Politicians, Bureaucrats, and the Battle for Credit

January 2, 2022

Abstract

Politicians may want to claim credit from bureaucrats when things go well and deflect blame when outcomes go awry. While many assume that this is a function of accountability concerns, another possibility is that principals are attempting to incentivize their largely insulated agents to produce better policy. We capture these incentives with a series of principal-agent models where bureaucrats have reputation concerns. In these models, the bureaucrat exerts effort to produce good policies and the politician comments on the bureaucrat's skill level to an interested audience. Without dynamic considerations, the politician can effectively induce bureaucratic effort by commenting about the bureaucrat. However, with multiple periods, the expectation of blaming and crediting may induce more bureaucratic effort today at the cost of weakening the bureaucrat's incentives tomorrow. Furthermore, if the politician values her reputation more than producing good policies she will find it harder to credibly reveal information about the bureaucrat's skill level and responsibility and, hence, to motivate him. Nonetheless, our results imply that blaming and crediting may be socially beneficial.

1 Introduction

Policy outcomes go wrong. In response, leaders sometimes are quick to attribute blame to underlings, while in other instances they stay quiet or accept responsibility. In the political realm a recent high profile example of the former was Donald Trump's condemnation of the Center for Disease Control (CDC) – and his predecessor, despite being in office for more than three years – in the midst of the coronavirus pandemic.[1] For an example of the latter, consider Barack Obama's public acceptance of responsibility when a December 2009 effort to blow up an airplane exposed weaknesses in American counter-terrorism efforts, after which in January 2010 he opined that, "I am less interested in passing blame than I am in learning from and correcting these mistakes to make us safer, for ultimately, the buck stops with me (...) when the system fails it is my responsibility."[2]

More generally, the phenomenon of leaders sometimes blaming subordinates spans political and economic environments.[3] A routine if rather jaundiced view of realizations of blaming is that leaders are engaging in self-aggrandizing behavior in a very base way. They appeal to an interested, if perhaps naive, audience by deflecting blame when they can credibly do so, which leaves open the question of why leaders in other instances abstain from blaming and even accept responsibility. Alternatively, it is conventionally thought there is a considerable downside associated with those at the top of a hierarchy blaming. Overstating credit and pointing blame is seen as the mark of a bad leader, as in Beer et al. (2011), or even a destructive leader, as expounded by Craig and Kaiser (2012), with a common aphorism

[1]For instance, on March 13, 2020 Trump tweeted "For decades the @CDCgov looked at, and studied, its testing system, but did nothing about it. It would always be inadequate and slow for a large scale pandemic, but a pandemic would never happen, they hoped. President Obama made changes that only complicated things further." See https://twitter.com/realDonaldTrump/status/1238410044263333894.

[2]See this report by *Politico* for more https://tinyurl.com/politico-blame.

[3]We study blaming *and* crediting but may only mention blaming for ease of exposition.

being that “Great [Good] Leaders take the Blame and Pass on the Credit.”[4] From this perspective, blaming's routine observation might be surprising, suggesting that such leaders just cannot help themselves from pointing fingers or we have missed some underappreciated aspects of blaming. Given the latter possibility, in this paper we examine aspects of blaming that are relatively understudied. We do so by developing a series of formal models involving strategic interactions between a politician (*she*), bureaucrat (*he*), and non-strategic observer – the public, voters, organized interests, or media. In our proposed theory, delegation is fixed but the principal reveals information about the bureaucrat to the public which in turn can affect the agent's effort. We analyze this in static and dynamic environments to examine the effort–inducing effects of blaming. Importantly, we view blaming or crediting purely as communication between the strategic principal and the non-strategic observer.

In our setup, the principal is typically an elected representative with a greater ability to communicate to the outside world than the agent. However, while the agent is the expert delivering the policies that the principal wishes to execute, he is also often insulated from her. More broadly, three elements of well-established public bureaucracies would seem fundamental for understanding the structure in which bureaucrats operate and how politician behavior might be conditioned: (1) bureaucrats are a diverse lot and may vary greatly in their abilities;[5] (2) removing most public sector bureaucrats is generally difficult, so shirking by reducing effort can be a viable option (Horn 1995);[6] and (3) bureaucratic effort and successful policy choices should be positively related (Carpenter 2001). Jointly, these features suggest that, in contrast to accountability avoidance, blaming via a politician's raised voice may serve a socially beneficial purpose too: Getting bureaucrats to work harder, despite

[4]See, e.g., https://hbr.org/2011/10/why-good-leaders-pass-the-cred.

[5]For evidence in the U.S., see, e.g., the 2003 *Volcker Commission Report*.

[6]For example, in the U.S., the Merit Systems Protection Board serves to protect civil servant job security. Additionally, public sector wage compression – a common issue – may further incentivize shirking.

most not being directly incentivized due to institutional reasons and yet possessing career concerns. Notwithstanding intuitive reasons to believe that it is important, as the principal's strategic use of blaming given delegation is not well-studied, we aim to isolate the effects of politician blaming on agency effort.

Our analysis begins with a one-period model where the politician is only interested in policy outcomes and is strategically interacting with a bureaucrat with reputation concerns. Here we show that the politician can induce bureaucratic effort by using blaming as a communication tool.[7] Given that politician-bureaucrat interactions are usually dynamic in nature, we add a second period to this game to account for the politician's consideration of the long-run effects of affecting the bureaucrat's reputation. Afterwards, we incorporate another layer where the politician also cares about his own reputation, balancing it with the desire for successful policy outcomes.

We find that while blaming can be effective if the principal and agent are playing a one-shot game, this is not necessarily the case if there is a second period. Initially the expectation of blaming and crediting may induce greater bureaucratic effort, and a higher likelihood of positive policy outcomes. That said, the bureaucrat's incentives to exert effort are dampened tomorrow once the politician shifts blame or gives credit. Put differently, while the politician benefits if the public learns about the bureaucrat's type, this learning effect hampers the politician's future effort extraction possibilities. Further, blaming's effectiveness is conditioned by the extent to which the politician weighs her reputation over outcomes; the more she values her reputation over producing good policies the harder she will find credibly revealing information about the bureaucrat's skill level and responsibility. Nonetheless,

[7]We start with bureaucrats caring about reputation to highlight how bureaucratic reputation plays out in a very stark environment, recognizing that assuming politicians value policy but not reputation is unintuitive in most instances. That said, these assumptions are not too restrictive and based on empirical patterns. In addition to simplifying the analysis, the assumptions help us highlight our key mechanism of crediting and blaming. We do, however, extend our model including the politician's reputation concerns.

our analysis indicates that, within limits, politician blaming and crediting can actually be socially beneficial.

We proceed in multiple steps in developing this largely novel perspective on blaming. After commenting on the relevant literature, we study the one-period model where the threat of blaming or the reward of crediting incentivizes the agent to exert more effort by altering the expected payoffs of policy success and/or failure. We then extend our analysis to a two-period setting. Subsequently, we include politician reputation concerns. We conclude by discussing the implications of our results.

2 Literature Review

By examining blaming as communication between a politician and an outside actor who evaluates the bureaucrat, we contribute to two different streams of literature and provide directions for future work in the experimental strand of the literature. That said, studying delegation is a matter of long-standing interest due to its importance and pervasive nature in hierarchies of all kinds.[8] However, analyses of the intersection of blame shifting and delegation are less developed. Specifically, much discussion in political science of blame via delegation starts with Weaver (1986),[9] which purports to apply prospect theory (Kahneman and Tversky 1979) to voters. We differ from this literature by studying an often overlooked

[8]Holmström (1980) is the canonical formal analysis in this area. Applications to the devolution of effort by elected representatives to bureaucrats are also widespread. While Gailmard and Patty (2012) provides an excellent review of this area, suffice it to say that this literature spans an array of topics, including delegated discretion (Epstein and O'Halloran 1994); principal-agent preference alignment (Bendor and Meirowitz 2004); the institutional context of the bureaucracy (Volden 2002); and delegation's effects on implementation of policies without being held accountable (Fox and Jordan 2011).

[9]See also Pierson (1994), which emphasizes blaming bureaucrats as a means for political avoidance of voter unhappiness with social welfare cutbacks — for a more contemporary discussion see Hood (2011), which does acknowledge that the risk of blame also may serve to keep people honest and compliant.

aspect — the effects of blaming or crediting on bureaucratic effort.

Blame shifting theories of delegation in the literature focus on the implications of the decision to delegate and its welfare consequences (Bartling and Fischbacher 2012), the principal's incentives to pander to the public (Ely and Välimäki 2003), uncertainty regarding who the decision maker was — after failure — that affects reelection (Almendares 2012), strategic hiring of agents who can be easily blamed *ex-ante* (Glazer and Segendorff 2005), and differences in the allocation of authority by a principal with reputation concerns (Tamada and Tsai 2018). By contrast, we depart from most existing work, and produce many of our key theoretical insights, by modeling and analyzing the cost and benefit of being blamed for policy failure in both static and dynamic settings. The closest theoretical work to ours is by Pei (2018), who views delegation as a means for inducing noisy public signals about the actors' types.

Additionally, by explicitly accounting for the bureaucrat's reputation concerns we contribute to the stream of literature that has focused on the importance of bureaucratic reputation. Despite their insulation, empirical scholars of administrative behavior in the developed world routinely conclude that bureaucrats care a good deal about their reputation with relevant audiences;[10] typically this is reflected by given bureaucrats desires to seek alternative careers outside the bureaucracy, be known for their expertise, move to a favorable location and/or be promoted within the bureaucracy (Alesina and Tabellini 2005; Iyer and Mani 2012; Xu 2018; Khan, Khwaja and Olken 2019; Bertrand et al. 2020).[11] For our purposes, such concerns for a

[10]For thorough description of these reputation concerns, see Carpenter (2001, 2010) and Carpenter and Krause (2012); see also Boon and Salomonsen (2020) and Bertelli and Busuioc (2021) for detailed reviews of the extant literature.

[11]In some places these incentives for developing a favorable reputation have been strongly institutionalized. For example, in Japan, amakudari or "descent from heaven," is the institutionalized practice of Japanese senior bureaucrats retiring to high-profile positions in the private and public sectors. For such bureaucrats, developing a favorable reputation has been crucial.

bureaucrat can be considered a function of one or more of the aforementioned factors and/or just the desire to do well in the job, raise budgetary allocations (in the spirit of performance–based budgeting) and be more favorably positioned should politicians and outside interests attempt to exert pressure or exercise oversight.

Finally, a fledgling empirical literature has tried to tackle the issue of blaming and crediting experimentally. For instance, Bartling and Fischbacher (2012), analyzing dictator games, shows that blame can be shifted successfully, providing a strong motive to delegate. Marvel (2014) uses a survey experiment to study the aftermath of the Boston Marathon bombings, and demonstrates that administrators — driven by their greater credibility compared to elected officials — are especially effective in influencing citizens opinions on whom to blame. Finally, Ruder (2014, 2015) shows that delegation to agencies cannot provide unconditional protection from escaping public blame. However, most of this literature does not focus on the communication aspects of blaming, notably how politicians may employ the threat of blame to induce bureaucratic effort and elicit better outcomes. Therefore, given the nature of bureaucrats and the structure of bureaucracy, communication effects of blaming — particularly in a dynamic context with endogenous costs and benefits — are potentially of considerable relevance and deserving of further empirical, particularly experimental, investigation.

As evidenced in this section, most analyses of bureaucratic blaming have highlighted politicians looking to avoid responsibility. Typically, normative assessments of such behavior are negative, some emphasizing accountability being undermined and others highlighting negative implications for both bureaucratic behavior and quality. Little of this literature focuses on the communication aspects of blaming, notably how politicians may employ the threat of blame to induce bureaucratic effort and elicit better outcomes. Our analysis fills in this notable gap and, in doing so, provides implications that can be tested experimentally.

3 Model

Our series of models are designed to resolve three main tensions. The first involves understanding how a politician influencing the reputation of a bureaucrat affects the latter's incentives to exert effort. Another entails evaluating how a politician influencing the reputation and incentives of a bureaucrat today affects her ability to do so tomorrow. The third is ascertaining how a politician with her own reputation concerns can manage the bureaucrat's reputation and subsequent effort incentives. We assess the first tension by presenting a baseline setup where dynamic tensions and politician reputation concerns are absent and then add complexities to investigate each additional tension.

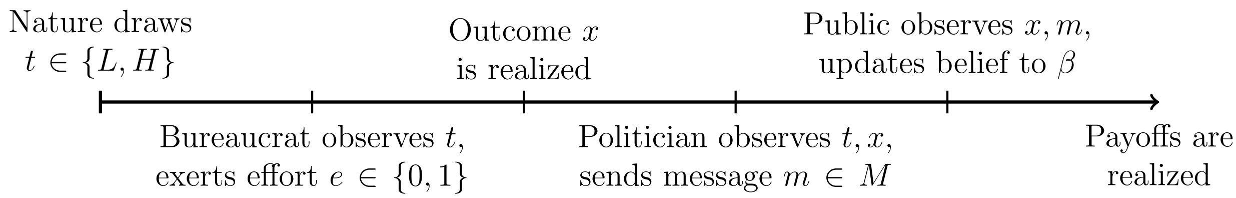
3.1 Setup

We consider a model with a politician, bureaucrat, and non-strategic public that observes the politician-bureaucrat interaction. In this baseline model we assume that there is uncertainty about the bureaucrat's talent, $t \in \{L, H\}$, which is either low or high. After observing his talent, the bureaucrat decides whether to exert effort or shirk. Based on the bureaucrat's effort choice and his talent, an outcome results. If the bureaucrat shirks, the policy outcome always fails, while if the bureaucrat exerts effort, the policy succeeds with probability p_L or p_H depending on the bureaucrat's talent t. We assume that $0 < p_L < p_H < 1$ so that, conditional on effort, highly talented bureaucrats are more likely to produce good outcomes. After the outcome, the politician sends a public message based on the bureaucrat's talent and whether the outcome was good or not. Finally, the game ends with the public observing the outcome and the politician's message, updating its beliefs, and payoffs are realized.

Figure 1 illustrates the game's timing, which is summarized as follows:

1. Nature draws the bureaucrat's talent t, which is H with probability $\beta_0 \in (0, 1)$ and L with complementary probability, where β_0 is the bureaucrat's *prior reputation*.
2. The bureaucrat observes his talent and chooses whether to exert effort, $e \in \{0, 1\}$. If

Figure 1: Model Timing



the bureaucrat shirks ($e = 0$) the policy fails ($x = f$), and if he exerts effort ($e = 1$) there is a success ($x = s$) with probability p_t and a failure otherwise.

3. Nature generates a policy success ($x = s$) or failure ($x = f$).
4. The politician observes the bureaucrat's talent and the policy outcome (but not effort) and sends a cheap talk message $m \in \{m^-, m^0, m^+\}$. Message m^0 is interpreted as silence, m^- as negative and m^+ as positive about the bureaucrat's talent.
5. The public observes the policy outcome and the politician's message, and subsequently updates its beliefs about the bureaucrat's talent.
6. The game ends and payoffs are realized.

Given outcome x and message m, the posterior probability that the bureaucrat is of the high type is denoted $\beta_{x,m} \in [0, 1]$. Also, let the cost of exerting effort equal $c < p_L$. We assume the bureaucrat's payoff is

$$u_B(\beta, e) = \mathbb{I}_{\{\beta_{x,m} > 1/2\}} - ec. \tag{1}$$

The first term in the equation means that the bureaucrat wants his reputation to exceed the threshold of $\frac{1}{2}$, whereas the second indicates that effort is costly. In the appendix we define the payoff that the bureaucrat obtains at the threshold $\frac{1}{2}$ as either 0 or 1, to ensure equilibrium existence under some parameter values. We label reputations exceeding the threshold as *good reputations* and those that do not as *bad reputations*. By contrast, the politician's payoff simply depends on the final outcome — 1 with policy success and 0 otherwise.

