Non-monotonic Disclosure in Policy Advice*

Anna Denisenko† Catherine Hafer‡ Dimitri Landa§

Sep, 2025

Abstract

The strategic context of bureaucratic advice to policymakers often takes the form of a disclosure game in which the relevant bureaucracy has an ideal policy interior to the policymaker's action space. We characterize conditions under which this game has sequential equilibria in which the sender adopts a non-monotonic disclosure strategy, implying partial disclosure. Further, multiple sequential equilibria exist under some conditions, including a fully revealing equilibrium and multiple partially revealing equilibria that vary in extent of disclosure. We show that these equilibria are strictly rankable both by actors' welfare and by their robustness to belief perturbations. Further, for sender preferences that are sufficiently close to the expected value of the state, (1) the most robust equilibrium is partially revealing and (2) set of states that the sender discloses becomes larger as the divergence in sender's and receiver's ex ante preferences increases.

1 Introduction

Effective policy-making requires expert information, for which policymakers must often rely on bureaucratic agencies. Because agencies have their own policy preferences, the problem of strategic disclosure – how to effectively motivate agencies to disclose information available to them without undermining policy-making – is a first-order concern in understanding incentives in policy-making.

While the literature on disclosure has developed important insights that shed light on the strategic logic of disclosure, existing models of disclosure assume away a crucial feature present in many settings: experts/agents/bureaucrats have an ideal action they would like to see implemented. Indeed, in many instances, these ideal actions are state-independent – for example, bureaucrats are often described as being strongly biased in favor of the status

*We thank Steve Callander, Wioletta Dziuda, Anthony Fowler, Amanda Friedenberg, Kun Heo, Roger Myerson, Keith Schnakenberg, and participants in the Virtual Formal Theory Workshop, Stanford GSB PE Workshop, NYU PE Workshop, the Political Economy Winter Workshop at Bocconi University, the Harris PE Workshop and Harris Political Economy of American Democracy Conference for valuable feedback on earlier versions of this paper.

†Post-Doctoral Researcher, Harris School of Public Policy, University of Chicago, 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

quo (not the least, because they are often residual claimants on costs of policy changes – see Kaufman (1981); McCarty (2004)). The presence of ideal actions belies the standard motivating examples of the disclosure literature, in which the sender always wants a higher action (e.g., selling more cars).

We study a model of disclosure that departs from the Milgrom (1981) canonical setting in focusing on senders with such preferences. We show that it generates a host of predictions that substantially depart from the conventional prediction of the "unraveling" logic of disclosure whereby an informed agent has an incentive to disclose her information to avoid the decision-maker inferring the worst possible state from non-disclosure (Grossman, 1981; Milgrom, 1981). While the studies of disclosure have consistently assumed monotonicity of sender preferences, we show that what the unraveling result needs is a sender with a most-preferred action that is sufficiently far from the expected value of the distribution of states. Monotonicity is not necessary for the existence of the full-disclosure equilibrium, but given non-monotonicity, if the sender's most-preferred action is within a specified interval around the expected value of the state, that can fundamentally change the nature of disclosure. Specifically, it can give rise to equilibria in which unraveling stops before being complete. We delineate two types of partial disclosure equilibria that can be sustained in such a case: the *guarded* equilibrium, where the sender reveals relatively less information, and the *expansive* equilibrium, where it reveals relatively more. We show that only guarded and the full disclosure equilibria are *belief-stable* in the sense of robustness to small perturbations in players' beliefs, and that, as the ideal action of the sender moves close to the expected value, the extent of such robustness is highest in the guarded equilibrium. Starkly, our analysis demonstrates that in this equilibrium, disclosure increases with ex ante preference divergence between the sender and the receiver, contrasting sharply with the canonical prediction on the effects of preference divergence on communication in the cheap-talk signaling context.

The rest of the paper is organized as follows. In Section 2, we present an illustrative example that helps motivate our theoretical framework. Section 3 formalizes the model and characterizes the equilibria under different conditions. Section 4 discusses the concept of belief-stable equilibria and its implications. In Section 5, we extend our analysis by generalizing to broader settings and explore the robustness of our results. Section 6 develops the core idea in several directions, including in the context with direct disclosure-contingent benefit to the sender.

2 Connection to the Literature

A key result in the communication games with verifiable messaging is that all private information is revealed in equilibrium (Grossman (1981), Milgrom (1981), Milgrom (2008)). Following the initial formulation of this result, subsequent works have studied conditions under which the unraveling logic of full disclosure remains intact in a variety of formal and substantive environments. An important review of the disclosure games literature to-date in Milgrom (2008). A key lesson from this literature has been the general robustness the unraveling prediction.

