Verifiable Advice to a Biased Policymaker[*]

Anna Denisenko[†] Catherine Hafer[‡] Dimitri Landa[§]

August 28, 2023; DRAFT: Comments Welcome

Abstract

We develop a model of verifiable communication between a biased policymaker and a bureaucratic agency that has a preference for maintaining the status quo. We show that, in the absence of additional utility pressures on the agency, an increase in the policymaker's bias, which increases the distance between the policymaker's and the agency's ideal points, leads the agency to disclose more information. A key intuition for this result is that, in equilibrium, the lack of a revealing message from the agency functions as a signal to the policymaker that credibly compels her to choose more radical policies, with the agency being forced to reveal in order to hold back the policy radicalism. We also show that, while introducing the possibility of a utility bonus for revelation results in increased agency revelation, it can reverse the positive effect of the policymaker's bias on revelation and hinder the disclosure of additional information to more biased policymakers. Finally, we demonstrate that the higher bias of the policymaker exacerbates the asymmetry in the agency's revelation strategy, creating the appearance of ideological conflict with the policymaker.

Introduction

Elected officials coming from extreme ends of the political spectrum frequently invoke the language of "deep state" to complain about the resistance of civil servants and other state bureaucrats to their efforts to move the policy away from the status quo. The complaints are, of course, not entirely unfounded, since bureaucrats tend to have preferences for upholding the status quo (Peter; Ginsburg and Huq), and operate with a certain degree of

[*]We thank Steve Callander, Matthew Stephenson, and participants in the Virtual Formal Theory Workshop and the NYU PE Workshop for valuable feedback on earlier versions of this paper.
[†]Ph.D. student, Wilf Family Department of Politics, NYU, e-mail: ad4205@nyu.edu
[‡]Associate Professor, Wilf Family Department of Politics, NYU, e-mail: catherine.hafer@nyu.edu
[§]Professor, Wilf Family Department of Politics, NYU, e-mail: dimitri.landa@nyu.edu

informational and executive autonomy. But the interactions between them and the policy-makers take place in a strategic environment, and the preference for different policy should not unproblematically be assumed to translate into choices.

We develop a strategic model of communication and policy-making that sheds new light on the effects of policy-maker bias on strategic information revelation from an agency. In the game we analyze, a bureaucratic agency with conservative – i.e., status-quo favoring – preferences has access to superior information about the state of the world than does the policy-maker. It chooses whether to reveal this information (send a verifiable message) to the policy-maker who prefers to tailor the policy choice to the state of the world but has a known directional bias. We detail the implications of two distinct pathways by which leader's bias affects the agency's revelation strategy: First, and more intuitively, greater leader bias makes the agency more reluctant to share information that reinforces the leader's bias and would prompt the leader to choose more extreme policy; however, by the same token, it is more willing to share information contrary to the leader's bias that might temper the leader's choice. Second, and more subtly, the leader's bias alters the default policy that the leader chooses in the absence of information revelation, which, in turn, affects the agency's decision to reveal.

Our first key result shows that when the agency's incentives to reveal information are induced by policy consequences, greater bias of the policy-maker strictly increases information revelation from the agency. To see the basic logic, suppose, without loss of generality, that the policy-maker has a right-ward bias. The agency will, then, want to reveal to her states of the world that pull the policy back to the left. It is straightforward that the states that would be revealed would include leftward states that the agency would never reveal to a more neutral policy-maker. But, and for the same reason, they will include some relatively right states as well: conditional on no revelation, the policy-maker anticipates that the state must be fairly far to the right, and the agency will wish to reveal its signal to the policy-maker to convince her that her guess is, in fact, too extreme.

We then study how the relationship between the policy-makers' bias and equilibrium revelation depends on two factors that may complicate the communication environment: the introduction of office-based utility for the agency (which may be manipulated by the policy-maker to help induce revelation) and the possibility of the agency's having state-dependent preferences with varying degrees of status-quo oriented conservatism. We show that the central comparative static– that greater bias of the policy-maker leads to greater revelation from the agency– continues to hold in the presence of these factors for substantial parts of the parameter space. However, the possibility of office benefits for the agency leads to another surprising result: While greater such benefits directly promote revelation in equilibrium, they undermine the indirect effect of leader bias via chosen default policy, so that when the agency's office benefits are large enough, the agency's informational incentive to reveal its signal disappears because the policy-maker's expectation of what the state must be conditional on no revelation becomes sufficiently close to the policy chosen in the absence of revelation. The policy-maker bias has a positive effect on agency revelation when the agency's preferences are sufficiently conservative in valuing the status quo, but it flips when the agency's preferences become sufficiently responsive to the state.

Among other implications, the overall logic of the results offers a novel perspective on policy-makers' complaints of a "deep-state." Bureaucracies play a number of roles in the policy-making process, including providing information to policy-makers. Focusing on the informational role of bureaucracies and setting aside implementation, our analysis suggests that leaders with well-defined policy biases should, indeed, expect to receive asymmetric patterns of bureaucratic advice, tilted against their ideologically preferred direction. However, despite that asymmetry, the information a more biased policy-maker receives from a status-quo oriented bureaucracy may be better, not worse, than the information that a less biased policy-maker would receive. In other words, as far as the provision of information is concerned, complaints about the "deep state" should be taken with a fair amount of skepticism.

Connection to the Literature

Delegation and communication within hierarchies has been a focus of a substantial body of political economy scholarship (for reviews, see Gailmard and Patty (2012) and Sobel (2013)). One branch of this scholarship models communication as "cheap talk" in which bureaucrats' potential messages are not directly constrained by their information (Crawford and Sobel (1982); Gilligan and Krehbiel (1989); Austen-Smith (1990); Austen-Smith (1993)). A key finding of this literature is that divergence in the actors' preferences curtails communication, and successful communication at all occurs only when the advisor's and the leader's preferences are sufficiently aligned. An important exception relevant for our analysis is Callander (2008), which studies an expert bureaucrat's advice to a legislator in an environment in which the bureaucrat's expertise is endogenously acquired and the legislator cannot fully recover the bureaucrat's private information from the advice. Callander shows that, in the absence of an institutionalized commitment to implement the received advice, greater divergence in primitive preferences between bureaucrat and legislator sometimes induces greater voluntary delegation of policy-making powers from the legislator to the bureaucrat. This suggests a certain affinity with our result that greater preference divergence spurs more information revelation. The mechanisms producing these results are, however, very different. [1]

A second branch of this scholarship models the communication of verifiable information. A key result in this literature is that all private information is revealed in equilibrium (Milgrom (1981), Milgrom (2008)). In the model we study, the agency's messages consist of hard evidence, but nonetheless the agency does not disclose everything. Since Milgrom (1981), important subsequent works study conditions under which the unraveling logic of full disclosure does not hold. This includes Shin (1994), which shows that the leader will not be able to infer perfectly the advisor's private information in the event of "no news" when the advisor's knowledge is imperfect; see also Wolinsky (2003). Dziuda (2011) shows that, in a setting where the fixed expert's preferences are different and unknown to the decision-maker, there is never full disclosure, but the expert offers pros and cons for the advocated alternative in order to pool with the honest/non-strategic type. While in our model, there is never full disclosure either, the unraveling logic is an important element of the mechanism underlying our results: because in the absence of disclosure, the policy-maker assumes that the state must be too extreme for the bureaucrat, the policy-maker chooses a more extreme policy, which lessens the incentives to withhold information.

Add discussion of Bhattacharya and Mukherjee (2013) and of Lipnowski and Ravid (2020).

Denisenko, Hafer and Landa (2022) study the transmission of verifiable information between a sender of known competence with preference for the status-quo and a neutral receiver who wishes to match the state of the world. They show that less competent senders have stronger incentives to reveal their information to the leader than more competent ones, inducing a trade-off between the quality of advice the sender receives and the likelihood of receiving it. In contrast with that model, we abstract away from variation in sender competence and introduce the bias of the receiver to focus our analysis on its effects on information transmission.

The General Environment

We analyze a strategic interaction between a Leader (she) and an Agency (it). The Leader wishes to implement a policy that will match the state of the world and accomplish her partisan agenda; these goals may be in tension. The Leader does not directly observe the state of the world and must, instead, obtain relevant policy information via the recommendation released by the Agency. The timeline of the game is as follows:

[1] A different exception, farther afield, is Battaglini (2002), which shows that the receiver may obtain full revelation when there are multiple senders.

1. Nature determines the state of the world $w \in \mathbf{R}$, where w is a draw from a standard normal distribution $N(0, 1)$.

2. The Advisor observes a signal s about the state of the world w, $s = w + \varepsilon$. The variable ε represents random noise drawn from a standard normal distribution, $\varepsilon \sim N(0, 1)$.

3. The Advisor chooses whether to send (verifiable) message $m = s$ to the Leader, $m \in \{s, \varnothing\}$.

4. The Leader observes message m and decides which policy $a \in \mathbf{R}$ to implement.

The Leader wants to choose policy a to best satisfy her partisan bias b and to best match the state of the world. These goals are in tension:

$$U_L(a|b) = -(w - a)^2 + b \cdot a. \tag{1}$$

When bias b is positive the Leader has a *rightward bias*, which implies that she benefits when she implements policies to the right of the status quo. When b is negative, the Leader, instead, exhibits a *leftward bias* and is tempted to implement left-leaning policies. In what follows, we assume that $b > 0$, meaning the Leader exhibits a rightward bias. (The case $b < 0$ is symmetric to this one.)

The Agency's inclinations differ from those of the Leader. We assume the agency to be conservative in its preference and dislike any policy change, with increasing marginal losses. The Agency's utility is

$$U_A(m|b) = -a(m)^2. \tag{2}$$

We will refer to the Agency's message $m = s$ as the *revealing* message. Conditional on receiving such a message from the Agency, the Leader rationally updates her belief about the state of the world $E[w|m] = \frac{m}{2}$ and, thus, implements optimal policy $a^*(m) = \frac{m}{2} + \frac{b}{2}$. Likewise, the Leader updates her belief about w when she does not receive the revealing message ($m = \emptyset$); we denote the optimal default policy the Leader implements after message $m = \emptyset$ as $d^*(b)$.