3.2 Equilibrium

Our major concern involves perfect Bayesian equilibria that are ex-ante optimal for the politician. A strategy profile consists of (i) a mapping from the bureaucrat's talent to a mixed effort level $e : \{L, H\} \to \Delta\{0, 1\}$, and (ii) a messaging strategy for the politician, which is a mapping from the bureaucrat's talent and the outcome to a mixture over the set of messages, i.e., $\pi : \{L, H\} \times \{s, f\} \to \Delta\{m^-, m^0, m^+\}$, where $\Delta(\cdot)$ is the set of all probability distributions of a set. We require that these strategies are sequentially rational given the beliefs about the bureaucrat's talent, which in turn are determined by Bayes' rule. Moreover, the public's belief about the bureaucrat's talent is a mapping from the set of outcomes and messages to a probability $\beta : \{s, f\} \times \{m^-, m^0, m^+\} \to [0, 1]$. Finally, a *politician-optimal* equilibrium maximizes the weighted average of the probability of success of both types, subject to sequential rationality and Bayesian updating wherever possible. That is, we seek PBE that have the largest values of $(1 - \beta_0)e_L p_L + \beta_0 e_H p_H$, which is the ex-ante probability of a good policy outcome. The selection of sender-optimal equilibria is in line with a related literature of cheap talk models (Lipnowski and Ravid 2020).

3.3 Discussion

We want to know to what extent (1) the bureaucrat can be incentivized to exert effort, and (2) the politician's signaling strategy induces such effort. Before describing the results, three key aspects of the model require explication.

First, it is assumed is that the politician (but not the public) knows the bureaucrat's talent. Several possibilities can motivate this assumption, including the politician observing past bureaucrat achievements that the public does not witness or her working closely with the bureaucrat on related projects. Analogous to people inside an organization knowing each other better than those outside, such occurrences make it reasonable to assume that the politician is better informed about the bureaucrat relative to the public.

Second, the meaning of the politician's message is derived endogenously. This meaning is an interpretation of the message relative to saying nothing. If saying nothing following a policy success leads to a reputation β, then the politician can provide further information through her messaging strategy by inducing reputations β' and β'' such that $\beta' < \beta < \beta''$. As long as $\beta_{x,m} < \frac{1}{2}$, the politician's message hurts the bureaucrat's reputation, and vice versa if $\beta_{x,m} > \frac{1}{2}$. Given this endogenous derivation of meaning, our main interest lies in understanding the reputation that the politician optimally induces with her messages to extract the maximum aggregate effort.

Third, reputation concerns take a threshold form. That is, the bureaucrat's payoff is higher if his reputation exceeds a certain threshold compared to when it does not. For the sake of exposition, we assume that this threshold is $\frac{1}{2}$. However, this is without loss of much generality; a different threshold would only shift equilibrium conditions without substantively changing the paper's core message. A more salient generalization would be to have reputation payoffs taking a different form. For example, one could consider reputation payoffs $V(\beta)$ that are an increasing function of β. Appendix D.1 provides some further calculations and discussion of this version of the model. Importantly, if the reputation payoffs are linear, e.g., $V(\beta) = \beta$, then the politician's management of the bureaucrat's reputation with messages cannot increase effort levels. To incentivize the low type to exert effort, reputation payoffs $V(\beta)$ to the bureaucrat must take a convex form, so that there is more to gain from achieving a good reputation over obtaining a bad one.

4 Baseline Results

We begin by describing equilibria where the politician is restricted to remaining silent, after which we compare and contrast equilibria when the politician actively sends messages to influence the bureaucrat's reputation. Regardless, an important aspect of equilibria is how the bureaucrat's reputation is shaped by the outcome and the politician's message.

4.1 Benchmark: Effort under Politician Silence

In the benchmark case, the politician always remains silent and the public only updates its beliefs based on the policy outcome. Bureaucratic incentives to exert effort are completely driven by avoiding a bad reputation and gaining a good one. This section now shows how an outcome that is sufficiently informative about the bureaucrat's talent suffices to induce bureaucratic effort. We study this formal condition by asking for which initial reputations β_0 the bureaucrat exerts effort regardless of his talent.

If both bureaucrat types are expected to exert effort, the public updates upward after a success (β_{s,m^0}) and downward following a failure (β_{f,m^0}).[12] The bureaucrat exerts effort if and only if a failure pushes the public's belief below the reputation threshold of $\frac{1}{2}$ and a success pushes it above the threshold, i.e., if the condition $\beta_{f,m^0} < \frac{1}{2} < \beta_{s,m^0}$ holds, where one inequality may be weak. Given that the cost of effort is sufficiently low, both types are willing to exert effort under the above condition. Alternatively, if this condition does not hold, both types shirk in equilibrium. Proposition 1 describes equilibrium effort levels as a function of the bureaucrat's initial reputation β_0. In other words, the bureaucrat, no matter his type, exerts effort if and only if his initial reputation falls in an intermediate range so that success and failure are sufficiently informative to provide a reputation incentive.

Proposition 1. *In politician-optimal equilibria with politician silence, both bureaucrat types exert effort for intermediate initial reputations* $\beta_0 \in \left[\frac{p_L}{p_H+p_L}, \frac{1-p_L}{2-p_H-p_L}\right]$ *and shirk otherwise.*

4.2 Effort with Politician Signaling

Active signaling by the politician may be fruitful when both types of bureaucrat shirk. The politician, knowing the bureaucrat's talent, can incentivize the bureaucrat. If the bureaucrat's initial reputation is too high so that he shirks, then the politician can punish failure by

[12] Formally, given that both types are expected to exert effort, the belief following a success is $\beta_{s,m^0} = \frac{p_H\beta_0}{p_H\beta_0+p_L(1-\beta_0)}$ and following a failure is $\beta_{f,m^0} = \frac{(1-p_H)\beta_0}{(1-p_H)\beta_0+(1-p_L)(1-\beta_0)}$.

strategically revealing information. More precisely, the politician can use a messaging strategy so that one message reveals more negative information about the bureaucrat's talent than the other. In that case, if the policy fails the bureaucrat would receive a lower payoff. Similarly, if the bureaucrat's initial reputation is too low to incentivize effort, then the politician can reward success by strategically revealing information and sometimes sending a message that induces a reputation that exceeds the threshold.

We focus on equilibria where both types of the bureaucrat exert effort. There are two cases, where either the bureaucrat starts off with a reputation above or below the threshold. When the bureaucrat initially has a good reputation and an insufficiently informative failure occurs, a negative message may push the bureaucrat's reputation below the threshold. This occurs when the politician is more likely to send message m^- after a failure following a low type than a high type. The larger is the relative likelihood of m^- conditional on the low type than the high type, the more informative this is about the bureaucrat's type. Specifically, for m^- to lead to a bad reputation, we require that the belief following a failure and message m^- satisfies $\beta_{f,m^-} \leq \frac{1}{2}$, so that the bureaucrat's reputation suffers. For example, if m^- were only to be sent after the low type fails, then the bureaucrat is guaranteed to have a bad reputation following a failure and message m^-.

In the other case, the bureaucrat starts off with a very low reputation β_0, so that a success on its own does not lead to a good reputation. Then, following a success, the politician may send message m^+ more often if the bureaucrat is of the high type. This pushes the bureaucrat's reputation upward, and the more often m^+ is sent following the high type than the low type, the higher this becomes. In this case, the relevant constraint must be that $\beta_{s,m^+} \geq \frac{1}{2}$ for message m^+ to induce a good reputation following a success.

Besides the condition on the politician's messages leading to a bad or good reputation, these messages must satisfy other conditions as well. That is, after a failure, message m^- must be sent sufficiently often for both types and, similarly, after a success, message m^+ must be sent often enough for both types. In general, for a bureaucrat of type t to be willing to exert effort, we require the following inequality to be met, as long as m^+ generates a good reputation and m^- a bad one after both successes and failures. After comparing the payoff of exerting effort and shirking, a type t bureaucrat exerts effort if:

$$\pi(m^+|s,t) - \pi(m^+|f,t) \geq \frac{c}{p_t},$$

where $\pi(m^+|x,t)$ is the probability that the politician sends message m^+ given an outcome $x \in \{f, s\}$ and type $t \in \{L, H\}$. This means that incentives to exert effort are stronger (i) the more likely a success leads to a good reputation, and (ii) the *less* likely a failure leads to a good reputation.

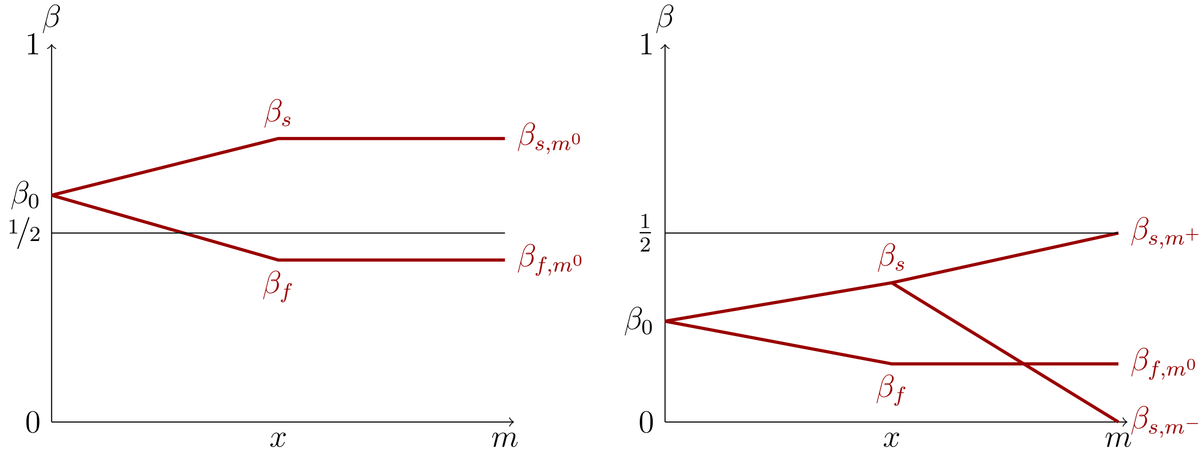
Figure 2 illustrates how the politician may use messages to induce more effort in cases where silence would yield an equilibrium where the bureaucrat shirks. The left panel shows how success and failure are already sufficiently informative to induce effort from the bureaucrat, as a success pushes his reputation above the threshold and a failure below. The right panel, however, shows that this no longer is true. The issue is that success does not push the bureaucrat's reputation above the threshold. Hence, the politician needs to send messages so that success is actually rewarded with a good reputation. The figure illustrates an example of beliefs β_{s,m^+} and β_{s,m^-} that could achieve this.

The politician's signaling strategy cannot guarantee that both bureaucrat types exert effort. The proposition below also describes equilibria where the low type mixes while the high type exerts effort, and where both types shirk. The politician can induce more effort by providing an additional reward for effort after success, and an additional cost after failure.[13]

Proposition 2. *In politician-optimal equilibria with active politician signaling, effort levels for both bureaucrat types are as follows:*

[13]We note that politician-optimal equilibria are not unique, as different signaling strategies may generate the same effort levels in equilibrium.

Figure 2: Blaming, Crediting, and Induced Reputations

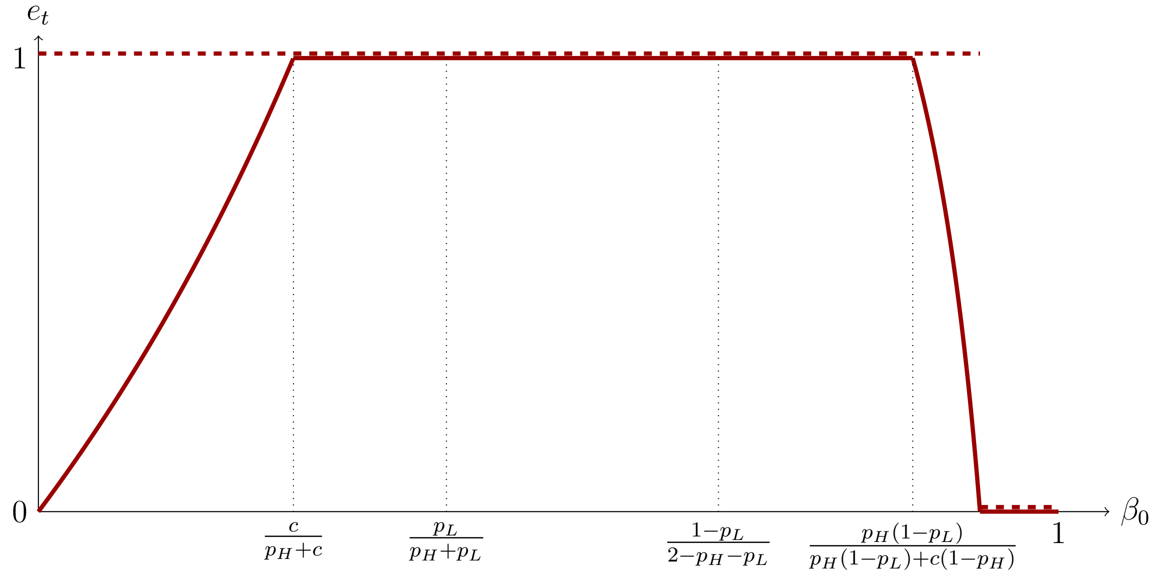


Notes: The x-axis denotes the two stages of belief-updating, where first the public updates based on the outcome x, and then based on the message m. The y-axis illustrates the bureaucrat's reputation β. On the left, failure and success are sufficiently informative to incentivize both types of the bureaucrat. On the right, success does not lead to a good reputation if both types exert effort, so P sends m^- and m^+ following a success to induce a bad and good reputation respectively.

- *if* $\beta_0 \in \left(0, \frac{c}{p_H+c}\right)$, *the low type mixes* $e_L^* = \frac{\beta_0 p_H}{(1-\beta_0)c}$ *and the high type exerts effort,*
- *if* $\beta_0 \in \left[\frac{c}{p_H+c}, \frac{p_H(1-p_L)}{p_H(1-p_L)+c(1-p_H)}\right]$, *both types exert effort,*
- *if* $\beta_0 \in \left(\frac{p_H(1-p_L)}{p_H(1-p_L)+c(1-p_H)}, \frac{p_H}{p_H+c-p_Hc}\right)$, *the low type mixes* $e_L^* = \frac{p_H-\beta_0(p_H+c-p_Hc)}{(1-\beta_0)p_Hp_L}$ *and the high type exerts effort,*
- *if* $\beta_0 \in \left[\frac{p_H}{p_H+c-p_Hc}, 1\right)$, *both types shirk.*

Figure 3 displays how effort levels differ for each initial reputation β_0. The main takeaway is that a politician forced to remain passive is less able to extract effort from the bureaucrat. The politician's ability to successfully incentivize the bureaucrat through signaling depends crucially on the bureaucrat's initial reputation and the different success rates of both types in producing good outcomes. Comparing the equilibria of Propositions 1 and 2 shows that for initial reputations close to the threshold of $\frac{1}{2}$ there is no benefit of signaling. If the bureaucrat has a low reputation, the politician rewards success with a good reputation with message m^+ to ensure that the bureaucrat's benefit of exerting effort increases. Analogously, for high reputations, the issue is that a failure is insufficiently informative about the bureaucrat's

Figure 3: Equilibrium Effort as a Function of the Bureaucrat's Reputation



Note: The solid line indicates the effort level of the low type (e_L), and the dashed line of the high type (e_H), assuming that $c = \frac{1}{4}, p_L = \frac{1}{2}$, and $p_H = \frac{3}{4}$. The cut-points on the x-axis indicate the relevant equilibrium constraints on the prior β_0 from Propositions 1 and 2.

talent. Hence, the politician sometimes sends a negative message about the bureaucrat following a failure. When the bureaucrat's reputation is too high, neither type can be incentivized and the bureaucrat shirks.

5 Extension: Dynamic Reputation Management

We now extend the model to allow for a repeated interaction. That is, the bureaucrat and politician engage in effort and signaling over two periods. The main goal of this extension is evaluating how the importance of future effort extraction affects the politician's ability to incentivize the bureaucrat in the present. In particular, the bureaucrat's reputation is updated after the first period, depending on the policy outcome and politician's message. Based on this bureaucrat's reputation, it may be more or less difficult for the politician to get the bureaucrat to exert second period effort. At the same time, there is a strategic effect that disciplines the politician's signaling behavior. The reason is that if the politician has two messages at her disposal that she wishes to send in equilibrium, both messages must induce the same future payoff.

5.1 Setup

We extend the baseline model to a two-period game. Denote first– and second–period effort choices as e_1 and e_2, and first– and second–period beliefs as β_1 and β_2. The bureaucrat obtains his payoff at the end of each period and earns

$$u_B(\beta, e) = \mathbb{I}_{\{\beta > \frac{1}{2}\}} - ec,$$

where $e \in \{e_1, e_2\}$ and $\beta \in \{\beta_1, \beta_2\}$ for period 1 and 2. Analogous to the one–period model, the politician solely cares about policy successes, i.e.,

$$u_P(x_1, x_2) = x_1 + x_2,$$

where x_1 and x_2 denote policy success or failure in the first and second period respectively. The politician obtains his payoff at the end of the game. As before, we study equilibria that are optimal for the politician and maximize the probability of success, weighted equally over the two periods.