To create the possibility for equilibrium non-disclosure, the sender must not be fully informed (Dye (1985), Jung and Kwon (1988), Shin (1994)). Alternatively, the receiver must be uncertain about sender's preferences (Wolinsky (2003), Dziuda (2011)). Callander, Lambert and Matouschek (2021) find that incomplete disclosure can also be supported in a (stylized multi-dimensional) setting in which the sender can not only provide a direct recommendation but also referential information that can influence the decision-maker's beliefs about other options.

The analysis of the incentives to disclose has focused on settings where the sender's revelation is monotonic in the state. In contrast to the preference for always "higher" or always "lower" policies (e.g., selling more cars in Milgrom's seminal example), preference satiation, which corresponds to interior ideal points on the range of possible policy alternatives, makes it possible to have interior boundaries on revelation intervals and non-monotonic revelation preferences, which we show to be supportable in equilibrium. Preference satiation is a standard feature of preferences in political economy contexts, which often work with the spatial model of preferences, where they are represented as ideal points. Seidmann and Winter (1997) provide a generalization of Milgrom's classic setting to environments with objective functions that are concave in actions. The key assumption in their study is that sender's utility is more state-dependent than the receiver's – the opposite of the assumption we maintain in this paper. Denisenko, Hafer and Landa (2024) study the transmission of verifiable information from an unbiased sender with a fixed and known ideal point, and focuses on the effects of sender competence on information transmission.

Delegation and communication within hierarchies is a focus of a substantial body of political economy scholarship (for reviews, see Gailmard and Patty (2012) and Sobel (2013)). The dominant approach to modeling communication in this literature has been 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 comparative static result that is at the core of this literature is the opposite of what we provide below for the setting with verifiable messaging – viz., that divergence in the actors' preferences curtails communication, and successful communication at all occurs only when the advisor's and the policymaker's preferences are sufficiently aligned. An important exception 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 may not be able to 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 disclosure. The mechanisms producing these results are, however, very different. (Battaglini (2002) and Aybas and Callander (2023) identify policy-relevant settings in which cheap-talk communication fully favors, respectively, the receiver, and the sender.)

3 Uniform Prior and State-independent Preferences

We begin with a relatively simple setting that illustrates the key features of equilibrium results. We later generalize the model with respect to both the distribution of the state and the functional form of utilities.

Assume there is an Agency and a Policymaker, where the Agency possesses information relevant to the Policymaker's current political agenda. The Agency has discretion over whether to share this information with the Policymaker, who then chooses a policy $p \in \mathbb{R}$ based on the received message. The Policymaker aims to implement a policy p that aligns with the true state of the world, ω. When the Agency discloses information, it does so truthfully.

The Policymaker's utility function is

$$u_P(p) = -(\omega - p)^2.$$

Thus, the Policymaker's objective is to set policy $p = \omega$. We begin with a simplified model where the Agency's preferences are state-independent; its utility is maximized when the implemented policy matches its most preferred policy, denoted by i. This allows us to isolate the key mechanisms before generalizing to state-dependent Agency preferences. The Agency's utility function is

$$u_A(p) = -(i - p)^2.$$

Note that while i captures only the Agency's preferences, the absolute value $|i|$ measures the divergence in the two actors' ex ante preferences. The Policymaker's ex ante preferred policy is $p = \mathbb{E}[\omega]$, while the Agency's is $p = i$, with the distance between these reflecting the alignment of their preferences.

The state of the world, ω, is drawn from a continuous distribution with cumulative distribution function (cdf) $F(\cdot)$ and probability density function (pdf) $f(\cdot)$ over a support Ω. For initial tractability and clarity of exposition, in the special case we assume ω is uniformly distributed on $\Omega := [-1, 1]$, i.e., $f(\omega) = 1/2$ for $\omega \in [-1, 1]$ and zero otherwise. The Agency observes ω and decides whether to disclose it to the Policymaker. The Agency's information is verifiable and it can send one of two messages: $m \in \{\omega, \varnothing\}$. The message $m = \varnothing$ is commonly understood to be not intrinsically informative.[1]

The timing of the game is as follows:

1. Nature draws the state of the world ω.
2. The Agency observes ω and chooses a message m.
3. The Policymaker observes m and selects a policy $p \in \mathbb{R}$.