The Agency, in turn, chooses its revelation strategy optimally, taking into account the Leader's policy response. The following proposition characterizes the Agency's and the Leader's equilibrium strategies

Proposition 1. *In equilibrium, the Agency reveals the signal it observes if and only if*

$s \in [\underline{s^*}(\cdot), \overline{s^*}(\cdot)]$, *where*

$$\begin{aligned}\underline{s^*}(b) &\equiv -2 \cdot \underbrace{d^*(b)}_{\substack{Indirect\ Effect\\ of\ Bias\ on\ s^*}} - \underbrace{b}_{\substack{Direct\ Effect\\ of\ Bias\ on\ s^*}}, \\ \overline{s^*}(b) &\equiv 2 \cdot \underbrace{d^*(b)}_{\substack{Indirect\ Effect\\ of\ Bias\ on\ s^*}} - \underbrace{b}_{\substack{Direct\ Effect\\ of\ Bias\ on\ s^*}},\end{aligned} \tag{3}$$

and $d^*(b)$ *solves*

$$d = b/2 + \frac{1}{\sqrt{\pi}} \cdot (e^{-(b/2-d)^2} - e^{-(b/2+d)^2}). \tag{4}$$

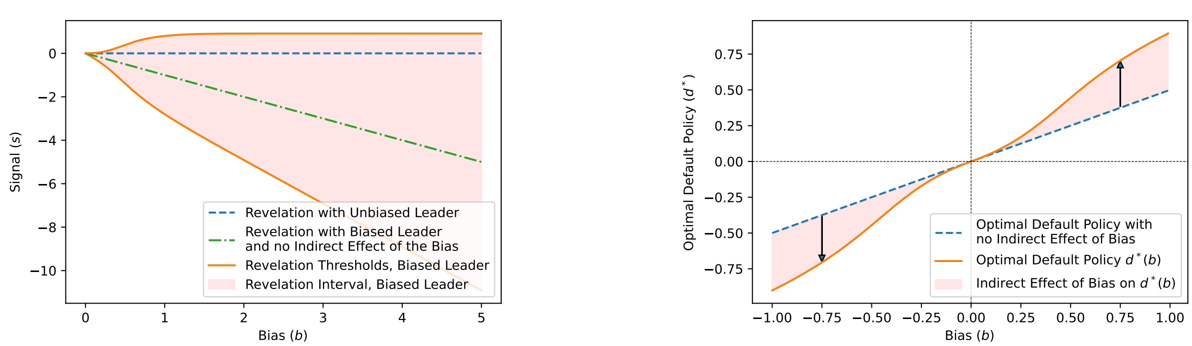
The Leader implements policy $a^*(m) = \frac{m}{2} + \frac{b}{2}$ *if the Agency reveals* s *and chooses policy* $d^*(b)$ *when the Agency does not* ($m = \emptyset$)*.*

Proof. Appendix A. □

The Leader's bias exerts two effects on the Agency's revelation strategy. The *direct effect* of the bias, deriving from the Leader's policy response to the revelation of s, shifts the Agency's revelation thresholds in the direction opposite to the Leader's bias.

The second, *indirect*, effect of the Leader's bias on the Agency's strategy occurs via the Leader's policy choice when she does not receive a revealing message, $d^*(b)$. Figure 1a shows the decomposition of the impact of the Leader's bias on the Agency's strategy as a function of bias b. The dashed horizontal line in Figure 1a represents a benchmark of the Agency's revelation with the unbiased Leader. The dash-dotted line depicts the shift of the Agency's revelation interval driven by the direct effect of the bias, holding fixed the default policy. Finally, the solid orange curves represent the equilibrium revelation thresholds, reflecting both the direct and indirect effects of the bias.

Figure 1: Impact of the Leader's Rightward Bias ($b > 0$) on the Agency's Revelation Strategy.



(a) The Agency's Revelation Strategy as a Function of the Leader's Bias (b).

(b) The Optimal Default Policy as a Function of the Leader's Bias (b).

In the absence of revelation, the Leader chooses default policy $d^*(\cdot)$ that solves

$$d \;=\; \underbrace{b/2}_{\substack{Direct\ Effect\\ of\ Bias\ on\ d^*}} \;+\; \underbrace{\frac{1}{\sqrt{\pi}}\left(e^{-(b/2-d)^2} - e^{-(b/2+d)^2}\right)}_{\substack{Indirect\ Effect\\ of\ Bias\ on\ d^*}}. \tag{5}$$

Lemma 1. *Optimal default policy $d^*(\cdot)$ always exceeds $b/2$ when the Leader's bias is positive and is less than $b/2$ otherwise.*

Proof. See Appendix B. □

The optimal default policy the Leader sets is similarly affected by the Leader's bias. We can separate two distinct effects. Keeping the Agency's revelation strategy fixed, the higher the Leader's bias, the more she benefits when she implements policies co-aligned with her bias. Thus, in the absence of the revealing signal, the default policy the Leader chooses shifts in the direction of her bias. At the same time, a biased Leader should infer that the information the Agency is concealing is more likely to be co-aligned with her bias than not. This further encourages the Leader to adopt policies that confirm her bias in the absence of information. To distinguish between the two outlined mechanisms by which the bias affects the optimal default policies, we denote the former one the *direct effect of bias on d^** and the second one the *indirect effect of bias on d^** in equation 8. Figure 1b depicts the optimal default policy (d^*) as a function of the Leader's bias when only the *direct effect of bias on d^** is included and when both effects are included. Because both effects of the bias on the optimal default policy are co-aligned, we can state the following result:

Proposition 2. *The optimal default policy (d^*), which the Leader implements when the Agency does not reveal s, increases in the Leader's bias (b).*

Proof. See Appendix C. □

Overall, the higher is the Leader's policy bias, the more extreme policies the Leader chooses in the absence of an informative recommendation (message) from the Agency. Because of that, the marginal opportunity cost of information revelation decreases in the Leader's bias, which encourages the Agency to conceal less from the Leader.

The following two propositions summarize the impact of the Leader's bias on the Agency's revelation strategy.

Proposition 3. *For any value of the Leader's bias (b),*

1. *the Agency reveals more information on the side opposite the Leader's bias than on the side co-aligned with the Leader's bias: for any $b > 0$, $|\underline{s}^*(b)| \geq |\overline{s^*}(b)|$;*

2. *the asymmetry in the Advisor's revelation increases in the Leader's bias: for any* $b > 0$, $|\underline{s^*}(b)| - |\overline{s^*}(b)|$ *increases in* b.

Proof. See Appendix D. □

The direct and indirect effects of the Leader's bias on the revelation thresholds are mutually reinforcing when it comes to the threshold opposite the bias (e.g., the impact of the rightward bias on the left-leaning signals) but are in tension for the threshold on the same side as the bias.

For a Leader with a rightward bias ($b > 0$), when it comes to the left-leaning (negative) signal, the direct effect of an increase in the bias shifts the left threshold leftward, and the indirect effect pushes it further leftward. However, when it comes to the right-leaning (positive) signals, the direct and indirect effects of the bias run counter to each other – the direct effect shifts the right threshold to the left, while the indirect effect pushes it to the right. Note that, as the Agency seeks to counteract the biased Leader's preference, the biased Leader always gets relatively less advice urging policies consistent with her bias and relatively more advice urging policies contrary to her bias, and this asymmetry worsens as the bias increases. While it might appear that the Agency systematically seeks to thwart the Leader's agenda or has preferences diametrically opposed to her (cue the former President's complaints about the "deep state"), a different interpretation here is that it is the Leader's own bias that leads to the lack of even-handedness in the Agency's revelation to the Leader.

Proposition 4. *The higher is the Leader's bias* (b)*, the more informative communication between the Leader and the Agency becomes.*

Proof. See Appendix E. □

It is important to note that lack of revelation remains communicative and informs the Leader's inferences about the state of the world. Proposition 4, thus, highlights that the Leader's bias improves the informativeness of communication between the Leader and the Agency both when revelation occurs and when the Agency conceals its information. This result might seem surprising as it runs counter to general predictions of the cheap-talk literature: when communication is costless and unverifiable, a greater gap between the sender's and the receiver's preferences inevitably lowers the quality of advice (Crawford and Sobel 1982). We demonstrate that, in contrast, verifiable information settings may produce the opposite result. Despite the highlighted tension between the direct and indirect effects of the bias on the revelation strategy, the Leader's bias encourages the Agency to reveal strictly more signals when the Agency is driven by the desire to retain the status-quo policy.

To disentangle the impact of the Leader's bias on the Agency's revelation, let us begin by assuming the Leader to be unbiased. In equilibrium, the Agency never reveals to the unbiased Leader signals different from the status quo: $\overline{s}^*(b=0) = \underline{s}^*(b=0) = 0$. The unbiased Leader, correspondingly, believes that, in the absence of revelation, the expected value of the state w equals the status quo and implements it as a default policy: $d^*(b=0) = 0$. Holding the Agency's revelation strategy fixed, a higher Leader's bias does not alter the Leader's posterior beliefs but shifts the optimal default policy away from the status quo. Given that the default policy does not equal the status quo, the Agency, then, benefits from revealing to the Leader some information that counteracts her bias, with higher bias inducing more extreme advice. The skewed revelation affects the inferences the Leader draws from the absence of advice ($m = \emptyset$), shifting the optimal default policy, which, in turn, causes the Agency to begin revealing signals co-aligned with the Leader's bias to avoid drastic policy changes.

