ISSN 2279-9362

Collegio Carlo Alberto

The Perils of Friendly Oversight

Dino Gerardi
Edoardo Grillo
Ignacio Monzón

No. 630
December 2020

Carlo Alberto Notebooks

www.carloalberto.org/research/working-papers

The Perils of Friendly Oversight

Dino Gerardi* Edoardo Grillo† Ignacio Monzón‡

Abstract

In democratic societies, politicians craft reform proposals which are often under the scrutiny of external authorities. Authorities themselves have their own agendas and may be in favor or against the reform under their scrutiny. We study how the authority's agenda affects the likelihood that a reform is approved and its quality, which both depend on a politician's unobservable and costly effort. We show that an authority in favor of a reform can be detrimental towards its approval. This happens when it is easy to incentivize effort and the status quo alternative is not too attractive.

Keywords: information transmission, moral hazard, oversight, persuasion

JEL Classification: D82, D83

*University of Turin and Collegio Carlo Alberto. Email: dino.gerardi@carloalberto.org
†Collegio Carlo Alberto. Email: edoardo.grillo@carloalberto.org
‡University of Turin and Collegio Carlo Alberto. Email: ignacio@carloalberto.org

1 Introduction

Following James Madison's idea that "ambition must be made to counteract ambition" (Hamilton et al., 2008), the separation of powers and a well-functioning system of checks and balances are two of the main building blocks of constitutional democracies.

Authorities with a high level of professionalism are important actors in the system of checks and balances (Ackerman, 2000). The oversight of these authorities is particularly relevant when the technical details of the issues at hand require an expert opinion (e.g., the Congressional Budget Office, CBO, or the Center for Disease Control, CDC). Ultimately, such opinion ought to improve the quality of policies that are implemented. However, authorities may have their own agenda. Then, endowing the authority with monitoring power may lead to political stalemate and inefficiencies. This problem is exacerbated when, as it is currently the case, political polarization is high (see Ranney, 1951 for an early account of this issue and Thurber and Yoshinaka, 2015 for a more recent discussion).

We study how the alignment or misalignment of interests between a politician and an oversight authority affects the likelihood that reforms are approved and their quality.

We present a simple environment. A constituency (e.g., the electorate or the legislature) must decide whether to approve a reform proposed by a politician (e.g., the executive power or the agenda setter within the legislature) or to reject it, maintaining a *status quo*. The reform can be either of high quality or of low quality. The constituency prefers the reform to the status quo only if the reform is of high quality. The expected quality of the reform depends positively on the costly and unobservable effort of the politician. The politician wants the reform to be approved as it is part of her agenda.

The constituency receives a report issued by an authority. The authority has specific expertise and can thus distinguish between reforms of high and low quality. In line with the literature on persuasion (Kamenica and Gentzkow, 2011), the authority first commits to a reporting strategy, then learns the quality of the reform, and finally issues a report. The commitment assumption captures that governance and procedures of the authority are set in advance.

We consider three types of authorities. First, a *truthful authority* always reports the true quality of the reform.[1] Second, a *friendly authority* is strategic and wants the

[1]We do not model the truthful authority as a player as we find convenient to take its perfectly informative reporting strategy as given. Of course, such truthful reporting strategy can be rationalized by assuming that the authority shares the same preferences as the constituency.

reform to be approved. Third, an *unfriendly authority* is also strategic but wants the reform to be rejected.

To simplify the exposition, we introduce an inactive player to the game: the politician's party. Similarly to the politician, the party wants the reform to be approved, but differently from her, the party does not internalize the cost of effort. Also to better describe the intuition, we let the *approval threshold* denote the (exogenous) expected quality of the reform that makes the constituency indifferent between approval and rejection.

In our main result we characterize the party's preferences over strategic authorities and show how a friendly authority can actually be detrimental to the party. To this goal, we first identify necessary and sufficient conditions on the cost of effort under which the party's payoff with a friendly authority is monotonic in the approval threshold. Intuitively, the party's payoff increases (decreases) in the approval threshold whenever the marginal cost of effort is relatively flat (steep). When the party's payoff is monotonic, the party's preferences over authorities is simple. If the party's payoff is increasing, then the party prefers the unfriendly authority when the approval threshold is low and the friendly authority when the approval threshold is high. Instead, if the party's payoff is decreasing, the party always prefers the friendly authority. Finally, we provide sufficient conditions for the party to find the friendly authority detrimental in the case when the party's payoff is non-monotonic.

To understand the driving force behind the main result, notice the type of authority impacts the party's payoffs through two channels. First, it directly affects the informativeness of the reports available to the constituency. Second, it modifies the politician's incentives to exert effort. Both the informativeness of the report and the level of effort exerted by the politician affect the decision of the constituency and thus the party's payoffs.

For a fixed level of effort, the likelihood of approval is higher with the friendly than with the unfriendly authority. A friendly authority inflates the report: with positive probability it claims that the reform is of high quality even when it is not. Instead, an unfriendly authority claims with positive probability that the reform is of low quality even when it is of high quality. Then, holding the level of effort fixed, the party would be better off with a friendly authority.

The type of authority also affects the level of effort that the politician exerts in equilibrium. On the one hand, when facing the friendly authority, the politician takes advantage of its favorable reporting strategy by exerting lower effort. On the other hand,

when facing the unfriendly authority, the politician must work hard to compensate its unfavorable reporting strategy. To sum up, the politician exerts a higher level of effort with an unfriendly authority than with a friendly one.

The party prefers the unfriendly to the friendly authority whenever the benefits from a higher level of effort outweigh the cost from an unfavorable reporting strategy. This happens whenever the approval threshold is low and the marginal cost of effort is relatively flat. A low approval threshold has two effects. First, it makes the politician exert a low level of effort when facing a friendly authority. Second, it prevents the unfriendly authority from tampering too often with reports. Then, whenever the marginal cost of effort is relatively flat an unfriendly authority makes the politician exert a high level of effort, which results in a high probability of approval.

We also study how a a truthful authority affects the politician's effort and the party's payoff. The main message extends to this case: compared to the truthful authority, a friendly authority is detrimental both for the quality and for the approval probability.

1.1 Related Literature

This paper focuses on a setting in which the approval probability of the reform (hence the players' payoffs) depends on the politician's costly and unobservable effort. The paper is thus related to the literature on moral hazard pioneered by Hölmstrom (1979). Within this extensive literature, our paper is closest to models in which incentive provision occurs through retention/approval binary choices, rather than through compensation contracts. Papers that study such retention rules often focus on political settings (see, for instance, Austen-Smith and Banks 1989, Banks and Sundaram 1993, 1998, Duggan and Martinelli 2017, Ashworth et al. 2017 and the references therein). This paper departs from this literature by introducing an informed third party (the authority) that issues reports concerning the quality of the reform and by studying how its objectives affect incentive provision and final outcomes.[2]

In our paper, the authority has information concerning the quality of the reform and commits to a reporting strategy. This links the paper to the literature on Bayesian persuasion (Kamenica and Gentzkow, 2011; see Kamenica, 2019 for a review on this topic). Among the papers on Bayesian persuasion, there are few that focus on the interaction between persuasion and moral hazard: Boleslavsky and Cotton (2015); Boleslavsky and

[2]Georgiadis and Szentes (2020) study optimal monitor design when the decision maker can pay a cost to directly acquire information.

Kim (2020); Zapechelnyuk (2020); Rodina (2020); Rodina and Farragut (2020).[3] In particular, Rodina (2020), Rodina and Farragut (2020) and Zapechelnyuk (2020) study the optimal information design to incentivize effort. Boleslavsky and Cotton (2015), instead, shows that inflated grading policies, despite decreasing the informativeness of grades, may increase the quality of graduate students if schools compete against each others by investing in education-enhancing technologies.[4] Finally, Boleslavsky and Kim (2020) extends concavification methods (Aumann et al., 1995; Kamenica and Gentzkow, 2011) to characterize optimal persuasion in settings with moral hazard. Differently from this literature, this paper does not take an information design approach.[5] Instead, it investigates how the preferences of an informed authority can affect the quality and the approval probability of reforms.[6]

Our paper is also related to literature that study the impact of strategic certification agency on the behavior of agents (Lizzeri, 1999; Albano and Lizzeri, 2001; Miklós-Thal and Schumacher, 2013; Bizzotto and Harstad, 2020). We differ from these paper as we focus on how the alignment of interests between politicians and oversight authorities impact the quality and the approval of political reforms.

