with enforcement $e^I \geq e_U$, the Bureaucrat chooses the myopic level of corruption, $k_1^*(e^I) = k^M(e^I)$, which guarantees re-election.

For intermediate levels of corruption, the objective function is

$$V(k_1) = (k_1 - \alpha(k_1)^2 e^I) + \delta\left[(1 - (\beta k_1 - \gamma))u^{B,M}(e^I) + (\beta k_1 - \gamma)V_B^C\right].$$

The first-order condition is:

$$\frac{\partial V}{\partial k_1} = 1 - 2\alpha k_1 e^I + \delta\left[-\beta u^{B,M}(e^I) + \beta V_B^C\right] = 0.$$

Solving for k_1 gives the unconstrained strategic level of corruption, $k_1^U(e^I)$:

$$2\alpha k_1 e^I = 1 - \delta\beta\left(u^{B,M}(e^I) - V_B^C\right)$$

$$k_1^U(e^I) = \frac{1 - \delta\beta\left(\frac{1}{4\alpha e^I} - \frac{1}{2\alpha\overline{e}}\right)}{2\alpha e^I} = \frac{4\alpha e^I\overline{e} + \beta\delta(2e^I - \overline{e})}{8\alpha^2(e^I)^2\overline{e}}.$$

This is the optimal choice provided it lies in the strategic interior, which occurs when the Incumbent's enforcement level e^I is such that the myopic choice would be neither in the safe zone nor in the saturation zone. That is, this solution applies for $e^I \in (e_L, e_U)$.

Therefore

$$k_1^*(e^I) = \begin{cases} k^M(e^I) & \text{if } e^I \leq e_L \quad \text{(Myopic Saturation);} \\ k_1^U(e^I) & \text{if } e^I \in (e_L, e_U) \quad \text{(Strategic Region);} \\ k^M(e^I) & \text{if } e^I \geq e_U \quad \text{(Myopic Safe Zone).} \end{cases}$$

F Proposition 5

The analysis assumes an interior solution where the strategic region is non-empty and contains the Bureaucrat's indifference point, i.e., $e_L < \overline{e}/2 < e_U$. The size of the collusion region is the length of the interval $[e_L, \overline{e}/2]$, given by $W_C = \overline{e}/2 - e_L$. The size of the sabotage region is the length of the interval $[\overline{e}/2, e_U]$, given by $W_S = e_U - \overline{e}/2$.

Substituting the definitions for e_L and e_U:

$$W_C = \frac{\overline{e}}{2} - \frac{\beta}{2\alpha(1+\gamma)}$$

$$W_S = \frac{\beta}{2\alpha\gamma} - \frac{\overline{e}}{2}$$

The comparative statics are found by taking the partial derivatives of W_C and W_S with respect to each parameter.

1. Collusion Region Size (W_C):
   (a) $\frac{\partial W_C}{\partial \bar{e}} = \frac{1}{2} > 0$. The collusion region expands with state capacity.
   (b) $\frac{\partial W_C}{\partial \alpha} = -\left(-\frac{\beta}{2\alpha^2(1+\gamma)}\right) = \frac{\beta}{2\alpha^2(1+\gamma)} > 0$. The collusion region expands with the inherent difficulty of graft.
   (c) $\frac{\partial W_C}{\partial \gamma} = -\left(\frac{\beta}{2\alpha} \cdot \frac{-1}{(1+\gamma)^2}\right) = \frac{\beta}{2\alpha(1+\gamma)^2} > 0$. The collusion region expands with the size of the safe zone.
   (d) $\frac{\partial W_C}{\partial \beta} = -\frac{1}{2\alpha(1+\gamma)} < 0$. The collusion region shrinks with better monitoring.
2. Sabotage Region Size (W_S):
   (a) $\frac{\partial W_S}{\partial \bar{e}} = -\frac{1}{2} < 0$. The sabotage region shrinks with state capacity.
   (b) $\frac{\partial W_S}{\partial \alpha} = -\frac{\beta}{2\alpha^2\gamma} < 0$. The sabotage region shrinks with the inherent difficulty of graft.
   (c) $\frac{\partial W_S}{\partial \gamma} = -\frac{\beta}{2\alpha\gamma^2} < 0$. The sabotage region shrinks with the size of the safe zone.
   (d) $\frac{\partial W_S}{\partial \beta} = \frac{1}{2\alpha\gamma} > 0$. The sabotage region expands with better monitoring.

