Kompromat Can Align Incentives But Ruin Reputations*

Ryan Hübert† Andrew T. Little‡

March 9, 2021

Abstract

Political leaders face many agency problems, such as managing subordinates who may not honestly report information. One potential solution to these problems is *kompromat*: the threat to release compromising information. Using a cheap talk model, we demonstrate how kompromat can improve communication, making both principal and agent better off. However, using kompromat to solve an agency problem generates two costs. First, its mere existence means it may leak inadvertently. Second, since kompromat works by threatening the reputation of subordinates, common knowledge that an organization uses kompromat might be costly even if it is never leaked. These possibilities may foreclose *all* communication from a subordinate who would have provided truthful information in the absence of kompromat.

Keywords: Formal Theory, Soviet and Post-Soviet Politics, Bureaucracy

*Authors are listed in alphabetical order. Many thanks to Jonathan Bendor, Benoit Crutzen, Tiberiu Dragu, Sean Gailmard, Scott Gehlbach, Anton Kazun, Monika Nalepa, John Patty, Melissa Samarin, Fiona Shen-Bayh, Konstantin Sonin, Henry Thompson, and audience members at WPSA 2019, EPSA 2019, the Higher School of Economics and Stanford GSB for helpful comments. We thank Otto Kienitz for excellent research assistance. All errors and omissions are our own.

†Assistant Professor, Department of Political Science, University of California, Davis.
Email: rhubert@ucdavis.edu. *Website*: http://www.ryanhubert.com/.

‡Assistant Professor, Department of Political Science, University of California, Berkeley.
210 Barrows Hall, Berkeley CA 94720 *Email*: andrew.little@berkeley.edu. *Website*: http://www.andrewtlittle.com/.

To keep kompromat on enemies is a pleasure. To keep kompromat on friends is a must.
– Yuliya Latynina

Political leaders rely on subordinates to implement their agendas. For example, in the early years of the Soviet Union, Joseph Stalin relied heavily on secret police organizations to consolidate the communist revolution and eliminate potential threats to centralized Soviet rule. These organizations were vital for the identification and repression of Soviet political enemies. During the Great Terror, Stalin's head of the secret police, Nikolai Yezhov, compiled lists of political enemies and monitored the progress of wide-scale purges occurring across the Soviet Union. As Gregory (2009) points out, Yezhov was an important source of information on the operations: "[Yezhov] met with Stalin eighty-nine times (about once every three days) between January 30 and November 17, 1938 [...] Yezhov was gathering statistics on a daily basis from his subordinates and would have been in possession of up-to-date figures" (p. 190).

A natural concern for Stalin was whether Yezhov actually provided him with accurate information. More generally, heads of agencies (like Yezhov) and lower-level bureaucrats may not always provide truthful information or undertake actions that align with their organization's or leader's objectives. And Soviet leaders often did question the motives of officials tasked with state repression. Stalin, for example, removed two different heads of the secret police, Genrykh Yagoda and Yezhov, each of whom were suspected of failing to enact his agenda.

Any employer can provide incentives by promoting those who do what they want and firing those who do not. However, when performance is difficult to monitor and the conflicts of interest are large, these levers may not be enough to keep subordinates in line. Particularly in more autocratic contexts, political leaders can resort to another extreme by arresting or even executing agents they believe to have crossed them. And, of course, Stalin and other dictators frequently go in this direction, as V.M. Molotov's wife Polina Zhemchuzhina (arrested) and Yezhov (executed) learned all too well.

Still, such methods are costly to the punishers in addition to those on the receiving end. Ex-

treme sanctions can cause leaders to lose talented and knowledgeable agents, and potentially deter similarly talented individuals from working for the government in the first place. In this paper we study a more intermediate—and, potentially, more effective—technology for principals to control agents: holding and threatening to release compromising material. We refer to this compromising material by its Russian name: *kompromat*.

This kind of kompromat was commonplace in the Soviet Union. As Gregory (2009) colorfully describes, Stalin famously held kompromat on his appointed heads of the secret police, some rumored to be hidden in a safe. The offending behavior ranged in seriousness from being bisexual or Jewish, to alcoholism, to having murdered a superior or also working for one of his enemies. Kompromat has also been used to keep agents in line in many other countries and time periods.

While kompromat can be used to manage a wide variety of agency problems (and, as discussed below, for other purposes), we illustrate our core arguments using a cheap talk game. To be concrete, we use a running example of authoritarian leaders (principals) relying on heads of secret police organizations (agents) for information about threats to the regime. An agent gets a noisy signal (e.g., whether an citizen or group of citizens is subversive) and sends a report to the principal, who takes an action (e.g., how punitive to be towards the citizen or group under surveillance). If the agent has a different preference over the principal's action (maybe he wants the principal to be more or less punitive towards the group), this may undermine truthful communication. However, if the principal can threaten to punish the agent by leaking kompromat if the agent is suspected of lying, this may make truthful communication possible. Further, our analysis also shows that the potential upsides of kompromat for an agent should outweigh the potential cost from leaked kompromat. Indeed, we argue that this is the nature of kompromat: for it to "work" in contexts with voluntary employment, it must make both the principal *and* the agent better off than if they choose not to work together.

If kompromat can be so effective, why is it pervasive in some political contexts but not others? We study two models that highlight different costs that arise when an organization uses kompromat

to improve communication between leaders and subordinates.

The core idea of the first model is that the mere existence of kompromat means it may leak. Even if the agent does exactly what the principal wants, leaks can happen for exogenous reasons or because the principal gets the mistaken impression than the agent lied. This lowers the value of employment and may make it impossible to find a wage that both parties find acceptable.

The second model is more technically original and highlights a problem more unique to kompromat (as opposed to other technologies like sanctioning/firing agents who provide poor information). If the harm of releasing kompromat is to damage one's reputation to outside observers, and those with kompromat are hired by an organization, then just being a part of the organization in the first place will harm one's reputation. This makes joining the organization costly, and diminishes the usefulness of kompromat by weakening the principal's ability to threaten to reveal the agent's corruption. As a result, it can be more challenging to to hire agents to work for an organization that is tainted by corruption. (Or just more tainted by corruption than average.) However, our analysis shows that this does not lead to a complete unraveling, and some agents are willing to accept some reputational cost to join the organization.

1 Prior research

There are many potential uses of kompromat. Using examples from Russia in the 1990s, Ledeneva (2006) provides a useful typology that distinguishes between three distinct functions: as a commodity, as a weapon, and as an instrument for informal persuasion. Put another way (and drawing on our epigraph), it is useful to differentiate kompromat on "enemies" from kompromat on "friends." Kompromat on enemies serves an intuitive purpose: in competitive environments like political campaigns and business, releasing damaging information about an opponent can be clearly beneficial. According to Ledeneva (2006), this tactic was so common in Russian campaigns in the 1990s that voters eventually became insensitive to all but the most scandalous revelations.

Our focus here is kompromat on friends. More formally, we study a relationship between actors with partially aligned preferences, and show how the ability of one to harm the other leads to better outcomes for both. The mechanism that drives our model is most in line with the persuasion function of kompromat, in the sense that kompromat allows the principal to "persuade" the agent to tell the truth. Or, perhaps a more natural interpretation of our model is that kompromat allows the *compromised party* (the agent) to persuade the *kompromat holder* (the principal) that the policy-relevant information he provides is valid and truthful.

Most existing work on the kind of kompromat we study is historical and qualitative. Gregory (2009) documents how kompromat was used by Stalin to manage his top security personnel and Vatlin (2016) demonstrates how Soviet leaders implementing the Great Terror preferred subordinates against whom they held kompromat. Gorlizki (2013) provides evidence that these practices continued after Stalin, although the nature of the kompromat changed. Darden (2001) uses evidence from leaked tapes from Ukrainian Prime Minister Leonid Kuchma to show how his regime used the threat of releasing kompromat to keep various agents in check. Outside of the (former) Soviet Union, the head of the secret police in Peru during Alberto Fujimori's rule, Vladimiro Montesinos, kept video recordings of politicians and others accepting bribes. These were later leaked and Fujimori's regime did not survive this revelation.