5.2 Analysis

Equilibrium behavior in the second period follows directly from Proposition 2. That is, for intermediate reputations β_1, both types are expected to exert effort, while if the bureaucrat's reputation is too low or high, then either the low type mixes and the high type exerts effort, or both types shirk. Denote $e_2(\beta_1)$ as the effort choices of both types in the second period as a function of the reputation β_1 that they start off with in the second period.

As before, it is useful to take an intermediate step prior to presenting the full analysis — specifically, investigating what happens if the politician remains silent in the first period while behaving as in Proposition 2 in the second. From Proposition 1 we know that the bureaucrat exerts effort if and only if $\beta_0 \in \left[\frac{p_L}{p_H+p_L}, \frac{1-p_L}{2-p_H-p_L}\right]$. This condition is necessary to ensure that a success pushes the bureaucrat's reputation above the reputation threshold and a failure below it. This means that if β_0 falls in this range and the bureaucrat generates a success, the bureaucrat's reputation in the second period increases to $\beta_1 = \beta_s$. If β_0 falls in the above range and there is failure, the bureaucrat's reputation decreases to $\beta_1 = \beta_f$. Alternatively, if $\beta_0 \notin \left[\frac{p_L}{p_H+p_L}, \frac{1-p_L}{2-p_H-p_L}\right]$, the bureaucrat shirks, and the public does not update its belief about the bureaucrat's reputation, i.e., $\beta_1 = \beta_0$ stays at the prior reputation.

The politician's signaling behavior depends crucially on the bureaucrat's initial reputation β_0 and the informativeness of the policy outcome. If $\beta_0 \in \left[\frac{p_L}{p_H+p_L}, \frac{1-p_L}{2-p_H-p_L}\right]$ and the bureaucrat exerts effort, the politician can remain silent after which second period effort is determined by the bureaucrat's reputation at that stage with $\beta_1 = \beta_s$ or $\beta_1 = \beta_f$. If a success or failure pushes the bureaucrat's reputation to an extreme, then $e_2(\beta_1) \neq 1$ and at least one type does not exert effort. Can the politician send signals in the first period to increase second-period effort? The immediate answer is negative, as information provision can only push the bureaucrat's reputation to be more extreme. Hence, if the bureaucrat's initial reputation is intermediate with $\beta_0 \in \left[\frac{p_L}{p_H+p_L}, \frac{1-p_L}{2-p_H-p_L}\right]$, then the politician cannot improve his welfare through first-period signaling.

Alternatively, if $\beta_0 < \frac{p_L}{p_H+p_L}$ or $\beta_0 > \frac{1-p_L}{2-p_H-p_L}$ then the bureaucrat shirks in the first period absent politician signaling. If the politician remains silent in the first period, then the bureaucrat's reputation remains constant with $\beta_1 = \beta_0$, after which second-period effort levels are as in Proposition 2. As in the one-period model, the politician can increase the probability of getting good outcomes, depending on the bureaucrat's initial reputation. An important difference with the one-period model, however, is that the politician faces a dynamic constraint in signaling. As alluded to earlier, any two messages the politician sends on the equilibrium path must generate the same second-period effort levels to ensure the politician is indifferent. There are three possibilities. First, the politician may send two messages after success or failure so that, following each message, both types are expected to exert second-period effort. Second, the two messages generate reputations β_1' and β_1'' so that, in the second period, the high type exerts effort and the low type mixes. Third, both messages may generate reputations so that both types shirk in the second period. To sum up, if different on-path messages are sent, they have to both induce reputations such that the bureaucrat of type t exerts the same amount of (mixed) effort e_t in the second stage.

We now investigate the possibility of effort in both periods. Three requirements must be met. The first is that the bureaucrat's prior reputation must still be sufficiently moderate, with $\beta_0 \in \left(\frac{c}{c+p_H}, \frac{p_H(1-p_L)}{p_H(1-p_L)+c(1-p_H)}\right)$. If this condition is not satisfied, then a success or failure pushes the bureaucrat's reputation to be too high or too low for there to be effort in the second period as well. The second requirement is that the politician's messages must not induce a reputation outside the above interval. The final condition is that both types of the bureaucrat must be incentivized to exert effort in the first period. These three requirements all point to the conclusion that the bureaucrat's reputation must still be sufficiently intermediate for him to exert effort. The more extreme is the bureaucrat's reputation, the harder it is to provide incentives through signaling. As in the one-period game, the more extreme is the bureaucrat's reputation the lower is the bureaucrat's effort level over both periods.

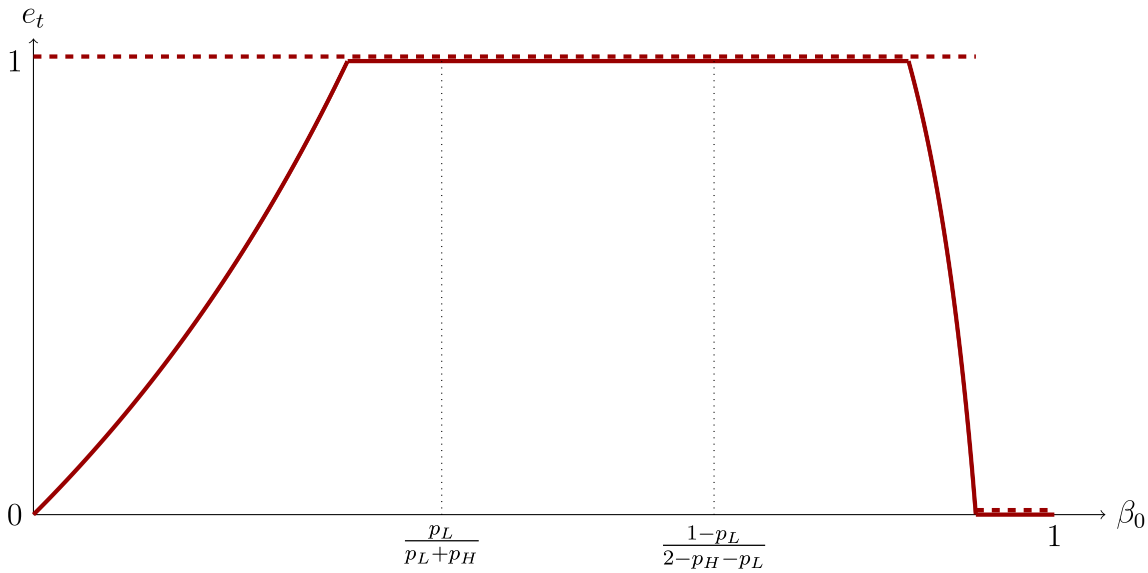
With these results we now turn to effort levels with politician messaging in the two-period game.

Proposition 3. *In politician-optimal equilibria in the two-period game, effort levels in the first period are as follows:*

- *If $\beta_0 < \frac{c(p_H+p_L-c)}{c(p_H+p_L-c)+p_H^2}$, then in period 1, the low type mixes and the high type exerts effort.*
- *If $\beta_0 \in \left[\frac{c(p_H+p_L-c)}{c(p_H+p_L-c)+p_H^2}, \bar{\beta}^*(c, p_L, p_H)\right]$,[14] the bureaucrat exerts effort.*

[14]The value $\bar{\beta}^*(c, p_L, p_H)$ is defined in the online appendix.

Figure 4: Equilibrium Effort in the First Period



Note: The solid line indicates the effort level of the low type ($e_t = e_L$), and the dashed line of the high type ($e_t = e_H$) in the first period, assuming that $c = \frac{1}{4}, p_L = \frac{1}{2}$, and $p_H = \frac{3}{4}$. The cut-points on the x-axis indicate the relevant equilibrium constraints on β_0 from Proposition 3.

- *If $\beta_0 \in \left(\bar{\beta}^*(c, p_L, p_H), \frac{p_H}{p_H+c-p_Hc}\right)$, the low type mixes and the high type exerts effort.*
- *If $\beta_0 \geq \frac{p_H}{p_H+c-p_Hc}$, the bureaucrat shirks.*

Second-period effort levels are determined by inputting β_1 in Proposition 2.

The proposition shows how the conditions under which the politician can extract effort from the bureaucrat become more stringent in a dynamic setting. While in the static setting the politician can freely choose how to punish or reward the bureaucrat by affecting his reputation, the politician has to be mindful of effects on future effort levels. As Proposition 2 has shown, if the bureaucrat's reputation moves further away from the reputation threshold of $\frac{1}{2}$, there is less aggregate effort from both bureaucrat types. As a result, providing stronger incentives by blaming and crediting in the present may dampen possibilities to extract effort in the future.

6 Extension: Politician Reputation Concerns

Having shown how a politician without reputational concerns can incentivize a bureaucrat by managing his reputation, our final extension delineates how a politician with her own reputational concerns can manage the bureaucrat's reputation and subsequent effort incentives. Adding politician reputation has strategic effects that are a product of her now having incentives to send messages to boost her own standing. It suffices to return to a one-period model but instead of there being asymmetric information about types, we now assume there is asymmetric information about who is responsible for the outcome. This is to highlight the strategic tension between providing incentives to the bureaucrat and the politician's concern for her reputation.

6.1 Setup

We now define the politician as having high talent $t_P = H$ with probability $\gamma_0 \in (0, 1)$ and low talent $t_P = L$ with remaining probability $1 - \gamma_0$. Importantly, neither politician or bureaucrat know their talent level, they only know who was responsible for the policy outcome. The politician's payoff is now a function of the outcome and the reputation that she obtains at the end.

$$U_P(x, \gamma) = x + \mathbb{I}_{\gamma > 1/2}.$$

We capture public updating on the politician's reputation by introducing uncertainty about whether the bureaucrat or the politician is responsible for the decision. With probability $\frac{1}{2}$ the bureaucrat exerted effort to determine the outcome, and with probability $\frac{1}{2}$ the politician is responsible for the outcome. Notice that the politician, given that she does not directly implement policy, makes no effort choice; her probability of success, p_H or p_L, depends on her talent. Both talents and responsibility are randomly drawn at the game's beginning.

6.2 Analysis

Given our setup, uncertainty about who is responsible for successes and failures results in the public only partially updating about a player's talent. Formally, define the bureaucrat's and politician's *ex ante* probability of success as:

$$g_s(\beta_0) := \hat{e}\left(p_H\beta_0 + p_L(1-\beta_0)\right),$$
$$h_s(\gamma_0) := p_H\gamma_0 + p_L(1-\gamma_0),$$

where $1 - g_s(\beta_0)$ and $1 - h_s(\gamma_0)$ are the respective probabilities of failure. Then beliefs following the good and bad outcome respectively, ignoring messages, are determined as:

$$\beta_s = \frac{\beta_0(p_H + h_s(\gamma_0))}{g_s(\beta_0) + h_s(\gamma_0)},$$
$$\beta_f = \frac{\beta_0(2 - p_H - h_s(\gamma_0))}{2 - g_s(\beta_0) - h_s(\gamma_0)}.$$

As in the previous models, the politician can send messages to affect these beliefs. Now formally define *blaming* and *crediting* in this new framework.

Definition 1. *The politician's message blames the bureaucrat if the policy fails and a message induces a lower reputation relative to the reputation induced if the politician remained silent.*

Definition 2. *The politician's message credits the bureaucrat if the policy succeeds and a message induces a higher reputation relative to the reputation induced if the politician remained silent.*

In other words, the politician's message affects the bureaucrat's reputation. Hence, after the public observes the outcome and updates its beliefs based on it, it may update further after observing the politician's message.

As before, there are conditions that determine whether the bureaucrat is willing to exert effort. Essentially, the bureaucrat must often enough obtain a good reputation following a success and a bad reputation often enough following a failure. This formal condition is similar as before but now the bureaucrat does not know whether he has low or high talent:

$$\pi(m^+|s, B) - \pi(m^+|f, B) \geq \frac{c}{g_s(\beta)}. \tag{2}$$

That is, conditional on being responsible for the decision, message m^+ must be sent often enough following a success and not too often following a failure. In addition, messages m^+ and m^- must generate a good and bad reputation for the bureaucrat respectively. Again, if the bureaucrat's initial reputation β_0 is sufficiently intermediate, then the politician does not need to send different messages to induce effort.

If the bureaucrat's initial reputation is too low, then the issue is that a success does not generate a good reputation. Similarly, with an initial reputation that is too high, a failure does not push the bureaucrat's reputation low enough. To generate maximal effort incentives, the politician is best off fully revealing who was responsible for the decision. This implies that if the bureaucrat faces the decision to exert effort, he knows that the public will infer that the bureaucrat is fully responsible for the outcome. In that case, the amount of updating that the public does will be at its possible maximum. A similar formal condition applies as before, where the bureaucrat's initial reputation must be sufficiently intermediate with $\beta_0 \in \left[\frac{p_L}{p_H+p_L}, \frac{1-p_L}{2-p_H-p_L}\right]$. Outside this range, the bureaucrat shirks.

Even if the bureaucrat's initial reputation is sufficiently moderate, however, the politician faces a constraint in signaling. The reason is that any message that gives credit to the bureaucrat will take credit away from the politician. Similarly, any message that blames the bureaucrat for a failure must come hand in hand with another message in equilibrium that requires the politician to take the blame. This concern does not constrain the politician if her reputation following a success and failure is far from the threshold. Closer to the threshold, however, the politician has little flexibility in sending messages to blame or credit the bureaucrat and induce effort.

The next proposition states how effort levels are determined in equilibrium.

Proposition 4. *In politician-optimal equilibria, effort levels are as follows*

- *If* $\beta_0 \in \left[0, \frac{p_L}{p_H+p_L}\right) \cup \left(\frac{1-p_L}{2-p_H-p_L}, 1\right]$ $(\mathcal{A}^-)$, *the bureaucrat shirks.*
- *If* $\beta_0 \in \left[\frac{p_L+h_s(\gamma_0)}{p_H+p_L+2h_s(\gamma_0)}, \frac{2-p_L-h_s(\gamma_0)}{4-p_H-p_L-2h_s(\gamma_0)}\right]$ $(\mathcal{B}^+)$, *then the bureaucrat exerts effort.*
- *If* $\beta_0 \in \left[\frac{p_L}{p_H+p_L}, \frac{p_L+h_s(\gamma_0)}{p_H+p_L+2h_s(\gamma_0)}\right)$, *then*

  1. *if* $\gamma_0 \leq \frac{p_L}{p_H+p_L}$ *or* $\gamma_0 \geq \frac{1}{2}$ $(\mathcal{C}^+)$, *the bureaucrat exerts effort*

  If $\gamma_0 \in \left(\frac{p_L}{p_H+p_L}, \frac{p_L+g_s(\beta_0)}{p_H+p_L+2g_s(\beta_0)}\right)$. *If the bureaucrat exerts effort, the politician gives credit to the bureaucrat.*
- *If* $\beta_0 \in \left[\frac{2-p_L-h_s(\gamma_0)}{4-p_H-p_L-2h_s(\gamma_0)}, \frac{1-p_L}{2-p_H-p_L}\right]$, *then*

  1. *if* $\gamma_0 \leq \frac{1}{2}$ *or* $\gamma_0 \geq \frac{1-p_L}{2-p_H-p_L}$ $(\mathcal{C}^+)$, *the bureaucrat exerts effort.*