G Proposition 6

Given k^M decreases in e^I, any non-monotonicity in $k_1^*(e^I)$ must occur within the strategic region, where $e^I \in (e_L, e_U)$.

Within this region, the level of corruption is given by the unconstrained strategic function, $k_1^U(e^I)$. We know that $k_1^U(e^I)$ is itself non-monotonic in e^I.

Thus, a non-monotonic relationship exists if and only if $e^{**} > e_L$. Substituting the expressions for e^{**} and e_L $\frac{\beta\delta\bar{e}}{\beta\delta+2\alpha\bar{e}} > \frac{\beta}{2\alpha(1+\gamma)}$, we get

$$2\alpha\delta(1+\gamma)\bar{e} - 2\alpha\bar{e} > \beta\delta \quad \Leftrightarrow \quad 2\alpha\bar{e}\left[\delta(1+\gamma) - 1\right] > \beta\delta$$

If $\delta(1+\gamma) - 1 \leq 0$, the left-hand side of the inequality is less than or equal to zero, while the right-hand side is strictly positive. The inequality cannot be satisfied.

When $\delta(1+\gamma) - 1 > 0$, we can divide both sides by this positive term to isolate $\bar{e}$

$$\bar{e} > \frac{\beta\delta}{2\alpha} \cdot \frac{1}{\delta(1+\gamma) - 1}$$

Both conditions are necessary and sufficient. If they hold, then $e^{**} > e_L$, which guarantees that there is an interval $(e_L, \min(e_U, e^{**}))$ where $\frac{dk_1^*}{de^I} > 0$, making the function non-monotonic. If either fails, then $e^{**} \leq e_L$, meaning that for all e^I in the strategic region (e_L, e_U), the derivative $\frac{dk_1^U}{de^I}$ is negative. In this case, the function $k_1^*(e^I)$ would be strictly decreasing over its entire domain.

H Proposition 7

The Voter's belief updating depends on the Bureaucrat's equilibrium strategy, $k_1^*(e^I)$, which, is a piecewise function of the Incumbent's type, e^I. This partitions the analysis into three distinct regions based on the thresholds $e_L := \frac{\beta}{2\alpha(1+\gamma)}$ and $e_U := \frac{\beta}{2\alpha\gamma}$.

By Bayes' rule, Voter's posterior expectation of the Incumbent's type conditional on observing a corruption signal is

$$\mathbb{E}[e^I|s_1 = 1] = \frac{\int_0^{\overline{e}} e^I \cdot \Pr(s_1 = 1|e^I) f(e^I) de^I}{\int_0^{\overline{e}} \Pr(s_1 = 1|e^I) f(e^I) de^I}. \tag{7}$$

where the prior density is $f(e^I) = 1/\overline{e}$ for $e^I \in [0, \overline{e}]$. The conditional probability, $\Pr(s_1 = 1|e^I)$, is determined by the Bureaucrat's equilibrium action, $p(k_1^*(e^I))$.