The political economy literature has long investigated the role of informed experts and how they interact with elected officials (see Gailmard and Patty, 2012, for a review of models on this topic.) In particular, Maskin and Tirole (2004); Alesina and Tabellini (2007, 2008); Iaryczower et al. (2013) (among others) study the allocation of heterogeneous political tasks between bureaucrats—who are equipped with superior technical expertise and/or independent of public approval—and politicians. On a different vein, Banks (1989); Huber et al. (2001); Bueno de Mesquita and Stephenson (2007) investigate how politicians can control and incentivize authorities to behave in their interests. Differently from both these literatures, we assume that the constituency relies on the report issued by the authority and we study how the reporting strategy affect the effort chosen by the politician.

[3]The interaction between effort and persuasion is also studied in Feng and Lu (2016); Zhang and Zhou (2016). However, these papers focus on contests rather than on settings with moral hazard. More in general, Bizzotto et al. (2019) study information design when the receiver of information exhibits moral hazard.

[4]Thanks to the complementarity between students' effort and schools' investment, the result holds true even when students react to more lenient grading policy.

[5]The paper also differs from Dragu et al. (2014) that characterize optimal checks and balances from a mechanism design perspective in a setting without uncertainty and moral hazard.

[6]Alonso and Câmara (2016) studies the role of persuasion in a political setting. Differently from us, it considers a two-player persuasion game in which one player designs how to release information to the other player and shows how different voting rules affect the equilibrium analysis.

Finally, our paper shows that biased experts may lead to inefficient approval of reforms. In particular, some high quality reforms will not be approved under an unfriendly authority. Although obvious differences in the setting, this establishes a link to other papers that study political gridlock (see Krehbiel, 1998; Binder, 1999; Ortner, 2017; Austen-Smith et al., 2019, and the references therein).

2 The model

A politician ("she") puts forward a policy reform. The reform can be either of high quality ($\omega = 1$) or of low quality ($\omega = 0$). The reform is of high quality with probability $e \in [0, 1]$ and of low quality with probability $1 - e$ where e denotes the politician's costly effort. The politician's effort level is not observable.

The constituency of the politician must decide whether to approve the reform ($a = 1$) or to reject it ($a = 0$). The constituency does not observe the quality of the reform. Instead, it must rely on a report from an oversight authority ("it"), which observes the quality ω of the reform. The authority may favor or oppose the reform and can bias its report to persuade the constituency into choosing the authority's preferred action. The authority commits to a reporting strategy *before* learning the quality ω. The reporting strategy is a mapping from the quality of the reform to reports. As the constituency has two actions (and the authority has commitment power), we assume without loss of generality that the authority has two available reports: $m \in \{0, 1\}$. Thus, the authority reporting strategy is summarized by a pair $\mu = (\mu_0, \mu_1)$, where μ_ω denotes the probability that the authority reports $m = 1$ when the quality is ω. We assume that $\mu_1 \geq \mu_0$, so $m = 1$ constitutes (weak) evidence that the reform is of high quality.[7]

The timing of the game is simple. First, the politician and the authority simultaneously choose the effort level and the reporting strategy. Then, the quality of the reform is determined and the authority issues its report. The constituency observes both the reporting strategy of the authority and its actual report. Finally, the constituency decides whether to approve or to reject the reform and payoffs are realized.

[7]For every reporting strategy (μ_0, μ_1) with $\mu_1 \geq \mu_0$ there is a strategically equivalent reporting strategy $(1 - \mu_0, 1 - \mu_1)$ in which the report $m = 0$ provides instead (weak) evidence that the reform is of high quality. We focus on reporting strategies with $\mu_1 \geq \mu_0$ to overcome this trivial source of multiplicity.

The politician obtains a payoff of 1 if the constituency approves the reform and a payoff of 0 otherwise. She also pays a cost of effort captured by the function $e \mapsto c(e) \in \mathbb{R}$. Thus, she obtains utility $u(a, e) = a - c(e)$.

Assumption *The cost function c is continuous, strictly increasing and strictly convex. Furthermore, $c(0) = 0$, $c'(0) < 1$, and $c'(1) > 1$ and for all $e \in [0, 1]$, $c'''(e) \geq 0$.*

The assumption $c'(0) < 1$ implies that it is possible to incentivize the politician to exert effort. On the other side, $c'(1) > 1$ implies that it is not optimal for the politician to exert maximal effort and guarantee a reform of high quality with certainty.[8] The assumption on the third derivative of the cost function simplifies the analysis, but our equilibrium characterization extends beyond it (see footnote 13 for details). The remaining assumptions on the cost function are standard.

The constituency obtains a payoff of 1 if it approves a high quality reform and a payoff of 0 if it approves a low quality reform. Instead, if the constituency rejects the reform, it obtains a payoff of $q \in (0, 1)$ from a status quo policy. The authority can be of two possible types: a friendly authority, which gets a payoff of 1 if the reform is approved and 0 otherwise, or an unfriendly authority, which gets a payoff of 1 if the reform is rejected and 0 otherwise. An inactive player in the game, a party ("he"), wants the reform to be approved, but does not pay any effort cost. Hence, the party obtains a payoff of 1 if the reform is approved and a payoff of 0 otherwise.

2.1 Equilibrium concept

We use Perfect Bayesian Equilibrium (PBE) as our solution concept. After observing the report from the authority, the constituency forms a belief about the quality of the reform.[9] We let $\pi(m)$ denote that the probability that the reform is of high quality given report $m \in \{0, 1\}$.

The constituency approves the reform if $\pi(m) > q$, rejects it if $\pi(m) < q$ and is indifferent if $\pi(m) = q$. We refer to q as the *approval threshold*. In equilibrium an indifferent constituency approves the reform if the authority is friendly and rejects it if the authority is unfriendly.[10]

[8] The main findings of the paper extend to the case in which $c'(1) \leq 1$.

[9] In a PBE, when the equilibrium level of effort $\tilde{e}$ belongs to the interval $(0, 1)$, off-path beliefs are computed according to Bayes rule, using $\tilde{e}$ as the prior probability of high quality and the observed reporting strategy of the authority. Footnote 11 describes the restrictions that PBE imposes on beliefs when the effort level takes an extreme value.

[10] This behavior of the constituency guarantees that the best response of the authority is non-empty.

Given the simplicity of the constituency's behavior, we focus hereafter on the behavior of the politician and of the authority, summarized by a pure strategy profile (e, μ). As the cost function is strictly convex and the authority has commitment, all equilibria are in pure strategies.

We start the analysis observing that our model has equilibria where the politician exerts zero effort and thus the reform is of low quality with certainty. This is true regardless of whether the authority is friendly or unfriendly. In these equilibria, there is no room for the authority to persuade the constituency.

Remark 1. Equilibrium with zero effort. *Both with a friendly and an unfriendly authority, there exist equilibria where the politician exerts zero effort and the reform is never approved.*

To see why Remark 1 holds true, suppose that the politician exerts zero effort: $e = 0$. For every reporting strategy and for every on-path message, the constituency rejects the reform. As a result, the authority is indifferent among all reporting strategies. Suppose that the authority chooses a reporting strategy (μ_0, μ_1) with either $\mu_0 > 0$, or $\mu_0 = \mu_1 = 0$. Then, given the behavior of the constituency, exerting zero effort is indeed optimal for the politician.[11]

Henceforth, we turn our attention to the interesting case where persuasion plays a role. We study equilibria with interior effort levels: $e \in (0, 1)$, which do not fully determine the quality of the reform.[12] We refer to these equilibria as *interior equilibria*.

We conclude the section characterizing the useful benchmark in which the authority is not strategic and truthfully reveals the quality of the reform. In this case, the constituency approves the reform if and only if the authority reports $m = 1$. Thus, the approval probability of the reform coincides with the politician's effort level.

Remark 2. Benchmark: truthful authority. *Assume that a non-strategic authority truthfully reveals the quality of the reform: $\mu = (0, 1)$. Then, the politician solves $\max_{e \in [0,1]} e - c(e)$, which yields $e^* = (c')^{-1}(1)$.*

[11] When $e = 0$, off-path beliefs are computed according to Bayes rule (using $e = 0$ as the prior and the observed reporting strategy) unless $\mu_0 = 0$ and $\mu_1 > 0$ and the report is $m = 1$. In this case, the constituency assigns probability one to the reform being of high quality.

[12] In our model there is no equilibrium in which the politician chooses an effort level equal to 1.

3 Results

3.1 Friendly authority

When facing a friendly authority, the politician in equilibrium exerts a level of effort lower than the approval threshold. Indeed, if the equilibrium effort was higher or equal than q, the authority could guarantee the approval of the reform with certainty using an uninformative reporting strategy (i.e., by choosing a reporting strategy μ with $\mu_0 = \mu_1$). Then the politician would be better off deviating and exerting zero effort.

To maximize the probability that the reform is approved, a friendly authority inflates the quality of the reform: it sends report $m = 1$ not only when the reform is of high quality, but also with some probability, when the reform is of low quality ($\mu_0 \in (0, 1)$ and $\mu_1 = 1$).

