DISCUSSION PAPER SERIES

DP16100

The Perils of Friendly Oversight

Dino Gerardi, Ignacio Monzon and Edoardo Grillo

ORGANIZATIONAL ECONOMICS

CEPR

ISSN 0265-8003

The Perils of Friendly Oversight

Dino Gerardi, Ignacio Monzon and Edoardo Grillo

Discussion Paper DP16100
Published 04 May 2021
Submitted 03 May 2021

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programmes:

- Organizational Economics

The Perils of Friendly Oversight

Abstract

In democratic societies, politicians craft reform proposals which are then subject to the scrutiny of external authorities. Politicians want their proposals approved and can work to improve their quality. Authorities have their own agendas: they may be in favor or against the reforms under their scrutiny. We study how the authority's agenda affects the likelihood that a reform is approved and its quality. We show that an authority in favor of a reform can be detrimental towards its approval. This happens when it is easy to incentivize the politician's work and the status quo alternative is not too attractive.

JEL Classification: D72, D73, D82, D83

Keywords: information transmission, moral hazard, oversight, Persuasion

Dino Gerardi - dino.gerardi@carloalberto.org
Collegio Carlo Alberto and University of Turin and CEPR

Ignacio Monzon - ignacio.monzon@carloalberto.org
Collegio Carlo Alberto and University of Turin

Edoardo Grillo - edoardo.grillo@carloalberto.org
Collegio Carlo Alberto

The Perils of Friendly Oversight*

Dino Gerardi† Edoardo Grillo‡ Ignacio Monzón§

May 3, 2021

Abstract

In democratic societies, politicians craft reform proposals which are then subject to the scrutiny of external authorities. Politicians want their proposals approved and can work to improve their quality. Authorities have their own agendas: they may be in favor or against the reforms under their scrutiny. We study how the authority's agenda affects the likelihood that a reform is approved and its quality. We show that an authority in favor of a reform can be detrimental towards its approval. This happens when it is easy to incentivize the politician's work and the status quo alternative is not too attractive.

Keywords: information transmission, moral hazard, oversight, persuasion

JEL Classification: D72, D73, D82, D83

*We thank Jacopo Bizzotto, Federico Boffa, Leandro Gorno, Juan S. Morales, Santiago Oliveros, and Amedeo Piolatto for valuable comments.

†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), a well-functioning system of checks and balances is one of the building blocks of modern democracies.

Authorities with a high level of professionalism are important actors in the system of checks and balances, as they often play a crucial role in the legislative process (Ackerman, 2000). Politicians (e.g., the executive power or agenda setters within the legislature) craft reform proposals that constituencies (e.g., the electorate or the legislature) then decide whether to approve or to reject. Politicians want their proposals to be approved as they are part of their agendas, while constituencies only want to approve high-quality reforms. Constituencies often lack the technical expertise to evaluate the consequences and the quality of complex proposals. They thus rely on the reports of authorities with specific expertise (e.g., the Congressional Budget Office, CBO, or specialized committees within the legislature).

The informed opinion of authorities ought to improve the quality of implemented reforms. However, authorities may have their own agendas. An authority that favors a reform could help the politician in getting the reform approved. Instead, an authority that opposes a reform could lead to political stalemate, particularly when political polarization is high (see Ranney, 1951 for an early account of this issue and Thurber and Yoshinaka, 2015 for a more recent discussion).[1]

We study how the alignment or misalignment of interests between a politician and an oversight authority affects the likelihood that reforms are approved. We show that an authority that favors a reform can be detrimental toward its approval.

Consider the following simple strategic environment. A politician (e.g., the executive) proposes a reform and a constituency (e.g., the legislature) then decides whether to approve it 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 politician always wants the reform to be approved. At the onset of the game, the politician exerts costly and unobservable effort to draft the reform. This effort determines the expected quality of the reform.

[1] The impact of ideological polarization on the legislative process is a growing concern. Finkel et al. (2020) write: "Political sectarianism also undermines the core government function of competence–of providing for and protecting the people. Members of Congress increasingly prioritize partisan purity over the sorts of compromises that appeal to a large proportion of the population, a tendency that creates legislative gridlock."

The constituency does not observe the quality of the reform, but must instead rely on a report from an authority (e.g., the legislative committee). The authority has specific expertise and can thus distinguish between reforms of high and of low quality. The authority establishes procedures to issue the report. For example, it decides the composition and ideological bias of those providing testimony during hearings. Similarly, it can submit more or less detailed reports about the consequences of the reform. Typically, these procedures are public: interested parties have access to the minutes of the hearings, observe the specific enquiries of the authority, and know who is selected to provide testimony. We model these features by assuming that the authority chooses a *reporting strategy*: a mapping from the quality of the reform to a set of reports. We also assume that the constituency observes both the reporting strategy and the actual report issued by the authority.

We consider three types of oversight authorities. First, a *truthful authority* always reports the true quality of the reform. Second, a *friendly authority* is strategic and wants the reform to be approved. Third, an *unfriendly authority* is also strategic but wants the reform to be rejected. The type of the authority is given and known to all players. The authority that evaluates the reform may have been appointed by and represent the interests of the party of the politician (friendly authority) or of the opposition (unfriendly authority). In some cases the constituency may rely on non-partisan authorities like the CBO (truthful authority).[2]

The party of the politician is an inactive player. 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. Thus, the party's payoff is equal to the probability that the reform is approved. To simplify the exposition, we let the *approval threshold* denote the status quo's payoff. When the expected quality of the reform is equal to the approval threshold, the constituency is indifferent between approving and rejecting the reform.

In our main result we characterize the party's preferences over strategic authorities and show how the friendly authority can actually be detrimental to the party, that is, it can reduce the approval probability of the reform. To understand the driving force behind our main result, note that the type of authority affects the party's payoff through two channels. First, it directly affects the informativeness of the re-

[2]Even nominally non-partisan authorities are often accused of being politically biased. See, for instance, the accusation from the Trump administration against the NIAID director Dr. Fauci during the COVID-19 pandemic (https://cutt.ly/qvR05xz).

ports 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 payoff.

For a fixed level of effort, the likelihood of approval is higher under the friendly than under the unfriendly authority. The friendly authority inflates its report: with positive probability it claims that the reform is of high quality even when it is not. Instead, the unfriendly authority downplays the quality of the reform: with positive probability it claims 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 under the friendly authority.

The type of authority also affects the level of effort that the politician exerts in equilibrium. When facing the friendly authority, the politician takes advantage of its favorable reporting strategy by exerting lower effort. When instead 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 under the unfriendly authority than under the friendly one.

The party prefers the unfriendly to the friendly authority whenever the benefit from a higher level of effort outweighs the lower approval probability due to an unfavorable reporting strategy. This happens when the following two conditions hold. First, the approval threshold is relatively low. Second, the marginal cost of effort is relatively flat. A low approval threshold prevents the unfriendly authority from tampering too often with reports, which incentivizes the politician to exert effort. Both the less severe reporting strategy and the higher level of effort benefit the party. A low approval threshold also allows the friendly authority to inflate its report often, which disincentivizes the politician from exerting effort. While the more favorable reporting strategy benefits the party, the lower level of effort hurts him. When the marginal cost of effort is flat, the politician's level of effort is highly responsive to changes in the reporting strategy: a small decrease in the approval threshold leads to a large decrease in the level of effort under the friendly authority and to a large increase under the unfriendly authority. Together, the two conditions above thus guarantee that the favorable effect of the friendly authority's reporting strategy on the party's payoff does not prevail. The party's payoff under the unfriendly authority is then higher than under the friendly one.

Finally, we also study how 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, the friendly authority is also detrimental to the party.