The total probability of observing a corruption signal

$$\begin{aligned}\Pr(s_1 = 1) &= \frac{1}{\overline{e}} \left[\int_0^{e_L} p(k^M(e^I)) de^I + \int_{e_L}^{e_U} p(k_1^U(e^I)) de^I + \int_{e_U}^{\overline{e}} p(k^M(e^I)) de^I \right] \\ &= \frac{1}{\overline{e}} \left[\int_0^{e_L} 1 \, de^I + \int_{e_L}^{e_U} (-\gamma + \beta k_1^U(e^I)) \, de^I + \int_{e_U}^{\overline{e}} 0 \, de^I \right].\end{aligned}$$

Solving these integrals

$$\Pr(s_1 = 1) = -\frac{1}{\overline{e}} \cdot \frac{\beta(\alpha\delta\overline{e} + (\beta\delta + 2\alpha\overline{e}) \log(\frac{\gamma}{1+\gamma}))}{4\alpha^2\overline{e}}. \tag{8}$$

The numerator for the posterior expectation is

$$\mathbb{E}[e^I \cdot \mathbf{1}_{\{s_1=1\}}] = \frac{1}{\overline{e}} \left[\int_0^{e_L} e^I \cdot 1 \, de^I + \int_{e_L}^{e_U} e^I(-\gamma + \beta k_1^U(e^I)) \, de^I + \int_{e_U}^{\overline{e}} e^I \cdot 0 \, de^I \right].$$

Solving this

$$\mathbb{E}[e^I \cdot \mathbf{1}_{\{s_1=1\}}] = \frac{1}{\overline{e}} \cdot \frac{\beta^2(\beta\delta + \alpha\overline{e} + \alpha\delta\overline{e}\gamma(1+\gamma) \log(\frac{\gamma}{1+\gamma}))}{8\alpha^3\overline{e}\gamma(1+\gamma)}. \tag{9}$$

Combining them

$$\mathbb{E}[e^I|s_1 = 1] = \frac{\mathbb{E}[e^I \cdot \mathbf{1}_{\{s_1=1\}}]}{\Pr(s_1 = 1)} = -\frac{\beta \left(\beta\delta + \overline{e} \left(\alpha + \alpha\delta\gamma(1+\gamma) \log(\frac{\gamma}{1+\gamma})\right)\right)}{2\alpha\gamma(1+\gamma) \left(\beta\delta \log(\frac{\gamma}{1+\gamma}) + \overline{e} \left(\alpha\delta + 2\alpha \log(\frac{\gamma}{1+\gamma})\right)\right)}. \tag{10}$$

The derivation for the posterior belief conditional on observing no corruption signal follows symmetrically.

$$\mathbb{E}[e^I|s_1 = 0] = \frac{\mathbb{E}[e^I \cdot \mathbf{1}_{\{s_1=0\}}]}{\Pr(s_1 = 0)} = -\frac{\beta^3\delta + \overline{e} \left(\alpha\beta^2 - 4\alpha^3\gamma(1+\gamma) + \alpha\beta^2\delta\gamma(1+\gamma) \log(\frac{\gamma}{1+\gamma})\right)}{2\alpha\gamma(1+\gamma) \left(\beta^2\delta \log(\frac{\gamma}{1+\gamma}) + \overline{e} \left(\alpha(4\alpha + \beta\delta) + 2\beta\alpha \log(\frac{\gamma}{1+\gamma})\right)\right)}. \tag{11}$$

For the simple retrospective voting rule to be a Perfect Bayesian Equilibrium (PBE), two conditions of sequential rationality must be met by the Voter.

$$\mathbb{E}[e^I|s_1 = 1] < \mathbb{E}[e^C] = \frac{\overline{e}}{2}$$

and

$$\mathbb{E}[e^I|s_1 = 0] \geq \mathbb{E}[e^C] = \frac{\overline{e}}{2}.$$

Note that the Voter's posterior belief after observing $s_1 = 1$ has a support that is strictly bounded above by e_U. Formally, $\mu(e^I|s_1 = 1) = 0$ for all $e^I \geq e_U$. This implies that the posterior expectation, $\mathbb{E}[e^I|s_1 = 1]$, must also be strictly less than e_U

$$\mathbb{E}[e^I|s_1 = 1] = \frac{\int_0^{e_U} e \cdot p(k_1^*(e)) \cdot \frac{1}{\overline{e}} de}{\int_0^{e_U} p(k_1^*(e)) \cdot \frac{1}{\overline{e}} de} < e_U.$$