Proposition 1 characterizes the interior equilibrium under a friendly authority. Its proof, as well as all the other proofs, are reported in Appendix A.

Proposition 1. Friendly authority. *Under a friendly authority, there exists a unique interior equilibrium* $\left(e^F, \mu^F\right)$ *with*

$$c'(e^F) = 1 - \mu_0^F \quad \textit{and} \tag{1}$$

$$\mu^F = (\mu_0^F, \mu_1^F) = \left(\frac{1-q}{q}\frac{e^F}{1-e^F}, 1\right). \tag{2}$$

The constituency approves the reform if and only if it observes report $m = 1$. The politician then maximizes $e + (1 - e)\mu_0 - c(e)$. Equation (1) reflects the politician's first order condition. Relative to a truthful authority, a friendly authority decreases the politician marginal benefit of effort by μ_0^F. Since we are considering interior equilibria ($e^F \in (0, 1)$) $\mu_0^F > 0$ and thus the politician slacks off: $e^F < e^*$. Equation (2) instead describes optimal persuasion: the authority chooses a reporting strategy that makes the constituency indifferent between approving or rejecting the reform after receiving report $m = 1$, that is, it sets $\pi(1) = q$. In this way, it maximizes the approval probability of the reform without undermining the credibility of report $m = 1$.

In equilibrium, the net marginal benefit of effort, $1 - \mu_0 - c'(e)$ with $\mu_0 = \frac{1-q}{q}\frac{e}{1-e}$, must be equal to zero. This net marginal benefit is positive when the effort is close to zero, and negative when the effort is close to one. Moreover. it is also strictly decreasing in the level of effort as both $\mu_0 = \frac{1-q}{q}\frac{e}{1-e}$ and c' are strictly increasing in e. Hence, an interior equilibrium exists and it is unique. Intuitively, the uniqueness comes from

the fact that an increase in the effort level would allow the authority to the send report $m = 1$ more often when the reform is of low quality. This would reduce the marginal benefit of effort and destroy the incentive to work harder.

In equilibrium, the effort of the politician is a function of the approval threshold q. When q increases, the constituency requires a higher posterior belief $\pi(1)$ to approve the reform. Because $\pi(1)$ is decreasing in μ_0, the authority must limit its misreporting: μ_0^F goes down. The drop in μ_0^F increases the politician's marginal benefit of effort and yields an increase in e^F.

An increase in q has thus two effects on the equilibrium approval probability of the reform: it lowers the approval probability through the drop in μ_0^F, but it also increases it through the raise in e^F. The convexity of the cost function implies that the overall effect of these two forces on the equilibrium payoff of the politician is negative. Interesting, whereas always hurting the politician, an increase in q may benefit the party. To see why, recall that the party's equilibrium payoff is equal to the approval probability of the reform: $e^F + (1 - e^F)\mu_0^F$. Then, if the approval threshold changes marginally, the party's equilibrium payoff changes by:

$$\left[(1-\mu_0^F)\frac{de^F}{d\mu_0^F} + (1-e^F)\right]\frac{d\mu_0^F}{dq} = \left[c'(e^F)\frac{de^F}{d\mu_0} + (1-e^F)\right]\frac{d\mu_0^F}{dq}$$

where the equality follows from equation (1). As discussed above, μ_0^F is decreasing in q. Hence, the party's utility increases if and only if the square bracket on the right-hand side of the previous equality is negative. The sign of the square bracket, in turn, depends on the value of $de^F/d\mu_0$, which, by equation (1), is a function of the steepness of the cost function at e^F.

We summarize these comparative statics in the following lemma that plays an important role in the analysis of Section 4.

Lemma 1. Comparative static with friendly authority. *In the interior equilibrium with friendly authority, e^F is increasing in q, the politician's equilibrium payoff is strictly decreasing in q, and the party's equilibrium payoff is strictly increasing in q if and only if $c'(e^F) - (1 - e^F)c''(e^F) > 0$.*

Lemma 1 states that the payoff of the party in the interior equilibrium with friendly media can increase or decrease with the approval threshold q. Figure 1 provides examples of this comparative static for four specific cost functions. In all cases, the cost function guarantees that $e^* = \frac{1}{2}$. Nonetheless, the equilibrium payoff of the party is respectively

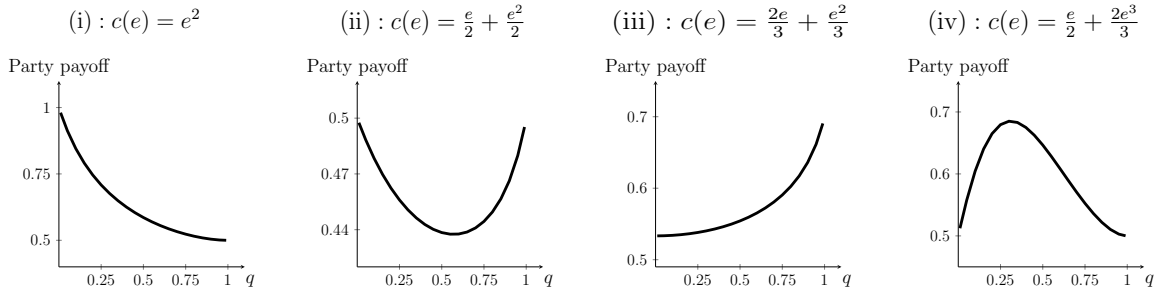

Figure 1: Party payoff as a function of q under a friendly authority.

decreasing (case i), U-shaped (case ii), increasing (case iii), or inversely U-shaped with respect to q (case iv). Obviously, these cases are not exhaustive as the party's payoff ought not to be strictly quasi-concave, nor quasi-convex.

3.2 Unfriendly authority

When the authority is unfriendly, the politician has to work hard to achieve the approval of the reform. In particular, the level of effort in an interior equilibrium must exceed the approval threshold q. If this was not the case, the authority could guarantee the rejection of the reform by adopting an uniformative reporting strategy ($\mu_0 = \mu_1$). However, with this reporting strategy, the optimal level of effort for the politician would be zero.

To minimize the probability that the reform is approved, the authority downplays the quality of the reform: it sends report $m = 0$ not only when the reform is of low quality, but also, with some probability, when the reform is of high quality. Proposition 2 characterizes the set of interior equilibria under an unfriendly authority.

Proposition 2. Unfriendly authority. *Under an unfriendly authority, there exists $q^\dagger \in (0, e^*]$ such that interior equilibria exist if and only if $q \leq q^\dagger$. When $q < q^\dagger$, there are two interior equilibria $\left(e^{N,\ell}, \mu^{N,\ell}\right)$ and $\left(e^{N,h}, \mu^{N,h}\right)$ with $e^{N,\ell} < e^{N,h}$ and for $k \in \{\ell, h\}$*

$$c'(e^{N,k}) = \mu_k^{N,1} \quad \textit{and} \tag{3}$$

$$\mu^{N,k} = \left(\mu_0^{N,k}, \mu_1^{N,k}\right) = \left(0, \frac{e^{N,k} - q}{e^{N,k}(1-q)}\right). \tag{4}$$

When $q = q^\dagger$ the two equilibria collapse into one.

Similarly to the case of the friendly authority, the constituency approves the reform after report $m = 1$ and rejects it after $m = 0$. However, the unfriendly authority downplays the quality of the reform and sends report $m = 0$ not only when the reform is

of low quality, but also when it is of high quality ($\mu_1 < 1$). Thus, the politician's payoff is $e\mu_1 - c(e)$ and her first order condition is given by equation (3). Since $\mu_1 < 1$, also in this case, the politician exerts exerts an effort level lower than e^*. Equation (4), instead, describes the equilibrium reporting strategy. Under optimal persuasion, the authority makes the constituency just indifferent between accepting or rejecting the reform after $m = 0$, that is it chooses $\mu_1^{N,k}$ so that $\pi(0) = q$.

Combining equations (3) and (4) the necessary and sufficient condition for an interior equilibrium is that the politician's net marginal benefit of effort (under the optimal reporting strategy of the authority) is equal to zero:

$$\frac{e-q}{e(1-q)} - c'(e) = 0. \tag{5}$$

To see why an interior equilibrium does not exist for high values of the approval threshold ($q > q^\dagger$), let us first recall that the equilibrium level of effort must be between q and e^*. Trivially, this rules out interior equilibria when q is above e^*. Furthermore, when q is below but sufficiently close to e^*, the marginal cost of effort in the range $(q, 1)$ is close to one, while the marginal benefit of effort ($\frac{e-q}{e(1-q)}$) is bounded away from one. Hence, it is not possible to satisfy equation (5).