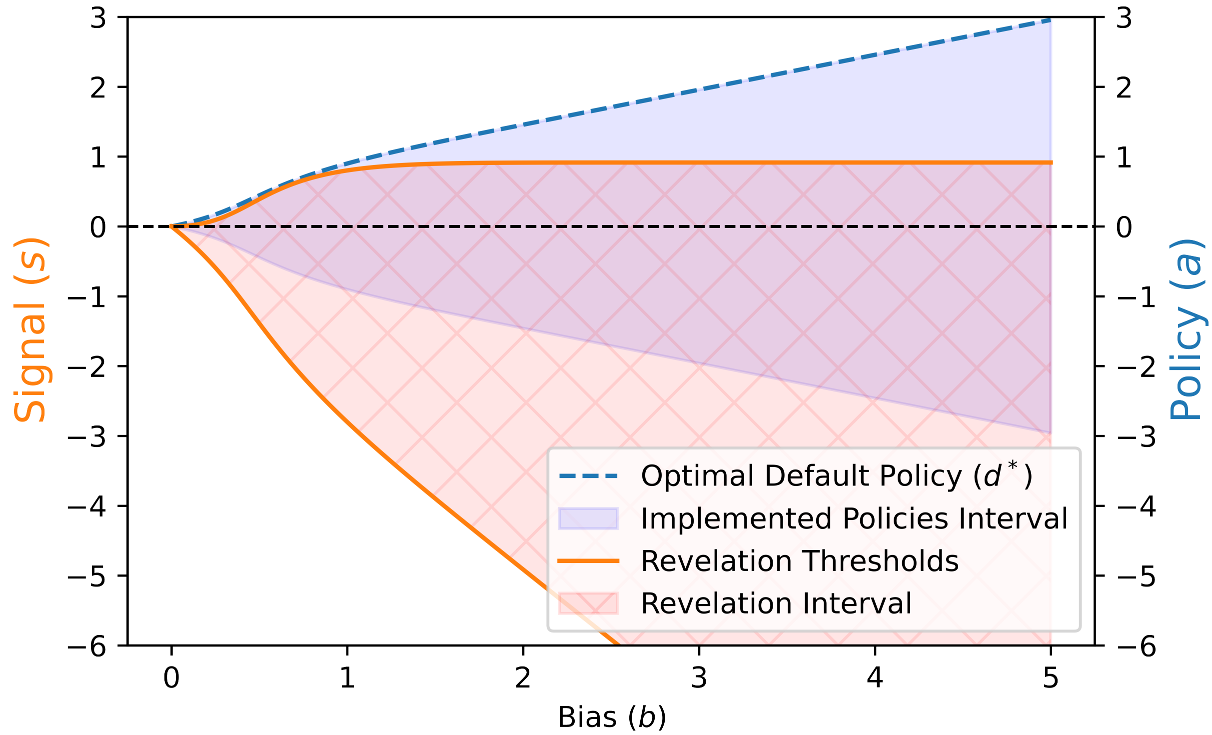
Figure 2 depicts the Agency's revelation strategy (solid orange curves), the range of policies the Leader implements when the Agency reveals s (solid blue shaded area), and the default policy (dashed blue curve) she chooses when she receives no message ($m = \emptyset$). The Leader's bias, thus, allows her to elicit more advice from the Agency despite not directly rewarding revelation or punishing the lack thereof.

As described above, in this model, the Agency is restricted to messages $m = s$ and $m = \emptyset$. It is natural to ask, then, whether the Agency might wish to reveal some information about s without revealing s precisely.

but also that the players share a rich common language and that the Agency can fully convey any received signal to the Leader. However, exogenous constraints might hinder the Agency's ability to communicate the specific details of the information it holds. For instance, language barriers between the sender and the receiver can limit the sender's ability to fully comprehend the intended message (Blume and Board, 2013). Alternatively, the Agency may only have a coarse understanding of the signals it observes, preventing it from sending messages it does not fully comprehend (Hagenbach and Koessler, 2020). Finally, external constraints on the Agency-Leader communication, which might result from the nature of the message exchange, can constrain actors to employ a language insufficiently expressive to capture all details of the Agency's understanding.

Against this background, it is natural to wonder whether Proposition 4 would be robust to expanding the Agency's action set to allow for the possibility of strategies that could partially obscure the Agency's signal. Here we provide an intuition for why the proposition is, in fact, robust. (We provide a formal argument in the Appendix – see Proposition 9.) Suppose that the signal space is partitioned into M (s.t. $M > 1$) convex intervals and

Figure 2: Agency's Equilibrium Revelation Thresholds ($\overline{s^*}(b),\ and\ \underline{s^*}(b)$), the Implemented Policies Interval ($a^*(m)$), and the Default Policy ($d^*(b)$) as a Function of the Leader's Bias (b)



that the Agency can send a message m to the Leader, such that $m \in \{\varnothing, m_1, m_2, ..., m_M\}$: a message $m = \varnothing$ represents a lack of revelation, while a message $m \neq \varnothing$ indicates that the Agency's signal s falls within a specific interval. Assume that this partition is common knowledge.

Observe, first, that when the Agency has the freedom to choose a partition, it cannot commit to anything other than the most precise partition available: When the Leader receives a message indicating that the Agency's signal falls within a particular partition, she forms a belief about the most likely state of the world and devises a policy that accounts for her bias and her belief. Given that $M > 1$, at least one message revealed by the Agency won't implement the Agency's most preferred policy and, instead, will result in a policy more extreme than what the Agency desires. However, this implies that there will always be a signal within the corresponding interval that the Agency would prefer to report directly, rather than sending the coarse message.

Office Benefit

Given the value to the Leader of the Agency's information, it is natural to consider the possibility that the Leader might induce revelation through rewards and punishments. In particular, the Leader might impose costs on the Agency for failing to reveal its information, including, but not limited to, higher levels of congressional scrutiny or the reduction of the Agency budgets (Balla 1998). To examine the effects of such inducements on equilibrium revelation, in this section we assume that the Agency receives a benefit R, which we call *office benefit*, when it shares its private information with the Leader. The Agency's utility, then, becomes

$$U_A(m|b,R) = \begin{cases} -a(m)^2 + R & \text{if } m \neq \varnothing \\ -a(m)^2 & \text{if } m = \varnothing. \end{cases} \tag{6}$$

The presence of the office benefit R does not affect the inferences the Leader makes when the Agency reveals its signal. Upon observing message m, the Leader implements policy $a^*(m) = \frac{m}{2} + \frac{b}{2}$. We denote the optimal default policy the Leader implements in the absence of revelation as $d^*(b,R)$. The next proposition characterizes the equilibrium strategy profile for the Agency and the Leader:

Proposition 5. *In the equilibrium of the game with office benefit R, the Agency reveals signal s when $s \in [\underline{s^*}(b,R), \overline{s^*}(b,R)]$ and conceals it otherwise, where*

$$\begin{aligned} \underline{s^*}(b,R) &\equiv -2 \cdot \sqrt{R + \underbrace{d^*(b,R)^2}_{\substack{\text{Indirect Effect} \\ \text{of the Bias on } s^*}}} - \underbrace{b}_{\substack{\text{Direct Effect} \\ \text{of the Bias on } s^*}}, \\ \overline{s^*}(b,R) &\equiv 2 \cdot \sqrt{R + \underbrace{d^*(b,R)^2}_{\substack{\text{Indirect Effect} \\ \text{of the Bias on } s^*}}} - \underbrace{b}_{\substack{\text{Direct Effect} \\ \text{of the Bias on } s^*}} \end{aligned} \tag{7}$$

and the optimal default policy $d^*(b,R)$ solves

$$d = \underbrace{b/2}_{\substack{\text{Direct Effect} \\ \text{of Bias on } d^*}} + \frac{1}{\sqrt{\pi}} \cdot \underbrace{\left(e^{-(b/2-\sqrt{R+d^2})^2} - e^{-(b/2+\sqrt{R+d^2})^2}\right)}_{\substack{\text{Indirect Effect} \\ \text{of Bias on } d^*}}. \tag{8}$$

The Leader implements policy $a^*(m) = \frac{m}{2} + \frac{b}{2}$ if she observes a revealing message, $m = s$, and chooses policy $d^*(b,R)$ otherwise.

Proof. See Appendix G. □

Note that 7 and 8 evaluated at $R = 0$ yield the equilibrium of the main model, characterized in Proposition 1.

Proposition 6.

1. *The higher the office benefit R, the more informative communication between the Leader and the Agency becomes.*

2. *The Agency's revelation strictly increases in the Leader's bias when*

$$d^*(b,R)^2 \cdot (4(\frac{\partial d^*(b,R)}{\partial b})^2 - 1) > R. \tag{9}$$