1.1 Related literature

In our paper, the approval probability of a reform depends on the politician's costly and unobservable effort and affects players' payoffs. Our work 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 Ashworth 2012 and Duggan and Martinelli 2017 for surveys of this literature). We depart from this literature by introducing an informed third party (the authority) who issues a report about the quality of the reform. This allows us to study how the authority's objectives affect incentive provision and outcomes.[3]

There is by now a large literature that studies Bayesian persuasion (Kamenica and Gentzkow, 2011; see Kamenica, 2019 for a review on this topic). In our paper, the authority knows the quality of the reform (but not the level of effort exerted by the politician) and chooses a reporting strategy that is observable to the constituency. A growing line of research within the Bayesian persuasion framework focuses on the interaction between persuasion and moral hazard: Boleslavsky and Cotton (2015), Boleslavsky and Kim (2020), Zapechelnyuk (2020), Rodina (2020), and Rodina and Farragut (2020).[4] Boleslavsky and Cotton (2015) show that inflated grading policies, despite decreasing the informativeness of grades, may increase the quality of graduating students if schools compete against each other.[5] Rodina (2020), Rodina and Farragut (2020) and Zapechelnyuk (2020) study the optimal information design to incentivize effort. Finally, Boleslavsky and Kim (2020) extends concavification methods (Aumann et al., 1995 and Kamenica and Gentzkow, 2011) to characterize optimal persuasion in settings with moral hazard. Our paper differs from this line of research, in

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

[4] Feng and Lu (2016) and Zhang and Zhou (2016) study the interaction between effort and persuasion in contests. Bizzotto et al. (2019) study information design when the receiver is subject to moral hazard.

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

that it does not take an information design approach.[6] Instead, we investigate how the preferences of an informed authority affect its reporting strategy and the likelihood of approval of reforms.[7]

Our paper is also related to the literature on certification agencies (Lizzeri, 1999 Albano and Lizzeri, 2001, Miklós-Thal and Schumacher, 2013, and Bizzotto and Harstad, 2020). This literature focuses on market environments where third parties certify the products of sellers.

There is a broad literature in political economy that investigates the role of informed experts and how they interact with elected officials (see Gailmard and Patty, 2012, for a review). Maskin and Tirole (2004), Alesina and Tabellini (2007, 2008), and Iaryczower et al. (2013) study the trade-offs associated with the allocation of political tasks to bureaucrats—who are equipped with superior technical expertise and who are not elected—rather than to politicians. Banks (1989), Huber et al. (2001), and Bueno de Mesquita and Stephenson (2007) investigate how politicians can control and incentivize authorities to behave in the politicians' interests. Differently from both these literatures, we assume that the constituency relies on the report issued by an authority and we study how the reporting strategy affects the likelihood of the approval of a reform.

In Taylor and Yildirim (2011), like in our paper, the behavior of a reviewer affects the effort exerted by an agent. Taylor and Yildirim compare blind evaluations—i.e., evaluations in which the identity of the agent is kept secret—against informed evaluations. We instead focus on how the preferences of the reviewer affect outcomes.

Finally, our paper shows that biased experts may lead to inefficient approval of reforms. Our paper is thus linked to the literature on political gridlock (see Krehbiel, 1998, Ortner, 2017, Austen-Smith et al., 2019, and the references therein). Our work highlights the role of information transmission as a source of political stalemate.

[6] Our paper also differs from Dragu et al. (2014), who characterize optimal checks and balances from a mechanism design perspective in a setting without uncertainty.

[7] Alonso and Câmara (2016), Bardhi and Guo (2018), Chan et al. (2019) and Heese and Lauermann (2020) study the role of persuasion in political settings. These papers study how a politician can persuade an electorate made by multiple receivers into approving a policy. In our model, there is only one receiver, the quality of the policy under consideration is determined endogenously, and the sender is a third party with its own preferences.

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.[8]

The timing of the game is simple. First, the politician chooses the level of effort. Second, without observing the level of effort, the authority chooses the reporting strategy. Third, the quality of the reform is determined and the authority issues its report. Fourth, 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.

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, $c(0) = 0$, $c'(0) < 1$, $c'(1) > 1$, and $c'''(e) \geq 0$ for all $e \in [0, 1]$.*

The assumption $c'(0) < 1$ implies that it is possible to incentivize the politician to exert some effort. Next, the assumption $c'(1) > 1$ implies that it is never optimal for

[8]For every reporting strategy (μ_0, μ_1) with $\mu_1 > \mu_0$ there is a strategically equivalent reporting strategy $(1 - \mu_0, 1 - \mu_1)$ in which the report $m = 0$ provides instead 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 to exert maximal effort and guarantee a reform of high quality.[9] The assumption on the third derivative of the cost function simplifies the analysis, but our equilibrium characterization extends beyond it (see footnote 18 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. If the constituency instead rejects the reform, it obtains a payoff of $q \in (0,1)$ from a status quo policy.

The authority is of one of three possible types: a friendly authority, which gets a payoff of 1 if the reform is approved and 0 otherwise, an unfriendly authority, which gets a payoff of 1 if the reform is rejected and 0 otherwise, and a non-strategic truthful authority that always reveals the true quality of the reform.[10]

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.[11] We let $\pi(m)$ denote the probability that the reform is of high quality when the authority issues a 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.[12]

Given the simplicity of the constituency's behavior, we focus hereafter on the behavior of the politician and of the authority. When the authority chooses its reporting strategy, it does not know the politician's level of effort. Thus, we model the interaction between the politician and the authority as a static game and summarize their

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

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

[11] In a PBE, off-path beliefs are computed according to Bayes rule when the equilibrium level of effort is interior: $e \in (0,1)$. Footnote 13 describes the restrictions that the solution concept imposes on beliefs when the effort level takes an extreme value.

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

behavior with a pure strategy profile (e, μ). As the cost function is strictly convex and the authority commits to a reporting strategy, all equilibria are in pure strategies.

3 Equilibrium analysis

3.1 Preliminary results

We start the analysis by highlighting that there are always 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 under the friendly and the 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. 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 (μ_0, μ_1). If the authority chooses a reporting strategy with either $\mu_0 > 0$, or with $\mu_0 = \mu_1 = 0$, then exerting zero effort is indeed optimal for the politician.[13]

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.[14] We refer to these equilibria as *interior equilibria*.

We next characterize the useful benchmark of the equilibrium under the truthful authority. In this case, the authority reports the quality of the reform truthfully and 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 $\max_{e \in [0,1]} e - c(e)$ by choosing $e^* = (c')^{-1}(1)$.*

[13] If $e = 0$, the observed reporting strategy has $\mu_0 = 0$ and $\mu_1 > 0$, and the report is $m = 1$, then the constituency assigns probability one to the reform being of high quality. For all other cases with $e = 0$, off-path beliefs can be computed according to Bayes rule.

[14] The assumption $c'(1) > 1$ guarantees that in equilibrium the politician never chooses $e = 1$.

3.2 Friendly authority

The friendly authority wants the approval of the reform. Thus, it inflates the quality of the reform: the friendly authority 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. The politician takes advantage of this friendly oversight and slacks off: she exerts a level of effort lower than the approval threshold q.[15] Proposition 1 characterizes the interior equilibrium under the friendly authority. We report its proof, as well as all other proofs, in Appendix A.

PROPOSITION 1. FRIENDLY AUTHORITY. *Under the 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 only when 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 the truthful authority, the friendly authority decreases the politician's marginal benefit of effort by μ_0 and thus the politician slacks off and chooses a level of effort lower than e^*. Equation (2) 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, the authority sends report $m = 1$ as often as possible, while still making it credible enough to induce the approval of the reform.