The type of kompromat we study is conceptually closest to blackmail. Schelling (1956) provides an early and influential account of how making threats can help with bargaining, which Ellsberg (1968) applies to blackmail. In a more modern game-theoretic treatment, Schwarz and Sonin (2007) analyze a model in which an aggressor can extract all surplus from the victim by making threats even though those threats impose a cost on the aggressor. Similarly, Dal Bó, Dal Bó, and Di Tella (2006) study how private actors can use threats and bribes to get public agents to do their bidding.

In contrast, others have argued that that kompromat and other forms of blackmail can do more than redistribute surplus—it can mitigate Pareto inefficiencies induced by agency problems. Gam-

betta (2009) argues that this kind of "self-inflicted" blackmail helps criminals build trust with one another.[1] While our model formalizes the logic for how self-inflicted blackmail can be Pareto improving, it also highlights two important limitations of using kompromat to establish trust.

First, many accounts of kompromat or blackmail implicitly assume that a relationship has already commenced. Consider the canonical example from Schelling (1956) in which a kidnapped person provides his kidnapper with some kompromat that allows the kidnapper to release the prisoner without fear that the prisoner will go to the police. In this example, the prisoner has no choice about whether to be in the relationship with the kidnapper; he has already been kidnapped. However, in many situations, such as the bureaucratic employment context we study here, the compromised party must first choose whether to engage with someone who could potentially blackmail them. Our model shows that it matters whether the compromised party has this choice. Specifically, if participation in a relationship is voluntary, then the possibility that kompromat may be leaked can deter principals and agents from forming Pareto improving relationships.

Second, if kompromat is used to manage relationships in an organization, then this may cause the organization to obtain a reputation for being staffed by compromised people. Prior research has documented how an organization's penchant for attracting corrupt (or otherwise "bad") types can affect who will be willing to join that organization in the future (e.g., Tirole 1996; Caselli and Morelli 2004; Klašnja, Little, and Tucker 2018).[2] In our model, all agents (whether corrupt or not) are harmed when their organization becomes tainted by a reputation for corruption. A surprising insight from this model is that in order for kompromat to "work," the organization must be able to attract sufficient numbers of agents with *no* kompromat.

We focus on a specific, and common, principal-agent problem in organizations: the ability of biased subordinates to misrepresent their information (see, for example, Crawford and Sobel 1982;

1. Gambetta (2009) distinguishes this kind of blackmail from standard blackmail in which the blackmailer illicitly or coercively obtains compromising information on the victim. In that case, the victim is typically worse off.

2. Other work points out that if entering the public sector entails getting threatened, less able people will enter (e.g., Dal Bó, Dal Bó, and Di Tella 2006).

Dessein 2002; Gailmard and Patty 2013; Schnakenberg 2015). Our core model is similar in spirit to Patty and Penn (2019), who study a model where a biased agent can send a cheap talk signal to a principal. They demonstrate how costly investments in "loyalty" to their organization by the principal and agent can expand the scope for truthful communication. While the mechanism in our model is different, a core idea motivating our theory is that agents may wish to pay a cost (by way of leaked kompromat) in order to more effectively communicate with the principal. However, for this cost to provide incentives for truthfulness, the principal in our model has some ability to verify whether the agent lied (as in Austen-Smith and Wright 1992).

While our insights are applicable across a wide range of environments, kompromat may be particularly useful in authoritarian or autocratic contexts where more formalized institutions may not exist to mitigate principal-agent problems.[3] Its use in authoritarian regimes also presents unique challenges for democratic transitions. Some formerly authoritarian countries have engaged in "lustration," the practice of revealing political leaders' past collaboration with the previous regime in order to remove the possibility of using this as kompromat. While lustration removes the possibility of blackmail and makes voters better off (Ang and Nalepa 2018), uncompromised politicians may not enact lustration if the taint of collaboration with the previous regime harms their opponents' electoral prospects (Nalepa and Sonin 2019). Even though lustration may improve the welfare of voters in a newly democratic context, our theory suggests it also comes at a cost if removing the possibility of kompromat worsens agency problems between a policymaker and her subordinate.

More broadly, our model contributes to a growing literature on principal-agent relationships in autocracies. Several studies focus on trade-offs between loyalty and competence (Egorov and Sonin 2011; Zakharov 2016). Our reputation model focuses on how the interaction between a lack of bias (similar to loyalty) and corruption affects employability. Other papers study commitment problems in the absence of credible institutions can exacerbate principal-agent problems (Dragu

3. However, even in highly institutionalized environments, the institutions themselves may not alleviate (and may even exacerbate) agency problems (see, for example Patty and Turner, forthcoming).

and Polborn 2013; Rundlett and Svolik 2016; Tyson 2018), or tools which may improve principal-agent relationships in autocracies such as electoral fraud (Gehlbach and Simpser 2015) or mass purges (Montagnes and Wolton 2019). Closest to our motivating example, Dragu and Przeworski (2019) present a model of agency problems within authoritarian security services, though with a focus on varieties of moral hazard not directly related to our analysis.

2 The core agency problem

Our two models of kompromat build on a simplified cheap talk game adapted from Galeotti, Ghiglino, and Squintani (2013). In this section, we analyze the cheap talk game (without kompromat) to serve as a benchmark. A *principal* (P, pronoun she) relies on information gathered by an organization that is headed by an *agent* (A, pronoun he). The agent provides advice to the principal based on whatever information the organization is able to collect. The principal needs accurate information to make a policy decision. However, if the agent is biased, he might have an incentive to lie to the principal about the information the organization has obtained.

We represent the information that the principal needs as an unknown state of the world, $\theta \in \Theta = [0, 1]$. The principal and the agent have a common prior belief at the outset that θ is distributed according to the probability density function $f(\theta)$. The agent then receives some information, $s \in \{0, 1\}$, which provides information about the state θ and hence would improve the principal's information about θ. After observing s, the agent provides advice to the principal. He can tell the principal what he knows, or he can lie about it. Finally, after receiving this (potentially bad) advice, the principal makes a policy decision $x \in \mathbb{R}$.

The principal wants to make a policy decision that corresponds to the state of the world: $x = \theta$. However, the agent is biased and would prefer the principal to make a "higher" policy decision: $x = \theta + b$, where we assume without loss of generality that $b \geq 0$.[4] Therefore, b represents how

4. What matters is the absolute magnitude of the bias. It is easy to check that if $b < 0$ the constraint for truthful reporting is the same as what we subsequently find, with $|b|$ replacing b.

biased the agent is. We formally represent these preferences with quadratic loss utilities centered around each player's ideal policy decision:

$$u_A = -(x - (\theta + b))^2 \qquad\qquad u_P = -(x - \theta)^2$$

We model the agent's expertise, represented by his information s, in a flexible manner. After observing s, the agent forms a posterior about θ. Define $\tilde{\theta}_s \equiv \mathbb{E}[\theta|s]$ as the average of this posterior belief. To capture the idea that s contains meaningful information about θ, we assume that the agent's posterior belief about θ upon observing signal $s = 1$ is strictly higher than upon observing $s = 0$: $\tilde{\theta}_1 > \tilde{\theta}_0$. Let π be the probability that the agent learns that $s = 1$.