Proof. See Appendix H. □

Figure 3: Impact of the Office Benefit R on the Agency's Revelation Thresholds

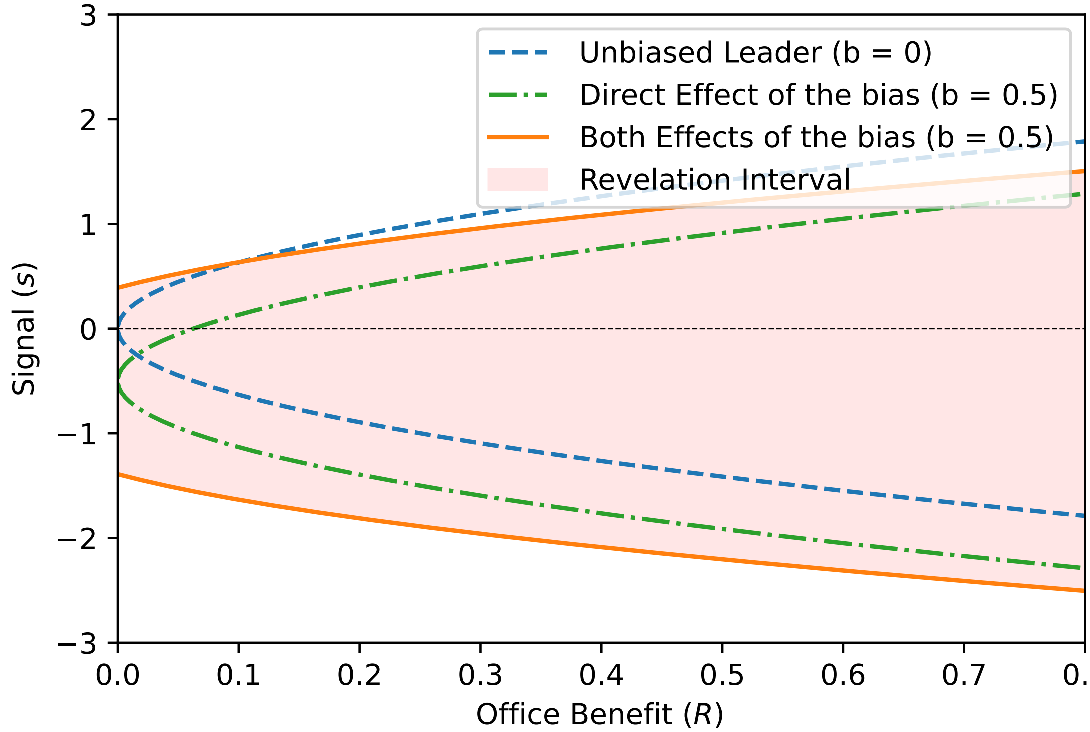


Figure 3 depicts the revelation thresholds the Agency adopts as the office benefit R increases. Higher office benefit induces the Agency to disclose more information to the Leader, widening the revelation interval.

The impact of the the Leader's bias on the Agency's optimal conduct is akin to that observed in the baseline model. The *direct* effect of the bias shifts revelation in the opposite direction of the bias, while its *indirect* effect expands the extent of disclosure. The (rightward) bias always decreases the lower bound of revealed signals, i.e., it increases the disclosure of signals that run counter to the Leader's bias. Nevertheless, contrary to Proposition 4, the Leader's bias has a non-monotonic effect on the upper bound of the revelation interval.

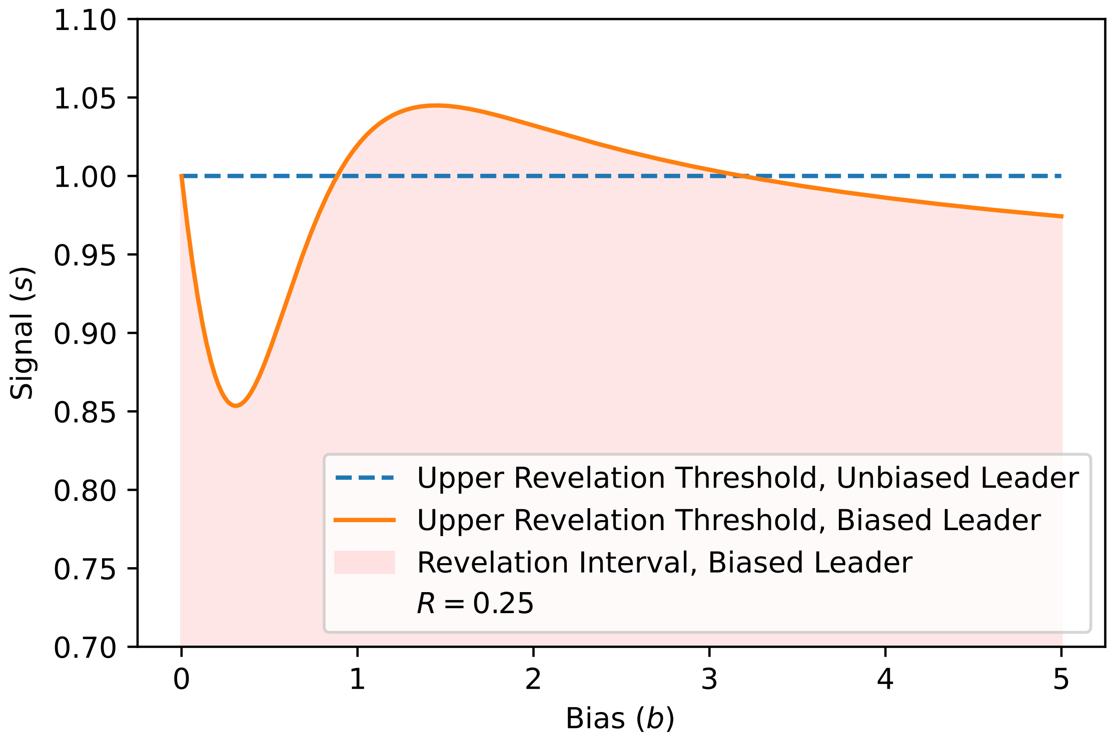
The non-monotonic effect of the Leader's bias on the Agency's revelation thresholds is a result of the trade-off between the policy benefits and the office benefit. When the office benefit is present, the Agency's revelation serves a dual purpose. First, as in the baseline model, the Agency discloses information when such disclosure yields favorable policy outcomes, as perceived by the Agency, compared to the expected outcomes under the default policy. Second, the Agency reveals information to achieve the office benefit R.

In equilibrium, when the Leader's bias is small, the Agency discloses information primarily to obtain the office benefit R. The revelation thresholds, thus, remain balanced around the status quo. (The dashed line in Figure 3 illustrates the revelation interval as a function of R for $b = 0$.) This near symmetry implies that, in the absence of disclosure, the Leader's posterior belief corresponds to an expected value that neighbors the status quo. As the Leader's bias increases, however, the policies he implements with and without revelation from the Agency shift away from the status quo in the direction of the Leader's bias. Importantly, because the Agency's revelation is driven by its pursuit of the fixed gain R and its (quadratic) utility from policy, the farther the policies shift from the Agency's ideal point, the higher the office benefit it requires to compensate it for revelation of the marginal signal. Thus, for low values of bias b, and increase in b discourages the revelation of information co-aligned with the Leader's bias, decreasing the upper revelation threshold.

Note, however, that the same factors that discourage revelation of information co-aligned with the Leader's bias simultaneously encourage revelation of signals that oppose the bias. Higher bias, thus, induces asymmetry in the Agency's revelation strategy. Because of that asymmetry, concealment shifts the default policy further away from the status quo since the Leader updates in the direction of his bias when $m = \emptyset$. Now, as the Leader's bias increases, the Agency reveals information not only to gain the office benefit R, but also to prevent the realization of the default policy. The upper revelation threshold, thus, begins to increase in the Leader's bias b.

Lastly, it is important to note that the posterior the Leader forms, and, thus, the indirect effect of the bias on the default policy, is bounded. Even when the Agency only reveal signals that oppose the Leader's bias, the state of the world conditional on the lack of revelation never exceeds the threshold $\frac{1}{\sqrt{\pi}}$. At the same time, the direct effect of the bias is unbounded – it shifts implemented policies away from the status quo, decreasing the office benefit's value for the Agency and thus decreasing the Agency's incentives to reveal information. Figure 4

Figure 4: Upper Revelation Threshold as a Function of the Leader's Bias



shows the Agency's revelation strategy as a function of the Leader's bias b in the presence of a positive office benefit. The solid orange line represents the Agency's upper bound on signals it will reveal. As described above, it first decreases in bias, then increases, and, finally, decreases in bias once again. In contrast, the lower revelation threshold strictly decreases in the Leader's bias.

The dashed blue line in Figure 4 highlights the important benchmark of the Agency's revelation strategy with an unbiased Leader, for the same office benefit. It should be noted that in certain circumstances, leaders with higher bias are privy to more informative communication than their unbiased counterparts. This disparity arises due to the Agency's efforts to circumvent the abrupt policy changes that follow failure to reveal s, much as in the baseline model. However, when the office benefit is significant, and the Agency's revelation serves to earn this benefit, mere policy objectives alone may not be compelling enough to encourage more disclosure. The next proposition characterizes conditions under which a more biased Leader will observe strictly more information than would an unbiased Leader.

Proposition 7. *There exists a threshold $R^*(b)$ such that the Agency reveals strictly more* *information to the biased Leader than to the unbiased one when $R \in [0, R^*(b))$.*

Proof. See Appendix I. $\square$

Figure 8 depicts the difference between the upper revelation threshold chosen by the Agency communicating with a biased Leader ($b = 1$) and the upper revelation threshold chosen by the Agency when the Leader is unbiased. For bias b, this disparity is

$$I(b, R) \equiv \overline{s^*}(b, R) - \overline{s^*}(b = 0, R).$$

When the office benefit is sufficiently low ($R < R^*(b)$), the Agency reveals strictly more information to a biased leader than to an unbiased one. However, as the office benefit (R) increases, the Agency is prompted to reassess the importance of its policy objectives vis-à-vis the value of the office benefit. When R is sufficiently high, greater bias shifts implemented policies away from the status quo, on the margin, and this deviation from the status quo prompts the Agency to demand a commensurately higher reward for disclosure.