An interior equilibrium exists and it is unique. To see why, let e^F be an equilibrium level of effort. Equations (1) and (2) together imply that e^F must satisfy

$$1 - \frac{1-q}{q}\frac{e^F}{1-e^F} - c'(e^F) = 0. \tag{3}$$

The left hand side of equation (3) is strictly decreasing in e^F, strictly positive at $e^F = 0$ and strictly negative at $e^F = e^*$. Then, e^F exists and is unique. The intuition behind equilibrium uniqueness is simple: a *higher* level of effort allows the authority to send

[15] If the equilibrium level of effort was greater or equal than q, the constituency would approve the reform regardless of the report. The politician would then choose zero effort. See the proof of Proposition 1 for details.

report $m = 1$ more often when the reform is of low quality. This in turn reduces the marginal benefit of effort, which would then lead to a *lower* level of effort instead.

The equilibrium level of effort increases with the approval threshold q. A higher threshold q forces the friendly authority to increase the informativeness of the report $m = 1$: the authority must limit its misreporting and so μ_0^F decreases. The drop in μ_0^F increases the politician's marginal benefit of effort and thus leads to an increase in e^F. Finally, as a higher approval threshold q reduces μ_0^F, the politician's equilibrium payoff also decreases.

Interestingly, while an increase in the approval threshold always hurts the politician, it may benefit the party. To see how, consider the effect of a higher approval threshold q on the party's payoff. A higher approval threshold decreases the distortion of the friendly authority, which has two countervailing effects on the approval probability. First, it decreases the likelihood of approval of low-quality reforms. Second, it forces the politician to work harder: she exerts a higher level of effort. The degree of convexity of the cost function determines how much the politician increases her level of effort. When the degree of convexity is not too high (see the definition of mildly convex cost functions below), the overall effect of a higher q on the approval probability is positive: the higher level of effort more than compensates the less lenient oversight. Thus, the party is better off with a higher approval threshold q.

To understand the effect of the convexity of the cost function on the party's payoff, recall that the party obtains a payoff equal to: $e^F + (1 - e^F)\mu_0^F$ in equilibrium. Then, the effect on the party's payoff of a marginal increase in q is equal to

$$\left[(1 - \mu_0^F)\frac{de}{d\mu_0} + (1 - e^F)\right]\frac{d\mu_0^F}{dq} = \left[c'(e^F)\frac{de}{d\mu_0} + (1 - e^F)\right]\frac{d\mu_0^F}{dq}, \tag{4}$$

where the equality follows from equation (1) and $de/d\mu_0$ is the slope of the politician's best response. As μ_0^F decreases with q, the party's payoff increases if and only if the square bracket on the right-hand side of equation (4) is negative. Again by equation (1), $de/d\mu_0$ depends on the convexity of the cost function: $de/d\mu_0 = -1/c''(e^F)$.

LEMMA 1. COMPARATIVE STATICS WITH FRIENDLY AUTHORITY. *In the interior equilibrium under the friendly authority, the level of effort e^F strictly increases with q and the politician's payoff strictly decreases with q. Moreover, at any given q, the party's payoff strictly increases with q if $c'(e^F) - (1 - e^F)c''(e^F) > 0$, strictly decreases with q if $c'(e^F) - (1 - e^F)c''(e^F) < 0$ and is constant otherwise.*

Lemma 1 shows that the local degree of convexity of the cost function determines how the approval threshold q affects the party's payoff. The cost function $c(e) = -\alpha \log(1-e)$ with $0 < \alpha < 1$ acts as a knife-edge case: the condition $c'(e^F) - (1 - e^F)c''(e^F) = 0$ always holds true. When $c(e) = -\alpha \log(1-e)$, the effect of a higher e^F exactly compensates that of a lower μ_0^F. As a result, the party's payoff is constant for all $q \in (0,1)$.

We then define two classes of cost functions for which the effect of q on the party's payoff is monotonic. First, a cost function is *highly convex* if it is convex relative to $-\log(1-e)$. Second, a cost function is *mildly convex* if $-\log(1-e)$ is convex relative to it.[16]

DEFINITION. HIGHLY CONVEX AND MILDLY CONVEX COST FUNCTIONS. *We say that a cost function c is highly convex if there exists a strictly convex and increasing function h such that $c(e) = h(-\log(1-e))$ for all $e \in [0, (c')^{-1}(1)]$. Similarly, we say that a cost function c is mildly convex if there exists a strictly convex and increasing function h such that $-\log(1-e) = h(c(e))$ for all $e \in [0, (c')^{-1}(1)]$.*

A cost function c is mildly (highly) convex if $c''(e)/c'(e) < (>)\ 1/(1-e)$ for all $e \in [0, (c')^{-1}(1)]$ (see footnote 16). When the cost function is mildly convex, the local condition $c'(e^F) - (1 - e^F)c''(e^F) > 0$ always holds true. Lemma 1 then implies that the party's payoff strictly increases with q. The blue thick curve in Figure 1 depicts the party's payoff when $c(e) = (2e + e^2)/3$, which is mildly convex. When instead the cost function is highly convex, the local condition $c'(e^F) - (1 - e^F)c''(e^F) < 0$ always hold true. Then, by Lemma 1, the party's payoff strictly decreases with q. The red dashed curve in Figure 1 depicts the party's payoff when $c(e) = e^2$, which is highly convex. Finally, the black dotted line corresponds to the knife-edge case of $c(e) = -\log(1-e)/2$. In this case, the party's payoff is constant in q and it is equal to e^*.

[16] A function f is convex relative to a function g if $f = h(g)$ for some increasing and strictly convex function h. For twice differentiable (strictly increasing and convex) functions the following definition is equivalent: f is convex relative to g in $[a,b]$ whenever $f''(x)/f'(x) > g''(x)/g'(x)$ for all $x \in [a,b]$, that is if the (absolute) value of the Arrow-Pratt coefficient of f is larger than the one of g at all points (see Mas-Colell et al., 1995, Proposition 6.C.2, for the proof of the equivalence between these two definitions).

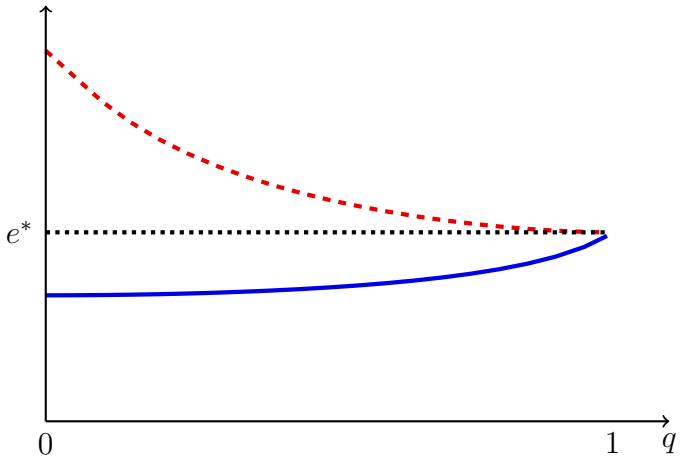


Figure 1: Party's payoff as a function of q under the friendly authority.

Notes: The red dashed curve depicts the party's payoff when $c(e) = e^2$. The blue thick curve depicts the party's payoff when $c(e) = (2e + e^2)/3$. The black dotted line depicts the party's payoff when $c(e) = -\log(1-e)/2$. Under all three cost functions $e^* = 1/2$.

3.3 Unfriendly authority

The unfriendly authority wants the rejection of the reform. Thus, it downplays the reform's quality: the unfriendly authority 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. In response, the politician works hard: her level of effort in an interior equilibrium exceeds the approval threshold q.[17] Proposition 2 characterizes the set of interior equilibria under the unfriendly authority.