We interpret $\tilde{\theta}_1 - \tilde{\theta}_0 \equiv C$ as a measure of the *organization's capacity* to collect good information about θ. For example, suppose a secret police organization only has access to a sparse network of unreliable informants. Receiving information from an unreliable source that a group is plotting against the regime will not change the agent's beliefs very much; formally, this implies that $\tilde{\theta}_1$ is not very different from $\tilde{\theta}_0$. In contrast, if the organization has cultivated a highly professionalized network of deeply embedded informants, then the secret police may be able to collect very good information about a potentially subversive group, so that $\tilde{\theta}_1$ is much higher than $\tilde{\theta}_0$ and hence C is high. The agent's ability to provide good information to the principal is therefore directly tied to the organization's capacity to generate this information.[5]

Building on the examples from the introduction, a simple way to interpret this model is that the agent is the head of a secret police organization, which collects information on behalf of a political leader (the principal). The secret police organization is tasked with gathering information on a potentially subversive individual or group. This could include the gathering of kompromat on actors outside the model, but when we explicitly introduce kompromat in the following sections, we will exclusively focus on kompromat that the principal has on the agent.

5. In the benchmark model it would also be natural to think of C as being driven by the ability of the agent, but this complicates the analysis in subsequent sections.

To avoid any dual interpretations of kompromat, suppose a signal of $s = 1$ corresponds to learning the group is more likely to be dangerous to the principal, and $s = 0$ means less likely to be dangerous. The principal wants to be more punitive towards dangerous groups. The bias represents whether the head of the secret police would prefer to be more or less punitive toward individuals or groups than the principal.

Our solution concept is Perfect Bayesian Equilibrium (hereafter, "equilibrium"). As in any cheap talk game, there is always a "babbling" equilibrium wherein the agent gives the same advice regardless of s, and the principal makes the same policy choice for all m. In the main text, we focus on the possibility of a *truthful* equilibrium where $m = s$. See Section A of the Supporting Information (p. SI-1) for a discussion of other possible (partially) informative equilibria.

In a truthful equilibrium, $m = s$, and given the quadratic loss utility the principal's best response is to make a policy decision $x^*(m) = \tilde{\theta}_m$. A truthful equilibrium exists if A has no incentive to lie about s, regardless of whether he learns $s = 0$ or $s = 1$. To assess the agent's incentives, we first use a standard "mean-variance" decomposition of quadratic loss utility:

$$\begin{aligned}\mathbb{E}_\Theta[u_A] &= \int_\Theta -(x - \theta - b)^2 dF(\theta|s) \\ &= -(x - (\tilde{\theta}_s + b))^2 - V. \end{aligned} \tag{1}$$

where $V \equiv \int_\Theta (\theta - \tilde{\theta}_s)^2 dF(\theta|s)$.[6] The agent's expected utility reveals that there are two components to his "loss." The first term in (1) is the squared distance between the policy choice and his best guess about his ideal policy. The second is the residual variance in the belief about θ, V. The V term does not depend on the policy choice and will drop out of future calculations. As a result, the agent will give whatever advice induces the principal to make a policy decision that is closer to $\tilde{\theta}_s + b$.

6. An intermediate calculation here is that since $\tilde{\theta}_s = \mathbb{E}_\theta[\theta|s]$, $-(x - b - \theta)^2 = -((x - b - \tilde{\theta}_s) - (\theta - \tilde{\theta}_s))^2 = -(x - b - \tilde{\theta}_s)^2 - (\theta - \tilde{\theta}_s)^2 + 2(x - b - \tilde{\theta}_s)(\theta - \tilde{\theta}_s)$, and when integrating over θ the third term becomes $2(x - b - \tilde{\theta}_s) \int_\Theta (\theta - \tilde{\theta}_s) dF(\theta|s) = 0$.

Let $m_s \equiv m(s)$ denote the message the agent sends after receiving a signal of s, representing the advice he offers the principal after observing s. The agent who observes $s = 1$ prefers that the principal make a policy decision $\tilde{\theta}_1 + b$. If the agent provides truthful advice in this situation, $m_1 = 1$, then the principal's policy decision will be $\tilde{\theta}_1$. If the agent deviates and offers untruthful advice, $m_1 = 0$, then the policy choice is $\tilde{\theta}_0$. Since $\tilde{\theta}_0 < \tilde{\theta}_1 < \tilde{\theta}_1 + b$, when the agent observes $s = 1$, he always has an incentive to offer truthful advice ($m_1 = 1$).[7] Intuitively, a secret police agent who prefers to be more punitive toward a potentially subversive individual or group will never have an incentive to lie when he learns that they are plotting an attack.

However, if the agent receives a signal of $s = 0$ (e.g., the group is not plotting an attack), he may now have an incentive to lie and report $m_0 = 1$ (e.g., the group *is* plotting an attack) to induce a more punitive action. After observing that signal, the agent's ideal policy is $\tilde{\theta}_0 + b$, and so there must be sufficient incentive for the agent to give truthful advice. We refer to his incentive compatibility constraint as the "truth-telling constraint," which reduces to:

$$\begin{aligned} -(\tilde{\theta}_0 - (\tilde{\theta}_0 + b))^2 - V &\geq -(\tilde{\theta}_1 - (\tilde{\theta}_0 + b))^2 - V \\ b &\leq \frac{\tilde{\theta}_1 - \tilde{\theta}_0}{2} = \frac{C}{2}. \end{aligned} \tag{TC}$$

The right hand side of the second line is the maximum level of bias where it is possible for the agent to provide truthful advice in this baseline cheap talk setting. The agent prefers to tell the truth after observing $s = 0$ if his ideal policy decision ($\tilde{\theta}_0 + b$) lies closer to $\tilde{\theta}_0$ than to $\tilde{\theta}_1$. This will be true if and only if b is less than half the distance between $\tilde{\theta}_0$ and $\tilde{\theta}_1$. Substantively, the agent can "get away" with being more biased if the organization has a higher capacity (C) to detect subversive individuals or groups. In the following sections, the inequality in (TC) will serve as an important benchmark to assess how kompromat affects the agent's incentive to give truthful advice.

7. Formally, he prefers to send $m_1 = 1$ to $m_1 = 0$ if $-(\tilde{\theta}_1 - (\tilde{\theta}_1 + b))^2 \geq -(\tilde{\theta}_0 - (\tilde{\theta}_1 + b))^2$, which always holds since $\tilde{\theta}_1 - \tilde{\theta}_0 > 0$.

3 Kompromat with "accidental" leaks

We now introduce kompromat into the baseline model by assuming the principal has some kompromat on the agent. The amount of kompromat is fixed and exogenous in this section.

Formally, let $\kappa \geq 0$ represent the cost to the agent of having his kompromat leaked. So κ represents the amount or magnitude of potential compromising information that the principal has. Leaking may be costly because it affects his reputation, financial security, or physical safety.

Modeling kompromat in a satisfying way requires several additions to the sequence of the baseline model. First, since working for the principal may now be costly for the agent (i.e., if kompromat is publicly released), we explicitly model the players' incentives to commence a relationship with each other. We model this sequentially, where the principal decides whether to appoint the agent to the organization $a_P \in \{0, 1\}$ and the agent decides whether to accept the appointment $a_A \in \{0, 1\}$.[8] The easiest interpretation of this stage is that, at the outset of the game, the principal needs to hire a new agent to fill a particular job. Alternatively, we could imagine the status quo is that the agent is already working his current job and so $a_P = 1$ corresponds to keeping the agent (with $a_P = 0$ meaning the agent gets fired) and $a_A = 1$ means to continue working with the principal (with $a_A = 0$ meaning quitting).

Second, the principal has to decide whether to leak the kompromat. She may condition this decision on whether she suspects that the agent lied to her. To make things simple, we assume that after making the policy choice, the principal directly learns the signal observed by the agent. One way to interpret this is that the agent's information is "verifiable at a later date." If the principal observes the outcome of her decision is inconsistent with the advice she received from the agent, she will know that the agent did not tell the truth.