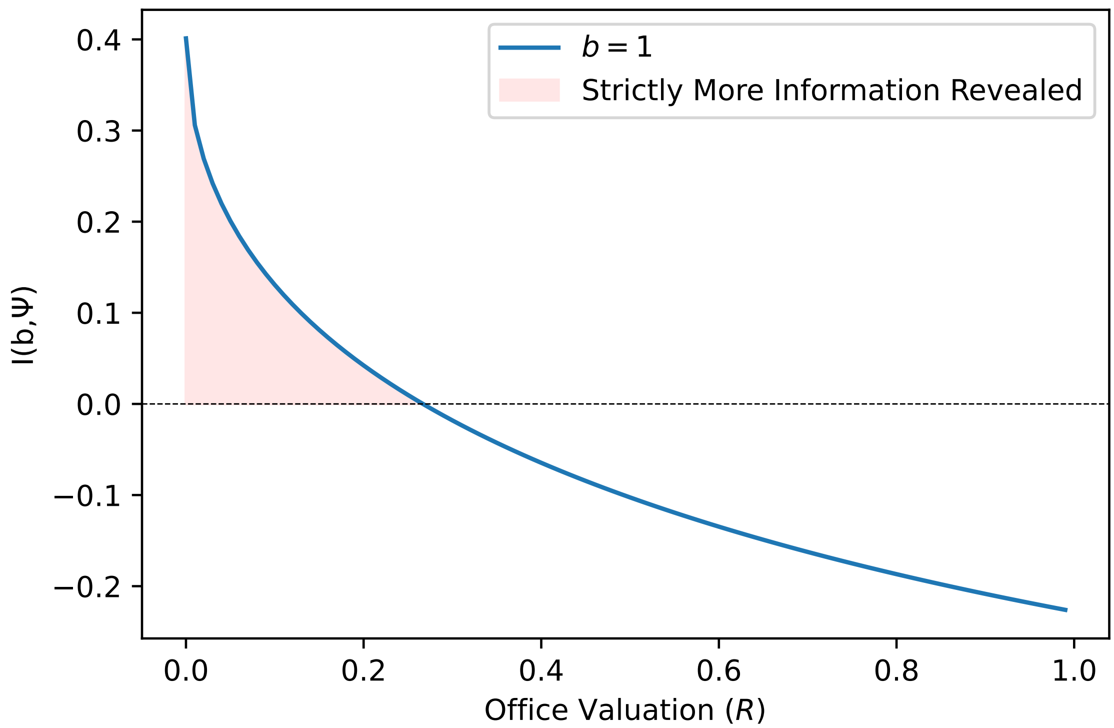
While the presence of the office benefit directly improves communication between the Leader and the Agency, as seen in Figure 3, the revelation of certain signals requires no reward. For instance, the orange dashed-shaded area in Figure 6 shows signals the Agency reveals when $R = 0$. The blue solid-shaded areas show the communication improvement that committing to directly rewarding revelation can purchase (assuming that revelation of any signal is rewarded).

Importantly, revelation under this conditional reward scheme does not differ from the one with unconditional reward. While in the context of our model it does not affect the Agency's nor the Leader's utilities, if the cost of the reward were to enter the Leader's objective function, conditional revelation presents a clear Pareto-improvement as it offers the same information at a lower cost.

Robustness to State-Dependence of Agency's Preferences

In the baseline model, two factors contribute to the misalignment of preferences between the Agency and the Leader. First, the Agency is assumed to be extremely conservative and, thus, to prioritize the maintenance of the status quo, whereas the Leader favors policies that adapt to the current state of the world. Second, in contrast to the Leader, the Agency is assumed to be free of bias relative to the state of the world, and so lacks a vested interest in implementing partisan (biased) policies.

Figure 5: Difference between Upper Revelation Threshold Introduced by the Agency with the Biased Leader ($\overline{s^*}(b, R)$) and the Agency with the Unbiased Leader ($\overline{s^*}(b = 0, R)$) as a Function of the Agency's Office Valuation (R)



In this section, we relax the first assumption, allowing the Agency's preferences to be state-dependent, to a degree. Formally, we assume the Agency's ideal policy is a weighted average of the state w and the status quo (0). In other words, the Agency wants to strike a balance between maintaining the status quo and aligning implemented policies with the state of the world. The Agency's utility is

$$U_A(a|c) = -(a - (1 - c) \cdot w)^2, \tag{10}$$

where c ($c \in [0, 1]$) measures the Agency's conservatism.

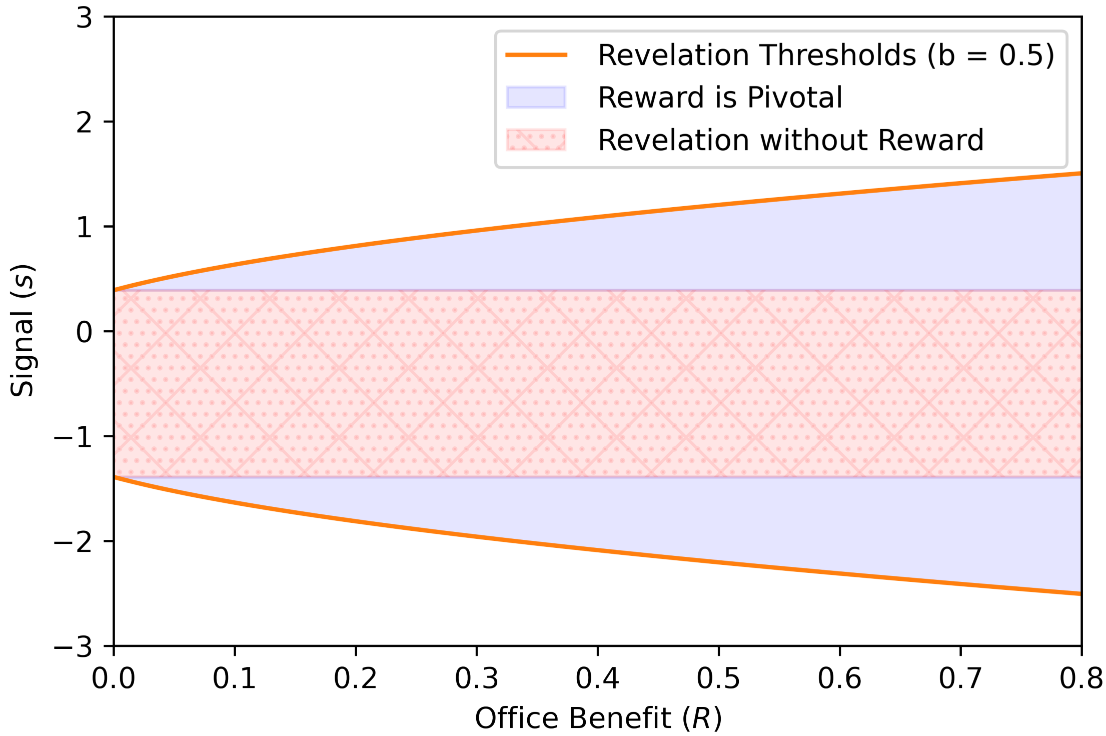
We obtain the following result:

Proposition 8. *The Agency's revelation*

1. *strictly decreases in the Agency's conservatism, and*

2. *strictly increases in the Leader's bias when $c \geq 1/2$ and strictly decreases in the Leader's*

Figure 6: Blue shaded area depict areas where revelation reward improves communication between the Leader and the Agency. The orange shaded area show the region where revelation occur regardless of the reward



bias otherwise.

Proof. See Appendix J. □

As the Agency becomes less conservative, i.e., as its preferences better align with those of the Leader, the Agency discloses more information to the Leader. Figure 7 depicts the Agency's revelation strategy as a function of its conservatism.

When the Agency discloses information opposite to the Leader's bias, the Leader implements policy less extreme than that which would fully match the state of the world. This policy moderation benefits the Agency, which balances matching the state of the world and maintaining the status quo. Lower conservatism, which corresponds to the Agency placing greater importance on matching the state of the world, encourages the revelation of information opposing the Leader's bias. The solid curve in Figure 7 shows how the extent of co-alignment between the Agency and the Leader influences the degree to which the Agency is incentivized to disclose information.

However, when the signal aligns with the Leader's partisan bias, the Leader, if informed of the signal, implements a policy more extreme than that which would match the expected state of the world. In this case, the Agency discloses information because the adoption of the endogenously determined default policy would be even worse.

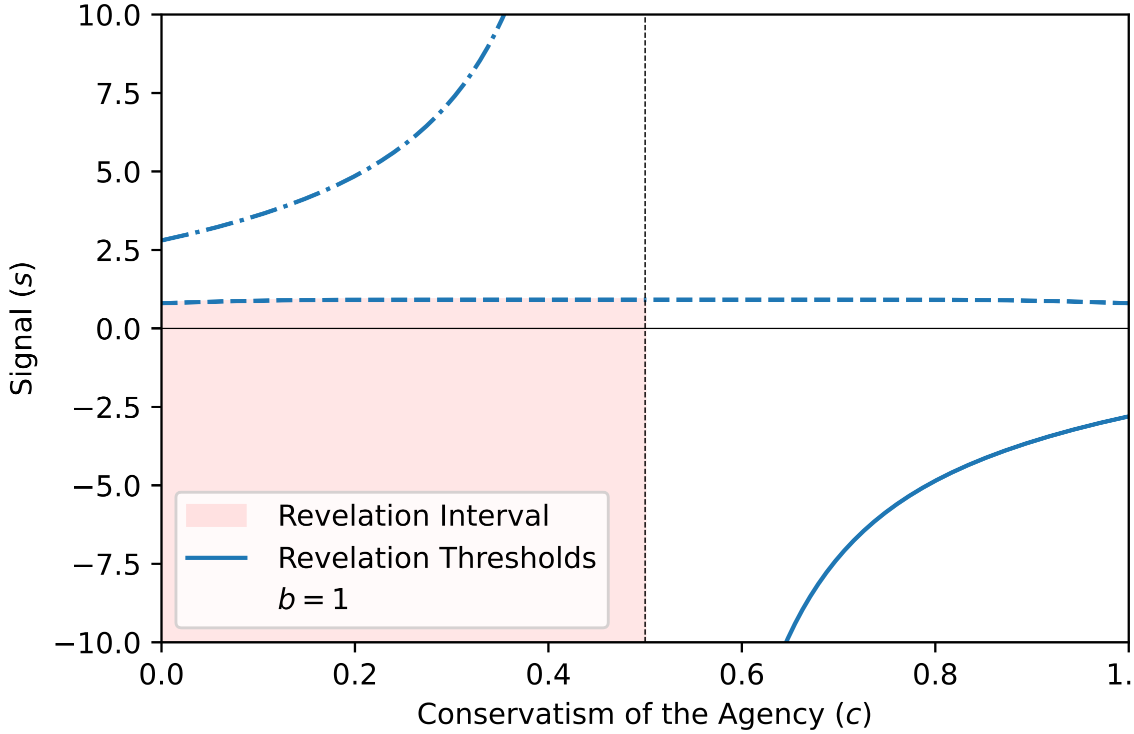
The revelation strategy of the Agency with respect to signals that align with the Leader's bias is motivated by two factors. First, the Agency discloses information when the default policy is excessively extreme given the signal received. The dashed curve in Figure 7 illustrates how the Agency's conservatism influences its willingness to disclose information to counteract the implementation of extreme default policies. When $c > 1/2$, a decrease in the level of conservatism leads to an increase in the disclosure of information opposing the Leader's bias, which further drives the default policy in the direction of the Leader's bias, which further encourages the Agency to disclose information co-aligned with the Leader's bias.