For replacement to be rational, we need $\mathbb{E}[e^I|s_1 = 1] < \overline{e}/2$. Since $\mathbb{E}[e^I|s_1 = 1]$ is bounded above by the constant e_U, while the Challenger's expected quality $\overline{e}/2$ grows linearly with state capacity, there must exist a threshold for $\overline{e}$ above which this condition is always met. A sufficient condition is

$$e_U \leq \frac{\overline{e}}{2} \implies \overline{e} \geq 2e_U$$

For any $\overline{e} > \frac{\beta}{\alpha\gamma}$, it is rational for the Voter to replace an Incumbent who generates a corruption signal.

The Voter's posterior expectation after observing $s_1 = 0$ is

$$\mathbb{E}[e^I|s_1 = 0] = \frac{\int_{e_L}^{e_U} e \cdot (1 - p(k_1^*(e)))de + \int_{e_U}^{\overline{e}} e \cdot 1 de}{\int_{e_L}^{e_U} (1 - p(k_1^*(e)))de + \int_{e_U}^{\overline{e}} 1 de}.$$

As $\overline{e} \to \infty$, the expectation $\mathbb{E}[e^I|s_1 = 0]$ approaches the average value of an incumbent in the range $[e_U, \overline{e}]$

$$\lim_{\overline{e}\to\infty} \mathbb{E}[e^I|s_1 = 0] = \frac{\frac{1}{2}(\overline{e}^2 - e_U^2)}{\overline{e} - e_U} = \frac{\frac{1}{2}(\overline{e} - e_U)(\overline{e} + e_U)}{\overline{e} - e_U} = \frac{\overline{e} + e_U}{2}$$

For retention to be rational, we need $\mathbb{E}[e^I|s_1 = 0] \geq \overline{e}/2$. In the limit, this condition becomes:

$$\frac{\overline{e} + e_U}{2} \geq \frac{\overline{e}}{2} \implies \overline{e} + e_U \geq \overline{e} \implies e_U \geq 0$$

Since $e_U > 0$, this condition is always satisfied for a sufficiently high state capacity $\overline{e}$. Thus, when $\overline{e} > \frac{\beta}{\alpha\gamma}$, both conditions are satisfied simultaneously.

I Proposition 8

The myopic level of corruption, $k^M(e)$, is implicitly defined by the first-order condition (FOC) of the Bureaucrat's static maximization problem

$$\pi'(k^M(e)) - c_k(k^M(e), e) = 0$$

The second-order condition (SOC) for a maximum requires that the objective function be locally concave, which is satisfied by our assumptions

$$\frac{\partial^2 u_B}{\partial k^2} = \pi''(k) - c_{kk}(k, e) < 0$$

where the strict inequality holds for an interior solution ($k > 0$), as $\pi'' \leq 0$ and $c_{kk} \geq 0$, and at least one must be strict for the problem to be well-behaved.

To find the effect of an increase in the enforcement level e on the optimal choice of corruption k^M, we apply the Implicit Function Theorem to the FOC. Let the FOC be denoted by $F(k, e) = \pi'(k) - c_k(k, e) = 0$. The derivative is given by:

$$\frac{dk^M}{de} = -\frac{\partial F/\partial e}{\partial F/\partial k} = -\frac{-c_{ke}(k, e)}{\pi''(k) - c_{kk}(k, e)} = \frac{c_{ke}(k, e)}{\pi''(k) - c_{kk}(k, e)}$$

The denominator is negative from the SOC. The numerator is non-negative by the assumption that $c_{ke} \geq 0$. For the result to be strictly decreasing as stated in the proposition, we require the standard assumption that $c_{ke} > 0$. This implies that an increase in enforcement strictly increases the marginal cost of corruption, thus reducing the Bureaucrat's optimal choice of k. Therefore:

$$\frac{dk^M}{de} < 0$$

State capacity, $\bar{e}$, affects corruption indirectly through its effect on the enforcement level, $e = \min\{\theta^P, \bar{e}\}$. We can find the effect of a change in $\bar{e}$ on k^M using the chain rule:

$$\frac{dk^M}{d\bar{e}} = \frac{dk^M}{de} \cdot \frac{de}{d\bar{e}}$$

From part (1), we know that the first term, $\frac{dk^M}{de}$, is strictly negative.