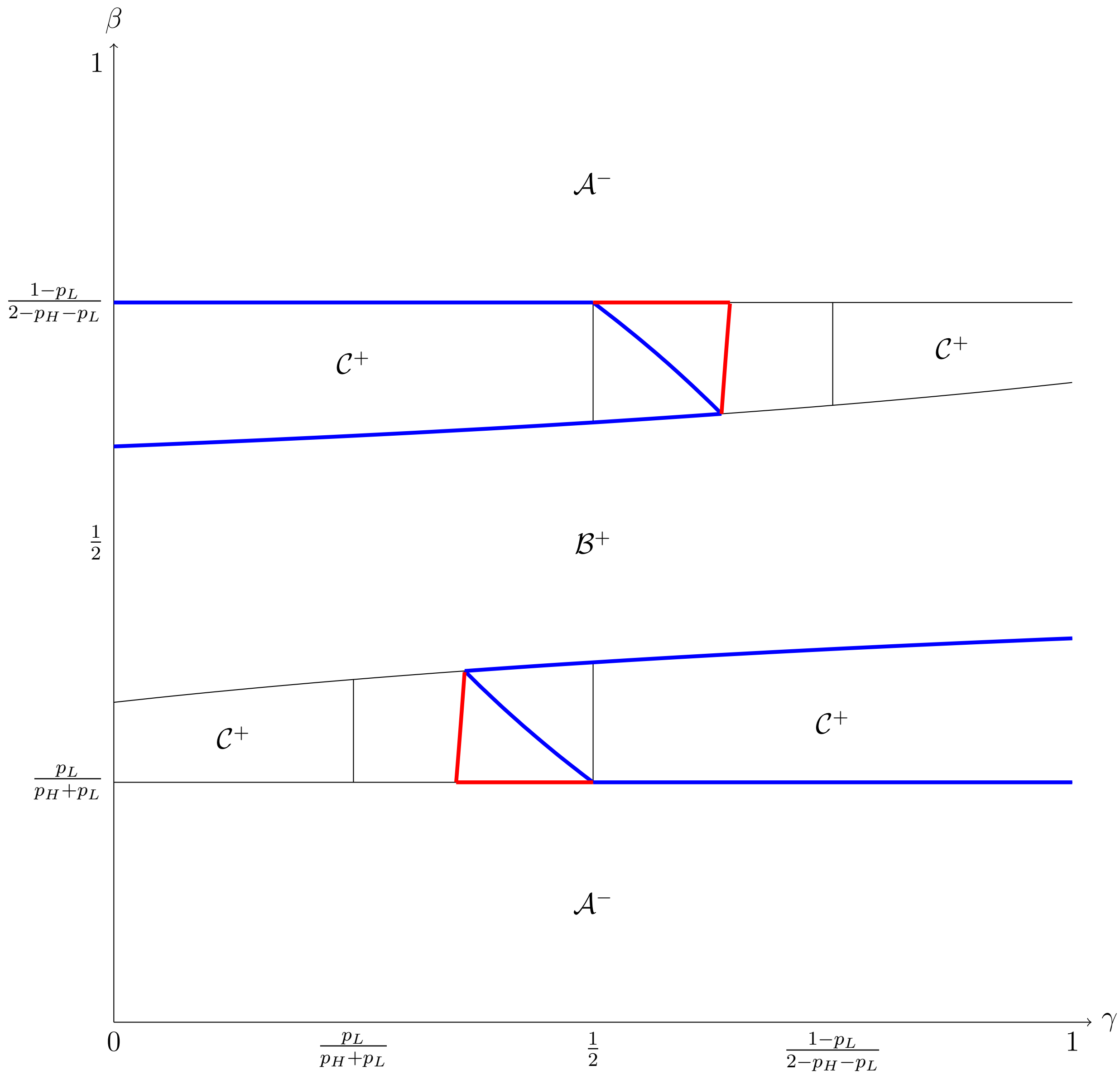
  If $\gamma_0 \in \left(\frac{1}{2}, \frac{1-p_L}{2-p_H-p_L}\right)$, *then we need that* β_0 *is s.t. ...*

The results demonstrate that with politician reputation concerns and the absence of politician messaging, the bureaucrat's reputation must be closer to the threshold relative to our earlier analyses of previous variations of our model for him to exert effort. Given that the public is now uncertain about who is responsible for a positive or negative outcome, the public now updates upward or downward about both the politician and bureaucrat type after a success or failure, resulting in a higher or lower reputation. However, the impact on the bureaucrat's reputation is dampened compared to a world where politicians are only outcome–oriented. As outcomes are less informative regarding the bureaucrat, the conditions under which the bureaucrat exerts effort without reputation management are stricter, as his reputation must fall closer to the threshold. As such, without communication there is a wider range of bureaucratic reputation where shirking will occur.

7 Applications

With our theoretical results in hand, we now briefly discuss several examples from Dutch and Australian politics. Our results help see blaming and crediting in a different light than has

Figure 5: Conditions for Equilibria with Effort



Note: Assumptions: $p_H = \frac{3}{4}$ *and* $p_L = \frac{1}{4}$*. The shaded region is the range of* (γ_0, β_0) *such that equilibria with effort exist.*

been the norm, taking into account its effect on bureaucratic performance. The examples illustrate how politicians may shift the blame to advisers, take the blame for failed policies, or even make efforts to protect bureaucrats from criticism.

Australian Covid vaccine rollout. Australia experienced a delay in the distribution of a Covid vaccine, potentially leading to many more cases.[15] In response, a blame game ensued between Prime Minister Scott Morrison and his immunization advisers. Morrison emphasized that the failed policy outcome resulted from his advisers being excessively cautious and incorrectly assessing that Covid case numbers would remain low. The advisers, in turn, defended themselves, saying the criticism was unfair and that they gave advice with the available evidence at that point in time. This seeming attempt to avoid blame was blunted by an impartial public health expert who attributed blame to both leadership and advisers.[16] This example illustrates that while politicians can blame bureaucrats or advisers this can be disputed, highlighting that the blame game is often a form of cheap talk. In addition, that Scott Morrison deflected and shifted the blame indicates that he was less worried about incentivizing future effort by his underlings to acquire evidence and give advice relative to the electoral consequences of his assuming responsibility.

Dutch childcare benefits scandal. In the Netherlands a scandal erupted in 2018 over the distribution of childcare benefits. Allegations of fraud were made by the Tax and Customs Administration, claims that would eventually prove unfounded. When the allegations were initially made, many MPs responded by criticizing senior civil servants in the Dutch tax au-

[15] https://www.theguardian.com/world/2021/jul/15/not-time-for-blame-game-former-atagi-representative-says-morrisons-astrazeneca-comments-unfair? (accessed on July 29, 2021).

[16] According to Bill Bowtell — adjunct professor at UNSW and public health expert — the federal government had "completely failed to reduce risk by diversifying supply in July/August last year" (...) "The blame is on Scott Morrison, Greg Hunt, and the people who wrongly advised the government that there was no rush, no race and, as a result, no vaccine."

thority for their alleged failures. In the aftermath of it coming to light that the tax authority withheld information from Parliament, Prime Minister Rutte praised his departed tax authority head and thanked him for his services. Further, when aggravated MPs subsequently pointed out that the failures harmed many citizens and criticized the officials involved as 'revolving door senior bureaucrats,' Rutte rose to the defense of these bureaucrats: "Revolving door top level bureaucrats! That is not the way to talk about our senior civil servants. (...) They cannot defend themselves. By talking about them this way you make them outlawed."[17] Rutte's response illustrates a leader with not just a sensitivity about bureaucrat effort necessarily but a broader concern for the protection of civil servants from criticism to ensure their willingness to continue working in the bureaucracy. Indeed, on April 26th, the government sent a letter to the Dutch Parliament emphasizing that acts of individual civil servants should not be detailed. The letter posited a principle that civil servants cannot be discussed in the political debate, as government officials and ministers are ultimately responsible and, while Parliamentarians can call on these officials and ministers to question them and hold them accountable, civil servants cannot defend themselves. In addition, a committee on social benefits concluded that failures were a function of structural deficiencies rather than individuals with bad intentions. This example illustrates that government officials may exert effort to shield bureaucrats from attack so that they can continue to advise politicians in confidence, trumping other considerations such as transparency, etc.

Dutch relaxation of lockdown restrictions. In another Dutch example, Covid restrictions were found to have been relaxed too soon, causing an unanticipated growth of cases. In response, Prime Minister Rutte apologized in a press conference for relaxing the restrictions, deeming it an 'error of judgment.' Minister of Health De Jonge also acknowledged that it was an error to allow young people having taken the Janssen vaccine to obtain access to places such as dance clubs before requiring them to wait two weeks for full vaccine effectiveness.

[17]Prime Minister Mark Rutte (January 2020) in response to questions of Socialist Party MP Renske Leijten on the Dutch childcare benefits scandal (Translated from Dutch).

In turn, having noticed that the number of cases quickly rose due to the new Delta variant of the Covid virus, the government imposed additional restrictions. This example illustrates how politicians may willingly accept blame, even though there are others to potentially point a finger at (in this case it being publicly known that an independent body of experts was providing advice). Unlike the Australian case, the Dutch Prime Minister did not shift any blame. Besides being potentially proper in its own right, another possibility is that Dutch government officials did not want to hurt the reputation of its advisers, ensuring that they still have proper incentives to work hard and provide accurate advice afterwards.

8 Discussion and Conclusion

When asking experts or voters what makes good political leaders, words like honesty, integrity, trustworthiness, and character are typical responses. Nonetheless, politician blaming of bureaucratic underlings when something goes awry is not uncommon. Interpretations of such accusations have typically been to view the politician as not embodying such characteristics but, rather, trying to avoid responsibility. If apt, this description suggests that being a good leader is not key for political selection and that voters and their surrogates are either insincere in saying what they care about or are easily misled.

Alternatively, we provide a rationale for blaming and credit claiming that is more consistent with positive leadership. Assuming that politicians want to get policy right, everything being equal, and that more bureaucratic effort raises success probability, the ability to blame may advantage politicians. Specifically, in our theory of blaming and blame–shifting a politician's ability to criticize can, under some conditions, increase the likelihood of effort-induced good policy. Given that the public is observing, the specter of blame may spur a bureaucrat to work harder despite our assuming that delegation cannot be adjusted and firing the bureaucrat is not an option — potentially strong sanctions that are not part of the principal's arsenal. The threat of blame alone may spur the reputation–sensitive bureaucrat on, as such criticism plays an informational role for a judgmental public in given circumstances. Politicians may not be avoiding accountability but instead providing the public with an ability to update about bureaucrats that, in turn, can produce better outcomes. Conversely there are instances, as in the real world, where the leader does not blame or provide credit in equilibrium.

However, even in this world, blaming is no panacea for inducing effort. Bureaucrats knowing that they will be scapegoated following a failure no matter how hard they work can lack incentive to work and shirk instead. Further, if politicians are themselves motivated by reputation concerns, particularly if their reputations matter far more to them than policy and if bureaucrats are not so insulated that politicians can be reasonably attributed responsibility, their abilities to promise credibly that they will not turn on bureaucrats and share blame themselves with a bad outcome can be dampened.

Broadly, our analysis contrasts with previous non-formal and formal considerations of blaming. Concerning the former, we show that an emphasis on politician blaming need not require non-standard utility functions, such as those associated with prospect theory, that are cited as key in much non-formal research. Given that blaming may be a product not just of efforts to elude responsibility but to induce bureaucratic effort, it may be unsurprising that there appears to be more politician emphasis on blaming than credit claiming. Further, empirical efforts to estimate when we witness blaming may be undermined by the possibility of two difficult to distinguish data-generating processes, one involving accountability and the other bureaucratic reputation. As implied, theoretically we show that blaming can be an equilibrium outcome of a model not relying principally on politician accountability — although such features can potentially be integrated.

Hence, our analysis contrasts with the near exclusively negative view of politician blaming. In bringing to the fore politicians wanting to motivate their agents via communication, we show that blame can serve a function that is analogous to the principal's ability to adjust an individual's wages and hire or fire as in principal-agent approaches in the theory of the firm. In turn, this recognition directs attention toward other considerations, such as whether politicians should be routinely condemned for criticizing their underlings. Our findings indicate that allowing a policy–motivated politician to blame her underlings might be a positive thing. Dissuading politicians from doing so, for example by fostering norms where politicians criticizing the bureaucracy are sanctioned by negative responses by voters, may induce less bureaucratic effort in the aggregate and the wrong choice being made more frequently. This contrasts with claims, dating back to Fiorina (1977, 1982)'s canonical works, viewing the ability to blame the bureaucracy as exclusively negative and indicating that blaming should be condemned.

As discerning with field data whether real world blaming is driven by accountability or communication is fraught with difficulties, it might be edifying to consider these alternatives in an experimental setting. One could also complement our theoretical analysis with survey experiments to ascertain how the public views blaming to assess whether it impacts bureaucratic reputation and politician approval.

References

Alesina, Alberto and Guido Tabellini. 2005. Why do Politicians Delegate? Technical report National Bureau of Economic Research.

Almendares, Nicholas. 2012. "Blame-Shifting, Judicial Review, and Public Welfare." *Journal of Law & Politics* 27(2):239–321.

Bartling, Björn and Urs Fischbacher. 2012. "Shifting the Blame: On Delegation and Responsibility." *The Review of Economic Studies* 79(1):67–87.

Beer, Michael, Russell A Einstat, Nathaniel Foote, Tobias Fredberg and Flemming Norgreen. 2011. *Higher Ambition: How Great Leaders Create Economic and Social Value*. Harvard Business Press.

Bendor, Jonathan and Adam Meirowitz. 2004. "Spatial Models of Delegation." *American Political Science Review* 98(2):293–310.

Bertelli, Anthony M and Madalina Busuioc. 2021. "Reputation-Sourced Authority and the Prospect of Unchecked Bureaucratic Power." *Public Administration Review* 81(1):38–48.

Bertrand, Marianne, Robin Burgess, Arunish Chawla and Guo Xu. 2020. "The Glittering Prizes: Career Incentives and Bureaucrat Performance." *The Review of Economic Studies* 87(2):626–655.

Boon, Jan and Heidi Houlberg Salomonsen. 2020. Public Sector Organizations and Reputation. In *The Handbook of Public Sector Communication*, ed. Vilma Luomaaho and Mara-José Canel. John Wiley & Sons pp. 215–227.

Carpenter, Daniel P. 2001. *The Forging of Bureacuratic Autonomy: Reputations, Networks, and Policy Innovation in Executive Agencies, 1862-1928*. Princeton University Press.

Carpenter, Daniel P. 2010. *Reputation and Power: Organizational Image and Pharmaceutical Regulation at the FDA*. Princeton University Press.

Carpenter, Daniel P and George A Krause. 2012. "Reputation and Public Administration." *Public Administration Review* 72(1):26–32.

Craig, S Bartholomew and Robert B Kaiser. 2012. Destructive Leadership. In *Oxford Handbook of Leadership*, ed. Michael G Rumsey. Oxford University Press.

Ely, Jeffrey C and Juuso Välimäki. 2003. "Bad Reputation." *The Quarterly Journal of Economics* 118(3):785–814.

Epstein, David and Sharyn O'Halloran. 1994. "Administrative Procedures, Information, and Agency Discretion." *American Journal of Political Science* 38(3):697–722.

Fiorina, Morris P. 1977. *Congress: Keystone of the Washington Establishment*. Yale University Press.

Fiorina, Morris P. 1982. “Legislative Choice of Regulatory Forms: Legal Process or Administrative Process?” *Public Choice* 39(1):33–66.

Fox, Justin and Stuart V Jordan. 2011. “Delegation and Accountability.” *Journal of Politics* 73(3):831–844.

Gailmard, Sean and John W Patty. 2012. “Formal Models of Bureaucracy.” *Annual Review of Political Science* 15:353–377.

Glazer, Amihai and Bjorn Segendorff. 2005. “Credit Claiming.” *Economics of Governance* 6(2):125–137.

Holmström, Bengt. 1980. On the Theory of Delegation. Technical report Discussion Paper.

Hood, Christopher. 2011. *The Blame Game: Spin, Bureaucracy, and Self-Preservation in Government*. Princeton University Press.

Horn, Murray J. 1995. *The Political Economy of Public Administration: Institutional Choice in the Public Sector*. Cambridge University Press.

Iyer, Lakshmi and Anandi Mani. 2012. “Traveling Agents: Political Change and Bureaucratic Turnover in India.” *Review of Economics and Statistics* 94(3):723–739.

Kahneman, Daniel and Amos Tversky. 1979. “Prospect Theory: An Analysis of Decision under Risk.” *Econometrica* 47(2):263–291.

Khan, Adnan Q, Asim Ijaz Khwaja and Benjamin A Olken. 2019. “Making Moves Matter: Experimental Evidence on Incentivizing Bureaucrats through Performance-Based Postings.” *American Economic Review* 109(1):237–270.

Lipnowski, Elliot and Doron Ravid. 2020. “Cheap Talk with Transparent Motives.” *Econometrica* 88(4):1631–1660.

Marvel, John D. 2014. “The Boston Marathon Bombings: Who’s to Blame and Why it Matters for Public Administration.” *Public Administration Review* 74(6):713–725.

Pei, Harry Di. 2018. “A Blame Shifting Theory of Delegation.” *Unpublished manuscript* .

Pierson, Paul. 1994. *Dismantling the Welfare State? Reagan, Thatcher and the Politics of Retrenchment*. Cambridge University Press.