PROPOSITION 2. UNFRIENDLY AUTHORITY. *Under the 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 $\left(e^\ell, \mu^\ell\right)$ and $\left(e^h, \mu^h\right)$ with $e^\ell < e^h$ and $\mu_1^\ell < \mu_1^h$. Moreover, for $k \in \{\ell, h\}$,*

$$c'(e^k) = \mu_1^k \quad \textit{and} \tag{5}$$

$$\mu^k = \left(\mu_0^k, \mu_1^k\right) = \left(0, \frac{e^k - q}{e^k(1-q)}\right). \tag{6}$$

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

[17] If the equilibrium level of effort was positive and lower than q, the constituency would reject the reform regardless of the report. The politician would then choose zero effort. See the proof of Proposition 2 for details.

As under the friendly authority, the constituency approves the reform after report $m = 1$ and rejects it after report $m = 0$. However, unlike in that case, the unfriendly authority only sends report $m = 1$ when the quality is high (and with probability $\mu_1 < 1$). Thus, the politician's payoff is $e\mu_1 - c(e)$. Equation (5) reflects the politician's first order condition. Since $\mu_1 < 1$, the politician exerts an effort level lower than e^*. Equation (6) describes the optimal persuasion strategy. The authority chooses a reporting strategy that makes the constituency indifferent between accepting or rejecting the reform after $m = 0$, that is, it sets $\pi(0) = q$.

There may be multiple equilibria under the unfriendly authority. This is because a higher level of effort *increases* the marginal benefit of effort. An increase in the politician's level of effort limits how much the authority can downplay the quality of the reform. When the level of effort is higher, the authority must increase μ_1 to remain credible. Hence, a higher effort level *increases* the marginal benefit of effort. This is in contrast to the case of the friendly authority, whose response to a higher level of effort *decreases* the marginal benefit of effort.

An interior equilibrium level of effort e^k must satisfy the following necessary and sufficient condition:

$$\frac{e^k - q}{e^k(1-q)} - c'(e^k) = 0, \tag{7}$$

which results from equations (5) and (6). The term $(e - q)/[e(1 - q)]$ captures the politician's marginal benefit of effort, which increases with e.

Figure 2 depicts the left hand side of equation (7) as a function of the effort level e. The red dashed curve corresponds to a low approval threshold: $q_1 < q^\dagger$. The blue thick curve corresponds to a high approval threshold: $q_2 > q^\dagger$. The blue thick curve lies below the red dashed curve as a higher approval threshold decreases the marginal benefit of effort. For any q, the left hand side of equation (7) is negative at both $e = 0$ and $e = 1$. It is also strictly concave.[18] Then, for low enough values of the approval threshold, there are two interior equilibria: a *low-effort equilibrium* (e^ℓ, μ^ℓ) and a *high-effort equilibrium* (e^h, μ^h). The red dashed curve in Figure 2 illustrates this: its intersection with the x-axis pins down e^ℓ and e^h. If instead $q > q^\dagger$, no interior equi-

[18] This results from the concavity of the marginal benefit of effort and from the assumption $c'''(e) > 0$ for all $e \in [0, 1]$. Even without this assumption on the third derivative, there would still be *at least* two equilibria and they would satisfy equations (5) and (6).

librium exists (the unique equilibrium is the zero-effort one). The lack of intersection of the blue thick curve with the x-axis illustrates this.[19]

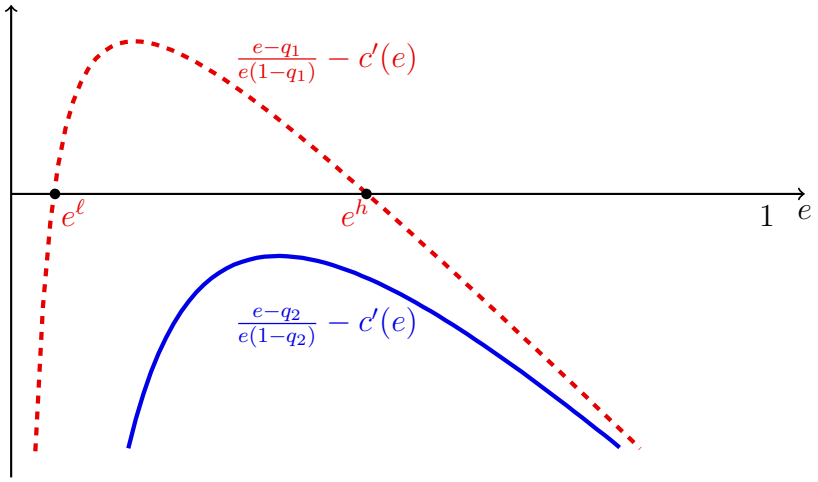


Figure 2: Multiple equilibria under the unfriendly authority

Notes: The red dashed curve depicts the left hand side of equation (7) with $q_1 = 0.05$. The blue thick curve depicts the left hand side of equation (7) with $q_2 = 0.2$. The cost of effort is $c(e) = e^2$ for both curves.

While e^ℓ increases with the approval threshold, e^h decreases with it. To see why, note that a higher q lowers the marginal benefit of effort. The concavity of the left hand side of equation (7) implies that, to preserve the equality, the equilibrium effort must increase in the low-effort equilibrium and decrease in the high-effort one.

The authority's distortion is more pronounced in the low-effort equilibrium than in the high-effort one: $\mu_1^\ell < \mu_1^h$. Moreover, μ_1^ℓ increases with the approval threshold and μ_1^h decreases with it. Hence, the effect of changes in q on the politician's payoff and on the party's payoff (respectively, $e^k\mu_1^k - c(e^k)$ and $e^k\mu_1^k$) is straightforward. We summarize these results in Lemma 2.

LEMMA 2. COMPARATIVE STATICS WITH UNFRIENDLY AUTHORITY. *Under the unfriendly authority, the effort level, the politician's payoff, and the party's payoff all increase with q in the low-effort equilibrium and decrease with q in the high-effort equilibrium.*

We conclude the section characterizing the players' ranking over the low-effort and the high-effort equilibrium under the unfriendly authority.

[19] When $q = q^\dagger$, there is only one equilibrium: the unique maximum of the left-hand side of equation (7) is equal to 0.

LEMMA 3. RANKING OF EQUILIBRIA WITH UNFRIENDLY AUTHORITY. *The politician, the party and the constituency all prefer the high-effort equilibrium to the low-effort equilibrium. The authority, instead, exhibits the reversed ranking.*

Lemma 3 follows from the fact the unfriendly authority tampers reports less often in the high-effort equilibrium than in the low-effort one.

4 Ranking over authorities

In this section we present the agents' ranking over different types of authorities. We explain how a friendly authority can be detrimental to the party.

4.1 The politician and the party

The constituency approves the reform with positive probability whenever the equilibrium level of effort is positive. As a result, both the politician and the party strictly prefer any interior equilibrium to any equilibrium with zero effort. In what follows we thus focus on interior equilibria only.

The type of authority affects the approval probability of the reform through two channels. First, authorities distort their reports. The friendly authority inflates the quality of the reform by sending report $m = 1$ not only when the reform is of high quality, but also with probability $\mu_0^F \in (0,1)$ when the reform is of low quality. On the contrary, the unfriendly authority downplays the quality of the reform by sending report $m = 0$ not only when the reform is of low quality, but also with probability $\mu_1^k \in (0,1)$ when the reform is of high quality. Holding the effort of the politician constant, these distortions imply that the approval probability is higher under the friendly authority. Second, because of the authorities' distortion, the politician works harder under the unfriendly authority than under the friendly one: $e^F < q < e^\ell < e^h$. Holding the authority's reporting strategy constant, this channel implies that the approval probability is higher under the unfriendly authority.