Let $\mu : \Omega \to [0, 1]$ represent the Agency's disclosure strategy, where $\mu(\omega)$ is the probability the Agency discloses state ω. The Policymaker's beliefs, conditional on observing message m, are represented by the pdf β. Consistent with verifiable information, $\beta(\omega|m = \omega) = 1$ and $\beta(\omega|m = \hat{\omega} \neq \omega) = 0$. From Bayes' Rule,

$$\beta(\omega|\varnothing; \mu) = \frac{(1 - \mu(\omega))f(\omega)}{\int_\Omega (1 - \mu(\hat{\omega}))f(\hat{\omega})d\hat{\omega}}.$$

[1] The Agency's message space is restricted, excluding vague but truthful messages, which provides the most challenging setting for our results, as we discuss later.

The Policymaker's optimal strategy $p^* : \Omega \cup \varnothing \to \mathbb{R}$ maximizes the Policymaker's utility given her beliefs conditional on m.

The following lemma summarizes the Policymaker's optimal policy-implementation strategy:

Lemma 1. *The Policymaker's optimal policy $p^*(m)$ is*

$$p^*(m) = \begin{cases} m \text{ if } m \neq \varnothing, \\ \mathbb{E}[\omega|\varnothing; \mu^*(.)] \text{ if } m = \varnothing, \end{cases} \tag{1}$$

where $\mu^*(.)$ denotes the Policymaker's conjecture about the Agency's disclosure strategy.

The Agency discloses its information to the Policymaker when withholding it would result in a policy farther from the Agency's preferred policy than the one the Policymaker would implement if fully informed about the state. The following lemma describes the Agency's optimal disclosure strategy, given the Agency conjectures that the Policymaker's strategy is of the form specified in Lemma 1, with $x := p^*(\varnothing)$.

Lemma 2. *The Agency's optimal disclosure strategy is*

$$\mu^*(\omega) = \begin{cases} 1 \text{ if } \omega \in M(x,i), \\ 0 \text{ if } \omega \in N(x,i), \end{cases} \tag{2}$$

where $M(x,i) := [i - \sqrt{(x-i)^2}, i + \sqrt{(x-i)^2}] \cap \Omega$ *and* $N(x,i) := \Omega \backslash M(x,i)$.[2]

Although some signals are never disclosed, the Policymaker infers in equilibrium that when she receives $m = \varnothing$, the state fell outside the disclosure interval, and she updates her beliefs about the state absent disclosure accordingly. In every equilibrium, after observing ω, the Agency follows an optimal disclosure strategy $\mu^*(\omega)$, anticipating that absent disclosure the Policymaker will select the optimal policy, denoted

$$x^* := \mathbb{E}[\omega|\varnothing; \mu^*(.)].$$

In the next proposition, we characterize all disclosure strategies supported in Sequential Equilibrium (SE); for the remainder of the paper, "equilibrium" will mean Sequential Equilibrium.

Proposition 1. *1. For all $i \in \Omega$, a full disclosure strategy can be supported in equilibrium, with $x^* = x_F$ and the disclosure interval $M(x^*, i) = \Omega$, where*

$$x_F := \begin{cases} 1 & \text{if } i \leq 0, \\ -1 & \text{if } i \geq 0. \end{cases}$$

[2]Both $m = \omega$ and $m = \varnothing$ are optimal if $\omega = i - \sqrt{(x^* - i)^2}$ or $\omega = i + \sqrt{(x^* - i)^2}$. Henceforth, we will assume the Agency discloses when indifferent, with no substantive effect on our results.

2. *For $i \in [-\frac{1}{4}, \frac{1}{4}]$, two partial disclosure strategies can be supported in equilibrium, with $x^* \in \{x_E, x_G\}$ and the disclosure interval $M(x^*, i) \subset \Omega$, where*

$$x_E := \frac{1}{2}\left(2 \cdot i - sign(i) - sign(i) \cdot \sqrt{1 - 4 \cdot |i|}\right);$$

$$x_G := \frac{1}{2}\left(2 \cdot i - sign(i) + sign(i) \cdot \sqrt{1 - 4 \cdot |i|}\right).$$