Finally, we allow for some possibility that the kompromat on the agent leaks even if the principal

8. The sequential structure allows us to rule out a pathological Nash equilibrium in which both players choose $a_J = 0$ solely because they expect the other will. Moreover, the order of moves does not affect the equilibrium choices.

decides not to leak it herself. There is ample evidence of this risk, including the Kuchma and Fujimori examples we discuss above. These leaks are "accidental" from the perspective of the principal, but may be initiated by other actors who have access to the information. Put another way, we are assuming that by joining the organization, some individuals beyond the principal will gain access to the kompromat on the agent, and these actors may choose to leak it for some reason outside of the agency relationship we study. Formally, we assume that if the principal does not leak the kompromat herself, there is still a probability $\nu \in (0, 1)$ that it leaks anyway.

In Section C of the Supporting Information (p. SI-4) we relax the assumption that the message is perfectly verified with an extension where the principal gets a noisy signal of whether the agent's advice was truthful. The core conclusions of this extended model are nearly identical to those we present here; leaking after a "misread" signal ends up having the same effect in equilibrium as what we call accidental leaks.

The sequence of moves is as follows.

1. P chooses whether to offer an appointment ($a_P = 1$) or not ($a_P = 0$), and if $a_P = 1$ the agent can accept ($a_A = 1$) or not ($a_A = 0$). If $a_P = 0$ or $a_A = 0$, the game ends with reservation utilities $(\overline{u}_A, \overline{u}_P)$. If $a_A = a_P = 1$, then:
2. A privately observes signal $s \in \{0, 1\}$, and gives advice in the form of a message $m \in \{0, 1\}$.
3. P observes m and chooses a policy $x \in \mathbb{R}$.
4. P observes a validation signal $s_v = s$, and chooses whether to leak $l \in \{0, 1\}$.
5. If kompromat is not leaked by P, it leaks exogenously with probability ν.

The utility functions in this section are:

$$u_A = a(w - (x - (\theta + b))^2 - (l + (1 - l)\nu)\kappa + (1 - a)\overline{u}_A$$
$$u_P = a(-w - (x - \theta)^2) + (1 - a)\overline{u}_P,$$

where $a = a_P a_A$.[9] If the agent is appointed to the organization ($a = 1$), then principal pays the agent a wage w, The agent's policy utility is the same as in the baseline cheap talk model. Finally, the agent pays a cost κ when kompromat is leaked. We place more structure on the reservation utilities $\overline{u}_P$ and $\overline{u}_A$ below, when we examine the principal's incentive to appoint the agent and the agent's incentive to accept the appointment.

Some comments on how we model kompromat To create a simple and clear model where kompromat may or may not be able to induce honest communication, we abstract away from several issues. Before proceeding we highlight three that are particularly relevant.

First, we do not model where the kompromat comes from. Doing so captures in a reduced-form manner the fact that different potential agents carry with them different histories of illegal or embarrassing behavior. However, it is often the case that kompromat does not yet exist at the time of hiring, but then the principal either tacitly allows or explicitly encourages corrupt behavior which can later be held over the agent (see Darden 2001, for examples of Leonid Kuchma engaging in this behavior). Alternatively, kompromat can be "manufactured", e.g., by planting evidence (see Ang and Nalepa 2018).

Second, we do not model why the public release of kompromat is costly for the agent. In some cases, leaked kompromat generates a straight-forward legal cost, as one may be prosecuted for the behavior unearthed or revealed. The question of why and when certain material is embarrassing is murkier (see Gambetta 2009; Gorlizki 2013). For example, being bisexual or Jewish may have been damaging in early 20th century Russia, but would be less so in other contexts.

Third, we model kompromat as a "one-way street": the principal has it on the agent, but not vice versa. In some contexts the agent could have kompromat on the principal as well. In Section D of the Supporting Information (p. SI-11), we analyze an alternative model with "bilateral" kompromat, with similar substantive conclusions.

9. Technically if $a_P = 0$ the agent does not make a choice, but a is set to zero here regardless.

Fourth, we assume the principal is indifferent between leaking and not. Given this indifference, there are other equilibria with different leaking strategies, an issue common in models of blackmail (Schwarz and Sonin 2007). However, just as it is common to focus on the most informative equilibrium in cheap talk games (as we do), we also restrict attention to the leaking strategy which makes truthful communication as easy as possible.

Truth-telling constraint Starting at the end, the principal is indifferent between leaking kompromat and not. As a result, any leaking strategy can be part of an equilibrium. The simplest way to give the agent an incentive to tell the truth is to leak kompromat whenever he lies, i.e., when $s_v \neq m$.

It is easy to check that an agent observing $s = 1$ has no incentive to lie, as in the baseline model. Given the principal expects truthfulness and uses this leaking strategy, an agent who gets a signal of $s = 0$ and tells the truth gets an expected policy payoff which is again includes the loss associated with the distance from her ideal $-(\tilde{\theta}_0 - (\tilde{\theta}_0 + b))^2$, the residual variance $-V$ (as derived in equation 1), and now an additional expected cost of kompromat (from an accidental leak) $\nu\kappa$. An agent who lies gets an expected policy payoff of $-(\tilde{\theta}_1 - (\tilde{\theta}_0 + b))^2 - V$, and has kompromat leaked with certainty, generating cost κ.

Combining, the truth-telling constraint when $s = 0$ is now:

$$-(\tilde{\theta}_0 - (\tilde{\theta}_0 + b))^2 - V - \nu\kappa \geq -(\tilde{\theta}_1 - (\tilde{\theta}_0 + b))^2 - V - \kappa,$$

which reduces to

$$b \leq \frac{C}{2} + \left(\frac{1-\nu}{2C}\right)\kappa. \tag{TC1}$$

The first term in the right-hand side of (TC1) is the maximum level of bias with no kompromat, which we derived in the core agency model. The second term is positive, so the truth-telling con-

straint is easier to satisfy than the truth-telling constraint without kompromat.

Participation constraint Now we analyze when and whether the principal and the agent will find it worthwhile to work with one another.

Both players' payoffs from the agent's appointment depend on the wage that the principal pays the agent, $w \geq 0$. Rather than explicitly model a wage bargaining process, we ask if there exists a wage where both prefer the agent's appointment to their outside options. It is sequentially rational for both to accept employment if and only if this is true.

To keep the analysis relatively tidy, we make the following assumptions about what happens if $a = 0$. First, the agent finds employment elsewhere with an expected payoff of $\overline{w}_A \geq 0$.[10] Importantly for contexts where kompromat is pervasive, it is possible that the value of the outside option includes the possibility that the agent's next best employment option will involve a risk of leaked kompromat as well. Second, the principal appoints a different individual who will provide truthful advice at wage $\overline{w}_P \geq 0$.[11] Both of these outside wages may reflect the delay before finding matching with another job/employee. For example, because the principal can hire a different individual who will be truthful, her outside wage may reflect the increased cost associated with hiring a "better" agent.

We let y_J^T be the expected policy payoff for player J when the principal receives truthful advice and makes a policy decision accordingly.[12] Then, given the assumptions above, the players' reservation utilities (i.e., when $a = 0$) are

$$\overline{u}_A = \overline{w}_A - y_A^T \qquad\qquad \overline{u}_P = -\overline{w}_P - y_P^T$$

10. There may or may not be a risk of leaked kompromat in outside employment, but we assume this is already priced into the expected payoff $\overline{w}_A$.

11. Since we assume that the outside option for the principal involves hiring someone else who will give truthful advice, this must be better than appointing the agent if the agent lies. So, if there is no truthful equilibrium in the advice and policy decision stage of the game, the principal will never appoint the agent to head the organization.