Second, the Agency reveals information when the default policy is not extreme enough to account for necessary changes in policy. When $c < 1/2$, the Agency prioritizes matching the current state of the world over maintaining the status quo, and begins providing advice inducing more extreme policies, because, in some cases, the default policy is not sufficiently extreme to reflect adequately the realized state of the world. The dash-dotted curve in Figure 7 demarcates the revelation area in which the Agency discloses signals because the default policy is not sufficiently extreme. This area expands as the Agency's conservatism decreases, which then decreases the default policy when $c < 1/2$. Lowering the default policy, in turn, diminishes the possibility that the default policy is excessive for the state of the world, and which reduces the Agency's incentive to reveal information to counteract an excessive default policy.

The second part of Proposition 8 highlights that when the Agency's conservatism falls below a certain threshold ($c < 1/2$) the informativeness of the communication between the Leader and the Agency decreases in the Leader's bias, but increases when the Agency's conservatism is sufficiently high ($c > 1/2$). When the Agency is highly conservative ($c > 1/2$), the mechanism behind the bias's impact on revelation does not differ from that described in the baseline model. In the subsequent analysis, we focus on the case of low Agency conservatism.

When the Agency's conservatism is low ($c < 1/2$), the Agency reveals all signals inducing policies opposite to the Leader's bias. The impact of bias on revelation strategy is, therefore, only relevant when it concerns signals co-aligned with the Leader's bias. Figure 8 illustrates the Agency's revelation strategy as a function of the Leader's bias when $c = 0$. Low conservatism encourages the Agency to reveal signals co-aligned with the Leader's bias when the

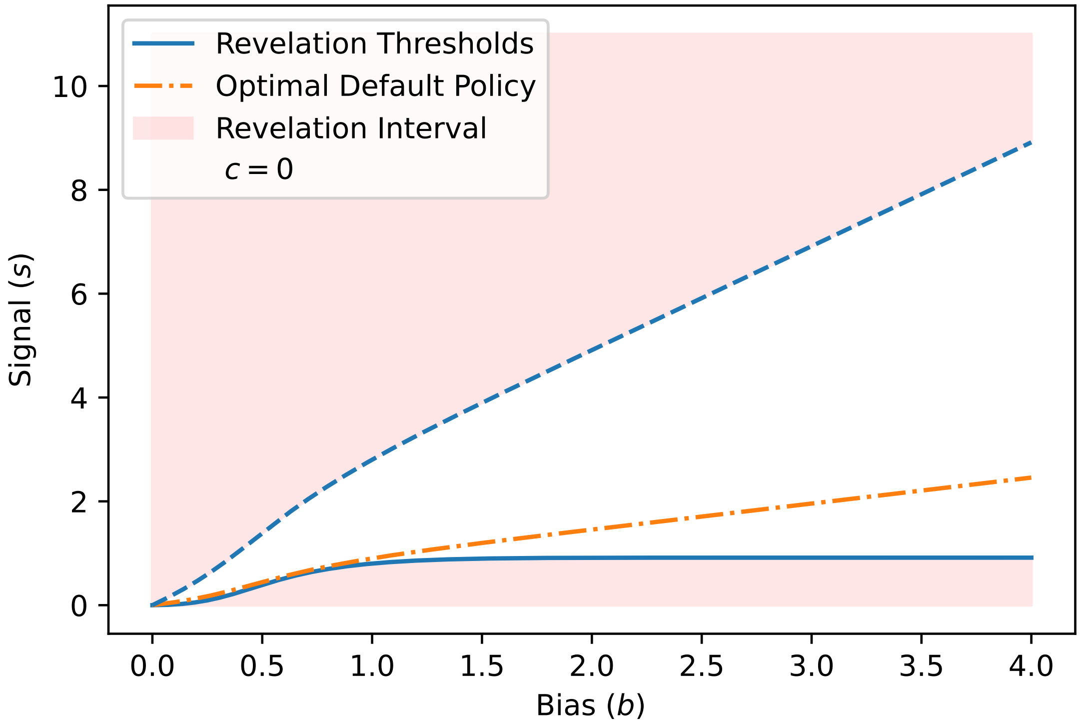
Figure 7: Agency's Revelation as a Function of its Conservatism



default policy is either excessively extreme given the observed signal (area below the solid curve in Figure 8) or not extreme enough (area above the dashed curve in Figure 8). Increasing Leader's bias induces implementation of policies more extreme than the Agency's ideal policy given the signal it observes. The Agency, thus, is less willing to reveal information that provokes more extreme policies than the default policy as the Leader's bias increases (the dashed curve in Figure 8 increases in bias b).

The change in the Agency's revelation strategy encourages the choice of a more extreme policy conditional on the lack of revelation, which has consequences like those described in the baseline model. In particular, increasing the default policy encourages the Agency to reveal more signals, namely those that produce policies less extreme than the default policy (the solid curve in Figure 8 increases in bias b). However, despite this increase, the communication between the Agency and the Leader becomes, overall, less informative as a function of bias when the Agency's conservatism is low, because the indirect effect of the bias on the default policy is bounded.

Figure 8: Agency's Revelation as a Function of the Leader's Bias (b), $c = 0$



Discussion

TO BE ADDED

A Equilibrium Characterization: Baseline Model

The Leader's optimal policy as a function of her beliefs about w is $E[w|m]+b/w$. When the Leader observes the revealing message $m=s$, she implements policy

$$a^*(m=s) = m/2 + b/2.$$

Let $d^*(b,R)$ represent the default policy the Leader implements when $m=\emptyset$. Therefore, the Agency reveals observed signal s if and only if

$$-(a^*(m=s))^2 > -(d^*(b,R))^2.$$

Substituting for $a^*(m=s)$ and isolating s,

$$-2 \cdot d^*(b) - b < s < 2 \cdot d^*(b) - b.$$

Given the agency's revelation strategy, the posterior expectation of the state of the world after $m=\emptyset$ is

$$\begin{aligned}\int_{-\infty}^{-2\cdot d^*(b)-b} x \cdot \phi(x)dx + \int_{2\cdot d^*(b)-b}^{\infty} x \cdot \phi(x)dx \\ = \frac{1}{\sqrt{\pi}} \cdot (e^{-(b/2-d)^2} - e^{-(b/2+d)^2}),\end{aligned}$$

where $\phi(\cdot)$ is the PDF of the standard normal distribution. Thus, the unique optimal default policy d the Leader implements after $m=\emptyset$ is implicitly defined by

$$d = b/2 + \frac{1}{\sqrt{\pi}} \cdot (e^{-(b/2-d)^2} - e^{-(b/2+d)^2}). \tag{11}$$

The solution to (11) is $d^*(b)$.

B Default Policy and Bias

The optimal default policy d^* solves

$$d = b/2 + \frac{1}{\sqrt{\pi}} \cdot (e^{-(b/2-d)^2} - e^{-(b/2+d)^2}). \tag{12}$$

Let's first consider case when $b > 0$. Then, $e^{-(b/2-d)^2} - e^{-(b/2+d)^2}$ exceeds zero when $d > 0$ and $e^{-(b/2-d)^2} - e^{-(b/2+d)^2}$ is less than zero otherwise. Let's assume instead that $d < 0$. Then,

$$\begin{aligned} &\frac{1}{\sqrt{\pi}} \cdot (e^{-(b/2-d)^2} - e^{-(b/2+d)^2}) \\ > &\frac{1}{\sqrt{\pi}} \cdot (e^{-(b/2-d)^2} - e^{-(0)^2}) = \frac{1}{\sqrt{\pi}} \cdot (e^{-(b/2-d)^2} - 1) \end{aligned} \tag{13}$$

We now will prove that

$$e^{-(b/2-d)^2} > (d - b/2) + 1. \tag{14}$$

Note that the LHS of 14 is always positive, while the RHS is only positive when d is sufficiently large and b is sufficiently small. Next note that $(d - b/2) + 1$ can only exceed the RHS when $(d - b/2) + 1$ is positive.

Because it is a monotone function, taking the logarithm of RHS and LHS of 14 will preserve the inequality, though its result is defined only when both sides (in particular, the RHS) are positive.

$$-(b/2 - d)^2 > \ln (d - b/2 + 1). \tag{15}$$

Next note that the derivative of the LHS of B wrt b always exceeds the derivative of RHS of B wrt b (and because both derivative are positive, it indicates that the RHS, when it exists, decreases faster than LHS) while LHS and RHS equal each other when b is equal to zero. Thus, the LHS of always exceeds the RHS of B. Next, because $(d-b/2)+1 > \sqrt{\pi}\cdot(d-b/2)+1$, it implies that d can never be negative, and, thus, d^* always exceeds $b/2$.

C Default Policy as a Function of Bias

Without loss of generality, let us assume that the Leader exhibits a right-ward bias $b > 0$. Recall that the optimal default policy $d^*(b)$ is characterized by (11).