Proof. See Appendix. □

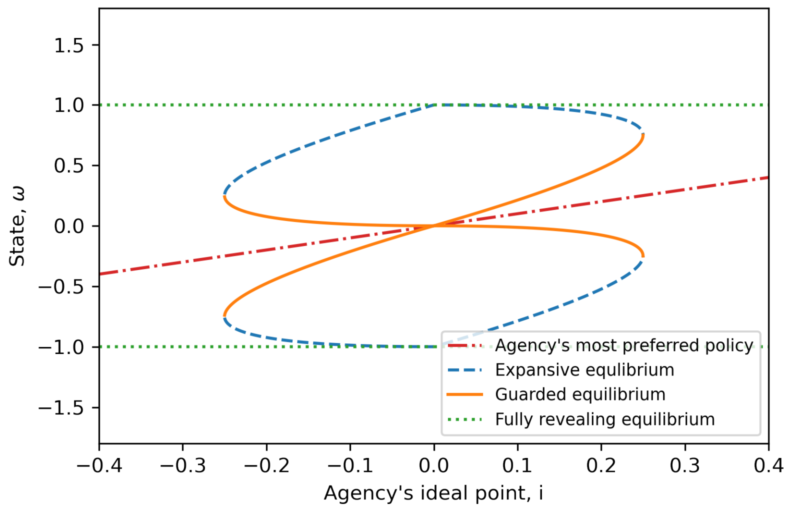

Figure 1: Agency's disclosure boundaries in fully revealing, guarded, and expansive equilibria as a function of Agency's ideal point i.

We will, for convenience, adopt the following terminology for this special case. We refer to the partial disclosure strategy characterized by x_E as the *expansive* disclosure strategy, and the partial disclosure strategy characterized by x_G as the *guarded* disclosure strategy. We term the equilibria in which these strategies are employed as the expansive partial equilibrium and the guarded partial equilibrium, correspondingly. This nomenclature is motivated by the difference in the extent of information revelation across equilibria: the Agency discloses a strictly broader set of states in the expansive equilibrium than in the guarded equilibrium. This feature is visually represented in Figure 1, which juxtaposes all equilibrium disclosure boundaries. As the figure demonstrates, the guarded disclosure interval is nested within the expansive disclosure interval, which, in turn, lies within the full disclosure interval for all values of i.[3]

The nestedness of the disclosure strategies has a direct implication for the Policymaker's welfare, summarized in the following corollary.

[3]While the nestedness of equilibrium disclosure sets is maintained across different prior distributions, the existence of at most two partial disclosure equilibria is a consequence of the uniform prior assumption.

Corollary 1. *For any given Agency's ideal point i, the Policymaker's ex ante expected utility is highest when there is full disclosure in equilibrium and lowest in the (partial disclosure) guarded equilibrium.*

The Policymaker trivially prefers more disclosure to less, and the full disclosure equips the Policymaker with complete information about the state, letting her select policies tailored best to her preferences. In contrast, partial disclosure leaves the Policymaker with residual uncertainty, leading to suboptimal policy choices and reduced utility. The welfare loss is larger when disclosure is smaller and, therefore, Policymaker's utility is lowest under the guarded equilibrium.

Proposition 1 establishes that for any given ideal point i, there are at most three pure-strategy equilibrium profiles. We now argue that mixed strategies are not sustainable in any equilibrium. For any state realization ω, and conditional on the policy implemented in the absence of disclosure, the Agency has a strict preference for disclosing states that produce a policy closer to its ideal point. Probabilistic disclosure is therefore never incentive compatible. Similarly, the Policymaker's optimal policy choice is also restricted to pure strategies. Were the Policymaker to randomize across policies following non-disclosure, this would induce a change in the agency's optimal disclosure interval. However, given any disclosure interval chosen by the Agency, there exists a unique optimal policy for the Policymaker in the absence of disclosure that constitutes a best response. Thus, neither player finds it optimal to employ mixed strategies in equilibrium.

As Figure 1 illustrates, partial disclosure is non-monotonic in realized states of the world. It is this non-monotonicity that prevents full unraveling: The Agency's unwillingness to disclose signals both too high and too low, relative to i, ensures that policy absent disclosure is not too extreme. Consequently, since the expected state conditional on non-disclosure is not too extreme, the Agency optimally chooses to withhold some states on either side of its ideal point.

The unique equilibrium where disclosure is monotonic in realized states and full, is the equilibrium where the Policymaker chooses policy x_F absent disclosure. Proposition 1 establishes sufficient and necessary conditions for full disclosure to be the unique equilibrium of the game. Specifically, when Agency's ideal point i, is sufficiently far from the mean of the prior distribution (or, alternatively, when ex ante preference divergence $|i|$ is sufficiently high) , the unique equilibrium involves full disclosure. Conversely, when i lies within the interval $[-\frac{1}{4}, \frac{1}{4}]$ (when $|i| < \frac{1}{4}$), multiple equilibria exist, including those with partial disclosure.