A revealed-preference argument shows that the politician prefers the friendly authority to the unfriendly one. The politician's payoff is $e^F + (1-e^F)\mu_0^F - c(e^F)$ under the friendly authority and $e^k\mu_1^k - c(e^k)$ under the unfriendly one. Then for every $k \in \{\ell, h\}$:

$$e^F + (1-e^F)\mu_0^F - c(e^F) > e^k + (1-e^k)\mu_0^F - c(e^k) > e^k\mu_1^k - c(e^k).$$

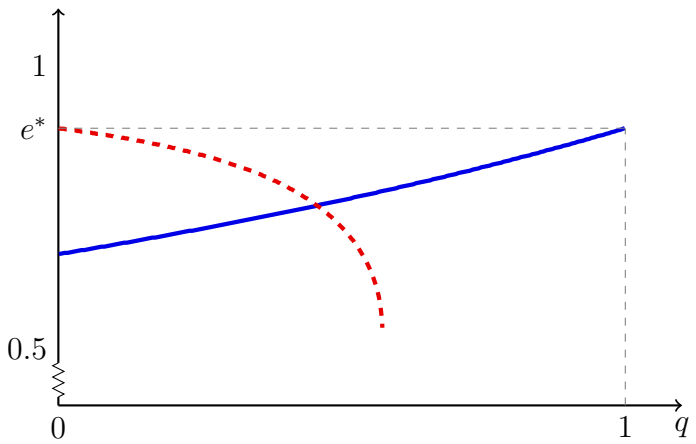


Figure 3: Party's payoff under the friendly and unfriendly authority

Notes: The blue thick curve depicts the party's payoff under the friendly authority. The red dashed curve depicts the party's payoff in the high-effort equilibrium under the unfriendly authority. The cost function is $c(e) = \frac{1}{3}\left[\frac{3}{4}e + \frac{1}{4}\left(-\log(1-e)\right)\right]$, so $e^* = \frac{8}{9}$.

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

The party's ranking over authorities does not necessarily coincide with that of the politician. Although both the party and the politician care about the approval probability of the reform, the party, unlike the politician, does not factor in the cost of effort.

Figure 3 illustrates that the party may prefer the unfriendly authority over the friendly one. The cost function is $c(e) = \frac{1}{3}\left[\frac{3}{4}e + \frac{1}{4}\left(-\log(1-e)\right)\right]$, which is mildly convex.[20] Thus, the party's payoff under the friendly authority (depicted with the blue thick curve) increases with q. Instead, the party's payoff under the unfriendly authority in the high-effort equilibrium (depicted with the red dashed curve) decreases with q. For low levels of the approval threshold, the party is strictly better off under the unfriendly authority than under the friendly authority. This corresponds to the range in Figure 3 where the red dashed curve lies above the blue thick curve.

To see how this ranking over authorities may emerge, note that the friendly authority does not have much room for persuasion when the approval threshold q is close to one. The friendly authority must behave almost like the truthful one, so the

[20]The cost function c is a weighted sum of $-\log(1-e)$ and a liner function. Hence, c is mildly convex, since $-\log(1-e)$ is convex relative to c.

politician's level of effort is close to e^*. The party thus obtains a payoff close to e^*. As the approval threshold *decreases*, the party's payoff under the friendly authority goes down: the negative effect from a lower level of effort dominates the positive effect from the friendly oversight. Note next that when the approval threshold q is close to zero the unfriendly authority behaves almost like the truthful one. This gives the party a payoff close to e^* in the high-effort equilibrium.

The party finds the friendly authority detrimental when the cost function is mildly convex and the approval threshold is low. A low approval threshold means that the unfriendly authority cannot tamper its reports too much, while the friendly authority can inflate them often. A mildly convex cost function, in turn, means that the level of effort is highly responsive to incentives: it decreases a lot when the oversight of the friendly authority becomes more lenient and it increases a lot when the oversight of the unfriendly authority becomes less severe. When the approval threshold is low (and the cost function is mildly convex), the party's payoff under the friendly authority is low because the (negative) effect on the level of effort offsets the (positive) effect of the favorable reporting strategy. At the same time, the party's payoff under the unfriendly authority is high thanks to the little tampering and its associated high level of effort. The party's is thus better off with the unfriendly authority.

When the cost function is highly convex, the level of effort does not react much to incentives. Then, the party always prefers the friendly authority. Indeed, while the approval probability under the unfriendly authority is bounded above by e^*, this probability is bounded below by e^* under the friendly authority.

PROPOSITION 3. PARTY'S PREFERENCES OVER INTERIOR EQUILIBRIA. *Whenever the cost function is mildly convex, there exists $\overline{q} \in [0, q^\dagger]$ such that the party prefers the high-effort equilibrium under the unfriendly authority to the interior equilibrium under the friendly one if and only if $q \geq \overline{q}$.*

Whenever the cost function is highly convex the party prefers the interior equilibrium under the friendly authority to the high-effort equilibrium under the unfriendly one for every $q \in (0, 1)$.

The party may even prefer the low-effort equilibrium under the unfriendly authority to the interior equilibrium under the friendly one. We next discuss how the party ranks the low-effort equilibrium under different configurations of the primitives. First, the party unambiguously prefers the friendly authority whenever either

the cost function is highly convex or $q > \overline{q}$.[21] Second, when the cost function is mildly convex and q is close to zero, the party again prefers the friendly authority. The politician's level of effort converges to zero as q approaches zero both under the friendly authority and in the low-effort equilibrium under the unfriendly one.[22] Nonetheless, the politician benefits from the oversight of the friendly authority, so the party prefers the friendly authority to the low-effort equilibrium under the unfriendly one.[23] Finally, for interior values of the approval threshold q, the party may rank the equilibrium under the friendly authority above or below the low-effort equilibrium under the unfriendly authority.

Figure 4 illustrates the party's ranking between the friendly authority and the low effort equilibrium under the unfriendly one with two examples. Figure 4a shows that with the cost function from Figure 3, the party always prefers the equilibrium under the friendly authority to the low-effort equilibrium under the unfriendly one. Figure 4b shows that with a similar cost function the order can be reversed: for an interior range of approval thresholds, the party prefers the low-effort equilibrium under the unfriendly authority to the equilibrium under the friendly one.

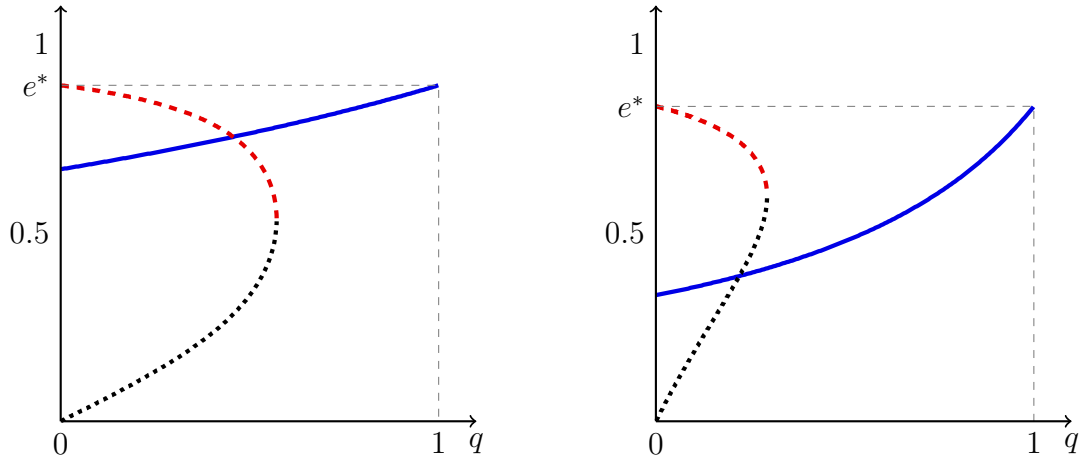
The degree of convexity of the cost function also guides the comparison between the friendly authority and the truthful one. When the cost function is mildly convex, the party prefers the truthful authority to the friendly one. The opposite is true when the cost function is highly convex. To see why, note first that the party's payoff under the friendly authority is equal to e^F when $q = 1$ and it is increasing (decreasing) in q if the cost function is mildly (highly) convex. Second, the party's payoff under the truthful authority is exactly e^* for all values of the approval threshold q (see Remark 2).