On the other hand, when q is not excessively high ($q \leq q^\dagger$), the net marginal benefit of effort is negative for e close to q and e close to e^*, while it is positive for intermediate values. Under the assumption that the third derivative of the cost function is positive, the net marginal benefit of effort is concave and there are exactly two equilibria.[13] A *low effort equilibrium* $(e^{N,\ell}, \mu^{N,\ell})$ in which the authority downplays more the quality of the reform, and a *high effort equilibrium* $(e^{N,h}, \mu^{N,h})$ in which the authority's distortion is more limited. This multiplicity disappears if and only if $q = q^\dagger$, in which case the two equilibria coincide.

The equilibrium multiplicity arises because an increase in effort limits the possibility of the authority to downplay the quality of the reform (it can send report $m = 0$ less often when the reform is of high quality). Hence, the marginal benefit of effort increases, strengthening the politician's incentives to work hard. This is in contrast with case of the friendly authority, where the authority response to an increase in effort weakens politician's response

[13] If the assumption on the third derivative fails, for low values of q there are *at least* two equilibria and they all satisfy equations (3) and (4).

The effort levels in the two interior equilibria under the unfriendly authority both depend on the approval threshold q. However, whereas $e^{N,\ell}$ increases with q, $e^{N,h}$ decreases with it. To see why, observe that an increase in q lowers the marginal benefit of effort μ_0^N for every effort level. To preserve equation (3), the equilibrium levels of effort must adjust. The concavity of the net marginal benefit of effort, $\frac{e-q}{e(1-q)} - c'(e)$, implies that the equilibrium effort increases in the low effort equilibrium and decreases in the high effort one.

Under the oversight of the unfriendly authority, the politician's and the party's payoff are respectively given by $e^{N,k}\mu_1^{N,k} - c(e^{N,k})$ and $e^{N,k}\mu_1^{N,k}$. Both these payoffs exhibit the same comparative static of the equilibrium effort level: they increase in q in the low effort equilibrium and decrease in q in the high effort one.

Lemma 2. Comparative static with unfriendly authority. *In the interior equilibria with unfriendly authority, the equilibrium effort of the politician, her equilibrium payoff, and the party's equilibrium payoff are all increasing in q in the low effort equilibrium and decreasing in q in the high effort equilibrium.*

Although under the unfriendly authority there may be two interior equilibria, the ranking of these equilibria is simple and reflects the differences in the level of effort: the politician, the party, and the constituency all prefer the high effort equilibrium to the the low effort equilibrium, while the opposite ranking holds true for the authority.

Lemma 3. Ranking of equilibria with unfriendly authority. *The politician, the party and the constituency all prefer equilibrium $\left(e_h^N, \mu_h^N\right)$ to equilibrium $\left(e_\ell^N, \mu_\ell^N\right)$. The authority, instead, exhibits the reversed ranking.*

4 Friendly versus unfriendly authority

Section 3 shows that interior equilibria with a positive level of effort exist together with the zero-effort equilibria characterized in Remark 1. In this section we investigate how different agents rank equilibria these various equilibria.

4.1 The politician and the party

In an interior equilibrium, the politician chooses a positive level of effort and the reform is thus approved with positive probability. As a result, both her and the party strictly prefer any interior equilibrium to any equilibrium with zero effort.

Henceforth, our analysis will focus on interior equilibria only. In these equilibria, the type of the authority affects the approval probability of the reform through two different channels. First, the authority distorts the reporting strategy. In particular, a friendly authority inflates the quality of the reform by sending report $m = 1$ also when the reform is of low quality, $\mu_0^F \in (0,1)$. On the contrary, the unfriendly authority downplays the reform's quality by sending report $m = 0$ also when the reform is of high quality, $\mu_1^{N,k} \in (0,1)$ for $k \in \{\ell, h\}$. Holding the effort of the politician constant, these distortions imply that the approval probability is higher under the friendly authority. Second, exactly because of the first channel, the politician works less hard under a friendly authority than under an unfriendly one: $e^F < q < e^{N,\ell} < e^{N,h}$. Holding the authority's reporting strategy constant, this second channel implies that the approval probability is higher under the unfriendly authority.

A simple revealed-preference argument implies that the politician prefers the interior equilibrium with the friendly authority to the ones with the unfriendly authority. Indeed, the equilibrium payoff of the politician under the friendly authority is $e^F + (1-e^F)\mu_0^F - c(e^F)$, while under the unfriendly authority it is $e^{N,k}\mu_1^{N,k} - c(e^{N,k})$. Then:

$$e^F + (1-e^F)\mu_0^F - c(e^F) > e^{N,k} + (1-e_{N,k})\mu_0^F - c(e^{N,k}) > e^{N,k}\mu_1^{N,k} - c(e^{N,k}),$$

where the first inequality follows from optimality of the politician's behavior and the second from $\mu_1^{N,k} \in (0,1)$. A similar argument also shows that the politician ranks the truthful authority strictly between the friendly and unfriendly authority.

Differently from the politician, the party cares only about the approval probability of the reform and does not factor in the cost of effort. As we show in Section 3.1, the party's payoff under a friendly authority is not necessarily monotonic in q. This prevents a complete characterization of the party's preferences over interior equilibria.

However, when the party's payoff with the friendly authority is monotonic in the approval threshold, we can characterize the party's ranking between the equilibrium under the friendly authority and the high-effort equilibrium under the unfriendly one. To this goal, define the function $e \mapsto g(e) = c'(e) - (1-e)c''(e)$. We can then state the following proposition.

Proposition 3. Party's preferences over interior equilibria.

I. *Suppose that for every $e \in [0, e^*]$, $g(e) \leq 0$. Then for every $q \in (0,1)$ the party prefers the interior equilibrium with the friendly authority to the high effort equilibrium with the unfriendly authority.*

II. *Suppose that for every* $e \in [0, e^*]$, $g(e) \geq 0$ *and* $g(e^*) > g(0)$. *Then there exists* $\overline{q} \in [0, q^\dagger]$ *such that the party prefers the interior equilibrium with the friendly authority to the high effort equilibrium with the unfriendly authority if and only if* $q \geq \overline{q}$.

To understand this result, let us first focus on the interior equilibrium with the friendly authority. The proof of Proposition 3 shows that, as q ranges in the interval $(0, 1)$, the level of effort in the interior equilibrium e^F spans the whole interval $(0, e^*)$. In addition, by Lemma 1, the derivative of the party's equilibrium payoff with respect to q has the same sign as $g(e^F)$. Thus, the conditions on $g(e)$ in Proposition 3 are necessary and sufficient for an increasing (case I) or decreasing (case II) party's equilibrium payoff.

In addition, as q goes to one, the party's payoff with the friendly authority converges to e^* independently of its shape. Indeed, when the approval threshold grows higher, the authority must limit its bias in the reporting strategy to ensure that the politician still approves the reform after report $m = 1$. Faced with a more demanding constituency and a less favorable reporting strategy, the politician must increase her effort. In the limit as q goes to 1, any attempt to inflate the quality of the reform would result in the constituency rejecting the reform after any message. Hence, in such limit, the authority plays a truthful reporting strategy and the politician chooses e^*.

Finally, consider the high-effort equilibrium with the unfriendly authority. By Lemma 2 the party's equilibrium payoff is decreasing in q. Furthermore, as q goes to zero, this payoff converges to e^*. To understand this last result, recall that the the unfriendly authority downplays the quality of the reform and this implies that the constituency's belief after report $m = 0$ is above 0. When q is close to 0, the constituency accepts reforms of low expected quality. Then, to allow for the rejection of the reform after $m = 0$, the authority must distort the reporting strategy very little. In the limit as q goes to 0, the authority adopts a truthful reporting strategy and the politican best responds to it choosing e^*.

Combining these observations, we conclude that the party ranks the high-effort equilibrium with the unfriendly authority above the equilibrium with the friendly authority if and only if both the following two conditions hold. First, the party's payoff with the friendly authority increases in q, and second q is not too high ($q \leq \overline{q}$).