12. While it does not matter for our calculations, for the sake of completeness this quantity is given by the average posterior variance in the agent's belief about θ, or $\pi\mathbb{E}_{\theta|s=1}[(\tilde{\theta}_1 - \theta)^2] + (1-\pi)\mathbb{E}_{\theta|s=0}[(\tilde{\theta}_0 - \theta)^2]$.

If the agent accepts the principal's appointment and there is an equilibrium where he provides truthful advice, he gets a wage w and expected policy payoff y_A^T. He also knows that kompromat may leak accidentally, generating expected cost $\nu\kappa$.

Combining, the agent's "participation constraint" to accept an appointment is

$$\begin{aligned} w - y_A^T - \nu\kappa &\geq \overline{w}_A - y_A^T \\ w &\geq \overline{w}_A + \nu\kappa \end{aligned} \tag{2}$$

For the principal, if the above condition is not met then the agent will not accept an appointment so either choice is sequentially rational. If the agent will accept an appointment, the utility for choosing $a_P = 1$ to an agent who will report truthfully given the rest of the equilibrium is $-w - y_P^T$. The principal's participation constraint is that this is higher than the outside option:

$$\begin{aligned} -w - y_P^T &\geq -\overline{w}_P - y_P^T \\ w &\leq \overline{w}_P \end{aligned} \tag{3}$$

Combining (2) and (3), there exists a mutually agreeable wage (or set of wages) if and only if:

$$\nu\kappa \leq \overline{w}_P - \overline{w}_A \equiv D \tag{PC1}$$

It is natural to assume that $\overline{w}_P > \overline{w}_A$ (or $D > 0$); if not, then both players would be better off pursuing their outside option even if they could have a truthful equilibrium with no kompromat. As the outside options for both players improve (higher $\overline{w}_A$, meaning the agent can find good employment elsewhere; lower $\overline{w}_P$ meaning the principal can find another person who will head the organization for less money), this window shrinks.

Equilibrium If both the truth-telling and the participation constraints are satisfied, then there exists an equilibrium where the agent is appointed and provides truthful advice. We can combine the two constraints as follows:

> **Proposition 1.** If there is no truthful equilibrium without kompromat ($b > C/2$), then there is a truthful equilibrium with kompromat if and only if
>
> $$\frac{2(b - C/2)C}{1 - \nu} \leq \kappa \leq \frac{D}{\nu} \tag{4}$$
>
> *Proof.* Follows immediately from rearranging (TC1) and (PC1). □

This inequality presents a simple condition for kompromat to be effective. It must be costly enough to induce agents to tell the truth (the truth-telling constraint), but not so costly that the risk of accidental leaks makes employment infeasible.

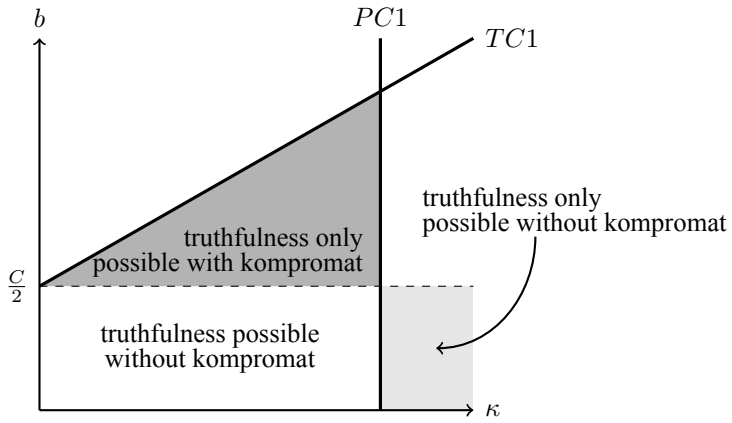

Figure 1: *In the model with accidental leaking, kompromat can increase the feasibility of truthful communication, but only if it is not so costly that it prevents the principal from appointing the agent or the agent from accepting the appointment*

In Figure 1, we plot the truth-telling and participation constraints for a specific constellation of parameter values. The dark gray region illustrates how kompromat can mitigate the agency problem between the principal and the agent. When there is the possibility of some leaked kompromat, this makes it possible for more biased agents to truthfully communicate with the principal, when they would not do so in the absence of kompromat. However, there can be *too much* kompromat. In the

region to the right of the participation constraint (and below the truth-telling constraint), the agent would be able to communicate honestly if employed by the principal. However, the deadweight loss from kompromat means that there is no mutually acceptable wage for both agents, so they do not engage in a relationship in the first place.

This latter observation highlights that there is an important caveat to the claim that "The harsher the punishment and the greater the likelihood of receiving it that can be triggered by the revelation of any one piece of compromising information, the greater the binding effect of *that* information" (Gambetta 2009, p. 69). Indeed, the idea that the "binding effect" of kompromat monotonically increases in its severity only applies in our model in the extreme situation in which the principal is perfectly able to detect when the agent lies and there is zero risk of accidental leaking (i.e., $\nu = 0$). In this special case, there is always an incentive to appoint the agent as long as $\overline{w}_P > \overline{w}_A$, and the potential inefficiency created by kompromat poses no problem. Once we move from an environment with involuntary participation (as in Gambetta 2009) to one with voluntary participation and accidental leaks, then very incriminating kompromat becomes counterproductive as it shuts down cooperation between the principal and agent. This is precisely what is depicted by the light gray area in Figure 1.

In Section B of the Supporting Information (p. SI-2), we discuss in more detail how the parameters of the model affect the relationship between the principal and the agent. Specifically, we shed light on the situations in which kompromat will be the most effective for mitigating the core agency problem. Two observations are worth noting here. First, the model suggests that kompromat will be especially effective for the lowest capacity organizations since these will be the ones where the agent is most tempted to lie. Second, it is generally in the interest of both actors to have as little kompromat as possible to induce truthful communication. For sufficiently biased agents, this means *some* amount of kompromat (enough to bind the truth-telling constraint). So if the amount of kompromat on the agent is an endogenous choice, he will want to produce enough to enable truth-telling, but no more than that.

4 Kompromat and reputation

In the model we analyze in the previous section, the agent faces a direct cost whenever damaging kompromat is leaked. However, we did not explicitly address why leaking kompromat is costly. In some cases, leaked kompromat can lead to negative legal ramifications and prosecution (for several historical and more recent examples, see Ledeneva 2006; Gregory 2009). However, the the damage from leaked kompromat can also come from the fact that it reveals embarrassing or unethical behavior. And if it is common knowledge that everyone who works for an organization is the subject of kompromat, an observer could infer that the agent has done embarrassing and unethical things from their employment itself.

This creates two potential problems for the previous analysis. First, to the extent that being appointed to the organization already makes the agent look shady, the incremental damage of having kompromat leaked is lessened. For example, suppose that everyone knows that a political leader keeps her secret police officials in line by threatening to publicly release dirt on them. Then, upon appointing a new head of the secret police, it becomes apparent to everyone that the agent must be extremely compromised. Taking this logic to the extreme, by simply appointing the agent, the damage is already done; leaking the agent's kompromat no longer poses a threat to him. If that is the case, then kompromat cannot be used to induce the agent to provide truthful advice. Second, if being appointed by the principal makes one look unethical (more so than other employment opportunities), this will lower the value of joining the organization in the first place. So, kompromat can be harmful even if it is never leaked.

In this section we present a model where these dynamics arise in equilibrium. However, our analysis shows that the damage to the agent's reputation does not completely preclude the possibility that a principal can use kompromat to keep the agent honest. Interestingly, our analysis suggests that kompromat is an effective tool only when the organization hires sufficient numbers of agents with no kompromat.