The second term, $\frac{de}{d\bar{e}}$, depends on the politician's type, θ^P:

$$\frac{de}{d\bar{e}} = \begin{cases} 1 & \text{if } \theta^P > \bar{e} \\ 0 & \text{if } \theta^P < \bar{e} \end{cases}$$

Combining these results, we find:

$$\frac{dk^M}{d\bar{e}} = \begin{cases} \frac{dk^M}{de} < 0 & \text{if } \theta^P > \bar{e} \\ 0 & \text{if } \theta^P < \bar{e} \end{cases}$$

Thus, an increase in state capacity weakly decreases the myopic level of corruption. It has no effect if the politician's type is already a binding constraint on enforcement ($\theta^P < \bar{e}$), but it strictly decreases corruption if the politician is constrained by state capacity ($\theta^P > \bar{e}$).

J Proposition 9

The Bureaucrat's problem in period 1 is

$$\max_{k_1 \geq 0} \left\{ u_B(k_1, e^I) + \delta \left[(1 - p(k_1)) u_B(k^M(e^I), e^I) + p(k_1) V_B^C \right] \right\}$$

The first-order condition for an interior solution is:

$$\frac{\partial u_B(k_1, e^I)}{\partial k_1} + \delta \frac{d}{dk_1}\left[(1-p(k_1))u_B(k^M(e^I), e^I) + p(k_1)V_B^C\right] = 0$$

$$\pi'(k_1) - c_k(k_1, e^I) - \delta p'(k_1)\left[u_B(k^M(e^I), e^I) - V_B^C\right] = 0 \tag{12}$$

Let $L(k_1, e^I) = \pi'(k_1) - c_k(k_1, e^I)$. The condition 12 can be rewritten as:

$$L(k_1, e^I) = \delta p'(k_1)\left[u_B(k^M(e^I), e^I) - V_B^C\right].$$

By definition, $k^M(e^I)$ is the unique value that satisfies $L(k^M(e^I), e^I) = 0$. Since the objective function is concave in k_1 (as $\pi'' - c_{kk} < 0$), the optimal choice k_1^* will be greater or less than $k^M(e^I)$ according to the sign of the right-hand side.

Let $g(e^I) := u_B(k^M(e^I), e^I)$. By the Envelope Theorem, the derivative of g with respect to e^I is

$$g'(e^I) = \frac{\partial}{\partial e^I}\left[\pi(k^M(e^I)) - c(k^M(e^I), e^I)\right] = -c_e(k^M(e^I), e^I) < 0.$$

Thus, $g(e^I)$ is a strictly decreasing function of e^I. Since V_B^C is a constant value determined by the distribution of Challenger types, and $g(e^I)$ is continuous and strictly decreasing over its domain, there exists a unique threshold e^* such that $g(e^*) = u_B(k^M(e^*), e^*) = V_B^C$.

We can now characterize the solution based on e^I.

1. When $e^I \leq e^*$, since $g(e^I)$ is decreasing, $e^I \leq e^*$ implies $g(e^I) \geq g(e^*) = V_B^C$. The retention value is positive. The right-hand side of the first-order condition is positive

$$L(k_1, e^I) = \delta p'(k_1)[g(e^I) - V_B^C] \geq 0$$

This implies $\pi'(k_1) \geq c_k(k_1, e^I)$. Given the concavity of the Bureaucrat's payoff function, this can only hold if $k_1^* \leq k^M(e^I)$. This is strategic collusion.