The level of effort e^* is also an upper bound for the party's payoff under the unfriendly authority. Thus, the party always rank the truthful authority above the unfriendly one.

[21] Proposition 3 implies that the party prefers $\left(e^F, \mu^F\right)$ to $\left(e^h, \mu^h\right)$ and Lemma 3 implies that the party prefers $\left(e^h, \mu^h\right)$ to $\left(e^\ell, \mu^\ell\right)$.

[22] We show this for the low-effort equilibrium under the unfriendly authority in the proof of Proposition 3. Under the friendly authority, this limit follows from rewriting equation (3) as $1 - c'(e^F(q))\left[1 - e^F(q)\right] = e^F(q)/q$. The left-hand side is bounded, so $\lim_{q\to 0} e^F(q) = 0$.

[23] The party's equilibrium payoff under the friendly authority is equal to e^F/q and converges to $1 - c'(0) > 0$ as q approaches 0. This follows from applying L'Hôpital Rule and exploiting (A-2). Instead, the party's payoff in the low-effort equilibrium converges to 0 as q approaches 0.



(a) $c(e) = \frac{1}{3}\left[\frac{3}{4}e + \frac{1}{4}\left(-\log(1-e)\right)\right]$ **(b)** $c(e) = \frac{2}{3}\left[\frac{9}{10}e + \frac{1}{10}\left(-\log(1-e)\right)\right]$

Figure 4: Party may prefer low-effort equilibrium under unfriendly authority to equilibrium under friendly authority.

Notes: The blue thick curve depicts the party's payoff under the friendly authority. The red dashed curve depicts the party's payoff in the high-effort equilibrium under the unfriendly authority. The black dotted curve depicts the party's payoff in the low-effort equilibrium under the unfriendly authority.

Finally, there are cost functions that are neither mildly convex nor highly convex. In Appendix B we provide sufficient conditions for an unambiguous ranking over authorities from the point of view of the party.

4.2 The constituency

The constituency is indifferent between any equilibrium with zero effort and the interior equilibrium under the friendly authority. Whenever the politician exerts zero effort, the constituency rejects the reform and obtains a payoff equal to the approval threshold q. In the interior equilibrium under the friendly authority, the constituency rejects the reform after receiving report $m = 0$ and obtains a payoff of q. When it instead receives report $m = 1$, the constituency is indifferent between approving the reform or rejecting it, so it also receives a payoff of q. Thus, the constituency obtains a payoff of q in the interior equilibrium under the friendly authority.

The payoff of the constituency in an interior equilibrium is larger under the unfriendly authority than under the friendly one. Under the unfriendly authority, the constituency approves the reform after receiving report $m = 1$. When instead it re-

ceives report $m = 0$, the constituency is indifferent between approving the reform or rejecting it. Thus, the constituency's equilibrium payoff is equal e^k, the payoff it would get by always approving the reform. As $e^k > q$ for both $k \in \{\ell, h\}$, the constituency prefers the unfriendly authority to the friendly one.

Finally, the constituency ranks the truthful authority over the unfriendly one (and thus also over the friendly one). Indeed, when the authority is truthful, the politician exerts a level of effort equal to e^* and the constituency observes a perfectly informative signal. Thus, the constituency obtains a payoff of $e^* + (1 - e^*)q$. Since $e^* > e^k$ for both $k \in \{\ell, h\}$, the constituency is better off under the truthful authority than under the unfriendly one.

REMARK 3. CONSTITUENCY'S PREFERENCES OVER INTERIOR EQUILIBRIA. *The constituency prefers the truthful authority to the unfriendly one and the unfriendly one to the friendly.*

The constituency's ranking over authorities is intuitive. The constituency wants to approve reforms of high quality. As such, it would like the politician to exert high effort and the reports to be informative. The truthful authority achieves both these goals. On the contrary, the friendly authority disincentives effort and can also result in the approval of low quality reforms. The unfriendly authority ranks in-between: although some high quality reforms are rejected, the reforms that are approved are always of high quality.

5 Conclusion

In modern democracies, oversight authorities play a relevant role in evaluating reform proposals. An expert authority has the technical know-how to evaluate intricate bill proposals, so it can inform the constituency about the quality of potential reforms. However, an oversight authority may be biased. It can have its own agenda, and thus it can use its expertise to steer the constituency's decisions towards its own interests. These interests may be aligned or misaligned with those of the proponents of the reforms.

We show that reforms can be approved less often under a *friendly* oversight than under an *unfriendly* one. To do so, we describe a simple environment. A politician brings forward a reform proposal. The quality of the reform depends positively on the politician's unobservable and costly effort. A constituency does not observe the

quality of the reform directly. Instead, an authority provides a report that contains information about such quality. The authority may want the reform to be approved (friendly authority) or to be rejected (unfriendly authority). The friendly authority is detrimental for the approval of the reform when two conditions hold. First, the alternative to the reform is not too attractive for the constituency. Second, incentivizing the politician's effort is not too difficult.

Our results speak to the debate on the importance and potential drawbacks of checks and balances. The alignment of interests between the proponent of a reform and the authority that evaluates it may result not only in worse reforms, but also in more frequent political gridlock.

Appendix

A Proofs

Proof of Proposition 1

We first show the auxiliary result that $e^F < q$. Suppose instead that there exists an equilibrium where the politician exerts a level of effort $e \geq q$. The authority can then guarantee the approval of the reform by choosing an uninformative reporting strategy, namely by setting $\mu_0 = \mu_1$. In such a case, the constituency approves the reform with certainty after either report. This in turn pushes the politician to choose $e = 0$, which violates the premise that $e \geq q$ (as $q > 0$).

We next obtain the authority's best response to an interior level of effort $e \in (0, q)$. Given the behavior of the constituency, standard results from the literature on persuasion imply that the authority's best response must be as follows. First, the authority reports $m = 1$ whenever the reform is of high quality. Second, the authority reports $m = 0$ with probability $\mu_0 \in (0, 1)$ when the reform is of low quality, where μ_0 solves $\frac{e}{e+(1-e)\mu_0} = q$. The authority's best response is thus the reporting strategy $\mu = \left(\frac{e(1-q)}{(1-e)q}, 1\right)$, which is well defined since $e < q$. This reporting strategy induces posterior beliefs $\pi(0) = 0$ and $\pi(1) = q$. The constituency then approves the reform after receiving report $m = 1$ and rejects it after receiving report $m = 0$.

We then solve for the politician's best response to an arbitrary reporting strategy $\mu = (\mu_0, \mu_1)$ with $\mu_1 \geq \mu_0$. The politician solves $\max_{e \in [0,1)} e\mu_1 + (1-e)\mu_0 - c(e)$, which

is a strictly concave problem (given the properties of c). The politician's best response thus satisfies $\mu_1 - \mu_0 \leq c'(e)$ with strict equality whenever $e > 0$.

The combination of the two best responses leads to equations (1) and (2) in the main text, which characterize an interior equilibrium. Moreover, equations (1) and (2) together imply that the following equality must hold in equilibrium:

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

The left-hand side of the previous expression decreases with e, while the right-hand side increases with e. Hence, the equality holds at most once. Moreover, the left hand side is larger than the right hand side at $e = 0$, while the right hand side is larger than the left hand side close to $e = 1$. Thus, the equilibrium exists and it is unique. □

Proof of Lemma 1

Equations (1) and (2) together imply that in the unique interior equilibrium the following equality must hold:

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

Because of the implicit function theorem, this implies that

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

where the inequality follows from $c'(e^F) < 1$. Hence, the equilibrium level of effort increases with q.