Ruder, Alex I. 2014. “Institutional Design and the Attribution of Presidential Control: Insulating the President from Blame.” *Quarterly Journal of Political Science* 9(3):301–335.

Ruder, Alex I. 2015. “Agency Design, the Mass Media, and the Blame for Agency Scandals.” *Presidential Studies Quarterly* 45(3):514–539.

Tamada, Yasunari and Tsung-Sheng Tsai. 2018. “Reputation versus Information: The Delegation Policy when the Principal has Reputational Concerns.” *Journal of Public Economic Theory* 20(3):367–389.

Volden, Craig. 2002. "A Formal Model of the Politics of Delegation in a Separation of Powers System." *American Journal of Political Science* 46(1):111–133.

Weaver, R Kent. 1986. "The Politics of Blame Avoidance." *Journal of Public Policy* 6(4):371–398.

Xu, Guo. 2018. "The Costs of Patronage: Evidence from the British Empire." *American Economic Review* 108(11):3170–98.

Supplementary Information for Politicians, Bureaucrats, and the Battle for Credit

Online Appendix

Table of Contents

A Proofs for Baseline Model

A.1 Building Blocks

At the game's final stage the public's belief about the bureaucrat's talent depends on policy outcome x and message m. The public has a conjecture about the bureaucrat's mixed effort level if he has talent t, denoted by $\hat{e}_t$. If a success is achieved with positive probability on the equilibrium path, then Bayes' rule implies that the public's posterior following success and message m equals

$$\beta_{s,m} = \frac{\hat{e}_H p_H \pi(m|s,H)\beta_0}{\hat{e}_H p_H \pi(m|s,H)\beta_0 + \hat{e}_L p_L \pi(m|s,L)(1-\beta_0)}.$$

Following a failure and message m, beliefs are determined analogously

$$\beta_{f,m} = \frac{(1-\hat{e}_H p_H)\pi(m|f,H)\beta_0}{(1-\hat{e}_H p_H)\pi(m|f,H)\beta_0 + (1-\hat{e}_L p_L)\pi(m|f,L)(1-\beta_0)}.$$

These beliefs are equivalent to the bureaucrat's reputation, and ultimately determine the bureaucrat's payoffs and incentives, relative to the threshold of $\frac{1}{2}$. The beliefs are also determined by the public's conjecture about the bureaucrat's effort strategy and we specify this throughout the proofs wherever necessary.

A.2 Proof of Proposition 1

Proof. We look for equilibria where the politician sends m^0 at every information set.

Case 1. Full effort. First, assume $e_H = e_L = 1$, which is the politician-optimal equilibrium if it exists. In equilibrium, the public's conjecture about the bureaucrat's effort is correct, i.e., $\hat{e}_L = \hat{e}_H = 1$. By Bayes' rule, beliefs following a success (β_{s,m^0}) and failure (β_{f,m^0}) are

$$\beta_{s,m^0} = \frac{p_H\beta_0}{p_H\beta_0 + p_L(1-\beta_0)}, \tag{3}$$

$$\beta_{f,m^0} = \frac{(1-p_H)\beta_0}{(1-p_H)\beta_0 + (1-p_L)(1-\beta_0)}. \tag{4}$$

Each bureaucrat type's incentive condition requires that exerting effort is a best response:

$$\begin{gathered} p_t \mathbb{I}_{\{\beta_{s,m^0}>\frac{1}{2}\}} + (1-p_t)\mathbb{I}_{\{\beta_{f,m^0}>\frac{1}{2}\}} - c \geq \mathbb{I}_{\{\beta_{f,m^0}>\frac{1}{2}\}}, \\ \mathbb{I}_{\{\beta_{s,m^0}>\frac{1}{2}\}} - \mathbb{I}_{\{\beta_{f,m^0}>\frac{1}{2}\}} \geq \frac{c}{p_t}. \end{gathered} \tag{5}$$

For this inequality to hold, we need that $\beta_{s,m^0} > \frac{1}{2} > \beta_{f,m^0}$, where one inequality may be weak. If it holds, then the LHS equals p_t, otherwise it equals 0. By assumption $0 < c < p_L < p_H$, and thus there exists an equilibrium with full effort if and only if $\beta_{s,m^0} > \frac{1}{2} > \beta_{f,m^0}$, where one inequality may be weak. After rearranging, a full effort equilibrium exists if:

$$\frac{p_L}{p_H + p_L} \leq \beta_0 \leq \frac{1 - p_L}{2 - p_H - p_L}. \tag{6}$$

Case 2: Mixing equilibria. If the bureaucrat mixes, he must be indifferent. Equation 5 shows that this cannot be true.

Case 3: Separating equilibria. If the high type exerts effort and the low type shirks, then $\beta_{s,m^0} = 1$. A failure must induce reputation $\beta_{f,m^0} \leq \frac{1}{2}$, otherwise the high type shirks. But if $\beta_{f,m^0} \leq \frac{1}{2}$, the low type earns a payoff of 0 while a deviation yields $p_L - c > 0$ and is profitable; a contradiction. Alternatively, if the low type exerts effort and the high type does not, then shirking leads to a higher expected reputation and saves the cost c, which must be profitable, which is also a contradiction. As a result, there exists no separating equilibrium.

Case 4: Shirking. Finally, the remaining case is where both bureaucrat types shirk. Both types earn a payoff of 0 if $\beta_0 < \frac{1}{2}$ and 1 if $\beta_0 \geq \frac{1}{2}$. Following a success, Bayes' rules does not apply because it is off the equilibrium path. As long as $\beta_{s,m^0} < \frac{1}{2}$, this is an equilibrium. Therefore, for all remaining β_0 not covered in Case 1, both types shirk. □

A.3 Proof of Proposition 2

Proof. Now we consider the existence of politician-optimal equilibria when the politician may send different messages. Each case verifies incentive compatibility conditions of both types and beliefs derived by Bayes' rule that generate incentive compatibility. Note that the politician's strategy is always incentive compatible because she is indifferent between all messages after effort choices are already made. Throughout, we assume that m^+ induces a weakly higher reputation than m^-, i.e., $\beta_{s,m^+} \geq \beta_{s,m^-}$ and $\beta_{f,m^+} \geq \beta_{f,m^-}$.

Case 1: Full effort. It is immediate that equilibria with full effort exist if $\beta_0 \in \left[\frac{p_L}{p_H+p_L}, \frac{1-p_L}{2-p_H-p_L}\right]$ as the politician can send m^0 at every information set. Thus, consider other β_0.

Case 1a: $\beta_0 < \frac{p_L}{p_H+p_L}$. Note that $\beta_f < \frac{1}{2}$ given $e_H = e_L = 1$. It is potentially incentive compatible for the low type to exert effort if $\beta_{s,m^+} \geq \frac{1}{2}$. The politician must ensure that (s, m^+) leads to a good reputation with a sufficiently high weight for each type. Rewarding failure does not increase incentives to exert effort, so we assume the politician remains silent following a failure.

Formally, the belief given (s, m^+) is

$$\beta_{s,m^+} = \frac{p_H \beta_0 \pi(m^+|s, H)}{p_H \beta_0 \pi(m^+|s, H) + p_L(1 - \beta_0)\pi(m^+|s, L)}. \tag{7}$$

To ensure $\beta_{s,m^+} \geq \frac{1}{2}$, P's messaging strategy must satisfy the following inequality:

$$p_H \beta_0 \pi(m^+|s, H) \geq p_L(1 - \beta_0)\pi(m^+|s, L). \tag{8}$$

To guarantee that this inequality holds for the widest possible parameters of β_0, we set $\pi(m^+|s,H) = 1$. This implies that the high type's incentive constraint is satisfied as

$$p_H\pi(m^+|s,H) - c = p_H - c \geq 0. \tag{9}$$

To ensure the low type exerts effort, we need

$$p_L\pi(m^+|s,L) - c \geq 0 \iff \pi(m^+|s,L) \geq \frac{c}{p_L}. \tag{10}$$

Recall that the necessary condition is that $\beta_{s,m^+} \geq \frac{1}{2}$ is as in equation 8. By substituting in $\pi(m^+|s,H) = 1$ and noting that the RHS is increasing in $\pi(m^+|s,L)$, we set $\pi(m^+|s,L) = \frac{c}{p_L}$, which is the minimum. Substituting this in yields

$$p_H\beta_0 \geq c(1-\beta_0) \iff \beta_0(p_H + c) \geq c \iff \beta_0 \geq \frac{c}{p_H + c}, \tag{11}$$

as required. This establishes that there exists a full effort equilibrium if $\frac{c}{p_H+c} \leq \beta_0 \leq \frac{p_L}{p_H+p_L}$.

Case 1b: $\beta_0 > \frac{1-p_L}{2-p_H-p_L}$. If both types exert effort this implies that, following shirking, (f,m^-) may not lead to a reputation that exceeds the threshold. The goal is to show that an equilibrium with full effort exists if $\beta_0 \leq \frac{p_H(1-p_L)}{p_H(1-p_L)+c(1-p_H)}$. A necessary condition for the existence of such an equilibrium is that $\beta_{f,m^-} \leq \frac{1}{2}$. By Bayes' rule:

$$\beta_{f,m^-} = \frac{\beta_0\pi(m^-|f,H)(1-p_H)}{\beta_0\pi(m^-|f,H)(1-p_H) + (1-\beta_0)\pi(m^-|f,L)(1-p_L)}. \tag{12}$$

Further, we require that $\beta_{f,m^-} \leq \frac{1}{2}$, i.e.,

$$\beta_0\pi(m^-|f,H)(1-p_H) \leq (1-\beta_0)\pi(m^-|f,L)(1-p_L). \tag{13}$$

To guarantee that this necessary condition holds for the widest possible parameters of β_0, we set $\pi(m^-|f,L) = 1$. This implies that the low type's incentive condition is met as

$$p_L + (1-p_L)(1-\pi(m^-|f,L)) - c \geq 1 - \pi(m^-|f,L) \iff p_L \geq c. \tag{14}$$

Further, the high type bureaucrat's incentive condition equals

$$p_H + (1-p_H)(1-\pi(m^-|f,H)) - c \geq 1 - \pi(m^-|f,H) \iff \pi(m^-|f,H) \geq \frac{c}{p_H}. \tag{15}$$

The goal is to discover for which β_0 this equilibrium exists. We thus set $\pi(m^-|f,H)$ to its minimum while still satisfying incentive compatibility for the high type, i.e., $\pi(m^-|f,H) =$

$\frac{c}{p_H}$. Hence, after substitution, we get the following simplification

$$\begin{aligned} c\beta_0(1-p_H) &\leq p_H(1-\beta_0)(1-p_L) \\ \beta_0\left(c(1-p_H)+p_H(1-p_L)\right) &\leq p_H(1-p_L) \\ \beta_0 &\leq \frac{p_H(1-p_L)}{c(1-p_H)+p_H(1-p_L)}. \end{aligned} \tag{16}$$

Hence, there exists a full effort equilibrium if $\beta_0 \in \left(\frac{1-p_L}{2-p_H-p_L}, \frac{p_H(1-p_L)}{p_H(1-p_L)+c(1-p_H)}\right]$.

Case 2: Low type mixing, high type effort. There are two cases:

Case 2a. $\beta_0 < \frac{c}{p_H+c}$**.** If the high type exerts effort and the low type mixes, a necessary condition is that $\beta_{s,m^+} \geq \frac{1}{2}$:

$$\begin{aligned} \frac{p_H\beta_0\pi(m^+|s,H)}{p_H\beta_0\pi(m^+|s,H)+e_Lp_L(1-\beta_0)\pi(m^+|s,L)} &\geq \frac{1}{2} \\ p_H\beta_0\pi(m^+|s,H) &\geq e_Lp_L(1-\beta_0)\pi(m^+|s,L). \end{aligned} \tag{17}$$

The goal is to maximize e_L subject to these conditions, so it is optimal to maximize the LHS and set $\pi(m^+|s,H)=1$, which immediately satisfies the high type's condition. Also, it is optimal to minimize the RHS parts that are not e_L, i.e., to minimize $\pi(m^+|s,L)$ as much as possible. Further, to ensure that the low type is indifferent, we need

$$p_L\pi(m^+|s,L)-c=0 \iff \pi(m^+|s,L)=\frac{c}{p_L}. \tag{18}$$

Note that we implicitly assume that the politician remains silent following a failure because the low type cannot be given stronger incentives if the politician speaks following a failure. Substituting this into equation 17 yields

$$p_H\beta_0 \geq e_Lp_L(1-\beta_0)\frac{c}{p_L} \iff e_L \leq \frac{p_H\beta_0}{c(1-\beta_0)}. \tag{19}$$

Hence, in a politician-optimal equilibrium, $e_L^* = \frac{p_H\beta_0}{c(1-\beta_0)}$ for all $\beta_0 \in \left(0, \frac{c}{p_H+c}\right)$.

Case 2b. $\beta_0 > \frac{p_H(1-p_L)}{p_H(1-p_L)+c(1-p_H)}$**.** Given that the low type mixes, he is indifferent between exerting effort and shirking. This generates the following equality, under the assumption that m^+ always leads to a good reputation and m^- to a bad one:

$$\begin{aligned} p_L\pi(m^+|s,L)+(1-p_L)\pi(m^+|f,L)-c &= \pi(m^+|f,L) \\ \pi(m^+|s,L) &= \pi(m^+|f,L)+\frac{c}{p_L}. \end{aligned} \tag{20}$$

Similarly, the high type exerts effort if

$$p_H \pi(m^+|s,H) + (1-p_H)\pi(m^+|f,H) - c \geq \pi(m^+|f,H) \tag{21}$$

$$\pi(m^+|s,H) \geq \pi(m^+|f,H) + \frac{c}{p_H}. \tag{22}$$

Following a success, we need that the two beliefs satisfy the following conditions, where one inequality may be weak:

$$\frac{p_H\beta_0\pi(m^-|s,H)}{p_H\beta_0\pi(m^-|s,H) + e_L p_L(1-\beta_0)\pi(m^-|s,L)} < \frac{1}{2} < \frac{p_H\beta_0\pi(m^+|s,H)}{p_H\beta_0\pi(m^+|s,H) + e_L p_L(1-\beta_0)\pi(m^+|s,L)}. \tag{23}$$

This implies the following inequalities, where one may be weak:

$$p_H\beta_0\pi(m^-|s,H) < e_L p_L(1-\beta_0)\pi(m^-|s,L) \tag{24}$$

$$p_H\beta_0\pi(m^+|s,H) > e_L p_L(1-\beta_0)\pi(m^+|s,L). \tag{25}$$

Similarly, following a failure, we need that the two beliefs satisfy the following condition, where one inequality may be weak:

$$\frac{(1-p_H)\beta_0\pi(m^-|f,H)}{(1-p_H)\beta_0\pi(m^-|f,H) + (1-e_L p_L)(1-\beta_0)\pi(m^-|f,L)} < \frac{1}{2} < \frac{(1-p_H)\beta_0\pi(m^+|f,H)}{(1-p_H)\beta_0\pi(m^+|f,H) + (1-e_L p_L)(1-\beta_0)\pi(m^+|f,L)}. \tag{26}$$

This implies the following inequalities, where one may be weak:

$$(1-p_H)\beta_0\pi(m^-|f,H) < (1-e_L p_L)(1-\beta_0)\pi(m^-|f,L) \tag{27}$$

$$(1-p_H)\beta_0\pi(m^+|f,H) > (1-e_L p_L)(1-\beta_0)\pi(m^+|f,L). \tag{28}$$

The goal is to maximize e_L subject to these conditions, while maintaining incentive compatibility for the high type and the four constraints on the beliefs. A first observation from equations 24 and 25 is that, by setting $\pi(m^+|s,H) = 1$, 24 is immediately met because $\pi(m^-|s,L) > 0$. Further, by also substituting in $\pi(m^+|s,H) = 1$, 25 can be rearranged to

$$e_L \leq \frac{p_H\beta_0}{p_L(1-\beta_0)\pi(m^+|s,L)}, \tag{29}$$

implying that e_L can be set to higher values if $\pi(m^+|s,H) = 1$. In addition, equation 28 can be immediately met with $\pi(m^+|f,L) = 0$. Then, by also substituting in $\pi(m^-|f,L) = 1$, equation 27 can be rearranged to

$$e_L \leq \frac{1-\beta_0-\beta_0(1-p_H)\pi(m^-|f,H)}{p_L(1-\beta_0)}. \tag{30}$$

From equation 29 we learn that it is optimal to minimize $\pi(m^+|s,L)$ to create the largest range for e_L and, by setting $\pi(m^+|s,L) = \frac{c}{p_L}$, we retain incentive compatibility for the low type. Further, from equation 30 we learn that it is optimal to minimize $\pi(m^-|f,H)$ to create the largest range for e_L. This minimum is $\pi(m^-|f,H) = \frac{c}{p_H}$ to retain incentive compatibility for the high type. After substituting this into equations 29 and 30, we get the following two equations:

$$e_L \leq \frac{p_H\beta_0}{p_L(1-\beta_0)\frac{c}{p_L}} = \frac{p_H\beta_0}{c(1-\beta_0)} \tag{31}$$

$$e_L \leq \frac{1-\beta_0-\beta_0(1-p_H)\frac{c}{p_H}}{p_L(1-\beta_0)} \tag{32}$$

We find that the second of these is more restrictive, generating an optimal mixed effort level $e_L^* = \frac{1-\beta_0-\beta_0(1-p_H)\frac{c}{p_H}}{p_L(1-\beta_0)}$ if $\beta_0 < \frac{p_H}{p_H+c-p_Hc}$, as stated in the proposition.