2. If $e^I > e^*$, then $g(e^I) < g(e^*) = V_B^C$. The retention value is negative. The right-hand side of the first-order condition is negative

$$L(k_1, e^I) = \delta p'(k_1)[g(e^I) - V_B^C] < 0.$$

This implies $\pi'(k_1) < c_k(k_1, e^I)$, which requires that $k_1^* > k^M(e^I)$. This is strategic sabotage.

K Proposition 10

The equilibrium level of corruption k_1^* is implicitly defined by the first-order condition (FOC) of the Bureaucrat's period 1 optimization problem:

$$F(k_1^*, e^I) \equiv \pi'(k_1^*) - c_k(k_1^*, e^I) - \delta p'(k_1^*)\left[u_B(k^M(e^I), e^I) - V_B^C\right] = 0$$

To find the effect of a change in e_I on k_1^*, we apply the Implicit Function Theorem, which states that $\frac{dk_1^*}{de_I} = -\frac{\partial F/\partial e^I}{\partial F/\partial k_1}$.

First, we compute the denominator, $\frac{\partial F}{\partial k_1}$.

$$\frac{\partial F}{\partial k_1} = \pi''(k_1^*) - c_{kk}(k_1^*, e^I) - \delta p''(k_1^*) \left[u_B(k^M(e^I), e^I) - V_B^C\right]$$

For k_1^* to be a maximum, this term, which is the second derivative of the Bureaucrat's objective function, must be negative. This is guaranteed by our assumptions that $\pi'' \leq 0$, $c_{kk} \geq 0$, and $p'' \leq 0$, which ensure that each component of the expression is non-positive (assuming the retention value term is not sufficiently negative to overcome the first two terms, a standard condition for stability). Thus, the sign of $\frac{dk_1^*}{de_I}$ is the same as the sign of the numerator, $\frac{\partial F}{\partial e^I}$.

Next, we compute the numerator by differentiating the FOC with respect to e_I:

$$\frac{\partial F}{\partial e^I} = -c_{ke}(k_1^*, e^I) - \delta p'(k_1^*) \frac{d}{de_I} \left[u_B(k^M(e^I), e^I)\right]$$

We use the Envelope Theorem to evaluate the derivative of the Bureaucrat's continuation value function:

$$\frac{d}{de_I} u_B(k^M(e^I), e^I) = \frac{d}{de_I} \left[\pi(k^M(e^I)) - c(k^M(e^I), e^I)\right] = -c_e(k^M(e^I), e^I)$$

Substituting this back into the expression for $\frac{\partial F}{\partial e^I}$ yields:

$$\frac{\partial F}{\partial e^I} = -c_{ke}(k_1^*, e^I) - \delta p'(k_1^*)[-c_e(k^M(e^I), e^I)] = -c_{ke}(k_1^*, e^I) + \delta p'(k_1^*) c_e(k^M(e^I), e^I)$$

This expression captures the two competing effects. The first term, $-c_{ke}$, is the direct deterrence effect. Since $c_{ke} \geq 0$, this term is non-positive. The second term, $\delta p'(k_1^*) c_e(k^M(e^I), e^I)$, is the strategic electoral effect. Since $\delta > 0$, $p' > 0$, and $c_e > 0$, this term is positive.

Combining these results, we have:

$$\frac{dk_1^*}{de_I} = -\frac{-c_{ke}(k_1^*, e^I) + \delta p'(k_1^*) c_e(k^M(e^I), e^I)}{\frac{\partial F}{\partial k_1}}$$

Since $\frac{\partial F}{\partial k_1} < 0$, the sign of the derivative is determined by the sign of the numerator, completing the proof.