Further, at $i = 0$, the Agency's optimal disclosure strategy is symmetric around zero, and any lack of disclosure should indicate to the Policymaker that the optimal policy in the absence of information is equal to the mean of the prior distribution. Yet, the Policymaker's beliefs about the average concealed state in both expansive and full equilibria are different from zero at $i = 0$. Full disclosure becomes possible in these cases because the Policymaker forms extreme beliefs, leading her to implement an extreme policy (-1 or 1) absent disclosure, thereby encouraging the Agency to fully disclose information, reinforcing the Policymaker's beliefs.

Note that as $|i|$ approaches zero, the Agency's incentives to disclose or withhold information become increasingly symmetric around the mean of the distribution of states. The Agency, being indifferent between right-leaning and left-leaning policies, does not prioritize the disclosure of one over the other. Thus, all else equal, as $|i|$ converges zero, the expected state in the absence of disclosure should converge to the mean of the prior distribution (zero), following the Agency's optimal strategy as established in Lemma 2 (with the Policymaker's strategy, x, held constant). While this anticipated state convergence is sustained in the guarded disclosure equilibrium, this comparative statics fails in equilibria characterized by x_F or x_E absent disclosure. In these latter cases, as the Policymaker continuously updates her beliefs about the state expecting the Agency to disclose (weakly) more information as preference misalignment decreases, this belief becomes self-reinforcing, realizing in greater disclosure from the Agency.

3.1 Comparative Statics

In this section, we examine how the game's parameters affect the players' decisions and the resulting outcomes. Our first result details how the Agency's ideal point, i, impacts the Policymaker's beliefs and choices in the absence of disclosure:

Proposition 2. *Increasing i, the difference between the Agency's ideal point and the Policymaker's ex ante expected ideal point,*

1. *has no effect on the policy implemented in the case of nondisclosure $x^* = x_F$, given $i \neq 0$, in the full disclosure equilibrium;*
2. *decreases $x^* = x_G$ on $i \in [-\frac{1}{4}, \frac{1}{4}]$ in the guarded disclosure equilibrium; and*
3. *increases $x^* = x_E$ on $i \in (-\frac{1}{4}, 0) \cup (0, \frac{1}{4})$ but discontinuously decreases it at $i = 0$ in the expansive disclosure equilibrium.*

Proof. See Appendix. □

Figure 2 illustrates the policy chosen by the Policymaker in the absence of disclosure, as a function of the Agency's ideal point i, across the three equilibria: the guarded disclosure equilibrium, the expansive disclosure equilibrium, and the full disclosure equilibrium. Notably, the comparative statics for the guarded and expansive equilibria present a stark contrast: the equilibrium policy absent disclosure decreases in the Agency's ideal point i in the former, while it increases with i (for $i \neq 0$) in the latter.

We, next, characterize the impact of ex ante preference divergence $|i|$ on the actors' communication.

Proposition 3. *Increasing the magnitude of the ex ante preference divergence of the Agency from the Policymaker, $|i|$,*

1. *has no effect on disclosure in the full disclosure equilibrium;*
2. *(weakly) decreases disclosure in the expansive equilibrium; and*
3. *increases disclosure in the guarded equilibrium.*

Proof. See Appendix. □

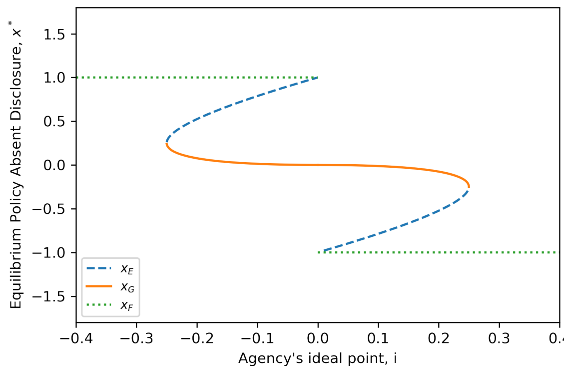

Figure 2: Policymaker's equilibrium policy selection in the absence of disclosure.