Case 3: Separating equilibrium. Suppose the high type exerts effort and the low type shirks. Then $\beta_{s,m^0} = 1$. Failure must lead to a belief $\beta_f \leq \frac{1}{2}$, otherwise the high type deviates. To ensure existence of this equilibrium under the widest circumstances, we ensure that the high type is indifferent between exerting effort and shirking:

$$p_H + (1-p_H)\pi(m^+|f,H) - c = \pi(m^+|f,H), \tag{33}$$

which implies $\pi(m^+|f,H) = \frac{p_H-c}{p_H}$ and $\pi(m^-|f,H) = \frac{c}{p_H}$. To ensure $\beta_{f,m^-} \leq \frac{1}{2}$:

$$\frac{(1-p_H)\beta_0\frac{c}{p_H}}{(1-p_H)\beta_0\frac{c}{p_H}+1-\beta_0} \leq \frac{1}{2} \iff \beta_0 \leq \frac{p_H}{p_H+c-cp_H}. \tag{34}$$

Finally, to ensure the low type has no profitable deviation, we set $\pi(m^+|f,L) = 0$, so that even if the low type exerts effort, this leads to an off-path information set, after which we set the belief equal to $\beta_{f,m^-} = 0$. This is an equilibrium if $\beta_0 \leq \frac{p_H}{p_H+c-p_Hc}$. This overlaps with sub-case 2b and therefore a separating equilibrium is never politician-optimal.

Case 4: High type mixing, low type shirking. Now suppose the high type mixes and the low type shirks. Again, success leads to a belief of $\beta_{s,m^0} = 1$. To ensure the high type is indifferent we need

$$\begin{aligned} p_H + (1-p_H)\pi(m^+|f,H) - c &= \pi(m^+|f,H) \\ \pi(m^+|f,H) &= \frac{p_H-c}{p_H}. \end{aligned} \tag{35}$$

This implies that $\pi(m^-|f,H) = \frac{c}{p_H}$, and the reputation following (f,m^-) is

$$\beta_{f,m^-} = \frac{(1-e_Hp_H)\frac{c}{p_H}\beta_0}{(1-e_Hp_H)\frac{c}{p_H}\beta_0+(1-\beta_0)\pi(m^-|f,L)}. \tag{36}$$

To ensure $\beta_{f,m^-} \leq \frac{1}{2}$, we require

$$(1 - e_H p_H)\frac{c}{p_H}\beta_0 \leq (1 - \beta_0)\pi(m^-|f, L). \tag{37}$$

To ensure this inequality holds for a wider range of parameters, we set $\pi(m^-|f, L) = 1$. Then, solving for e_H yields

$$1 - e_H p_H \leq (1 - \beta_0)\frac{p_H}{c\beta_0}. \tag{38}$$

Given that this equilibrium can only be politician-optimal if $\beta_0 > \frac{p_H}{p_H + c - cp_H}$, we make this assumption. Note, however, that the LHS above is decreasing in e_H. Hence, if the equilibrium exists for $e_H \in (0, 1)$, it would also exist for $e_H = 1$. But if $e_H = 1$, we have

$$1 - p_H \leq (1 - \beta_0)\frac{p_H}{c\beta_0} \iff \beta_0 \leq \frac{p_H}{p_H + c - cp_H}, \tag{39}$$

a contradiction. Hence, an equilibrium where the low type shirks and the high type exerts effort is never politician-optimal.

Case 5. $\beta_0 > \frac{p_H}{p_H+c-cp_H}$ If $\beta_0 > \frac{p_H}{p_H+c-cp_H}$ and both types shirk, both types' equilibrium payoff is 1. Neither type has a profitable deviation as equilibrium payoffs are at their maximum. □

B Proofs for Extension 1: Dynamic Reputation Management

B.1 Proof of Lemma 1

Before we prove the general statement of the proposition, we prove several lemmata that aid in the analysis. First, we prove what payoff the politician expects as a function of the induced posterior in the beginning of the second period and the bureaucrat's type.

Lemma 1. *The continuation value for the politician in the second period for each induced posterior $\beta_1 \in [0, 1]$ and type H and L respectively are as follows.*

$$V_P(\beta_1, H) = \begin{cases} p_H & \text{if } \beta_1 \in \left(0, \frac{p_H}{p_H+c-p_H c}\right] \\ 0 & \text{otherwise.} \end{cases} \tag{40}$$

$$V_P(\beta_1, L) = \begin{cases} \frac{\beta_1 p_H}{(1-\beta_1)c} p_L & \text{if } \beta_1 \in \left(0, \frac{c}{c+p_H}\right) \\ p_L & \text{if } \beta_1 \in \left[\frac{c}{c+p_H}, \frac{p_H(1-p_L)}{p_H(1-p_L)+c(1-p_H)}\right] \\ \frac{p_H+\beta_1(cp_H-c-p_H)}{(1-\beta_1)p_H p_L} p_L & \text{if } \beta_1 \in \left(\frac{p_H(1-p_L)}{p_H(1-p_L)+c(1-p_H)}, \frac{p_H}{p_H+c-p_H c}\right) \\ 0 & \text{otherwise.} \end{cases} \tag{41}$$

Proof. For each induced posterior β_1, the equilibrium strategy of the bureaucrat is as in

Proposition 2. Define $e(\beta_1, t)$ as the bureaucrat's mixed effort level as a function of β_1 and his type $t \in \{L, H\}$. Conditional on the bureaucrat's type t, the politician's continuation value for each β_1 equals $e(\beta_1, t)p_t$.

Consider first type H for $\beta_1 \in [0, 1]$.

- If $\beta_1 \in \left(0, \frac{p_H}{p_H + c - p_H c}\right]$, then type H exerts effort. This means the politician's continuation value is $V_P(\beta_1, H) = p_H$.
- Otherwise, type H shirks, and the politician's continuation value is 0.

Now consider type L for $\beta_1 \in [0, 1]$.

- If $\beta_1 \in \left(0, \frac{c}{c+p_H}\right)$, L mixes $e(\beta_1, L) = \frac{\beta_1 p_H}{(1-\beta_1)c}$, which means that the politician's continuation value is $\frac{\beta_1 p_H}{(1-\beta_1)c} p_L$.
- If $\beta_1 \in \left[\frac{c}{c+p_H}, \frac{p_H(1-p_L)}{p_H(1-p_L)+c(1-p_H)}\right]$, L exerts effort, and thus $V_P(\beta_1, L) = p_L$.
- If $\beta_1 \in \left(\frac{p_H(1-p_L)}{p_H(1-p_L)+c(1-p_H)}, \frac{p_H}{p_H+c-p_H c}\right)$, L mixes $e(\beta_1, L) = \frac{p_H + \beta_1(cp_H - c - p_H)}{(1-\beta_1)p_H p_L}$, which implies that $V_P(\beta_1, L) = \frac{p_H + \beta_1(cp_H - c - p_H)}{(1-\beta_1)p_H p_L} p_L$.
- Otherwise, L shirks, and the politician's continuation value is 0 for the remaining β_1.

This establishes the politician's continuation value for each $\beta_1 \in [0, 1]$ and both types. $\square$

B.2 Proof of Lemma 2

Lemma 2. *In the first period at every information set after success and failure, the politician either remains silent, or induces posteriors β_1' and β_1'', where $\beta_1' \neq \beta_1''$, such that both $V_P(\beta_1', H) = V_P(\beta_1'', H)$ and $V_P(\beta_1', L) = V_P(\beta_1'', L)$.*

Proof. In every equilibrium in the first period, the politician must be indifferent between every message and its induced resulting posterior reputation β_1. If not, then the politician cannot mix between two different messages. Thus, if no such two posteriors can be induced, the politician must remain silent and always induce the same posterior. $\square$

B.3 Proof of Lemma 3

Lemma 3. *In the first period, if the politician actively signals and induces different posteriors, then the following induced posteriors β_1' and β_1'' ($\beta_1' < \beta_1''$) are potentially feasible in equilibrium:*

- $\frac{c}{p_H+c} \leq \beta_1' < \beta_1'' \leq \frac{p_H(1-p_L)}{p_H(1-p_L)+c(1-p_H)}$,
- $\beta_1' \in \left(0, \frac{c}{p_H+c}\right)$ *and* $\beta_1'' \in \left(\frac{p_H(1-p_L)}{p_H(1-p_L)+c(1-p_H)}, \frac{p_H}{p_H+c-p_H c}\right)$ *such that* $e(\beta_1', L) = e(\beta_1'', L)$,
- $\frac{p_H}{p_H+c-p_H c} < \beta_1' < \beta_1'' \leq 1$,
- $\beta_1' = 0$ *and* $\frac{p_H}{p_H+c-p_H c} < \beta_1'' \leq 1$.

Otherwise, the politician remains silent.

Proof. Suppose the politician actively blames and/or gives credit and wants to induce β_1' and β_1'' where $0 \leq \beta_1' < \beta_1'' \leq 1$. By Lemma 2, we need that both (i) $V_P(\beta_1', H) = V_P(\beta_1'', H)$ and (ii) $V_P(\beta_1', L) = V_P(\beta_1'', L)$. Constraint (i.) implies that, by Lemma 1, there are three possibilities:

- $0 < \beta_1' < \beta_1'' \leq \frac{p_H}{p_H + c - p_H c}$,
- $\frac{p_H}{p_H + c - p_H c} < \beta_1' < \beta_1'' \leq 1$,
- $\beta_1' = 0$ and $\frac{p_H}{p_H + c - p_H c} < \beta_1'' \leq 1$.

Further, constraint (ii.) implies that, also by Lemma 1, there are four possibilities:

- $\frac{c}{p_H + c} \leq \beta_1' < \beta_1'' \leq \frac{p_H(1-p_L)}{p_H(1-p_L) + c(1-p_H)}$,
- $\beta_1' \in \left(0, \frac{c}{p_H + c}\right)$ and $\beta_1'' \in \left(\frac{p_H(1-p_L)}{p_H(1-p_L) + c(1-p_H)}, \frac{p_H}{p_H + c - p_H c}\right)$ such that $e(\beta_1', L) = e(\beta_1'', L)$,
- $\frac{p_H}{p_H + c - p_H c} < \beta_1' < \beta_1'' \leq 1$,
- $\beta_1' = 0$ and $\frac{p_H}{p_H + c - p_H c} < \beta_1'' \leq 1$.

Taken together, we observe that constraint (ii.) is more restrictive in all cases, which proves the claim of the Lemma. If the politician is not indifferent between the two induced posteriors, she must always send the same message, which is denoted by m^0. □

B.4 Proof of Proposition 3

Proof. We begin by analyzing equilibria as a function of the prior reputation $\beta_0 \in (0, 1)$.

Case 1: $\beta_0 \in \left[\frac{p_L}{p_H + p_L}, \frac{1-p_L}{2-p_H-p_L}\right]$. By Proposition 1, if the politician remains silent, there exists an equilibrium where the bureaucrat exerts effort in the first period. In the second period, either $\beta_1 = \beta_s$ or $\beta_1 = \beta_f$ after a first-period success and failure respectively, inducing effort as determined by Proposition 2, substituting in β_1 for β_0. The politician could only potentially do better if there is more effort in the second period. This is, however, not possible, because any active politician signaling can only spread out posteriors more, leading to less effort in the second period. As a result, for all such β_0, the equilibrium is that the bureaucrat exerts effort in the first period, and second-period effort levels are determined by Proposition 2.

Case 2: $\beta_0 < \frac{p_L}{p_H + p_L}$. In this case, $\beta_s < \frac{1}{2}$ if both types exert effort, and a success is not rewarded. Without politician signaling, the bureaucrat shirks in the first period and second-period effort is $e(\beta_0)$. The goal is to check how politician signaling may increase the politician's welfare. We sub divide case 2 into multiple cases.

Case 2a: Full Effort. This requires that after a success, m^+ is sent such that the bureaucrat obtains a good reputation. To maximize the probability of m^+ being sent for both types subject to $\beta_{s,m^+} \geq \frac{1}{2}$, it is optimal to set $\beta_{s,m^+} = \frac{1}{2}$ and $\beta_{s,m^-} = \underline{\beta}$. This creates the strongest incentives for the bureaucrat of both types to exert effort. Via Bayes' rule,

$$\frac{p_H \beta_0 \pi(m^+|s,H)}{p_H \beta_0 \pi(m^+|s,H) + p_L(1-\beta_0)\pi(m^+|s,L)} = \frac{1}{2}, \tag{42}$$

$$\frac{p_H \beta_0 \pi(m^-|s,H)}{p_H \beta_0 \pi(m^-|s,H) + p_L(1-\beta_0)\pi(m^-|s,L)} = \frac{c}{c+p_H}, \tag{43}$$

which implies that

$$\pi(m^+|s,H) = \frac{\beta_0 p_H^2 - c p_L(1-\beta_0)}{\beta_0 p_H (p_H - c)}, \tag{44}$$

$$\pi(m^+|s,L) = \frac{\beta_0(p_H^2 + c p_L) - c p_L}{(1-\beta_0)(p_H - c)p_L}. \tag{45}$$

Substituting in the above values for both bureaucrat types' effort conditions, we need

$$\pi(m^+|s,H) \geq \frac{c}{p_H} \iff \beta_0 \geq \frac{c p_L}{c(c - p_H + p_L) + p_H^2}, \tag{46}$$

$$\pi(m^+|s,L) \geq \frac{c}{p_L} \iff \beta_0 \geq \frac{c(p_H + p_L - c)}{c(p_H + p_L - c) + p_H^2}. \tag{47}$$

The low type's constraint on β_0 is the binding one. Hence, there exists an equilibrium where the bureaucrat exerts effort in the first period if $\beta_0 \in \left[\frac{c(p_H+p_L-c)}{c(p_H+p_L-c)+p_H^2}, \frac{p_L}{p_H+p_L}\right]$.

Case 2b. $\beta_0 < \frac{c(p_H+p_L-c)}{c(p_H+p_L-c)+p_H^2}$ **and** $e_H = 1$ **and** $e_L \in (0,1)$ **in period 1.** Alternatively, consider a next best possible equilibrium where in the first round $e_L \in (0,1)$ and $e_H = 1$. Then, via Bayes' rule, we have

$$\beta_{s,m^+} = \frac{p_H \beta_0 \pi(m^+|s,H)}{p_H \beta_0 \pi(m^+|s,H) + p_L(1-\beta_0) e_L \pi(m^+|s,L)}, \tag{48}$$

$$\beta_{s,m^-} = \frac{p_H \beta_0 (1 - \pi(m^+|s,H))}{p_H \beta_0 (1 - \pi(m^+|s,H)) + p_L(1-\beta_0) e_L (1 - \pi(m^+|s,L))}. \tag{49}$$

To guarantee the low type mixes while also giving strongest possible incentives, we have $\pi(m^+|s,L) = \frac{c}{p_L}$. Failure leads to a belief below the threshold. Success leads to a belief above the threshold as long as e_L is sufficiently low. Given a success belief β_s, it is optimal to set $\beta_{s,m^+} = \frac{1}{2}$ and $\beta_{s,m^-} = \frac{c}{p_H+c}$ so that there is effort by both types in the second period following a first period success. Solving this system of equations for every $\beta_0 \in \left(0, \frac{c(p_H+p_L-c)}{c(p_H+p_L-c)+p_H^2}\right)$ generates $\pi(m^+|s,H) = \frac{p_H}{p_H+p_L-c}$ and effort level $e_L = \frac{\beta_0 p_H^2}{(1-\beta_0)c(p_H+p_L-c)}$. Thus, there exists an equilibrium where

- In the first period, the low type mixes and the high type exerts effort.
- After a first period success, there is effort in the second period by both types.
- After a first period failure, effort levels are $e(\beta_f)$, where β_f is generated by $e_L =$

$\frac{\beta_0 p_H^2}{(1-\beta_0)c(p_H+p_L-c)}$.