The intuition behind these two conditions is as follows. Under a friendly authority, the politician has lower incentives to exert effort because she can rely on the favorable reporting strategy. Nonetheless, when the approval threshold is sufficiently high, the

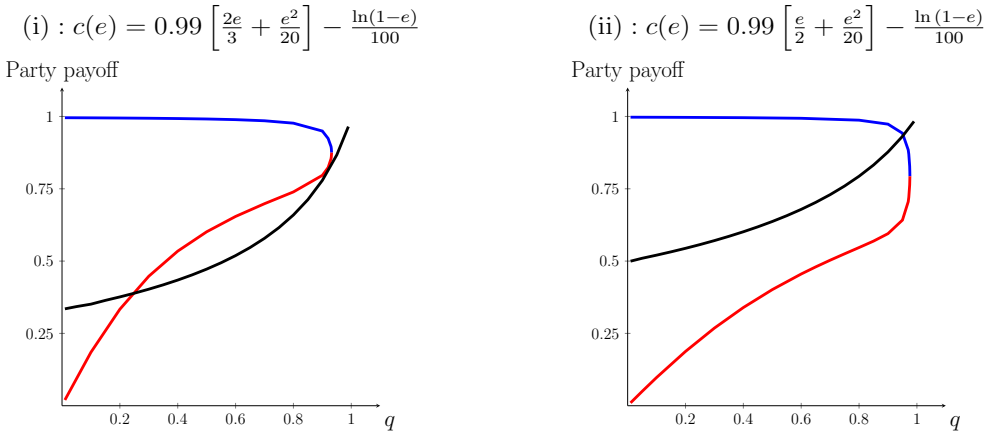

Figure 2: Party's payoff in the interior equilibrium with friendly authority (black), unfriendly authority and low effort (red), unfriendly authority and high effort (blue).

politician still exert high effort: since the constituency only approves reforms of high expected quality, the politician must supplement the friendly oversight with her own work. In this case ($q > \overline{q}$), the party unambiguously prefers the friendly authority: it yields a higher approval probability by inflating the reported quality of the reform and it does not excessively disincentivize effort. On the contrary, when the approval threshold is low ($q \leq \overline{q}$) and the authority is friendly, a politician may exert a low level of effort. This happens when the friendly oversight weakens a lot the politician's incentive to work, that is when, for low levels of effort, the marginal cost is high. These conditions also guarantee that the party's payoff under the friendly authority is increasing in q ($g(e) = c'(e) - c''(e)(1-e) \geq 0$ in the interval $[0, e^*]$). Under the same conditions, there is an equilibrium in which the politician exerts high effort with the unfriendly authority. Indeed, the unfavorable reporting strategy, rather that pushing the politician to slack off, incentivizes her to work hard.[14] As a result, the approval probability and thus the party's equilibrium payoff is higher than in the friendly-authority case.

We now turn to the comparison between the low-effort equilibrium with the unfriendly authority and the equilibrium with the friendly authority. In light of Lemma 3 and Proposition 3, this comparison is interesting only when $q < \overline{q}$ and $g(e) > 0$ on $(0, e^*)$.[15] When q is close to zero, the ranking of these equilibria is unambiguous: the

[14] As explained in Section 3.2, there is also an equilibrium in which the unfavorable oversight discourages effort. Below we discuss the party's ranking between the equilibrium under the friendly authority and the low-effort one under the unfriendly authority.

[15] In all other cases, the party prefers the friendly equilibrium to the high-effort equilibrium with the friendly authority and thus, a fortiori, to the low-effort one.

party's prefer the equilibrium with the friendly authority. Indeed, in the proof of Proposition 3 we show that the party's payoff converges to $1-c'(0) > 0$ with friendly authority and to 0 with the unfriendly authority (low-effort equilibrium). For higher values of q, both rankings are possible. Figure 2 provides an example of each case.

The logic we highlighted above also guides the comparison between the equilibrium with the friendly authority (when the party's payoff is monotonic in q) and the one with the truthful authority. Remark 2 implies that under such truthful authority, the approval probability of the reform, hence the party's equilibrium payoff, is equal to e^*. When incentivizing effort is not too costly ($g(e) \leq 0$), the party prefers the friendly authority for all values of the approval threshold. This is because the favorable reporting strategy increases the approval probability of the reform even when we account for the fact that the politician slacks off. Formally, the party's equilibrium payoff with the friendly authority is decreasing in q and, as discussed above, it converges to e^* as q converges to 1. The opposite is true when incentivizing effort is costly ($g(e) \geq 0$). In this case, the party would prefer the truthful authority as, under the friendly one, his payoff is bounded above by e^*.

Finally, since the effort level e^* is also an upper bound for the party's equilibrium payoff under the unfriendly authority, we can immediately conclude that the party prefers a truthful authority to an unfriendly one.

As we explain above, a characterization of the party's preferences is not possible when his payoff under the friendly authority is non-monotonic in q. In this case the ranking depends on the specification of the cost function. Nonetheless, it is possible to provide sufficient conditions for which the interior equilibrium with the friendly authority is either better or worse than both equilibria with the unfriendly authority. We provide these results in Appendix B.

4.2 The constituency

The constituency gets a payoff respectively equal to q if the reform is rejected, to 1 if the reform is approved and it is of high quality, and to 0 if the reform is approved and it is of low quality.

It is thus immediate to conclude that, when the politician exerts zero effort, the constituency payoff is certain and equal to q. The approval threshold q is also the payoff in the interior equilibrium with the friendly authority. In this case, the approval probability of the reform is $e^F + (1-\mu_0^F)(1-e^F) = e^F/q$. Furthermore, when the reform

is approved, the discussion in Section 3.1 shows that the constituency is indifferent between accepting the reform or rejecting it, which yields a payoff equal to q. Thus, the payoff of the constituency is equal to $(e^F/q)q + \left(1 - e^F/q\right) q = q$.

Under the unfriendly authority, only reforms of high quality get approved. Hence, the payoff of the constituency in an arbitrary interior equilibrium $(e^{N,k}, \mu^{N,k})$ with $k \in \{\ell, h\}$ is equal to $e^{N,k}\mu_1^{N,k} + (1 - e^{N,k}\mu_1^{N,k})q$. Substituting for $\mu_1^{N,k}$, this payoff simplifies to the effort of the politician, $e^{N,k}$, which we know to be greater than q. Intuitively, after report $m = 0$, Section 3.2 implies that, after report $m = 0$, the constituency is indifferent between rejecting the reform or accepting it, which yields an expected payoff equal to the posterior probability of the reform being of high quality: $e^{N,k}(1 - \mu_1^{N,k})/(1 - e^{N,k}\mu_1^{N,k})$. Hence the payoff of the constituency can be rewritten as $e^{N,k}\mu_1^{N,k} + e^{N,k}(1 - \mu_1^{N,k})(1 - e^{N,k}\mu_1^{N,k})/(1 - e^{N,k}\mu_1^{N,k}) = e^{N,k}$.

To sum up, the following remark holds.

Remark 3. *The constituency prefers the high-effort interior equilibrium with the unfriendly authority to the low effort-equilibrium with the unfriendly authority. Furthermore, the constituency prefers either of these equilibria to any other equilibrium with a strategic authority.*

Finally, the constituency always prefers the truthful authority to any strategic authority. To see this, recall that with a truthful authority, the politician's level of effort is equal to e^* and the constituency observes a perfectly informative signal. Thus, the constituency's equilibrium payoff is $e^* + (1 - e^*)q$. The previous remark together with the fact that $e^* > e^{N,h}$ imply that the truthful authority is the constituency's most preferred scenario.

5 Conclusion

A friendly oversight is detrimental for the quality of policy reforms. As a consequence, high quality reforms may be approved less often under the oversight of a friendly authority than under the oversight of an unfriendly one.

We showed these results in a model where an authority provides information to a constituency about the quality of a reform that depends positively on a politician's unobservable and costly effort. A friendly oversight is harmful when the alternative to the reform is not too attractive and when incentivizing effort is not too difficult.

Our results speak to the debate on the importance and potential drawbacks of checks and balances. The conflict of interest between the proponent of a reform and the authority that analyzes it may result not only in better reforms, but also in less frequent political gridlock.

Appendix

A Proofs

Proof of Proposition 1

Recall that, as discussed in the main text, the equilibrium level of effort of the politician must be strictly lower than q.

Consider the authority best response first. Suppose the politician is choosing an effort level $e \in (0, q)$. Given the behavior of the constituency, standard results from the literature on persuasion imply that the authority's best response is to always report $m = 1$ if the project is of high quality, and to report $m = 0$ with the probability $x \in (0, 1)$ solving $\frac{e}{e+(1-e)x} = q$ if the project is of low quality. In other words, the authority chooses reporting strategy $\mu = \left(\frac{e(1-q)}{(1-e)q}, 1\right)$. Such reporting strategy is well defined because in equilibrium $e < q$. Furthermore, it induces the following posterior beliefs: $\pi(0) = 0$, and $\pi(1) = q$. As a result, the constituency approves the reform if the report is $m = 1$ and rejects it if the report is $m = 0$.

Now, consider the best response of the politician. If the authority plays $\mu = (\mu_0, \mu_1)$, the politician solves: $\max_{e\in[0,1)} e\mu_1 + (1 - e)\mu_0 - c(e)$. Given the properties of c, the problem is strictly concave. Hence, the politician's best response satisfies $\mu_1 - \mu_0 = c'(e)$.