Let us denote the following function as F:

$$F \equiv d - b/2 - \frac{1}{\sqrt{\pi}} \cdot (e^{-(b/2-d)^2} - e^{-(b/2+d)^2}). \tag{16}$$

By the implicit function theorem

$$\partial_b d(\cdot) = -\frac{\partial_b F}{\partial_d F}. \tag{17}$$

We compute that

$$\partial_b F = -\frac{1}{2} - \frac{1}{\sqrt{\pi}} \cdot \big(e^{-(b/2-d)^2} \cdot (d - b/2) + e^{-(b/2+d)^2} \cdot (d + b/2)\big). \tag{18}$$

Because d^* exceeds $b/2$,

$$\partial_b F < 0. \tag{19}$$

At the same time,

$$\partial_d F = 1 - \frac{2}{\sqrt{\pi}} \cdot \big(- e^{-(b/2-d)^2}(d - b/2) + e^{-(b/2+d)^2}(d + b/2)\big). \tag{20}$$

Because for $b > 0$, the optimal default policy d^* exceeds $b/2$, it must be that

$$-\frac{2}{\sqrt{\pi}} \cdot \big(- e^{-(b/2-d)^2}(d - b/2) > 0$$

. Therefore,

$$\partial_d F > 1 - \frac{2}{\sqrt{\pi}} \cdot e^{-(b/2+d)^2}(d + b/2). \tag{21}$$

Note that $e^{-x^2}x$ reaches its maximum of $\frac{1}{\sqrt{2e}}$ at $x = \frac{1}{\sqrt{2}}$. Therefore $-\frac{2}{\sqrt{\pi}} \cdot e^{-(b/2+d)^2}(d + b/2)$ has a minimum value $-\frac{\sqrt{2}}{\sqrt{e\pi}}$, and thus for $d = d^*$, it must be that

$$\partial_d F > 1 - \frac{\sqrt{2}}{\sqrt{e \cdot \pi}} > 0. \tag{22}$$

Combining (17), (19), and (22), we obtain $\delta_b d(\cdot) > 0$.

D Impact of Bias on Agency's Strategy

D.1 Asymmetry in Revelation

Without loss of generality, let us assume that $b > 0$. Than $\overline{s^*}(b) = 2 \cdot d - b$ is less than $2 \cdot d + b = |\underline{s^*}(b)|$.

D.2 Asymmetry as a Function of Bias

Note that the difference between the range of revelation of positive signals and the range of revelation of negative ones increases as b increases:

$$\overline{s^*}(b) - |\underline{s^*}(b)| = 2 \cdot b$$

, where $2 \cdot b$ increases in b.

E Impact of Bias on Informativeness of Communication

Because the default policy increases in bias (Proposition 2), both the direct and the indirect effects of increasing bias on lower revelation thresholds are negative and thus the lower revelation threshold $\underline{s^*}(b)$ decreases in the Leader's bias.

Let us now consider upper revelation threshold. Without loss of generality, we assume $b > 0$. From Proposition **??**,

$$\overline{s^*}(b) = 2 \cdot d^*(b) - b.$$

The derivative of the upper threshold with respect to b is

$$\frac{\partial \overline{s^*}(b)}{\partial b} = 2 \cdot \frac{\partial d^*(b)}{\partial b} - 1. \tag{23}$$

Substituting from (**??**),

$$\begin{aligned} \frac{\partial \overline{s^*}(b)}{\partial b} &= \frac{\partial \big(\frac{2}{\sqrt{\pi}} \cdot (e^{-(b/2 - d^*(b))^2} - e^{-(b/2 + d^*(b))^2}) \big)}{\partial b} \\ = \frac{2}{\sqrt{\pi}} \cdot \big(& e^{-(b/2 + d^*(b))^2} \cdot (d^*(b) + b/2) \cdot (2 \cdot \frac{\partial d^*(b)}{\partial b} + 1) \\ & - e^{-(b/2 - d^*(b))^2} \cdot (d^*(b) - b/2) \cdot (2 \cdot \frac{\partial d^*(b)}{\partial b} - 1) \big). \end{aligned}$$

To establish the result by contradiction, let us assume that $\overline{s^*}(b)$ weakly decreases in b, and, thus, at some point $\hat{b}$, and, thus, at this point $2 \cdot \frac{\partial d^*(b)}{\partial b} - 1 \leq 0$. Then, at this point, $e^{-(b/2 - d^*(b))^2} \cdot (d^*(b) - b/2) \cdot (2 \cdot \frac{\partial d^*(b)}{\partial b} - 1) \leq 0$ because $e^{-(b/2 - d^*(b))^2}$ is always positive and $d^*(b)$ exceeds $b/2$. But then, the sign of the derivative is positive, which contradicts our initial assumption. Thus, upper threshold threshold always increases in b, and bias increases informativeness of communication.

F Coarse Partition of Signal Space

Proposition 9. *Suppose that the Agency's possible actions include messages corresponding to convex partitions of the signal space. Then an increase in the Leader's bias improves the informativeness of communication between the Leader and the Agency.*

Proof. Order the messages so that m_1 indicates the interval that includes $-\infty$ and m_M indicates the interval that includes $+\infty$. To simplify exposition, we assume that $E[w|m_i] \equiv$

m_i. The Agency, then, reveals a message m_i when

$$-d^* - b/2 < m_i < d^* - b/2, \tag{24}$$

and sends the message $m = \varnothing$ otherwise. For any two messages m_i and m_j such that $m_i < m_j$, if $-d^* - b/2 < m_i$ then $-d^* - b/2 < m_j$ and if $m_j < d^* - b/2$ then $m_i < d^* - b/2$. This implies that there will be upper and lower thresholds such that the Agency sends the message revealing interval containing s when it observes signals that belong to all intervals an interval within these thresholds and conceal its information otherwise.

The default policy the Leader sets equals her posterior belief about the state of the world upon observing $m = \emptyset$ plus her bias:

$$d^*(b) = E[w|m = \varnothing] + b/2.$$

From (24), the Agency reveals its information when

$$-E[w|m = \varnothing] - b < m_i < E[w|m = \varnothing].$$

Holding fixed the Leader's posterior, $E[w|m = \emptyset]$, suppose that the Leader's bias increases. Given $b > 0$, increasing b encourages the Agency to reveal more information on the side opposite of the Leader's bias. Returning to $E[w|m = \emptyset]$, note that because of the updated revelation strategy of the Agency, the Leader will form different beliefs about the state of the world conditional on a lack of revelation. In particular, the Leader will shift her belief in the direction of her bias (counteracting the Agency's effort to shift revelation in the opposite direction). But, then, as $E[w|m = \varnothing]$ increases, the Agency will be encouraged to reveal still more information, as the lack of revelation results in still more extreme default policy implementation. □

G Equilibrium Characterization: Office Benefit

When the Leader observes a revealing message $m \neq \varnothing$, she implements policy $a^*(m \neq \varnothing) = m/2 + b/2$. Otherwise, the Leader implements default policy $d^*(b, R)$ characterized below. Given these policy choices, the Agency reveals its signal s if s is such that

$$-(a^*(m))^2 + R > -(d^*(b, R))^2.$$

Substituting for $a^*(m)$ and isolating s, we get

$$-2 \cdot \sqrt{R + d^*(b,R)^2} - b < s < 2 \cdot \sqrt{R + d^*(b,R)^2} - b. \tag{25}$$

Because of the agency's revelation strategy, the posterior expected value of the state of the world after the agency chooses $m = \emptyset$ is

$$\begin{aligned} \int_{-\infty}^{-2\cdot\sqrt{R+d^*(b,R)^2}-b} x \cdot \phi(x)dx + \int_{2\cdot\sqrt{R+d^*(b,R)^2}-b}^{\infty} x \cdot \phi(x)dx \\ = \frac{1}{\sqrt{\pi}} \cdot (e^{-(b/2-\sqrt{R+d^2})^2} - e^{-(b/2+\sqrt{R+d^2})^2}), \end{aligned} \tag{26}$$

where $\phi(\cdot)$ is PDF of standard normal distribution, and the only sequentially rational policy the Leader might implement is d solving

$$d = b/2 + \frac{1}{\sqrt{\pi}} \cdot (e^{-(b/2-\sqrt{R+d^2})^2} - e^{-(b/2+\sqrt{R+d^2})^2}) \tag{27}$$

Denote with $d^*(b,R)$ this solution.

H Effect of Office Benefit on informativeness of communication

First note that the upper revelation threshold increases in the office benefit R if and only if $(1 + 2 \cdot d \cdot \frac{\partial d^*(R,\cdot)}{\partial R})$ exceeds zero. To prove that it always exceeds zero, let us note that

$$\begin{aligned} \frac{\partial d^*(R,\cdot)}{\partial R} = \frac{1}{\sqrt{\pi}} \cdot \frac{1}{\sqrt{R+d^2}} \cdot (1 + 2 \cdot d \cdot \frac{\partial d^*(R,\cdot)}{\partial R}) \\ \cdot \big(e^{-(\sqrt{R+d^2}+b/2)^2} \cdot (\sqrt{R+d^2} + b/2) - e^{-(\sqrt{R+d^2}-b/2)^2} \cdot (\sqrt{R+d^2} - b/2)\big). \end{aligned} \tag{28}$$

Therefore,

$$\frac{\partial d^*(R,\cdot)}{\partial R} = \frac{\frac{1}{\sqrt{\pi}} \cdot \frac{1}{\sqrt{R+d^2}} \cdot \big(e^{-(\sqrt{R+d^2}+b/2)^2} \cdot (\sqrt{R+d^2} + b/2) - e^{-(\sqrt{R+d^2}-b/2)^2} \cdot (\sqrt{R+d^2} - b/2)\big)}{1 - 2 \cdot \frac{d}{\sqrt{R+d^2}} \cdot \frac{1}{\sqrt{\pi}} \cdot \big(e^{-(\sqrt{R+d^2}+b/2)^2} \cdot (\sqrt{R+d^2} + b/2) - e^{-(\sqrt{R+d^2}-b/2)^2} \cdot (\sqrt{R+d^2} - b/2)\big)}. \tag{29}$$

For simplicity, let us denote $\big(e^{-(\sqrt{R+d^2}+b/2)^2} \cdot (\sqrt{R+d^2}+b/2) - e^{-(\sqrt{R+d^2}-b/2)^2} \cdot (\sqrt{R+d^2}-b/2)\big)$ as E.