Case 3: $\beta_0 > \frac{1-p_L}{2-p_H-p_L}$. In the third case, if both types were to exert effort and the politician would remain silent, then $\beta_f > \frac{1}{2}$, and a failure is not punished. Absent politician signaling, the bureaucrat shirks in the first period and second-period effort is $e_2(\beta_0)$. Again, the goal is to check how politician signaling may increase the politician's welfare. We subdivide case 3 into multiple cases.

Case 3a: effort in period 1, silence after success Given effort, we have that $\beta_f > \frac{1}{2}$, and the goal is to create the widest range of β_0 such that both types are willing to exert effort. We need that both $\pi(m^-|f,H) \geq \frac{c}{p_H}$ and $\pi(m^-|f,L) \geq \frac{c}{p_L}$. Lemma 3 implies that there are three possibilities:

1. $\beta_{f,m^-} = 0$, $\beta_{f,m^+} \geq \frac{p_H}{p_H+c-p_H c}$. In this case, $\pi(m^+|f,H) = 1$, which contradicts the high type's incentive compatibility condition.
2. $\beta_{f,m^-} < \frac{c}{p_H+c}$, $\beta_{f,m^+} \in \left(\frac{p_H(1-p_L)}{p_H(1-p_L)+c(1-p_H)}, \frac{p_H}{p_H+c-p_H c}\right)$ such that second period effort following (f,m^-) and (f,m^+) is the same. The minimum value β_{f,m^-} can take is the LHS, which needs to be smaller than $\frac{c}{p_H+c}$

$$\frac{(1-p_H)\frac{c}{p_H}\beta_0}{(1-p_H)\frac{c}{p_H}\beta_0 + (1-p_L)(1-\beta_0)} \geq \frac{c}{p_H+c}, \tag{50}$$

   which is a contradiction.
3. $\beta_{f,m^-} \in \left[\frac{c}{p_H+c}, \frac{1}{2}\right]$ and $\beta_{f,m^+} \in \left[\frac{1}{2}, \frac{p_H(1-p_L)}{p_H(1-p_L)+c(1-p_H)}\right]$ such that after a first period failure, both types exert effort in the second period. We know that the lower bound for β_{f,m^-} forms no constraint as long as signaling strategies satisfy both types' incentive compatibility constraints. Thus, the first constraint is that $\beta_{f,m^-} \leq \frac{1}{2}$

$$(1-p_H)\pi(m^-|f,H)\beta_0 \leq (1-p_L)\pi(m^-|f,L)(1-\beta_0) \tag{51}$$

   The second constraint is $\beta_{f,m^+} \leq \frac{p_H(1-p_L)}{p_H(1-p_L)+c(1-p_H)}$.

$$\frac{(1-p_H)\pi(m^+|f,H)\beta_0}{(1-p_H)\pi(m^+|f,H)\beta_0 + (1-p_L)\pi(m^+|f,L)(1-\beta_0)} \leq \frac{p_H(1-p_L)}{p_H(1-p_L)+c(1-p_H)} \tag{52}$$

   We proceed by setting equation 51 with equality, which generates $\pi(m^-|f,H) = \frac{(1-p_L)\pi(m^-|f,L)(1-\beta_0)}{(1-p_H)\beta_0}$, and then substituting this into equation 52:

$$\frac{(1-p_H)(1-\pi(m^-|f,H))\beta_0}{(1-p_H)(1-\pi(m^-|f,H))\beta_0 + (1-p_L)(1-\pi(m^-|f,L))(1-\beta_0)} \leq \frac{p_H(1-p_L)}{p_H(1-p_L)+c(1-p_H)} \tag{53}$$

   For $\beta_0 \leq \frac{p_H(1-2p_L+p_L^2)}{p_H(1-2p_L+p_L^2)+c(1-2p_H+p_H^2)}$ there exists $\pi(m^-|f,L) \in (0,1)$ that satisfies this.

Case 3a: $\beta_s \leq \bar{\beta}$. In this case, there is a potential for there to be effort in both periods. To induce effort for the widest range of priors β_0, it is optimal to set $\beta_{f,m^-} = \frac{1}{2}$ and $\beta_{f,m^+} = \bar{\beta}$. Bayes' rule implies that

$$\frac{\beta_0 \pi(m^-|f,H)(1-p_H)}{\beta_0 \pi(m^-|f,H)(1-p_H) + (1-\beta_0)\pi(m^-|f,L)(1-p_L)} = \frac{1}{2}, \tag{54}$$

$$\frac{\beta_0 \pi(m^+|f,H)(1-p_H)}{\beta_0 \pi(m^+|f,H)(1-p_H) + (1-\beta_0)\pi(m^+|f,L)(1-p_L)} = \bar{\beta}. \tag{55}$$

This generates the following strategy:

$$\pi(m^+|f,H) = \frac{p_H(1-p_L)(1-p_L-\beta_0(2-p_H-p_L))}{\beta_0(1-p_H)(c(1-p_H)-p_H+p_Hp_L)}, \tag{56}$$

$$\pi(m^+|f,L) = \frac{c(1-p_H)(1-p_L-\beta_0(2-p_H-p_L))}{(1-\beta_0)(1-p_L)(c-cp_H-p_H(1-p_L))}. \tag{57}$$

For the high type to exert effort, we require

$$1-\pi(m^+|f,H) \geq \frac{c}{p_H} \iff \beta_0 \leq \frac{p_H^2(1-p_L)^2}{p_H^2(1-p_L)^2 - c^2(1-p_H)^2 + c(1-p_H)p_H(2-p_H-p_L)}. \tag{58}$$

For the low type to exert effort, we require

$$1-\pi(m^+|f,L) \geq \frac{c}{p_L} \iff \beta_0 \leq \frac{(1-p_L)(c^2(1-p_H)-cp_H)+(1+c)p_Hp_L-p_Hp_L^2)}{c^2(1-p_H)(1-p_L)+p_H(1-p_L)^2p_L-c(p_H-p_L)(1-p_Hp_L)}. \tag{59}$$

The low type's constraint is the binding one, which as a result determines the upper bound for β_0 to achieve effort in both periods. For the proposition, we define

$$\bar{\beta}^*(c,p_L,p_H) := \frac{(1-p_L)(c^2(1-p_H)-cp_H)+(1+c)p_Hp_L-p_Hp_L^2)}{c^2(1-p_H)(1-p_L)+p_H(1-p_L)^2p_L-c(p_H-p_L)(1-p_Hp_L)}. \tag{60}$$

Case 3b. $\beta_0 > \bar{\beta}^*(c,p_L,p_H)$**.** In this case, there is no equilibrium with effort in the first period. First, consider the potential for an equilibrium in which in the first period $e_H = 1$ and $e_L \in (0,1)$. There are two cases, where the politician remains silent following a success or not.

1. In the first case, the politician remains silent following a success. Then, to ensure that the low type mixes, we require $\pi(m^+|f,L) = 1 - \frac{c}{p_L}$. Also, to ensure that the high type exerts effort, we need $\pi(m^+|f,H) \leq 1 - \frac{c}{p_H}$. Further, we need that $\beta_{f,m^-} \leq \frac{1}{2}$. To

ensure greatest incentives to exert effort, we set $\beta_{f,m^-} = \frac{1}{2}$. This implies that

$$\frac{(1-p_H)(1-\pi(m^+|f,H))\beta_0}{(1-p_H)(1-\pi(m^+|f,H))\beta_0 + (1-ep_L)\frac{c}{p_L}(1-\beta_0)} = \frac{1}{2} \tag{61}$$

$$(1-p_H)(1-\pi(m^+|f,H))\beta_0 = (1-ep_L)\frac{c}{p_L}(1-\beta_0) \tag{62}$$

Further, we require that $\beta_{f,m^+} \leq \frac{p_H(1-p_L)}{p_H(1-p_L)+c(1-p_H)}$, for which it is optimal to set it with equality. This yields the following solutions:

$$\tag{63}$$

Consider a case where $\beta_{s,m^+} \geq \frac{p_H}{p_H+c-p_Hc}$. Then after (s,m^+) there is no effort in the second period. This implies that after (s,m^-) there also must be no second period effort, which further implies that $\beta_{s,m^-} = 0$, and $\pi(m^+|s,H) = 1$. Taken together, this case where $\beta_{s,m^+} \geq \frac{p_H}{p_H+c-p_Hc}$ occurs if

$$\frac{p_H\beta_0}{p_H\beta_0 + e_Lp_L(1-\beta_0)\pi(m^+|s,L)} \geq \frac{p_H}{p_H+c-p_Hc}. \tag{64}$$

To ensure incentive compatibility in the first period, we require

$$\begin{aligned} \pi(m^+|s,L) &= \pi(m^+|f,L) + \frac{c}{p_L}, \\ \pi(m^+|s,H) = 1 &\geq \pi(m^+|f,H) + \frac{c}{p_H}. \end{aligned} \tag{65}$$

1. If $\beta_0 \in \left[\frac{p_L}{p_L+c-cp_H}, \frac{p_H}{p_H+c-p_Hc}\right)$, then any $e_L \in [0,1]$ and $\pi(m^+|s,L)$ are feasible, as long as they satisfy incentive compatibility. Thus, the only constraints are on the beliefs following a failure, β_{f,m^-} and β_{f,m^+}. Then, to ensure that $\beta_{f,m^-} \leq \frac{1}{2}$, we need that

$$\frac{(1-p_H)\beta_0\pi(m^-|f,H)}{(1-p_H)\beta_0\pi(m^-|f,H) + (1-e_Lp_L)\pi(m^-|f,L)(1-\beta_0)} \leq \frac{1}{2} \iff \tag{66}$$

$$(1-p_H)\beta_0\pi(m^-|f,H) \leq (1-e_Lp_L)\pi(m^-|f,L)(1-\beta_0) \iff \tag{67}$$

$$e_L \leq \frac{\pi(m^-|f,L) - \beta_0(\pi(m^-|f,H)(1-p_H) + \pi(m^-|f,L))}{\pi(m^-|f,L)p_L(1-\beta_0)} \tag{68}$$

In addition, we require that $\beta_{f,m^+} \leq \frac{p_H(1-p_L)}{p_H(1-p_L)+c(1-p_H)} := \bar{\beta}$:

$$\frac{(1-p_H)\beta_0\pi(m^+|f,H)}{(1-p_H)\beta_0\pi(m^+|f,H)+(1-e_Lp_L)\pi(m^+|f,L)(1-\beta_0)} \leq \bar{\beta} \iff \tag{69}$$

$$\frac{(1-p_H)\beta_0\pi(m^+|f,H)(1-\bar{\beta})}{\bar{\beta}\pi(m^+|f,L)(1-\beta_0)} \leq 1-e_Lp_L \iff \tag{70}$$

$$\frac{1}{p_L} - \frac{(1-p_H)\beta_0\pi(m^+|f,H)(1-\bar{\beta})}{p_L\bar{\beta}\pi(m^+|f,L)(1-\beta_0)} \geq e_L \tag{71}$$

The goal is to create the widest range of e_L, so that we can set e_L as high as possible.

By similar arguments as before, it is optimal to set posteriors following a failure as $(\beta_{f,m^-}, \beta_{f,m^+}) = (\frac{1}{2}, \bar{\beta})$, where beliefs are also determined by mixed effort levels. Also, for there to be signaling following a success, both messages have to induce the same second period effort level. As $\beta_{s,m^-} \leq \frac{1}{2}$, would be necessary, and given that $\beta_{s,m^+} \geq \bar{\beta}$, this would imply $\beta_{s,m^-} \leq \frac{c}{p_H+c}$. To give the strongest incentives, let $\beta_{s,m^-} = 0$. Then $\pi(m^-|s,H) = 0$, and $\pi(m^+|s,H) = 1$. This implies that $\pi(m^+|f,H) \leq 1 - \frac{c}{p_H}$.

Case 4. $\beta_0 \geq \frac{p_H}{p_H+c-p_Hc}$. Suppose $\beta_0 \geq \frac{p_H}{p_H+c-p_Hc}$. Without active politician signaling, the bureaucrat shirks. If the politician sends messages, he must be indifferent between the two messages. Case 2b of the Proof of Proposition 2 shows, however, that there can be no equilibrium with effort under this condition, so the same must also be true with another period, given that it adds additional constraint for the politician. So for this range of β_0, both types shirk in period 1 and period 2.

C Proofs for Extension 2: Politician Reputation Concerns

C.1 Proof of Proposition 4

Case 1. $\beta_0 \in \left[0, \frac{p_L}{p_H+p_L}\right) \cup \left(\frac{1-p_L}{2-p_H-p_L}, 1\right]$. Suppose the bureaucrat exerts effort. The politician guarantees strongest incentives to exert effort by fully revealing who took the decision. If $\beta_0 < \frac{p_L}{p_H+p_L}$, this means that a success leads to posterior $\beta_s = \frac{\beta_0 p_H}{\beta p_H+(1-\beta_0)p_L}$, which, by rearranging does not lead to a belief that exceeds the threshold of $\frac{1}{2}$. Thus, in this case, the bureaucrat has a profitable deviation, and there is no equilibrium with effort (nor mixing). Likewise, if $\beta_0 > \frac{1-p_L}{2-p_H-p_L}$, then a failure leads to posterior $\beta_f = \frac{\beta_0(1-p_H)}{\beta_0(1-p_H)+(1-\beta_0)(1-p_L)}$, which by rearranging does not lead to a belief that is lower than the threshold of $\frac{1}{2}$. Hence, the bureaucrat shirks.

Case 2. $\beta_0 \in \left[\frac{p_L+h_s(\gamma_0)}{p_H+p_L+2h_s(\gamma_0)}, \frac{2-p_L-h_s(\gamma_0)}{4-p_H-p_L-2h_s(\gamma_0)}\right]$. In this case, suppose the politician remains silent. To ensure effort, we require that $\beta_s \geq \frac{1}{2} \geq \beta_f$ given silence. Then beliefs following a success and failure as follows: Via Bayes' rule, we know that

$$\beta_s = \frac{(\frac{1}{2}p_H + \frac{1}{2}h_s(\gamma))\beta_0}{(\frac{1}{2}p_H + \frac{1}{2}h_s(\gamma))\beta_0 + (\frac{1}{2}p_L + \frac{1}{2}h_s(\gamma))(1-\beta_0)} = \frac{(p_H + h_s(\gamma))\beta_0}{g_s(\beta_0) + h_s(\gamma_0)}. \tag{72}$$

Via a similar simplification, after a failure we have

$$\beta_f = \frac{(2 - p_H - h_s(\gamma))\beta_0}{2 - g_s(\beta_0) - h_s(\gamma)}. \tag{73}$$

Rearranging yields the result that if $\beta_0 \in \left[\frac{p_L + h_s(\gamma)}{p_H + p_L + 2h_s(\gamma)}, \frac{2 - p_L - h_s(\gamma)}{4 - p_H - p_L - 2h_s(\gamma)}\right]$, then there is effort.

Case 3. $\beta_0 \in \left[\frac{p_L}{p_H+p_L}, \frac{p_L + h_s(\gamma_0)}{p_H+p_L+2h_s(\gamma_0)}\right)$**.** Absent politician signaling, the bureaucrat would shirk. To ensure the bureaucrat exerts effort, we require that $\beta_{s,m^+} \geq \frac{1}{2}$ and $\pi(m^+|s,B) \geq \frac{c}{g_s(\beta_0)}$. First, with full information revelation after a success, the bureaucrat would exert effort. The politician faces no incentive constraints in full information revelation after a success if $\gamma_0 \leq \frac{p_L}{p_H+p_L}$ or $\gamma_0 \geq \frac{1}{2}$. The reason is that if the public knows the bureaucrat was responsible, the politician keeps reputation γ_0, while if the public knows the politician was responsible, the politician obtains reputation $\frac{\gamma_0 p_H}{\gamma_0 p_H + (1-\gamma_0)p_L}$. Under the above conditions, the politician faces no constraints in sending messages. That is, if $\gamma_0 \leq \frac{p_L}{p_H+p_L}$, then $\frac{\gamma_0 p_H}{\gamma_0 p_H + (1-\gamma_0)p_L} \leq \frac{1}{2}$, which means that the politician's messages give the politician an equal payoff associated with being below the reputation threshold. Similarly, if $\gamma_0 \geq \frac{1}{2}$, then the politician's messages give the politician an equal payoff associated being above the reputation threshold. However, if $\gamma_0 \in \left(\frac{p_L}{p_H+p_L}, \frac{1}{2}\right)$, then full information revelation is not possible as one politician reputation will be lower than the threshold, and one higher. The analysis differs if $\gamma_{s,m^0} < \frac{1}{2}$ and if $\gamma_{s,m^0} > \frac{1}{2}$.