Combining the two best responses, we conclude that an interior equilibrium is characterized by equations (1) and (2) in the main text. Putting these two expressions together, we can further see that a necessary condition for an equilibrium is

$$1 - \frac{1-q}{q}\frac{e}{1-e} = c'(e).$$

The left-hand side of the previous expression in increasing in e, while the right-hand side is decreasing in it. Hence, the two functions cross only once and the equilibrium is unique. □

Proof of Lemma 1

Equations (1) and (2) imply that in the interior equilibrium

$$\frac{q - e^F}{q(1 - e^F)} = c'(e^F) \tag{A-1}$$

Equation (A-1) is satisfied if and only if $c'(e^F)q(1-e^F) - q + e^F = 0$. Applying the implicit function theorem on this equality, we get

$$\frac{de^F}{dq} = \frac{1 - c'(e^F)(1-e^F)}{1 - c'(e^F)q + c''(e^F)q(1-e^F)} > 0, \tag{A-2}$$

where the inequality follows from the fact that $c'(e^F) < 1$. Hence, e^F is increasing in q.

First, consider the equilibrium payoff of the politician. It is given by

$$e^F + (1-e^F)\frac{e^F(1-q)}{(1-e^F)q} - c(e^F) = \frac{e^F}{q} - c(e^F).$$

If we take the derivative of this expression with respect to q, we obtain

$$\frac{de^F}{dq}\frac{1 - qc'(e^F)}{q} - \frac{e^F}{q^2} < \frac{1 - c'(e^F)(1-e^F)}{q} - \frac{e^F}{q^2} = 0,$$

where the inequality follows from inequality (A-2) and the equality follows from equation (A-1). Hence, the politician equilibrium payoff is decreasing in q.

Now, consider the equilibrium payoff of the party which is equal to the approval probability of the reform: $e^F + (1-e^F)\mu_0^F$. As discussed in the main text, the effect of a marginal change in the approval threshold on the equilibrium payoff of the party is:

$$\left[(1-\mu_0^F)\frac{de^F}{d\mu_0^F} + (1-e^F)\right]\frac{d\mu_0^F}{dq} = \left[c'(e^F)\frac{de^F}{d\mu_0} + (1-e^F)\right]\frac{d\mu_0^F}{dq}$$

Since μ_0^F is decreasing in q, the party payoff is strictly increasing in q if

$$c'(e^F)\frac{de^F}{d\mu_0} + (1-e^F) < 0,$$

and strictly decreasing in q if the opposite strict inequality holds. Applying the implicit function theorem on equation (1), we get $\frac{de^F}{d\mu_0} = -(c''(e^F))^{-1}$. Hence, the party payoff is strictly increasing in q if and only if

$$-\frac{c'(e^F)}{c''(e^F)} + (1-e^F) < 0.$$

Rearranging terms, we get the inequality in the statement of the proposition. □

Proof of Proposition 2

We argued in the main text that in an interior equilibria with an unfriendly authority, $e^N > q$.

Then, suppose there exists an equilibrium in which the politician chooses an effort level $e \in (q, 1)$. Given the behavior of the constituency, standard results in the persuasion literature imply that the authority best response is to never report $m = 1$ when the reform is of low quality, and to report $m = 1$ with the probability $x \in (0, 1)$ that satisfies $\frac{e(1-x)}{e(1-x)+1-e} = q$ if the reform is of high quality. In other words, the authority chooses the reporting strategy $\mu = \left(0, \frac{e-q}{e(1-q)}\right)$, which is well defined because in equilibrium the effort level of the politician must exceed the approval threshold q. Such reporting strategy induces the following posterior beliefs: $\pi(0) = q$ and $\pi(1) = 1$. Hence, the constituency approves the reform if and only if the authority reports message $m = 1$.

The best response of the politician to a reporting strategy $\mu = (\mu_0, \mu_1)$ is given (as in the case of the friendly authority) by the solution of $\mu_1 - \mu_0 = c'(e)$.

Combining the best responses of the politician and of the authority, we conclude that an equilibrium must satisfy equations (3) and (4) in the main text. In particular, equation (3) implies that effort level must be below e^*. Furthermore, the equilibrium must satisfy $\frac{e-q}{e(1-q)} - c'(e) = 0$. For every $q \in (0, 1)$, define

$$f(e; q) := \frac{e - q}{e(1 - q)} - c'(e) \tag{A-3}$$

This function represents the marginal benefit of effort net of its marginal cost. The properties of the cost function imply that $f(e; q)$ is continuous and strictly concave in e for every q. Furthermore, for every $q \in (0, 1)$, $f(q; q) = -c'(q) < 0$ and $f(e^*; q) < 0$. Only two cases are possible: either $f(e; q) \geq 0$ for some $e \in (q, e^*)$, or $f(\cdot; q)$ has an upper bound strictly below 0. In the latter case, $f(e; q) = 0$ admits no solution. In the former case the strictly concavity of f implies that $f(e, q) = 0$ has two solutions if $f(e; q) > 0$ for some $e \in (q, e^*)$ and only one solution otherwise.

Obviously, $f(e; q) \geq 0$ if and only if $\max_{e \in [q,1]} f(e, q) \geq 0$. Define

$$e^\dagger(q) = \arg \max_{e \in [q,1]} f(e, q)$$

(this is well defined because the function is strictly concave and the objective function is bounded above over the set $[q, 1]$). By the maximum theorem, $f(e^\dagger(q), q)$ is continuous

in q. Also, $\lim_{q\to 0} f(e^\dagger(q),q) > 0$, while $\lim_{q\to 1} f(e^\dagger(q),q) < 0$. Finally, $f(e^\dagger(q),q)$ is decreasing in q. To see this, first observe that $\frac{e-q}{e(1-q)}$ is strictly decreasing in q. Then, pick $q'' > q'$. Since the effort level under the unfriendly authority must exceed the approval threshold, the set of feasible effort levels under q' is a superset of the set of feasible effort levels under q''. Hence, $f(e^\dagger(q'),q') \geq f(e^\dagger(q''),q') > f(e^\dagger(q''),q'')$.

The previous properties on $f(e^\dagger(q),q)$ implies that we can find a value $q^\dagger \in (0,1)$ such that:

- if $q < q^\dagger$, there exist two interior equilibria $(e^{N,\ell},\mu^{N,\ell})$ and $(e^{N,h},\mu^{N,h})$ with $q < e^{N,\ell} < e^\dagger(q) < e^{N,h} < 1$;
- if $q = q^\dagger$, there exists a unique interior equilibrium (e^N,μ^N) with $e^N = e^\dagger(q) \in (q,1)$;
- if $q > q^\dagger$, there is no interior equilibrium.

Because $q^\dagger < e^{N,\ell} < e^{N,h} < e^*$, we also have that $q^\dagger < e^*$. □

Proof of Lemma 2

Recall from the proof of Proposition 2 that e_ℓ^N and e_h^N are the two roots of $f(e,q) = 0$, where $f(e;q)$ is defined by (A-3). Note that $f(e;q)$ is decreasing in q for all $e < 1$. Consider the lowest root of $f(e;q) = 0$ and denote it with $e_\ell^N(q)$. By the proof of Proposition 2, we know $f(e;q) < 0$ for every $e < e_\ell^N(q)$. Hence, if $q' > q$, we must have $f(e;q') < 0$ for every $e \leq e_\ell^N(q)$. We conclude that $e_\ell^N(q') > e_\ell^N(q)$. The comparative static for the highest root is similar. Let $e_\ell^N(q)$ denote the highest root of $f(e;q) = 0$. Observe that $f(e;q)$ is decreasing in q for all $e < 1$ and, by the proof of Proposition 2, $f(e;q) < 0$ for every $e > e_h^N(q)$. Hence, if $q' > q$, we have $f(e;q') < 0$ for every $e > e_2^N(q)$. We conclude, $e^N(q') < e_2^N(q)$.

Now, consider the equilibrium payoff of the politician when she chooses effort level e^N. This is equal to:

$$e^N\mu_1^N - c(e^N) = e^N c'(e^N) - c(e^N).$$

Taking the derivative of this expression with respect to q and simplifying, we get:

$$e^N c''(e^N)\frac{de^N}{dq} + \frac{de^N}{dq}c'(e^N) - c'(e^N)\frac{de^N}{dq} = e^N(q)c''(e^N(q))\frac{de^N(q)}{dq}.$$

Hence, the utility of the politician is increasing (respectively, decreasing) in q if the equilibrium effort level is increasing (respectively, decreasing) in it.

Finally, consider the equilibrium payoff of the party. Under the unfriendly authority, this is equal to $e^N \mu_1^N$. The derivative of this payoff with respect to q is equal to

$$e^N c''(e^N)\frac{de^N}{dq} + \frac{de^N}{dq}c'(e^N) = \left[e^N c''(e^N) + c'(e^N)\right]\frac{de^N}{dq},$$

whose sign is again determined by the sign of de^N/dq. □

Proof of Lemma 3

Recall that in the equilibria with unfriendly media, the approval probability of the reform in equilibrium is given $e\mu_1$. The ranking of equilibria for the party follows from observing that $e^{N,h} > e^{N,\ell}$ and $\mu_1^{N,h} > \mu_1^{N,\ell}$, so that $0 < e^{N,\ell}\mu^{N,\ell}, 1 < e^{N,h}\mu_1^{N,h}$. Obviously, the ranking of the authority is reversed with respect to the one of the party.