To formalize how reputational concerns affect the agent, we introduce a new player, who we call the "outsider." The outsider could be the voting public, other potential employers, or even friends and family. The outsider will form a belief about whether the agent is compromised based on (1) whether the agent decides to work for the organization and (2) any kompromat that is released. Let $\kappa = \kappa_H$ mean a compromised agent while $\kappa = \kappa_L$ means not compromised. We will loosely refer to a compromised agent as "corrupt," although other interpretations of being compromised are possible. The agent cares about the reputation that he has with the outsider, and so his utility declines as the outsider believes the agent is more likely to be corrupt. The sequence of the revised game is as follows:

1. Nature draws $b \geq 0$ and $\kappa \in \{\kappa_L, \kappa_H\}$.

2. P and A observe b and κ, and sequentially choose $a_P \in \{0, 1\}$ and $a_A \in \{0, 1\}$. If either chooses $a_J = 0$, the game ends with reservation utilities, and otherwise the game continues.

3. A privately observes signal $s \in \{0, 1\}$, and gives advice in the form of a message $m \in \{0, 1\}$.

4. P observes m and chooses a policy $x \in \mathbb{R}$.

5. P observes a validation signal $s_v = s$, and, if $\kappa = \kappa_H$, chooses to leak kompromat ($l = 1$) or not ($l = 0$).

6. An outside observer observes a, and if $l = 1$ observes κ.

There are three changes to the way we treat the release of kompromat. First, we assume that a leak can only happen when the agent is corrupt. We make this assumption to simplify the analysis since uncompromised agents are harder to threaten with kompromat. Second, we allow the principal to use a mixed strategy at the leaking stage, which will sometimes be important for the formal analysis. Finally, to focus more clearly on reputational costs, we remove the possibility of

accidental leaks.[13]

The bias and corruption level of the agent are still common knowledge among him and the principal. (In Section G the Supplemental Information (p. SI-19), we show that allowing the principal to be uncertain about the agent's bias at hiring shrinks the range of acceptable wages but does not alter the qualitative lessons of our model.)

However, in order to model the outsider's beliefs about the agents who go to work for the principal, we also place distributional assumptions on these parameters. The agent's bias b is drawn from a cumulative distribution function F. The main restriction we place on this distribution is that there are some agents who are biased enough to report honestly and some who are not. Formally, $F(C/2) \in (0, 1)$, where C is defined as above. To avoid dealing with extraneous cases we also assume that F is continuous and has no upper bound (i.e., $F(b) < 1$ for all $b \geq 0$).

Let q be the prior probability that an agent is a corrupt type, which is independent of the agent's bias b. We denote the outsider's posterior belief about whether the agent is corrupt by $\tilde{q}$. Given the agent cares about the outsider's assessment of his corruption, the players' utilities are now:

$$u_A = a(w - (x - \theta - b)^2 - r\tilde{q}) + (1 - a)\overline{u}_A \qquad u_P = a(-w - (x - \theta)^2) + (1 - a)\overline{u}_P$$

For the reservation utilities $\overline{u}_A$ and $\overline{u}_P$, we again assume that the principal will be able to find another agent to communicate truthfully at wage $\overline{w}_P$ (and that this is preferable to not hiring an agent who is truthful), and the agent can find employment elsewhere at wage $\overline{w}_A$. If the agent does not enter the organization, then let the outside observer's belief about his type be $\overline{q} \leq q$.[14] As in the previous model, we are not assuming that other organizations do not use kompromat; in fact,

13. In Section F the Supplemental Information (p. SI-17), we show that the reintroduction of accidental leaks would not alter our qualitative conclusions in the reputation model. Even though accidental leaks make the principal's own leaking strategy marginally less effective for inducing truth-telling, it does not alter the players' main strategic trade-offs nor does it substantially alter our formal analysis.

14. One natural assumption is that $\overline{q} = q$, so that the agent's decision to pursue outside options does not affect the outside observer's belief about how corrupt he is. Alternatively, it is also reasonable to assume that $\overline{q} < q$, so that the outside observer infers an agent is less likely to be corrupt if he does not join the organization. As the analysis is straightforward for any case here, we do not specify where $\overline{q}$ lies relative to q.

one can interpret $\overline{q}$ as the reputation of the "next best" job the agent could get.

Summarizing, the player's reservation utilities are

$$\overline{u}_A = \overline{w}_A - y_A^T - r\overline{q} \qquad\qquad \overline{u}_P = -\overline{w}_P - y_P^T$$

We continue to search for equilibria where those who are hired report truthfully. To focus our discussion, we will search primarily for the equilibrium that maximizes the probability that the agent is appointed to the organization. Given this goal, an important initial observation is that the principal will never appoint an agent who does not give truthful advice. (Again, this follows from the assumption that the outside option entails hiring an agent who *does* provide truthful advice.) So, finding the equilibrium that maximizes the probability that the agent is appointed boils down to checking when the agent has an incentive to provide truthful advice, and then whether the agent will join the organization.

Further, we restrict our analysis in the main text to equilibria where the agent's decision to join the organization is monotone in b. Formally, an agent of type κ_J joins the organization if and only if $b \leq \hat{b}_J$.[15] That is, if the agent is corrupt, there is a threshold of bias below which he joins the organization, and if he is non-corrupt, there is a (potentially different) threshold below which the he joins the organization. As shown in the proof of Proposition 2, there is always an equilibrium of this form which maximizes the probability that the agent is appointed.

Truth-telling constraint A non-corrupt agent cannot be threatened by the release of kompromat since the principal has no kompromat on non-corrupt types. So, the non-corrupt type will report truthfully if and only if $b \leq C/2$, as in the baseline cheap talk model without kompromat. By the assumption that $F(C/2) \in (0, 1)$, it is possible for a non-corrupt agent to be too biased to report truthfully (and, given the reservation utility assumptions, will never be appointed).

15. While the truth-telling constraint is always easier to meet for lower b, since the cheap talk game has multiple equilibria it is possible that, for example, the actors select a babbling equilibrium for some b_1 (even though truth-telling is possible), but would select a truthful equilibrium for some $b_2 > b_1$.

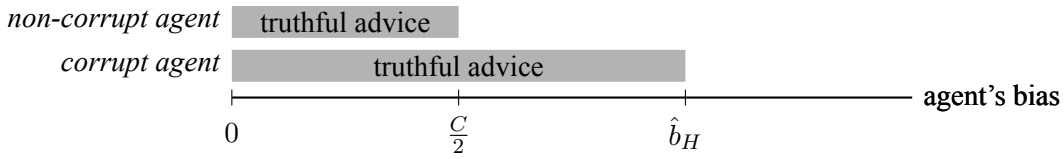

Figure 2: *Truth-telling in the candidate monotone equilibrium*

A corrupt agent with a bias below the threshold $C/2$ will also provide truthful advice even without the threat of kompromat. The important case to consider is a corrupt agent with a bias $b > C/2$ who is appointed, and who the principal can induce to report truthfully by threatening him with kompromat. Like the previous model, it is easier to induce a less biased agent to tell the truth. As a result, the monotone equilibrium which maximizes employment is one where a non-corrupt agent is hired if and only if $b \leq C/2$, and a corrupt agent is hired if and only if $b \leq \hat{b}_H$ for some $\hat{b}_H > C/2$. In an equilibrium like this, there is a $F(C/2)$ probability that a non-corrupt agent joins the organization, and a $F(\hat{b}_H)$ probability that a corrupt agent does the same. Figure 2 illustrates truth-telling in this candidate monotone equilibrium, which we characterize in the remainder of this section.