Case 3.1. $\gamma_{s,m^0} < \frac{1}{2}$**.** This case occurs if

$$\gamma_{s,m^0} := \frac{(p_H + g_s(\beta_0))\gamma_0}{h_s(\gamma_0) + g_s(\beta_0)} < \frac{1}{2} \iff \gamma_0 < \frac{p_L + g_s(\beta_0)}{p_H + p_L + 2g_s(\beta_0)}. \tag{74}$$

In this case, m^- will increase the politician's reputation. It can at most be $\frac{1}{2}$ to ensure politician incentive compatibility. In addition to this, $\pi(m^+|s,B)$ must at least be $\frac{c}{g_s(\beta)}$ to ensure bureaucrat incentive compatibility, and $\beta_{s,m^+} \geq \frac{1}{2}$ to give the bureaucrat a good reputation following (s, m^+).

To summarize, there are three conditions:

$$\frac{(p_H\pi(m^-|s,P) + g_s(\beta_0)\pi(m^-|s,B))\gamma_0}{h_s(\gamma_0)\pi(m^-|s,P) + g_s(\beta_0)\pi(m^-|s,B)} \leq \frac{1}{2} \tag{75}$$

$$\frac{(p_H\pi(m^+|s,B) + h_s(\gamma_0)\pi(m^+|s,P))\beta_0}{g_s(\beta_0)\pi(m^+|s,B) + h_s(\gamma_0)\pi(m^+|s,P)} \geq \frac{1}{2} \tag{76}$$

$$\pi(m^+|s,B) \geq \frac{c}{g_s(\beta_0)}. \tag{77}$$

There are two ranges of (γ_0, β_0) that we consider which span case 3.1 completely:

1. *Case 1.* $\gamma_0 \in \left(\frac{p_L}{p_H+p_L}, \frac{3p_Hp_L+p_L^2}{p_H^2+6p_Hp_L+p_L^2}\right]$ and $\beta_0 \in \left(\frac{p_L}{p_H+p_L}, \frac{p_L+h_s(\gamma_0)}{p_H+p_L+2h_s(\gamma_0)}\right)$.
2. *Case 2.* $\gamma_0 \in \left(\frac{3p_Hp_L+p_L^2}{p_H^2+6p_Hp_L+p_L^2}, \frac{\sqrt{p_Hp_L}-p_L}{p_H-p_L}\right)$ and $\beta_0 \in \left(\frac{\gamma_0 p_H-2p_L+3\gamma_0 p_L}{(p_H-p_L)(1-2\gamma_0)}, \frac{p_L+h_s(\gamma_0)}{p_H+p_L+2h_s(\gamma_0)}\right)$.

In both cases, the equations can be simultaneously met if

$$c \leq \frac{(1-2\beta_0)(g_s(\beta_0))(h_s(\gamma_0))(\beta_0(2\gamma_0-1)(p_H-p_L)-2p_L+\gamma_0(p_H+3p_L))}{2(\gamma_0-\beta_0(1-2\gamma_0))(p_H-p_L)((\gamma_0-1)p_L+\beta_0(\gamma_0(p_H-p_L)+p_L))}. \tag{78}$$

Case 3.2. $\gamma_{s,m^0} > \frac{1}{2}$**.** This occurs if

$$\gamma_s := \frac{(p_H+g_s(\beta_0))\gamma_0}{h_s(\gamma_0)+g_s(\beta_0)} > \frac{1}{2} \iff \gamma_0 > \frac{p_L+g_s(\beta_0)}{p_H+p_L+2g_s(\beta_0)}. \tag{79}$$

In this case, m^+ would decrease the politician's reputation and potentially bring it below the threshold of $\frac{1}{2}$. This generates a constraint, as the politician must find it incentive compatible to send both messages. There are now the following three conditions:

$$\frac{(p_H\pi(m^+|s,P)+g_s(\beta_0)\pi(m^+|s,B))\gamma_0}{h_s(\gamma_0)\pi(m^+|s,P)+g_s(\beta_0)\pi(m^+|s,B)} \geq \frac{1}{2}, \tag{80}$$

$$\frac{(p_H\pi(m^+|s,B)+h_s(\gamma_0)\pi(m^+|s,P))\beta_0}{g_s(\beta_0)\pi(m^+|s,B)+h_s(\gamma_0)\pi(m^+|s,P)} \geq \frac{1}{2}, \tag{81}$$

$$\pi(m^+|s,B) \geq \frac{c}{g_s(\beta_0)}. \tag{82}$$

These equations can be met if $\gamma_0 \in \left(\frac{\sqrt{p_Hp_L}-p_L}{p_H-p_L}, \frac{1}{2}\right)$ and $\beta_0 \in \left[\frac{p_L(1-\gamma_0)}{p_L(1-\gamma_0)+\gamma_0 p_H}, \frac{p_L+h_s(\gamma_0)}{p_H+p_L+2h_s(\gamma_0)}\right]$, in which case there is effort. Otherwise, if (i) $\gamma_0 \in \left[\frac{3p_Hp_L+p_L^2}{p_H^2+6p_Hp_L+p_L^2}, \frac{\sqrt{p_Hp_L}-p_L}{p_H-p_L}\right]$ or (ii) $\gamma_0 \in \left(\frac{\sqrt{p_Hp_L}-p_L}{p_H-p_L}, \frac{1}{2}\right)$ and $\beta_0 \in \left(\frac{p_L}{p_H+p_L}, \frac{p_L(1-\gamma_0)}{p_L(1-\gamma_0)+\gamma_0 p_H}\right)$, there is no effort.

Case 4. $\beta_0 \in \left(\frac{2-p_L-h_s(\gamma_0)}{4-p_H-p_L-2h_s(\gamma_0)}, \frac{1-p_L}{2-p_H-p_L}\right]$**.** In this case, absent politician signaling, failure does not lead to a bureaucrat reputation below the threshold. Hence, the politician must send message m^- sufficiently often following a failure to provide proper incentives to the bureaucrat. Message m^- increases the politician's reputation following a failure, and as in case 3, it must be the case that m^- and m^+ yield the same politician reputation, either both above or below the threshold. Now, if $\gamma_0 \leq \frac{1}{2}$, then a failure leads to politician reputation $\gamma_{f,m^0} < \frac{1}{2}$, and at most, m^- will push the politician's reputation to $\gamma_0 \leq \frac{1}{2}$. As a result, there exists an equilibrium with effort where the politician fully reveals who was responsible for the failure. Likewise, if $\gamma_0 \geq \frac{1-p_L}{2-p_H-p_L}$, then a fully revealing message will generate reputations $\frac{1}{2} \leq \gamma_{f,m^+} < \gamma_{f,m^-}$, which in turns always enables the politician to provide incentives to the bureaucrat. Hence, we focus on the case where $\gamma_0 \in \left(\frac{1}{2}, \frac{1-p_L}{2-p_H-p_L}\right)$ Again, the analysis differs depending on whether $\gamma_f < \frac{1}{2}$ or $\gamma_f > \frac{1}{2}$.

Case 4.1. $\gamma_f < \frac{1}{2}$. The fact that $\gamma_f < \frac{1}{2}$ implies that

$$\frac{((1-p_H)+(1-g_s(\beta_0)))\gamma_0}{(1-h_s(\gamma_0))+(1-g_s(\beta_0))} \leq \frac{1}{2} \iff 2\gamma_0(2-p_H-g_s(\beta_0)) \leq 2-(p_H\gamma_0+p_L(1-\gamma_0))-g_s(\beta_0) \tag{83}$$

$$\gamma_0(4-2p_H-2g_s(\beta_0)+p_H-p_L) \leq 2-p_L-g_s(\beta_0) \tag{84}$$

$$\gamma_0 \leq \frac{2-p_L-g_s(\beta_0)}{4-p_H-p_L-2g_s(\beta_0)}. \tag{85}$$

In this case, there are three conditions:

$$\frac{((1-p_H)\pi(m^-|f,P)+(1-g_s(\beta_0))\pi(m^-|f,B))\gamma_0}{(1-h_s(\gamma_0))\pi(m^-|f,P)+(1-g_s(\beta_0))\pi(m^-|f,B)} \leq \frac{1}{2}, \tag{86}$$

$$\frac{((1-p_H)\pi(m^-|f,B)+(1-h_s(\gamma_0))\pi(m^-|f,P))\beta_0}{(1-g_s(\beta_0))\pi(m^-|f,B)+(1-h_s(\gamma_0))\pi(m^-|f,P)} \leq \frac{1}{2}, \tag{87}$$

$$\pi(m^-|f,B) \geq \frac{c}{g_s(\beta_0)}. \tag{88}$$

These equations can be simultaneously solved if $\gamma_0 \in \left(\frac{1}{2}, \frac{1-p_L-\sqrt{1-p_H-p_L+p_Hp_L}}{p_H-p_L}\right)$ and $\beta_0 \in \left(\frac{2-\gamma_0p_H-2p_L+\gamma_0p_L}{4-p_H-2\gamma_0p_H-3p_L+2\gamma_0p_L}, \frac{1-\gamma_0-p_L(1-\gamma_0)}{1-\gamma_0p_H-p_L(1-\gamma_0)}\right]$.

Case 4.2. $\gamma_f > \frac{1}{2}$. In this case, there are the following three conditions:

$$\frac{((1-p_H)\pi(m^+|f,P)+(1-g_s(\beta_0))\pi(m^+|f,B))\gamma_0}{(1-h_s(\gamma_0))\pi(m^+|f,P)+(1-g_s(\beta_0))\pi(m^+|f,B)} \geq \frac{1}{2}, \tag{89}$$

$$\frac{((1-p_H)\pi(m^-|f,B)+(1-h_s(\gamma))\pi(m^-|f,P))\beta_0}{(1-g_s(\beta_0))\pi(m^-|f,B)+(1-h_s(\gamma_0))\pi(m^-|f,P)} \leq \frac{1}{2}, \tag{90}$$

$$\pi(m^-|f,B) \geq \frac{c}{g_s(\beta_0)}. \tag{91}$$

There are two ranges of (γ_0, β_0) that we consider which span case 4.2 completely:

1. *Case 1.* $\gamma_0 \in \left(\frac{1-p_L-\sqrt{1-p_H-p_L+p_Hp_L}}{p_H-p_L}, \frac{2-p_L-g_s(\beta_0)}{4-p_H-p_L-2g_s(\beta_0)}\right)$ *and* $\beta_0 \in \left(\frac{2-\gamma_0p_H-2p_L+\gamma_0p_L}{4-p_H-2\gamma_0p_H-3p_L+2\gamma_0p_L}, \frac{2-4\gamma_0+\gamma_0p_H+2p_L-3\gamma_0p_L}{p_H-2\gamma_0p_H-p_L+2\gamma_0p_L}\right)$.
2. *Case 2.* $\gamma_0 \in \left(\frac{4-3p_H-5p_L+3p_Hp_L+p_L^2}{8-8p_H+p_H^2-8p_L+6p_Hp_L+p_L^2}, \frac{1-p_L}{2-p_H-p_L}\right)$ *and* $\beta_0 \in \left(\frac{2-\gamma_0p_H-2p_L+\gamma_0p_L}{4-p_H-2\gamma_0p_H-3p_L+2\gamma_0p_L}, \frac{1-p_L}{2-p_H-p_L}\right)$.

In both, the equations can be met if and only if

$$c \leq \frac{(2\beta_0-1)p_L(-1+h_s(\gamma_0))(2-2p_L+\beta_0(p_L-p_H)+\gamma_0(-4+p_H+2\beta_0p_H+3p_L-2\beta_0p_L))}{2(-\gamma_0+\beta_0(2\gamma_0-1))(p_H-p_L)((-1+\gamma_0)(-1+p_L)+\beta_0(-1+\gamma_0(p_H-p_L)+p_L)}. \tag{92}$$

D Proofs for Robustness Extensions

D.1 Smooth Reputation Payoffs

Let the bureaucrat's payoff be $u_B(\beta, e) = \beta - ce$. Assume $p_H = 1$, $p_L \in (0, 1)$ and $c < p_L$.

Suppose both types exert effort and P remains silent. Note that failure leads to posterior $\beta_f = 0$. Note, also, that success leads to posterior:

$$\beta_s = \frac{p_H \beta_0}{p_H \beta_0 + p_L(1-\beta_0)} = \frac{\beta_0}{\beta_0 + p_L(1-\beta_0)}. \tag{93}$$

Then the high type exerts effort if

$$\frac{\beta_0}{\beta_0 + p_L(1-\beta_0)} - c \geq 0 \iff \beta_0 \geq c(\beta_0 + p_L(1-\beta_0)) \iff \beta_0(1 - c + cp_L) \geq cp_L, \tag{94}$$

i.e., if $\beta_0 \geq \frac{cp_L}{1-c+cp_L}$. Similarly, the low type exerts effort if

$$p_L \frac{\beta_0}{\beta_0 + p_L(1-\beta_0)} - c \geq 0, \tag{95}$$

which happens if $\beta_0 \geq \frac{cp_L}{p_L + cp_L - c}$. This is the more restrictive condition on β_0.

Now suppose that that P uses a messaging strategy. The interesting part is if $\beta_0 < \frac{cp_L}{p_L + cp_L - c}$ to see if there can be full effort in cases where there was no full effort with politician silence.

As failure can only happen following the low type, $\beta_f = 0$ regardless of the message. Following a success, P can use messages m^+ and m^-. Let $\pi(m^+|s, H) > \pi(m^+|s, L)$ so that the m^+ message is sent more often given the high type. Then, the low type exerts effort if

$$\begin{aligned} &\pi(m^+|s, L) \frac{\beta_0 \pi(m^+|s, H)}{\beta_0 \pi(m^+|s, H) + p_L(1-\beta_0)\pi(m^+|s, L)} + \\ &(1 - \pi(m^+|s, L)) \frac{\beta_0 \pi(m^-|s, H)}{\beta_0 \pi(m^-|s, H) + p_L(1-\beta_0)\pi(m^-|s, L)} \geq \frac{c}{p_L} \end{aligned} \tag{96}$$

However, this leads to a contradiction if $\beta_0 < \frac{cp_L}{p_L + cp_L - c}$. Hence, politician signaling does not generate extra effort. The same is true in comparing mixed strategy equilibria. The issue is that with linear payoffs, the expected reputation following success is at its maximum for the low type if there is no information provision. Thus, for information provision to be beneficial in providing incentives, we need that it is sufficiently convex. Let $V(\beta) - ce$ be the payoff of the bureaucrat now, where $V(\beta)$ is smooth and increasing in β. Put differently, for the low type to exert effort, it is necessary that

$$\pi(m^+|s, L) V(\beta_{s,m^+}) + (1 - \pi(m^+|s, L)) V(\beta_{s,m^-}) \geq \frac{c}{p_L}, \tag{97}$$

$$V(\beta_{s,m^-}) + \pi(m^+|s, L) \left(V(\beta_{s,m^+}) - V(\beta_{s,m^-}) \right) \geq \frac{c}{p_L}. \tag{98}$$

The goal is to maximize the LHS via $\pi(m^+|s,L)$ and $\pi(m^+|s,H)$. First note that because the bureaucrat's payoff is altered, the relevant condition is now that the prior β_0 must be so that

$$V(\beta_s) \geq \frac{c}{p_L} \tag{99}$$

for blaming and crediting to not be necessary to get effort. If $V(\beta_s) < \frac{c}{p_L}$, then there is no full effort. Note that P must provide information for there to be additional incentives. But P may not fully reveal the type so that m^- is always sent after the low type.

Conjecture. only send m^- following L type. Then simplifies to

$$\pi(m^+|s,L)V(\beta_{s,m^+}) \geq \frac{c}{p_L}. \tag{100}$$

This is first increasing in $\pi(m^+|s,L)$ and then decreasing in it. The more convex it is, the more increasing the LHS is.

□