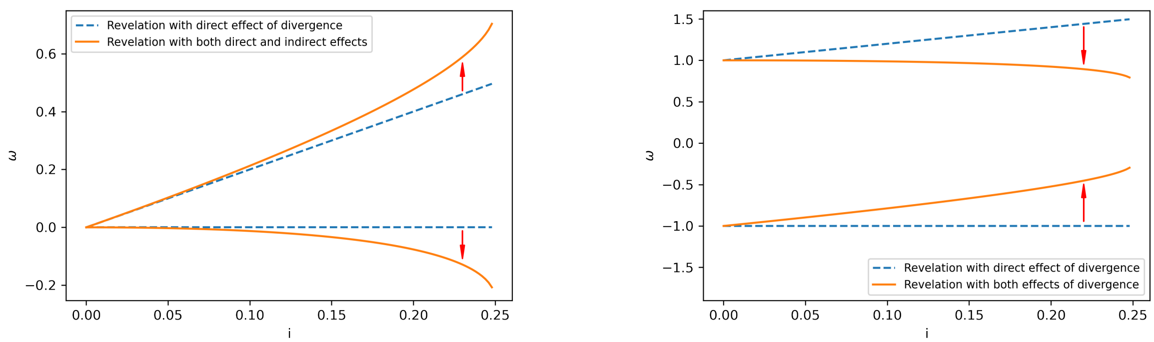

Figure 3: Direct and indirect effects of ex ante preference divergence on Agency's equilibrium disclosure in the guarded (left panel) and expansive (right panel) equilibria for $i \geq 0$.

The ex ante preference divergence has two effects on the Agency's disclosure strategy. First, there is a *direct effect*: Policymaker's strategy being fixed, the Agency discloses more information as divergence increases to obtain policy closer to its own, relatively more extreme, tastes. However, there is a secondary *indirect effect* of the ex ante divergence outlined in Proposition 2: the Agency needs to take into account how its changing disclosure strategy will affect the Policymaker's beliefs, and hence the policy she implements, after non-disclosure. Disclosure depends on both the direct and indirect effects of preference divergence.

Figure 3 shows direct and indirect effects in the guarded equilibrium in the left panel, and in the expansive equilibrium in the right panel. In the guarded equilibrium, the two effects reinforce each other, and so as ex ante preference divergence increases, communication becomes more informative. Greater divergence implies that the Agency's preferences become increasingly asymmetric relative to the mean of the prior distribution. This asymmetry shapes the Agency's incentives, encouraging it to prioritize disclosure of states on one side relative to the other, allowing the Policymaker to infer more about the state when information is withheld. Consequently, the policy choice in the absence of disclosure shifts away from the prior mean, further encouraging the Agency to disclose information.

In the expansive equilibrium, however, the direct and indirect effects are opposed. Thus, greater divergence in the Agency's preferences encourages the Agency to disclose more, which results in the disclosure interval being more symmetric and hence non-disclosure is less informative. The Policymaker, in the absence of disclosure, chooses policy closer to the prior mean, which, in turn, dampens the Agency's incentives to disclose. Here the direct and indirect effects of communication counteract each other. In this equilibrium, the indirect effect dominates the direct effect, leading to a decline in disclosure as preferences diverge.

4 Belief-Stable Equilibria

Given the different comparative static predictions in the different equilibria we describe, it is useful to consider more closely the relative stability of the corresponding equilibria. As we detail above, all three – the full disclosure, guarded, and expansive equilibria – are sequential equilibria, and all three satisfy standard action-perturbation refinement conditions. With this in mind, in this section, we present a different stability analysis, focusing on possible perturbations in beliefs, which is able to usefully differentiate between these equilibria.

We begin with the following definition:

Definition 1. *Consider an equilibrium strategy profile and system of beliefs* (σ, β) *and a perturbed system of beliefs* β_j^ε*. Let* σ^ε *be sequentially rational given the beliefs* $(\beta_j^\varepsilon, \beta_{-j})$*, and let* $\hat{\beta}_j^\varepsilon$ *be consistent with* σ^ε*. If there exists an* $\varepsilon > 0$ *such that, for every* β_j^ε *that satisfies* $|\beta_j^\varepsilon(y) - \beta_j(y)| < \varepsilon$*, condition* $|\hat{\beta}_j^\varepsilon(y) - \beta_j(y)| \leq |\beta_j^\varepsilon(y) - \beta_j(y)|$ *is satisfied for all decision nodes* y *assigned to* j*, then we say that equilibrium* (σ, β) *is* ***belief-stable for player j****. Equilibrium* (σ, β) *is* ***belief-stable*** *if it is belief-stable for every player* j*, and is* ***belief-unstable*** *otherwise.*