In an equilibrium of this form, the posterior belief that the appointed agent is corrupt (before any kompromat is leaked) is computed using Bayes' rule:

$$\tilde{q}_a \equiv \Pr(\kappa_H | a = 1) = \frac{qF(\hat{b}_H)}{qF(\hat{b}_H) + (1-q)F(C/2)}$$

Since $F(\hat{b}_H) > F(C/2)$, it follows that $q < \tilde{q}_a < 1$. Getting appointed to the organization increases the outside observer's belief that the agent is corrupt but not to certainty, since a non-corrupt agent also sometimes gets appointed. Importantly, this posterior belief is increasing in $\hat{b}_H$ since an increase in $\hat{b}_H$ makes it easier to appoint a corrupt agent relative to a non-corrupt agent.

Recall that in this section we allow the principal to use a mixed strategy when deciding whether to (purposefully) leak kompromat. Let the principal's mixed strategy probability of leaking be λ.

The truth-telling constraint becomes:

$$-(\tilde{\theta}_0 - (\tilde{\theta}_0 + b))^2 - r\tilde{q}_a - V \geq -(\tilde{\theta}_1 - (\tilde{\theta}_0 + b))^2 - r(\lambda + (1-\lambda)\tilde{q}_a) - V$$

$$b \leq \frac{C}{2} + \frac{\lambda r(1-\tilde{q}_a)}{2C} \tag{TC2}$$

Since the agent is hired if and only if he will provide truthful advice, then in equilibrium, the threshold $\hat{b}_H$ must be characterized by meeting (TC2) with equality. Since $\tilde{q}_a$ is increasing in $\hat{b}_H$, the right-hand side of (TC2) is decreasing in $\hat{b}_H$ while the left-hand side is increasing in $\hat{b}_H$. So, for a fixed λ, there is a unique threshold b^* such that (TC2) binds. Given this threshold strategy and the principal's leaking strategy, the agent has a strict incentive to provide truthful advice except when $b = b^*$. (But since b is drawn from a continuous distribution, the agent will have bias $b = b^*$ with probability zero.)

The equilibrium bias threshold b^* is increasing in λ. That is, a higher probability of leaking will enable the corrupt agent to join the organization even when he is more biased. Let $b^*(\lambda)$ be the value of $\hat{b}_H$ which solves (TC2) with equality for leaking probability λ, and $\tilde{q}_a(\lambda)$ be the corresponding belief about the corruption level of the appointed agent.

Participation constraints Unlike the model in the previous section, if the agent can be induced to give truthful advice to the principal, then he faces no risk of kompromat leaking. However, regardless of whether the agent is actually corrupt, there is a still a reputational cost of joining the organization when $\tilde{q}_a > \overline{q}$. Given the analysis above, the participation constraint for an agent who can report truthfully becomes:

$$w - y_A^T - r\tilde{q}_a(\lambda) \geq \overline{w}_A - y_A^T - r\overline{q}$$

$$w \geq \overline{w}_A + r(\tilde{q}_a(\lambda) - \overline{q}) \tag{5}$$

That is, the agent needs to be compensated above his outside option to accept the reputational cost of joining the organization. Of course, this cost will be higher as it becomes easier to hire an agent with a high bias since $\tilde{q}_a$ increases in $\hat{b}_H$. Moreover, since $\hat{b}_H$ increases in λ, then a principal that "uses kompromat more" (higher λ) will (correctly) earn a reptuation for cultivating more corrupt agents.

The principal's participation constraint remains $w \leq \overline{w}_P$. So, for there to be a wage that satisfies both participation constraints, the following condition must be satisfied:

$$r(\tilde{q}_a(\lambda) - \overline{q}) \leq D \tag{PC2}$$

where D is again equal to $\overline{w}_P - \overline{w}_A$.

Equilibrium If (PC2) is satisfied whenever the principal always releases kompromat after the agent lies (i.e., at $\lambda = 1$), then there is an equilibrium where the corrupt agent can be appointed as long as he can be successfully threatened with kompromat. In this equilibrium, a non-corrupt agent enters the organization when he is relatively unbiased. A corrupt agent enters the organization even if he is too biased to report truthfully without the threat of kompromat, since the principal now threatens to reveal that he is the corrupt type. The outside observer tends to think that the appointed agent is more likely to be corrupt, but they are not completely certain. Taken together, this ensures that the principal's threat to expose a corrupt agent is meaningful. The fact that being appointed to the organization hurts the agent's reputation does in fact lower the surplus associated with appointing the agent, but since D is sufficiently large, it is not so much as to preclude a range of mutually acceptable wages.

If (PC2) is not satisfied at $\lambda = 1$, then there is no equilibrium where all corrupt agents who can be threatened with kompromat become employed, because if they did no one would be willing to work for the organization in the first place (at a wage the principal is willing to offer). However, recall that as λ decreases, $b^*(\lambda)$ decreases, as does $\tilde{q}_a(\lambda)$. So, as long as $r(\tilde{q}_a(0) - \overline{q}) = r(q - \overline{q}) \leq D$,

there will be a critical value $\lambda^{\max}$ such that (PC2) is met with equality.[16]

Proposition 2. (i) If $D < r(q - \overline{q})$, there is no equilibrium where all those with $b \leq C/2$ are hired, and if $D = r(q - \overline{q})$ an equilibrium which maximizes employment involves $a^*(\cdot) = 1$ if and only if $b \leq C/2$ (independent of κ).

(ii) If $r(q - \overline{q}) < D < r(\tilde{q}_a(1) - \overline{q})$, then there is a monotone equilibrium which maximizes the probability of employment, with $a^*(\cdot) = 1$ for all agents with $b \leq C/2$ and some $\kappa = \kappa_H$ agents with $b > C/2$, and the principal leaks kompromat against agents who lie with probability $\lambda \in (0, 1)$.

(iii) If $D \geq r(\tilde{q}_a(1) - \overline{q})$, then there is a monotone equilibrium which maximizes the probability of employment, with $a^*(\cdot) = 1$ for all agents with $b \leq C/2$ and some $\kappa = \kappa_H$ agents with $b > C/2$, and involves leaking kompromat against agents who lie with probability $\lambda = 1$.

Proof. See Section E of the Supporting Information (p. SI-16). □

So, as long as there is enough surplus from appointing the agent, it is possible to hire some highly biased agents who will report truthfully when the principal can use kompromat. Further, if $q = \overline{q}$—i.e., the outsider makes no inference about the corruption of those who do not work for the organization—then any surplus ($D > 0$) will ensure that some additional, highly biased agents can be hired.

5 Conclusion

In this paper, we present a theory of kompromat that demonstrates how it can be used to ameliorate an agency problem caused by a principal's reliance on a biased agent for expert advice. We analyze two models that incorporate kompromat into a benchmark cheap talk model. The first model demonstrates how the possibility of leaked kompromat can enable a principal—such as an

16. In this case, there is a single wage that makes the principal and agent exactly indifferent about the agent's appointment. However, for any $\lambda < \lambda^{\max}$, there exists an equilibrium where the wage can be set to make both strictly prefer the agent's appointment. So, if we impose the requirement that the players *strictly* prefer the agent's appointment in an equilibrium, then if $D > r(q - \overline{q})$ there are monotone equilibria where kompromat expands the probability that the agent is appointed whenever the principal uses a leaking probability $\lambda \in (0, \lambda^{\max})$.

authoritarian leader—to get better advice from subordinates. Paradoxically, this benefits *both* the leader and the subordinate. However, there are limits: using "too much" kompromat can make it harder to hire subordinates in the first place. In the second model, we microfound the cost of "too much" kompromat to show how its widespread use in an organization can endogenously damage the reputation of agents working for the organization even when no kompromat is ever revealed.

Taken together, the models reveal a core trade-off: kompromat can make organizations run more effectively once employees are hired, but may make working for the organization less appealing in the first place. However, since our models build on a cheap talk framework, a potential limitation of the analysis is that it only considers kompromat that the principal might have on the agent. In many societies where kompromat is used, it is widespread: agents may have kompromat on principals, as well as on fellow agents. In evaluating the potential applicability of our theory, an obvious question is whether our core insights apply in such "bilateral" settings. In fact, they do.