Now, pick an arbitrary interior equilibrium (e^N, μ^N) under the unfriendly authority and recall that only reforms of high quality get approved. Thus, the equilibrium payoff of the constituency is:

$$e^N\mu_1^N + (1 - e^N\mu_1^N)q = \frac{e^N - q}{1-q} + \frac{1-e^N}{1-q}q = e^N.$$

We conclude that also the constituency prefers the high-effort equilibrium to the low-effort one.

To establish the ranking for the politician, observe that the convexity of c implies that for every $e_2 > e_1$

$$c(e_2) = c(e_1) + \int_{e_1}^{e_2} c'(s)ds < c(e_1) + (e_2 - e_1)c'(e_2). \tag{A-4}$$

Hence, exploiting again the convexity of c, we can conclude that the payoff of the politician in the equilibrium $\left(e^{N,h}, \mu^{N,h}\right)$ satisfies:

$$\begin{aligned} e^{N,h}c'\left(e^{N,h}\right) - c\left(e^{N,h}\right) &> e^{N,h}c'\left(e^{N,h}\right) - c(e^{N,\ell}) - (e^{N,h} - e^{N,\ell})c'(e^{N,h}) = \\ &= e^{N,\ell}c'\left(e^{N,h}\right) - c\left(e^{N,\ell}\right) > e^{N,\ell}c'\left(e^{N,\ell}\right) - c\left(e^{N,\ell}\right) > 0. \end{aligned}$$

Thus the politician prefers the equilibrium with the higher interior level of effort to the one with the lower interior level of effort. □

Proof of Proposition 3

We start characterizing the party's payoff in interior equilibria when the approval threshold q takes extreme values.

First, consider the interior equilibrium with the friendly authority. The party payoff is equal to the approval probability of the reform, which is given by:

$$e^F \mu_1^F + (1 - e^F)\mu_0^F = e^F + \frac{(1-q)e^F}{q} = \frac{e^F}{q}.$$

Note that, as q converges to 1, then μ_0 converges to 0 and the right-hand side of equation (1) converges to 1. We conclude that $\lim_{q\to 1} e^F = e^*$. Thus, the party's payoff in the interior equilibrium with the friendly authority converges to e^* as q converges to 1. Finally, suppose that $q \to 0$. Then, applying de l'Hôpital's rule to e^F/q and exploiting equation (A-2) in the proof of Lemma 1, we conclude that the party's payoff in the interior equilibrium under a friendly authority converges to $1 - c'(0) > 0$.

Now consider the interior equilibria with the unfriendly authority and recall that these equilibria are well defined if and only if $q \in 0, q^\dagger$. Consider an arbitrary interior equilibrium under the unfriendly authority: $(e^{N,k}, \mu^{N,k})$ with $k \in \{\ell, h\}$. In such equilibrium, the payoff of the party is equal to: $e^{N,k}\mu_1^{N,k}$. Equations (3) and (4) yield:

$$c'(e^{N,k})e^{N,k}(1-q) = e^{N,k} - q.$$

As q converges to 0, this equation admits two solutions: $e^{N,h} = e^*$ and $e^{N,\ell} = 0$. Furthermore, as $q \to 0$ and $e^{N,h} \to e^*$, we have that $\lim_{q\to 0} \mu_1^{N,h} = 1$, and that the party's payoff converges to e^*. On the contrary, because $\mu_1^{N,\ell}$ is bounded above by 1 and $e^{N,\ell}$ converges to 0 when $q \to 0$, the party's payoff in the low-effort equilibrium converges to 0 as q converges to 0.

Suppose that $g(x) \leq 0$ for all $x \in [0, e^*]$. Since $e^F \in (0, e^*)$ for every $q \in (0, 1)$, Lemma 1 implies that the party's payoff is decreasing in q. Hence the limits derived before imply that the party's payoff is bounded bounded below by e^*. On the other hand, by Lemma 2 the payoff of the party in the high-effort equilibrium with the unfriendly authority is decreasing in q. Thus, the previous limits also imply that such payoff is

bounded above by e^*. Combining these two bounds, the interior equilibrium with the friendly authority is always better than the interior high-effort equilibrium with the unfriendly authority.

Now suppose that for all $x \in [0, e^*]$, $g(x) \geq 0$ and $g(0) \neq g(e^*)$. In this case, Lemma 1 and the limits we derived above imply that the party's payoff in the interior equilibrium under the friendly authority is increasing in q and bounded below by $1 - c'(0)$. Since $c'(0) \in [0, 1)$, this expression is positive and the limits computed before (together with the monotonicity of the payoff with respect to q) imply that $1 - c'(0) < e^*$. As q converges to 0, e^* is also the limit of the party's payoff in the high-effort equilibrium with the unfriendly authority. We conclude that when q is close to 0, the high-effort equilibrium with the unfriendly authority is preferred to the friendly equilibrium, which is preferred to the low-effort equilibrium with the unfriendly authority. The statement of the proposition follows from the fact that Lemma 2 and $q^\dagger < 1$ imply that the payoff of the party in the high-effort equilibrium with the unfriendly authority decreases in q and it is constantly equal to 0 when $q > q^\dagger$. □

B Sufficient conditions for the ranking of equilibria

Proposition 3 in the main text characterizes the comparison of between the best interior equilibria from the point of view of the party in the special case in which the party's payoff under the friendly authority is monotonic in q. In this respect, the proposition provides condition under which the party is be better off under the unfriendly oversight. Figure 2 complements this analysis and shows that, for specific parametric values, the party may even prefer both interior equilibria with the unfriendly authority to the (unique) interior equilibrium with the friendly authority.

In this Section, we focus on interior equilibria and we provide sufficient conditions under which, *independently of the monotonicity of the party's utility*: the equilibrium with the friendly authority is worse than both equilibria with the unfriendly authority (Proposition B.1), or the equilibrium with the friendly authority is better than both equilibria with the unfriendly authority (Proposition B.2).

The result in Proposition B.1 is obtained defining a lower bound for the equilibrium effort under the friendly authority and an upper bound for the equilibrium effort under the unfriendly authority (z_1 and z_2 in the statement of the proposition). It then uses these quantities to bound the party's payoff under the two types of authority and it

provides a sufficient condition to guarantee that the party is better off with the unfriendly authority.

Proposition B.1. *Fix $q \leq q^{\dagger}$ and define $z_1 = \frac{q[1-c'(q)]}{1-qc'(q)}$ and $z_2 = q[1+(1-q)c'(q)]$. If*

$$z_2 c'(z_2) > \frac{1-c'(z_1)}{1-qc'(z_1)},$$

then the party is better off in any interior equilibrium with the unfriendly authority than in the interior equilibrium with the friendly authority.

Proof. Consider, the equilibrium payoff of the party in the interior equilibrium with the friendly authority: $e^F \mu_1^F + (1-e^F)\mu_0^F = e^F/q$. By equation A-1, we have:

$$\frac{e^F}{q} = (1-c'(e^F))\frac{1-e^F}{1-q}. \tag{A-1}$$

Recall that the specific value of μ_0^F guarantees that $\pi(1) = q$. Hence

$$e^F = \frac{\mu_0^F q}{1-q(1-\mu_0^F)} = \frac{\mu_0^F q}{1-qc'(e^F)},$$

where the second equality follows from equation (1). Plugging this expression for e^F into equation (A-1), we get

$$\frac{e^F}{q} = \frac{1-c'(e^F)}{1-qc'(e^F)}. \tag{A-1}$$

Observe that $z_1 \in (0,q)$. Define the function $v(x;q) = (1-x)/(1-qx)$ and observe that this function is decreasing in x. Suppose that $e^F < z_1$. Then, we would have:

$$\frac{e^F}{q} < \frac{z_1}{q} = v(c'(q);q) < v(c'(e^F);q) = \frac{1-c'(e^F)}{1-qc'(e^F)}.$$

This contradicts the characterization in Proposition 1. We conclude that $e^F \geq z_1$. Because $v(x;q)$ is decreasing in x, this implies that, once we fix q, the party's payoff in the interior equilibrium with the friendly authority is bounded above by

$$\overline{v}^F(q) = \frac{1-c'(z_1)}{1-qc'(z_1)}.$$

Now, consider the unfriendly authority and pick an arbitrary interior equilibrium (e^N, μ^N). By equation (3) the party's payoff in this equilibrium is $e^N \mu_1^N = e^N c'(e^N) =$

$(e^N - q)/(1-q)$. Observe that $z_2 > q$. Suppose that $e^N \in (q, z_2)$. Then, we would have:

$$\frac{e^N - q}{1-q} < \frac{z_2 - q}{1-q} = qc'(q) < e^N c'(e^N).$$

This contradicts the characterization in Proposition 2. We conclude that $e^N \geq z_2$ and thus the equilibrium payoff of the party is bounded below by

$$\underline{v}^N(q) = z_2 c'(z_2).$$

The statement of the proposition follows from the fact that the party is better off in any of the equilibria with the unfriendly authority than in the interior equilibrium with the friendly authority if $\underline{v}^N(q) > \overline{v}^F(q)$. □

Similarly to the previous result, also Proposition 2 hinges on bounds on the equilibrium effort levels that translate in bounds for the party's payoff. Compared to Proposition 2, however, these latter bounds are reversed: a lower bound on the party's payoff under the friendly authority and an upper bound under the unfriendly one.