Definition 2. *Let* ε_j^* *be the largest value* $\varepsilon > 0$ *such that, for every* β_j^ε *that satisfies* $|\beta_j^\varepsilon(y) - \beta_j(y)| < \varepsilon$*, condition* $|\hat{\beta}_j^\varepsilon(y) - \beta_j(y)| \leq |\beta_j^\varepsilon(y) - \beta_j(y)|$ *is satisfied for all decision nodes* y *assigned to* j*. We say* ε_j^* *is the* **extent of belief-stability of** (σ, β) **for player j.**

Intuitively *belief-stability* ensures that small perturbations in players' beliefs do not result in large deviations in their optimal strategies: in belief-stable equilibria, if players' beliefs depart slightly from equilibrium beliefs, the feedback they receive as they play the game reinforces the equilibrium beliefs. In contrast, if the equilibrium is belief-unstable, feedback will provoke larger and larger deviations from equilibrium beliefs.

The next proposition applies the concept of belief-stability to the sequential equilibria in the game.

Proposition 4.

1. *The guarded equilibrium is belief-stable if* $|i| \neq \frac{1}{4}$*.*

2. *The expansive equilibrium is belief-unstable.*

3. *The full disclosure equilibrium is belief-stable for all* $|i| \neq 0$*.*

4. *For* $|i| \leq \frac{1}{4}$*, the extent of belief stability of the fully disclosure equilibrium increases and of the guarded disclosure equilibrium decreases in* $|i|$*.*

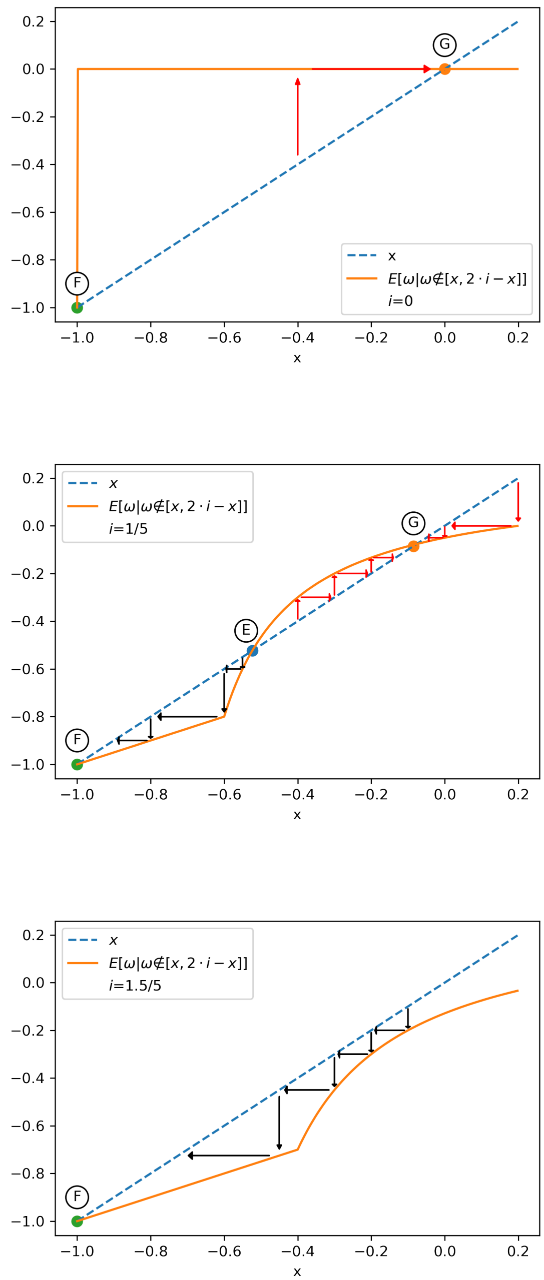

Figure 4: Policymaker's beliefs absent disclosure and the policy adopted in the absence of disclosure

Figure 4 illustrates three substantively different scenarios in this game: when $i = 0$, $i \in (0, 1/4]$, and $i > 1/4$. Let x-axes represent x, the policy adopted in the absence of disclosure. The blue dashed line is the 45° line, i.e., $y = x$. Given x, the Agency's optimal disclosure interval for $i \geq 0$ is $M(x, i) = [x, 2 \cdot i - x] \cap \Omega$. The orange solid line represents, then, the Policymaker's expected value of ω given non-disclosure and supposing the Agency uses this disclosure interval. Because the Policymaker's equilibrium policy choice x^* is such that $x^* = E[\omega | \omega \notin M(x^*, i)])$ in equilibrium, i.e., the blue dashed line and the orange solid line intersect, defining equilibrium x^*.