Let us first consider situation when E is below zero, then

$$\frac{\partial d^*(R,\cdot)}{\partial R} = \frac{\frac{1}{\sqrt{\pi}} \cdot \frac{1}{\sqrt{R+d^2}} \cdot E}{1 - 2 \cdot \frac{d}{\sqrt{R+d^2}} \cdot \frac{1}{\sqrt{\pi}} \cdot E} > -\frac{1}{2 \cdot d}, \tag{30}$$

and, thus, the upper revelation threshold increases in the office benefit R.

Now assume that E exceeds zero. Because

$$\begin{aligned}1 - 2 \cdot \frac{d}{\sqrt{R+d^2}} \cdot \frac{1}{\sqrt{\pi}} \cdot \big(e^{-(\sqrt{R+d^2}+b/2)^2} \cdot (\sqrt{R+d^2}+b/2) - e^{-(\sqrt{R+d^2}-b/2)} \cdot (\sqrt{R+d^2}-b/2)\big) \\ > 1 - 2 \cdot \frac{d}{\sqrt{R+d^2}} \cdot \frac{1}{\sqrt{\pi}} \cdot e^{-(\sqrt{R+d^2}+b/2)^2} \cdot (\sqrt{R+d^2}+b/2) \\ > 1 - 2 \cdot \frac{1}{\sqrt{\pi}} \cdot e^{-(\sqrt{R+d^2}+b/2)^2} \cdot (\sqrt{R+d^2}+b/2) \\ > 1 - \frac{\sqrt{2}}{\sqrt{e \cdot \pi}} > 0,\end{aligned} \tag{31}$$

when E exceeds zero, so does $\frac{\partial d^*(R,\cdot)}{\partial R}$ and, thus, $(1 + 2 \cdot d \cdot \frac{\partial d^*(R,\cdot)}{\partial R})$ exceeds zero and the upper revelation threshold increases in the office benefit.

I Office Benefit Unique Threshold

Note that when R equates to 0, $\overline{s^*}(b, R=0) > \overline{s^*}(b=0, R=0)$. When R, instead, converges to infinity, d converges to $b/2$, and $\overline{s^*}(b, R=\infty) < \overline{s^*}(b=0, R=\infty)$. Because both are continious, it implies that there will exceed a threshold R^* s.t., for all $R < R^*$, the Agency reveals strictly more information to the biased Leader than to the unbiased one.

J Agency's Revelation and Agency's Conservatism

WLOG, we assume $b > 0$. Let us note that the Agency reveals information to the Leader when

$$\begin{cases} c < 1/2,\ m \in (-\infty,\ 2 \cdot d^* - b] \cup [\frac{2 \cdot d^* + b}{1 - 2 \cdot c},\ +\infty), \\ c \geq 1/2,\ m \in [\frac{2 \cdot d^* + b}{1 - 2 \cdot c},\ 2 \cdot d^* - b], \end{cases}$$

where d^* solves

$$d = b/2 + \frac{1}{\sqrt{\pi}} \cdot (e^{-(b/2-d)^2} - e^{-\frac{(b/2+d)^2}{(1-2\cdot c)^2}}).$$

Let us denote as $F_1 \equiv d - b/2 - \frac{1}{\sqrt{\pi}} \cdot (e^{-(b/2-d)^2} - e^{-\frac{(b/2+d)^2}{(1-2\cdot c)^2}})$.

Note that the default function's derivative with respect to c is equal to

$$\frac{\partial d(c,\cdot)}{\partial c} = -\frac{\frac{\partial F_1(c,d,\cdot)}{\partial c}}{\frac{\partial F_1(c,d,\cdot)}{\partial d}},$$

where

$$\begin{aligned}\frac{\partial F_1(c,d,\cdot)}{\partial d} = 1 - \frac{2}{\sqrt{\pi}} \cdot (-e^{-(b/2-d)^2} \cdot (d-b/2) + e^{-(b/2+d)^2/(1-2\cdot c)^2} \cdot \frac{b/2+d}{(1-2\cdot c)^2}) \\ > 1 - \frac{2}{\sqrt{\pi}} \cdot e^{-(b/2+d)^2/(1-2\cdot c)^2} \cdot \frac{b/2+d}{(1-2\cdot c)^2} \\ > 1 - \frac{2}{\sqrt{\pi}} \cdot e^{-(b/2+d)^2} \cdot (b/2+d) \\ > 1 - \frac{\sqrt{2}}{\sqrt{e\cdot\pi}}.\end{aligned}$$

Therefore,

$$sgn(\frac{\partial d(c,\cdot)}{\partial c}) = -sgn(\frac{\partial F_1(c,d,\cdot)}{\partial c}).$$

Because

$$\frac{\partial F_1(c,d,\cdot)}{\partial c} = -\frac{(b+2\cdot d)^2 \cdot e^{-\frac{(b/2+d)^2}{(1-2\cdot c)^2}}}{(1-2\cdot c)^3 \cdot \sqrt{\pi}},$$

the optimal default policy d^* increases in c when $c < 1/2$ and decreases in c when $c \geq 1/2$.

Finally, when $c < 1/2$

$$\frac{\partial \frac{b+2\cdot d}{1-2\cdot c}}{\partial c} > \frac{\partial (2\cdot d - b)}{\partial c}.$$

Therefore, the Agency's revelation strictly decreases in the Agency's conservatism.

Let us now consider how Leader's bias affects the Agency's revelation for different levels of the Agency's conservatism. Because $\frac{\partial F(c,d,\cdot)}{\partial d} > 0$,

$$sgn(\frac{\partial d(b,\cdot)}{\partial b}) = -sgn(\frac{\partial F_1(b,d,\cdot)}{\partial b}).$$

$$\begin{aligned}\frac{\partial F_1(b,d,\cdot)}{\partial b} &= -1/2 + \frac{1}{\sqrt{\pi}} \cdot (e^{-(b/2-d)^2} \cdot (b/2-d)\cdot 1/2 - e^{-\frac{(b/2+d)^2}{(1-2\cdot c)^2}} \cdot (b/2+d)\cdot 1/2) \\ &= -1/2 - \frac{1}{\sqrt{\pi}} \cdot e^{-(b/2-d)^2} \cdot (d-b/2)\cdot 1/2 - \frac{1}{\sqrt{\pi}} \cdot e^{-\frac{(b/2+d)^2}{(1-2\cdot c)^2}} \cdot (b/2+d)\cdot 1/2 < 0.\end{aligned}$$

Therefore, the optimal default policy $d^*(b,\cdot)$ increases in the Leader's bias b.

The Agency reveals signal to the Leader when

$$\begin{cases} c < 1/2, \ m \leq 2 \cdot d^* - b \ or \ m \geq \frac{2 \cdot d^* + b}{1 - 2 \cdot c}, \\ c \geq 1/2, \ \frac{2 \cdot d^* + b}{1 - 2 \cdot c} \leq m \leq 2 \cdot d^* - b. \end{cases}$$

$\frac{2 \cdot d^* + b}{1 - 2 \cdot c}$ increases in b when $c < 1/2$ and decreases in b otherwise. Finally, because $Abs(\frac{\partial(2 \cdot d^* - b)}{\partial b}) < Abs(\frac{\partial \frac{2 \cdot d^2 + b}{1 - 2 \cdot c}}{\partial b})$, communication increases in bias when $c > 1/2$ and decreases in bias otherwise.

References

Austen-Smith, David. 1990. "Information transmission in debate." *American Journal of political science* pp. 124–152.

Austen-Smith, David. 1993. "Interested experts and policy advice: Multiple referrals under open rule." *Games and Economic Behavior* 5(1):3–43.

Battaglini, Marco. 2002. "Multiple referrals and multidimensional cheap talk." *Econometrica* 70(4):1379–1401.

Bhattacharya, Sourav and Arijit Mukherjee. 2013. "Strategic information revelation when experts compete to influence." *The RAND Journal of Economics* 44(3):522–544.

Blume, Andreas and Oliver Board. 2013. "Language barriers." *Econometrica* 81(2):781–812.

Crawford, Vincent P and Joel Sobel. 1982. "Strategic information transmission." *Econometrica: Journal of the Econometric Society* pp. 1431–1451.

Denisenko, Anna, Catherine Hafer and Dimitri Landa. 2022. "Competence and Advice.".

Dziuda, Wioletta. 2011. "Strategic argumentation." *Journal of Economic Theory* 146(4):1362–1397.

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

Gilligan, Thomas W and Keith Krehbiel. 1989. "Asymmetric information and legislative rules with a heterogeneous committee." *American journal of political science* pp. 459–490.

Hagenbach, Jeanne and Frédéric Koessler. 2020. "Cheap talk with coarse understanding." *Games and Economic Behavior* 124:105–121.

Lipnowski, Elliot and Doron Ravid. 2020. "Cheap talk with transparent motives." *Econometrica* 88(4):1631–1660.

Milgrom, Paul. 2008. "What the seller won't tell you: Persuasion and disclosure in markets." *Journal of Economic Perspectives* 22(2):115–131.

Milgrom, Paul R. 1981. "Good news and bad news: Representation theorems and applications." *The Bell Journal of Economics* pp. 380–391.

Shin, Hyun Song. 1994. "The burden of proof in a game of persuasion." *Journal of Economic Theory* 64(1):253–264.

Sobel, Joel. 2013. "Giving and receiving advice." *Advances in economics and econometrics* 1:305–341.

Wolinsky, Asher. 2003. "Information transmission when the sender's preferences are uncertain." *Games and Economic Behavior* 42(2):319–326.