For the second part, we analyze the effect of state capacity, $\bar{e}$. The parameter $\bar{e}$ only affects the FOC through the Challenger's expected utility, $V_B^C(\bar{e})$. By the Implicit Function Theorem, $\frac{dk_1^*}{d\bar{e}} = -\frac{\partial F/\partial \bar{e}}{\partial F/\partial k_1}$. Again, the sign is determined by $\partial F/\partial \bar{e}$.

$$\frac{\partial F}{\partial \bar{e}} = -\delta p'(k_1) \left[-\frac{dV_B^C}{d\bar{e}}\right] = \delta p'(k_1) \frac{dV_B^C}{d\bar{e}}$$

Recall that $V_B^C(\bar{e}) = \int_0^1 u_B(k^M(\min\{\theta^C, \bar{e}\}), \min\{\theta^C, \bar{e}\}) d\theta^C$. An increase in $\bar{e}$ weakly increases the enforcement level for any Challenger type $\theta^C \geq \bar{e}$ and thus weakly decreases the myopic utility u_B. Therefore, $\frac{dV_B^C}{d\bar{e}} \leq 0$. Since $\delta > 0$ and $p' > 0$, we have $\frac{\partial F}{\partial \bar{e}} \leq 0$. This implies that $\frac{dk_1^*}{d\bar{e}} \leq 0$.

L Proposition 11

The posterior of e^I after observing $s_1 = 1$ is

$$\mathbb{E}[e^I|s_1 = 1] = \int_0^{\bar{e}} e^I \cdot \mu(e^I|s_1 = 1) de_I$$

By Bayes' rule, $\mu(e^I|s_1 = 1) = \frac{\Pr(s_1=1|e^I)f(e^I)}{\Pr(s_1=1)}$, where $f(e^I)$ is the prior density and $\Pr(s_1 = 1|e^I) = p(k_1^*(e^I))$ and $\Pr(s_1 = 1) = \mathbb{E}[p(k_1^*(e^I))]$.

$$\mathbb{E}[e^I|s_1 = 1] = \frac{\int e^I \cdot p(k_1^*(e^I))f(e^I)de_I}{\mathbb{E}[p(k_1^*(e^I))]} = \frac{\mathbb{E}[e^I \cdot p(k_1^*(e^I))]}{\mathbb{E}[p(k_1^*(e^I))]}$$

Condition $\mathbb{E}[e^I|s_1 = 1] < \mathbb{E}[e^I]$ is equivalent to

$$\frac{\mathbb{E}[e^I \cdot p(k_1^*(e^I))]}{\mathbb{E}[p(k_1^*(e^I))]} < \mathbb{E}[e^I]$$

$$\mathbb{E}[e^I \cdot p(k_1^*(e^I))] - \mathbb{E}[e^I]\mathbb{E}[p(k_1^*(e^I))] < 0$$

The left-hand side is the definition of the covariance between e_I and $p(k_1^*(e^I))$. Thus, $\mathbb{E}[e^I|s_1 = 1] < \mathbb{E}[e^I] \iff \text{Cov}(e^I, p(k_1^*(e^I))) < 0$. Because $\mathbb{E}[e^I] = \Pr(s_1 = 0)\mathbb{E}[e^I|s_1 = 0] + \Pr(s_1 = 1)\mathbb{E}[e^I|s_1 = 1]$,

$$\mathbb{E}[e^I|s_1 = 0] - \mathbb{E}[e^I] = -\frac{\Pr(s_1 = 1)}{\Pr(s_1 = 0)}\left(\mathbb{E}[e^I|s_1 = 1] - \mathbb{E}[e^I]\right)$$

Thus, $\mathbb{E}[e^I|s_1 = 0] > \mathbb{E}[e^I]$ if and only if $\mathbb{E}[e^I|s_1 = 1] < \mathbb{E}[e^I]$, which happens if and only if $\text{Cov}(e^I, p(k_1^*(e^I))) < 0$.

References

Barro, Robert J. 1973. "The control of politicians: an economic model." *Public choice* pp. 19–42.

Bersch, Katherine, Sérgio Praça and Matthew M Taylor. 2017. "State capacity, bureaucratic politicization, and corruption in the Brazilian state." *Governance* 30(1):105–124.

Bertelsmann Stiftung. 2024. "BTI 2024: Serbia Country Report.".
URL: *https://bti-project.org/en/reports/country-report/SRB*

Cheeseman, Nic and Caryn Peiffer. 2022. "The curse of good intentions: Why anticorruption messaging can encourage bribery." *American Political Science Review* 116(3):1081–1095.