Proposition B.2. *Fix $q \leq q^\dagger$ and define $z_3 = q + (1-q)\frac{1-c'(q)}{1-qc'(q)}$. If $z_3 c'(z_3) > 1$, then the party is better off in the interior equilibrium with the friendly authority than in any of the equilibria with the unfriendly authority.*

Proof. Let $\chi \in (0,1)$ be the value the solution of the following equation: $\chi c'(\chi) = 1$. Note that $\chi$is well defined because c is increasing and $c'(1) > 1$. By construction (and by the convexity of c), we have: $c'(1) > c'(\chi) = 1/\chi > 1$.

Pick an arbitrary equilibrium with the unfriendly authority, (e^N, μ^N). It must be the case that $e^N < \chi$. Suppose by contradiction that $e^N \geq \chi$. By equation (3) in the main text, the equilibrium payoff of the party is $e^N \mu_1^N = e^N c'(e^N) > 1$, which establishes a contradiction because this payoff is equal to the probability of approval and it is thus bounded above by 1. The party's equilibrium payoff, $\frac{e^N - q}{1-q}$, is thus bounded above by

$$\overline{v}^N(q) = \frac{\chi - q}{1-q}.$$

Now consider the interior equilibrium with the friendly authority. Since $e^F < q$, we can follow the same steps of the proof of Proposition B.1 and conclude that the party's

payoff are bounded below by:

$$\underline{v}^F(q) = \frac{1 - c'(q)}{1 - qc'(q)}.$$

Hence, the party prefers the interior equilibrium with the friendly authority to any of the equilibria with the unfriendly authority if $\underline{v}^F(q) > \overline{v}^N(q)$ or equivalently if $z_3 > \chi$, where z_3 is the expression defined in the statement of the proposition. Because the function $x \mapsto xc'(x)$ is strictly increasing in x, $z_3 > k$ if and only if $z_3 c'(z_3) > 1$. □

References

Ackerman, Bruce, "The new separation of powers," *Harvard law review*, 2000, pp. 633–729.

Albano, Gian Luigi and Alessandro Lizzeri, "Strategic certification and provision of quality," *International economic review*, 2001, *42* (1), 267–283.

Alesina, Alberto and Guido Tabellini, "Bureaucrats or politicians? Part I: a single policy task," *American Economic Review*, 2007, *97* (1), 169–179.

_ **and** _ , "Bureaucrats or politicians? Part II: Multiple policy tasks," *Journal of Public Economics*, 2008, *92* (3-4), 426–447.

Alonso, Ricardo and Odilon Câmara, "Persuading Voters," *American Economic Review*, November 2016, *106* (11), 3590–3605.

Ashworth, Scott, Ethan Bueno de Mesquita, and Amanda Friedenberg, "Accountability and Information in Elections," *American Economic Journal: Microeconomics*, May 2017, *9* (2), 95–138.

Aumann, Robert J, Michael Maschler, and Richard E Stearns, *Repeated games with incomplete information*, MIT press, 1995.

Austen-Smith, David and Jeff Banks, *Electoral accountability and incumbency* 1989.

_ **, Wioletta Dziuda, Bård Harstad, and Antoine Loeper**, "Gridlock and inefficient policy instruments," *Theoretical Economics*, 2019, *14* (4), 1483–1534.

Banks, Jeffrey S., "Agency Budgets, Cost Information, and Auditing," *American Journal of Political Science*, 1989, *33* (3), 670–699.

Banks, Jeffrey S and Rangarajan K Sundaram, "Moral hazard and adverse selection in a model of repeated elections," *Political economy: Institutions, information, competition, and representation*, 1993, pp. 295–311.

Banks, Jeffrey S. and Rangarajan K. Sundaram, "Optimal Retention in Agency Problems," *Journal of Economic Theory*, 1998, *82* (2), 293 – 323.

Binder, Sarah A, "The dynamics of legislative gridlock, 1947-96," *American Political Science Review*, 1999, pp. 519–533.

Bizzotto, Jacopo and Bård Harstad, "The Choice of Certifier in Endogenous Markets," *Available at SSRN 3730508*, 2020.

_ , Eduardo Perez-Richet, and Adrien Vigier, "Information design with agency," *CEPR Discussion Paper No. DP13868*, 2019.

Boleslavsky, Raphael and Christopher Cotton, "Grading standards and education quality," *American Economic Journal: Microeconomics*, 2015, *7* (2), 248–79.

_ and Kyungmin Kim, "Bayesian Persuasion and Moral Hazard," *Working paper*, 2020.

Bueno de Mesquita, Ethan and Matthew C. Stephenson, "Regulatory Quality under Imperfect Oversight," *The American Political Science Review*, 2007, *101* (3), 605–620.

Dragu, Tiberiu, Xiaochen Fan, and James H Kuklinski, "Designing checks and balances," *Quarterly Journal of Political Science*, 2014, *9* (1), 45–86.

Duggan, John and César Martinelli, "The political economy of dynamic elections: Accountability, commitment, and responsiveness," *Journal of Economic Literature*, 2017, *55* (3), 916–84.

Feng, Xin and Jingfeng Lu, "The optimal disclosure policy in contests with stochastic entry: A Bayesian persuasion perspective," *Economics Letters*, 2016, *147*, 103–107.

Gailmard, Sean and John W. Patty, "Formal Models of Bureaucracy," *Annual Review of Political Science*, 2012, *15* (1), 353–377.

Georgiadis, George and Balazs Szentes, "Optimal monitoring design," *Econometrica*, 2020, *88* (5), 2075–2107.

Hamilton, Alexander, James Madison, and John Jay, *The federalist papers*, Oxford University Press, 2008.

Hölmstrom, Bengt, "Moral hazard and observability," *The Bell journal of economics*, 1979, pp. 74–91.

Huber, John D., Charles R. Shipan, and Madelaine Pfahler, "Legislatures and Statutory Control of Bureaucracy," *American Journal of Political Science*, 2001, *45* (2), 330–345.

Iaryczower, Matias, Garrett Lewis, and Matthew Shum, "To elect or to appoint? Bias, information, and responsiveness of bureaucrats and politicians," *Journal of Public Economics*, 2013, *97*, 230–244.

Kamenica, Emir, "Bayesian Persuasion and Information Design," *Annual Review of Economics*, 2019, *11* (1), 249–272.

_ **and Matthew Gentzkow**, "Bayesian Persuasion," *American Economic Review*, October 2011, *101* (6), 2590–2615.

Krehbiel, Keith, *Pivotal politics: A theory of US lawmaking*, University of Chicago Press, 1998.

Lizzeri, Alessandro, "Information revelation and certification intermediaries," *The RAND Journal of Economics*, 1999, pp. 214–231.

Maskin, Eric and Jean Tirole, "The Politician and the Judge: Accountability in Government," *American Economic Review*, September 2004, *94* (4), 1034–1054.

Miklós-Thal, Jeanine and Heiner Schumacher, "The value of recommendations," *Games and Economic Behavior*, 2013, *79*, 132–147.

Ortner, Juan, "A theory of political gridlock," *Theoretical Economics*, 2017, *12* (2), 555–586.

Ranney, Austin, "Toward a More Responsible Two-Party System: A Commentary," *American Political Science Review*, 1951, *45* (2), 488–499.

Rodina, David, "Information Design and Career Concerns," *Working paper*, 2020.

_ and John Farragut, "Inducing Effort Through Grades," *Working paper*, 2020.

Thurber, James A and Antoine Yoshinaka, *American gridlock: The sources, character, and impact of political polarization*, Cambridge University Press, 2015.

Zapechelnyuk, Andriy, "Optimal Quality Certification," *American Economic Review: Insights*, June 2020, *2* (2), 161–76.

Zhang, Jun and Junjie Zhou, "Information disclosure in contests: A Bayesian persuasion approach," *The Economic Journal*, 2016, *126* (597), 2197–2217.