To see why, consider a simple symmetric effort choice model with two co-equal agents. Each agent's effort is individually costly but provides a positive externality for the other agent. As a result, the collectively optimal effort level is higher than the effort chosen in a Nash equilibrium. Now introduce the possibility that both agents can release kompromat on the other if they do not put in sufficient effort. We construct an equilibrium in which the agents' leaking strategies allows them to attain collectively higher effort levels, which makes both agents better off (see Section D of the Supplemental Information, p. SI-11). However, if there are also accidental leaks, using kompromat again creates a trade-off: organizations that use more kompromat can get agents to exert effort at levels closer to what would be optimal, but also increase the expected loss due to accidental leaks. As in our main models, kompromat makes the players better off when leaks are rare. So, even though the underlying strategic interaction is quite different, the core trade-off in a bilateral setting would be similar to the models we examine above.

This trade-off has an important set of substantive and empirical implications. First, since kompromat is most effective when leaking is rare, this suggests that an empirical focus on *leaked* kom-

promat may paint a unrepresentative picture of its day-to-day use. Second, since kompromat is mutually beneficial, it suggests a mechanism by which an authoritarian leader can maintain support from a wide range of subordinates, even those who are not "true believers." Finally, since the benefits of kompromat accrue only if it exists but is hidden, this raises the stakes of lustration (and potentially the durability of an authoritarian regime). Not only does lustration cause costly kompromat to be released, it also removes an established tool for managing bureaucratic relationships.

So, why use kompromat and not other kinds of incentives like bonuses and sanctions, or attempting to instill a sense of "mission" among agents (Patty and Penn 2019)? While answering this question in a satisfying fashion would require explicitly modeling other possibilities, we provide two potential answers. First, kompromat could be a high-powered incentive that isn't too costly for principals since it is cheaper to threaten someone than to pay them a bonus. This is particularly so in countries with secret police and other institutions already gathering information on elites (and others). Second, kompromat seems to be more widespread in places with weaker private sectors, allowing the public sector to get away with having a reputation for utilizing compromising materials. Both of these ideas suggest it may be valuable to extend the modeling here to look at the more systemic use of kompromat, rather than how it can be used by a single organization.

A final promising topic for extensions is the origins of kompromat. From the principal's perspective, there is a relatively straightforward trade-off where (costly) collection of kompromat makes it easier to control agents. (Of course, the principal may require agents for the task of collecting kompromat itself, creating another agency problem which could be improved with kompromat on the kompromat collectors!) We have emphasized that for kompromat to work with voluntary employment, the agent must prefer working with a principal who has kompromat to their outside option. However, all else equal, agents would certainly prefer to be employed without this threat. A more comprehensive model could explore how taking actions that create compromising material confers some benefit to the agent by making them more employable for certain kinds of jobs.

References

Ang, Milena, and Monika Nalepa. 2018. "Can Transitional Justice Improve the Quality of Representation in New Democracies?" Forthcoming, World Politics.

Austen-Smith, David, and John R Wright. 1992. "Competitive lobbying for a legislator's vote." *Social Choice and Welfare* 9 (3): 229–257.

Caselli, Francesco, and Massimo Morelli. 2004. "Bad Politicians." *Journal of Public Economics* 88 (3–4): 759–782. ISSN: 0047-2727.

Crawford, Vincent P., and Joel Sobel. 1982. "Strategic Information Transmission." *Econometrica* 50 (6): 1431–1451.

Dal Bó, Ernesto, Pedro Dal Bó, and Rafael Di Tella. 2006. ""Plata o Plomo?": bribe and punishment in a theory of political influence." *American Political Science Review* 100 (1): 41–53.

Darden, Keith A. 2001. "Blackmail as a tool of state domination: Ukraine under Kuchma." *E. Eur. Const. Rev.* 10:67.

Dessein, Wouter. 2002. "Authority and Communication in Organizations." *Review of Economic Studies* 69 (4): 811–838.

Dragu, Tiberiu, and Mattias Polborn. 2013. "The administrative foundation of the rule of law." *The Journal of Politics* 75 (4): 1038–1050.

Dragu, Tiberiu, and Adam Przeworski. 2019. "Preventive Repression: Two Types of Moral Hazard." *American Political Science Review* 113 (1): 77–87.

Egorov, Georgy, and Konstantin Sonin. 2011. "Dictators and their viziers: Endogenizing the loyalty–competence trade-off." *Journal of the European Economic Association* 9 (5): 903–930.

Ellsberg, Daniel. 1968. *The Theory and Practice of Blackmail.* Technical report. Rand Coroporation.

Gailmard, Sean, and John W. Patty. 2013. *Learning While Governing: Expertise and Accountability in the Executive Branch.* Chicago: University of Chicago Press.

Galeotti, Andrea, Christian Ghiglino, and Francesco Squintani. 2013. "Strategic Information Transmission Networks." *Journal of Economic Theory* 148 (5): 1751–1769.

Gambetta, Diego. 2009. *Codes of the Underworld: How Criminals Communicate.* Princeton, NJ: Princeton University Press.

Gehlbach, Scott, and Alberto Simpser. 2015. "Electoral manipulation as bureaucratic control." *American Journal of Political Science* 59 (1): 212–224.

Gorlizki, Yoram. 2013. "Structures of Trust after Stalin." *The Slavonic and East European Review* 91 (1): 119–146.

Gregory, Paul R. 2009. *Terror by Quota: State Security from Lenin to Stalin (An Archival Study).* New Haven: Yale University Press.

Klašnja, Marko, Andrew T Little, and Joshua A Tucker. 2018. "Political corruption traps." *Political Science Research and Methods* 6 (3): 413–428.

Ledeneva, Alena V. 2006. *How Russia Really Works: The Informal Practices that Shaped Post-Soviet Politics and Business.* Cornell University Press.

Montagnes, B. Pablo, and Stephane Wolton. 2019. "Mass Purges: Top-Down Accountability in Autocracy." *American Political Science Review.*

Nalepa, Monika, and Konstantin Sonin. 2019. "How Does Kompromat Affect Politics? A Model of Transparency Regimes."

Patty, John W., and Elizabeth Maggie Penn. 2019. "Identity and Information in Organizations."

Patty, John W., and Ian R. Turner. Forthcoming. “Ex Post Review and Expert Policymaking: When Does Oversight Reduce Accountability?” *Journal of Politics.*

Rundlett, Ashlea, and Milan W Svolik. 2016. “Deliver the vote! micromotives and macrobehavior in electoral fraud.” *American Political Science Review* 110 (1): 180–197.

Schelling, Thomas C. 1956. “An essay on bargaining.” *The American Economic Review* 46 (3): 281–306.

Schnakenberg, Keith E. 2015. “Expert Advice to a Voting Body.” *Journal of Economic Theory* 160:102–113.

Schwarz, Michael, and Konstantin Sonin. 2007. “A Theory of Brinkmanship, Conflicts, and Commitments.” *Journal of Law, Economics, and Organization* 24 (1): 163–183.

Tirole, Jean. 1996. “A Theory of Collective Reputations (with Applications to the Persistence of Corruption and to Firm Quality).” *The Review of Economic Studies* 63 (1): 1–22.

Tyson, Scott A. 2018. “The Agency Problem Underlying Repression.” *The Journal of Politics* 80 (4): 1297–1310.

Vatlin, Alexander. 2016. *Agents of Terror: Ordinary Men and Extraordinary Violence in Stalin's Secret Police.* Madison, WI: University of Wisconsin Press.

Zakharov, Alexei V. 2016. “The Loyalty-competence trade-off in dictatorships and outside options for subordinates.” *The Journal of Politics* 78 (2): 457–466.