Note next that the equilibrium payoff of the politician is equal to

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

The derivative of the politician's payoff with respect to q is thus equal to

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

where the inequality follows from inequality (A-2) and the equality follows from equation (A-1). Hence, the politician's equilibrium payoff decreases with q.

Finally, equation (4) in the main text shows that the derivative of the party's payoff with respect to q is equal to

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

Since $d\mu_0^F/dq < 0$, the party's payoff strictly increases with q if $c'(e^F)\frac{de^F}{d\mu_0} + \left(1 - e^F\right) < 0$, and strictly decreases with 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's 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 first show the auxiliary result that $e^k > q$. Assume instead that there exists an equilibrium where the politician exerts a level of effort $e \in (0, q]$. The authority can then guarantee the rejection of the reform by adopting an uninformative reporting strategy, that is, by setting $\mu_0 = \mu_1$. In such a case the constituency rejects the reform with certainty after either report. This in turn pushes the politician to choose $e = 0$, which contradicts $e \in (0, q]$.

We next obtain the authority's best response to a level of effort $e \in (q, 1)$. Given the behavior of the constituency, standard results in the persuasion literature imply that the authority's best response must be as follows. First, the authority reports $m = 0$ when the reform is of low quality. Second, the authority reports $m = 1$ with probability $\mu_1 \in (0, 1)$ when the reform is of high quality, where μ_1 solves $\frac{e(1-\mu_1)}{e(1-\mu_1)+1-e} = q$. The authority's best response is thus the reporting strategy $\mu = \left(0, \frac{e-q}{e(1-q)}\right)$, which is well defined because $e > q$. This reporting strategy induces posterior beliefs $\pi(0) = q$ and $\pi(1) = 1$. The constituency approves the reform after receiving report $m = 1$ and rejects it after receiving report $m = 0$.

The politician's best response is identical to that under the friendly authority and satisfies $\mu_1 - \mu_0 \leq c'(e)$, with strict equality whenever $e > 0$.

The combination of the two best responses leads to equations (5) and (6) in the main text, which characterize an interior equilibrium. Equation (5) directly implies $e^k < e^*$.

The previous steps show that in any interior equilibrium (e^k, μ^k), the level of effort must satisfy $q < e^k < e^*$. Thus, whenever $q \geq e^*$ no interior equilibrium exists. We then assume that $q < e^*$ hereafter.

Next, for any $q \in (0, e^*)$ we define the auxiliary function $e \in [q, e^*] \mapsto f_q(e)$ as follows

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

The properties of the cost function imply that for every $q \in (0, e^*)$ f_q is continuous and strictly concave. Furthermore, both $f_q(q) = -c'(q) < 0$ and $f_q(e^*) = \frac{e^* - q}{e^* - qe^*} - 1 < 0$. Then, fixing q, either (i) $f_q(e) < 0$ for all $e \in (q, e^*)$, or (ii) $f_q(e) \geq 0$ for some $e \in (q, e^*)$. The existence of an equilibrium requires $f_q(e) = 0$ for some $e \in (q, e^*)$—see equation (7). In case (i), there are no interior equilibria. In case (ii), the concavity of f_q implies that exactly two solutions exists if $f_q(e) > 0$ for some $e \in (q, e^*)$ and only one solution exists otherwise.

An interior equilibrium exists if and only if $f^*(q) := \max_{e \in [q, e^*]} f_q(e) \geq 0$, where the maximum is well defined. Let $e^\dagger(q) = \arg\max_{e \in [q, e^*]} f_q(e)$. The maximizer is unique because f_q is strictly concave. The function $f^*(q)$ is continuous because of the maximum theorem and satisfies $\lim_{q \to 0} f^*(q) > 0$ and $\lim_{q \to e^*} f^*(q) = -1 < 0$. Finally, $f^*(q)$ is strictly decreasing: for any $q' > q$

$$f^*(q) = f_q\left(e^\dagger(q)\right) \geq f_q\left(e^\dagger(q')\right) > f_{q'}\left(e^\dagger(q')\right) = f^*(q'),$$

where the strict inequality holds since, for any given e, $f_{q'}(e) < f_q(e)$. Thus, there exists a unique $q^\dagger \in (0, 1)$ such that $f^*(q^\dagger) = 0$. Three cases are possible. First, if $q > q^\dagger$, then there is no interior equilibrium. Second, if $q = q^\dagger$, then there exists a unique interior equilibrium with level of effort $e^\dagger\left(q^\dagger\right)$. Third, if $q < q^\dagger$, then there exist two interior equilibria (e^ℓ, μ^ℓ) and (e^h, μ^h) with $q < e^\ell < e^\dagger(q) < e^h < e^*$. □

Proof of Lemma 2

Proposition 2 shows that for any $q < q^\dagger$ there exist exactly two interior equilibria, $\left(e^\ell, \mu^\ell\right)$ and $\left(e^h, \mu^h\right)$. The equilibrium levels of effort e^ℓ and e^h are the two roots of $f_q(e) = 0$, where the function f_q is as defined in (A-3) and $e^\ell < e^h$. So for any $q < q^\dagger$, we let $e^\ell(q)$ and $e^h(q)$ be such that $e^\ell(q) < e^h(q)$ and $f_q(e^k(q)) = 0$ for $k \in \{\ell, h\}$. First, note that $f_q(e) < f_q(e^\ell(q)) = 0$ for all $e < e^\ell(q)$. Second, note that if $q < q'$, then $f_{q'}(e) < f_q(e)$ for any e. Then, whenever $q < q' \leq q^\dagger$,

$$f_{q'}(e) < f_q(e) \leq f_q(e^\ell(q)) = 0 \qquad \forall e \in \left(0, e^\ell(q)\right]$$

Then, $e^\ell(q') > e^\ell(q)$. A similar argument shows that $e^h(q') < e^h(q)$.

In the interior equilibrium (e^k, μ^k), the politician's payoff is equal to

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

The derivative of the politician's payoff with respect to q is thus equal to

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

Hence, the politician's payoff increases with q in the low-effort equilibrium and decreases with q in the high-effort equilibrium.

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

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

Hence, once more, the party's payoff increases with q in the low-effort equilibrium and decreases with q in the high-effort equilibrium. □

Proof of Lemma 3

The party's payoff under the unfriendly authority in an interior equilibrium (e^k, μ^k), with $k \in \{\ell, h\}$, is equal to $e^k \mu_1^k$. The ranking of equilibria for the party follows from observing that $e^h > e^\ell$ and $\mu_1^h > \mu_1^\ell$, so $0 < e^\ell \mu_1^\ell < e^h \mu_1^h < 1$. The ranking for the authority is the opposite to that of the party.

Under the unfriendly authority, the constituency only approves reforms of high quality. Then, the constituency gets a payoff equal to

$$e^k\mu_1^k + (1 - e^k\mu_1^k)q = (1 - q)e^k\mu_1^k + q = e^k - q + q = e^k.$$

The constituency thus prefers the high-effort equilibrium to the low-effort one.

Finally, the politician's payoff is equal to $e^k\mu_1^k - c(e^k)$. The following inequalities show that the politician prefers the high-effort equilibrium to the low-effort one

$$e^h\mu_1^h - c(e^h) > e^\ell\mu_1^h - c(e^\ell) > e^\ell\mu_1^\ell - c(e^\ell),$$

where the first strict inequality follows from the optimality of the politician, and the second one from $\mu_1^h > \mu_1^\ell$. □

Proof of Proposition 3

Under the friendly authority, the equilibrium level of effort e^F increases with q and it is bounded above by e^*. Thus,

$$\lim_{q\to 1}\mu_0^F(q) = \lim_{q\to 1}\frac{1-q}{q}\frac{e^F(q)}{1-e^F(q)} = 0,$$

and so, by equation (1), $\lim_{q\to 1} e^F(q) = e^*$. We let $v^F(q)$ denote the party's payoff under the friendly authority:

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

Then, $\lim_{q\to 1} v^F(q) = e^*$.