Dawisha, Karen. 2015. *Putin's kleptocracy: who owns Russia?* Simon and Schuster.

Delgado-Vega, Álvaro and Ben Shaver. 2024. "A Model of Succession and Policy Concerns." *Available at SSRN 4981723* .

Fearon, James D. 1999. "Electoral accountability and the control of politicians: selecting good types versus sanctioning poor performance." *Democracy, accountability, and representation* pp. 55–97.

Fong, Christian and Keith Krehbiel. 2018. "Limited obstruction." *American Political Science Review* 112(1):1–14.

Freedom House. 2024. "Serbia: Freedom in the World 2024.".
URL: *https://freedomhouse.org/country/serbia/freedom-world/2024*

Gieczewski, Germán and Christopher Li. 2022. "Dynamic policy sabotage." *American Journal of Political Science* 66(3):617–629.

Groenendijk, Nico. 1997. "A principal-agent model of corruption." *Crime, law and social change* 27(3):207–229.

Heo, Kun and Elisa M Wirsching. 2024. Bureaucratic Sabotage and Policy Inefficiency. Technical report Working Paper.

Hirsch, Alexander V and Jonathan P Kastellec. 2022. "A theory of policy sabotage." *Journal of Theoretical Politics* 34(2):191–218.

Hoffmann, Leena Koni and Raj Navanit Patel. 2017. *Collective action on corruption in Nigeria: A social norms approach to connecting society and institutions*. Chatham House London.

Johnston, Michael. 2017. Why Do So Many Anti Corruption Efforts Fail? In *Transnational Financial Crime*. Routledge pp. 263–292.

Lanskoy, Miriam and Dylan Myles-Primakoff. 2018. "The rise of kleptocracy: Power and plunder in Putin's Russia." *Journal of Democracy* 29(1):76–85.

PARI. 2022. The State Capture Commission: A Bite-Sized Summary. Technical report Public Affairs Research Institute.
URL: *https://pari.org.za/wp-content/uploads/2022/09/PARI-Summary-The-Zondo-Commission-A-bite-sized-summary-v360.pdf*

Patty, John W. 2016. "Signaling through obstruction." *American Journal of Political Science* 60(1):175–189.

Reuters. 2018. "Nigeria's Buhari Launches Re-election Bid with Corruption Still in Focus.".
URL: *https://www.reuters.com/article/world/nigerias-buhari-launches-re-election-bid-with-corruption-still-in-focus-idUSKCN1NO07A/*

Reuters. 2025. "Ukraine curbs autonomy of anti-corruption agencies.".
URL: *https://www.reuters.com/en/ukraine-curbs-autonomy-anti-corruption-agencies-2025-07-22/*

Rose-Ackerman, Susan. 2013. *Corruption: A study in political economy.* Academic press.

Shleifer, Andrei and Robert Vishny. 1993. "Putin's kleptocracy: who owns Russia?".

South African Institute of International Affairs. 2024. "Commissions, Corruption and State Capture: Charting the Way Forward.".
URL: *https://saiia.org.za/wp-content/uploads/2024/05/SAIIA*$_S R_C$*ommissionsCorruptionState.pdf*

Transparency International Ukraine. 2025. "Stripping NABU and SAPO of Independence: Legal Analysis.".
URL: *https://ti-ukraine.org/en/news/stripping-nabu-and-sapo-of-independence-legal-analysis-of-amendments-to-draft-law-no-12414-2/*

U4 Anti-Corruption Resource Centre. 2023. "Nigeria: Corruption and Anti-Corruption.".
URL: *https://www.u4.no/publications/corruption-and-anti-corruption-efforts-in-nigeria.pdf*

Williams, Martin J. 2021. "Beyond state capacity: bureaucratic performance, policy implementation and reform." *Journal of Institutional Economics* 17(2):339–357.