In an interior equilibrium (e^k, μ^k) with $k \in \{\ell, h\}$ under the unfriendly authority, the party obtains a payoff $v^k(q) = e^k\mu_1^k$. Equation (7) is necessary and sufficient for an interior equilibrium under the unfriendly authority. We rewrite it as

$$c'\left(e^k(q)\right)e^k(q)(1-q) = e^k(q) - q.$$

Then $\lim_{q\to 0} e^\ell(q) = 0$ and $\lim_{q\to 0} e^h(q) = e^*$. This, together with equation (5) implies that $\lim_{q\to 0}\mu_1^\ell(q) = c'(0)$ and $\lim_{q\to 0}\mu_1^h(q) = 1$. Hence, $\lim_{q\to 0} v^\ell(q) = 0$ and $\lim_{q\to 0} v^h(q) = e^*$.

When the cost function is highly convex, the party ranks the friendly authority over the unfriendly one. To see why, note that the party's payoff under the friendly authority, $v^F(q)$, strictly decreases with q when the cost function is highly convex (see Lemma 1 and the definition of highly convex cost functions in terms of the Arrow-Pratt coefficient). Since $\lim_{q\to 1} v^F(q) = e^*$, the party's payoff is bounded below by e^*. On the contrary, the party's payoff $v^h(q)$ under the unfriendly authority is bounded above by e^*. Indeed, $v^h(q)$ strictly decreases with q (see Lemma 2) and $\lim_{q\to 0} v^h(q) = e^*$.

When the cost function is mildly convex, the ranking over authorities depends on the value of the approval probability q. To see why, note that $v^F(q)$ strictly increases with q when the cost function is mildly convex (see Lemma 1 and the definition of mildly convex cost functions in terms of the Arrow-Pratt coefficient). Moreover, $\lim_{q\to 1} v^F(q) = e^*$, so $v^F(q) < e^*$ for all $q \in (0,1)$. Instead, $\lim_{q\to 0} v^h(q) = e^*$. Then, $v^h(q) > v^F(q)$ for q close to zero. Finally, under the unfriendly authority, the party's payoff strictly decreases with q (see Lemma 2). Thus, the party ranks the high-effort equilibrium under the unfriendly authority over the interior equilibrium under the friendly one if and only if q is below a threshold $\bar{q}$. □

B Sufficient conditions to rank equilibria for arbitrary cost functions

Proposition 3 characterizes the party's preferred equilibrium and authority when the cost function is either mildly convex or highly convex. Some cost functions are neither mildly convex, nor highly convex. For arbitrary cost functions satisfying the assumptions in Section 2, Proposition B.1 and Proposition B.2 below provide sufficient conditions for the existence of an unambiguous ranking over authorities.

Proposition B.1 defines a lower bound for the equilibrium level of effort under the friendly authority and an upper bound for the equilibrium level of effort under the unfriendly authority. These are z_1 and z_2, respectively. These levels of effort provide bounds for the party's payoff under the two authorities.

PROPOSITION B.1. *Fix $q \leq q^\dagger$ and let $z_1 = \frac{q[1-c'(q)]}{1-qc'(q)}$ 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 under the unfriendly authority than in the interior equilibrium under the friendly authority.

Proof. The party's equilibrium payoff under the friendly authority is $e^F\mu_1^F + (1 - e^F)\mu_0^F = e^F/q$. By equation (3), we have:

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

The equilibrium value of μ_0^F is such 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). Using this expression into equation (B-1), we get

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

Note that $z_1 \in (0, q)$. Define the function $v(x; q) = (1 - x)/(1 - qx)$ and observe that $v(\cdot; q)$ is decreasing for every q. If $e^F < z_1$, we 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 equation (B-1), hence Proposition 1. We conclude that $e^F \geq z_1$. This implies that, fixing q, the party's payoff in the interior equilibrium under the friendly authority is bounded above by $\overline{v}^F(q) \equiv [1 - c'(z_1)]/[1 - qc'(z_1)]$.

Consider the unfriendly authority and pick an arbitrary interior equilibrium (e^k, μ^k). The party's equilibrium is $e^k\mu_1^k = e^k c'(e^k) = (e^k - q)/(1 - q)$. Note that $z_2 > q$. If $e^k \in (q, z_2)$, we have

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

This contradicts $e^k\mu_1^k = (e^k - q)/(1 - q)$. We conclude $e^k \geq z_2$. The party's payoff is thus bounded below by $\underline{v}^N(q) \equiv z_2 c'(z_2)$.

The party is better off under the unfriendly authority than under the friendly one if $\underline{v}^N(q) > \overline{v}^F(q)$. □

As in the previous result, also Proposition B.2 defines bounds on the equilibrium levels of effort under the two authorities. Differently from Proposition B.1, Proposition B.2 uses the bounds to show that the party's payoff under the friendly is higher than 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 under the friendly authority than in any interior equilibrium under the unfriendly authority.*

Proof. Let $\chi \in (0,1)$ be the solution of $\chi c'(\chi) = 1$. Since c is increasing and $c'(1) > 1$, this is well defined. By construction and by the convexity of c: $c'(1) > c'(\chi) = 1/\chi > 1$.

Pick an interior equilibrium under the unfriendly authority, (e^k, μ^k). If $e^k \geq \chi$, equation (5) implies that the party's equilibrium payoff is $e^k \mu_1^k = e^k c'(e^k) > 1$. This establishes a contradiction because the approval probability is bounded above by 1. Hence, we must have $e^k < \chi$. We conclude that the party's equilibrium payoff $e^k \mu_1^k = (e^k - q)/(1-q)$ is bounded above by $\overline{v}^N \equiv (\chi - q)/(1-q)$.

Now consider the interior equilibrium under the friendly authority. Recall that $e^F < q$. Following the same steps of the proof of Proposition B.1, we can conclude that the party's payoff are bounded below by $\underline{v}^F(q) \equiv [1 - c'(q)]/[1 - qc'(q)]$.

The party prefers the interior equilibrium under the friendly authority to any of the equilibria under the unfriendly authority if $\underline{v}^F(q) > \overline{v}^N(q)$ or equivalently if $z_3 > \chi$. 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, *113* (3), 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, "Electoral Accountability: Recent Theoretical and Empirical Work," *Annual Review of Political Science*, 2012, *15*, 183–201.

Aumann, Robert J, Michael Maschler, and Richard E Stearns, *Repeated Games with Incomplete Information*, MIT Press, 1995.

Austen-Smith, David, 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.

Bardhi, Arjada and Yingni Guo, "Modes of Persuasion toward Unanimous Consent," *Theoretical Economics*, 2018, *13* (3), 1111–1149.

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.

Chan, Jimmy, Seher Gupta, Fei Li, and Yun Wang, "Pivotal Persuasion," *Journal of Economic Theory*, 2019, *180*, 178–202.

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.

Finkel, Eli J., Christopher A. Bail, Mina Cikara, Peter H. Ditto, Shanto Iyengar, Samara Klar, Lilliana Mason, Mary C. McGrath, Brendan Nyhan, David G. Rand, Linda J. Skitka, Joshua A. Tucker, Jay J. Van Bavel, Cynthia S. Wang, and James N. Druckman, "Political Sectarianism in America," *Science*, 2020, *370* (6516), 533–536.

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.

Heese, Carl and Stephan Lauermann, "Persuasion and Information Aggregation in Large Elections," *Working paper*, 2020.

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.

Mas-Colell, Andreu, Michael D Whinston, and Jerry R Green, *Microeconomic Theory*, Oxford University Press, 1995.

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.

Taylor, Curtis R. and Huseyin Yildirim, "Subjective Performance and the Value of Blind Evaluation," *The Review of Economic Studies*, 02 2011, *78* (2), 762–794.

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.