Arrows in Figures 4 (a)-(c) show the direction of the best-response updating following the initial perturbation in the Policymaker's beliefs. For example, if in the expansive equilibrium, the Policymaker's $E[\omega|m^*(\omega) = \varnothing]$ shifts from $x = x_E$ to $x = x_E + \varepsilon$, the Agency responds by adjusting its disclosure strategy rightward, revealing less information. In turn, the Policymaker updates her beliefs such that $E[\omega|\omega \notin M(x_E + \varepsilon, i)]] > x_E + \epsilon$. In fact, the expansive equilibrium is belief-unstable and any perturbation of beliefs will cause adjustments that move behavior and beliefs farther from this equilibrium. Depending on the nature of the initial deviation, the game will converge to an equilibrium where the Policymaker's expectation of the state given non-disclosure is either x_G or x_F.[4] In contrast, small perturbations in the Policymaker's beliefs after non-disclosure in the guarded equilibrium will give rise to adjustments that lead back to to the guarded equilibrium.

Finally, while both the full disclosure equilibrium and the guarded disclosure equilibrium satisfy belief-stability, their robustness to belief perturbations exhibits opposing trends with ex ante preference convergence (decreasing $|i|$). As preferences ex ante align more closely, the extent of belief stability decreases in the full disclosure equilibrium, but it increases in the guarded equilibrium. Consequently, for sufficiently aligned preferences, partial disclosure emerges as the more belief-stable outcome. Figure 5 illustrates these belief-stability regions as a function of preference divergence (i).

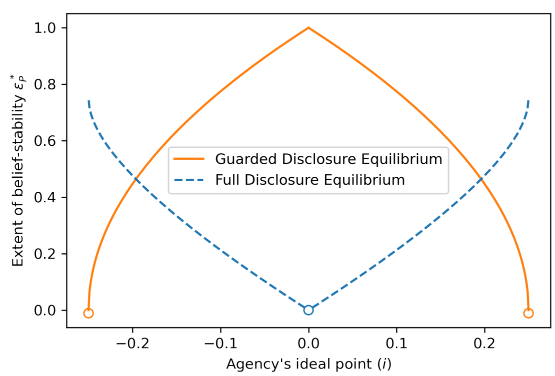

Figure 5: Extent of belief-stability for the guarded equilibrium and for the equilibrium with full disclosure as a function of the Agency's most preferred policy i.

5 A (More) General Model

Let the state space $\Omega \subseteq \mathbb{R}$ be a compact set, with the convex hull denoted by $Conv(\Omega) = [\underline{\Omega}, \overline{\Omega}]$.[5] The state of the world, $\omega \in \Omega$, is drawn from a continuous cumulative distribution function F with a corresponding probability density function f, where $E[\omega] = 0$.

[4] Given that $|i| \notin \{0, 1/4\}$.

[5] While compactness of Ω is not strictly required for the existence of partial disclosure equilibria, it is important to ensure that the concept of full disclosure equilibrium is well-defined. Without compactness, full disclosure may not be attainable.

The Policymaker's von Neumann-Morgenstern (vN-M) utility function $u_P(p;\omega)$ is C^1, is strictly concave in p, and the Policymaker's ideal policy is $p^P(\omega) := \arg\max_p u_P(p;\omega) = \omega$.[6] The Agency's vN-M utility function $u_A(p;\omega,\alpha,i)$ is C^1, is strictly concave in p, and the Agency's ideal policy given by $p^A(\omega,\alpha,i) := \arg\max_p u_A(p;\omega,\alpha,i) = \alpha \cdot p^P(\omega) + (1-\alpha)\cdot i$, where $i \in \mathbb{R}$ is the Agency's *preference (additive) bias* and $\alpha \in [0,1]$ is the Agency's *preference state dependence*. A higher α indicates that the Agency's preferences are more closely aligned with those of the Policymaker, while a lower α indicates a stronger weight for the Agency's bias i. Let $p_0^P := \arg\max_p \mathbb{E}[u_P(p;\omega)]$ be the Policymaker's ex ante optimal policy.

We analyze the equilibrium disclosure strategies, which fall into three categories based on the disclosure set, $M \subseteq \Omega$. In a *Full Disclosure Equilibrium (FDE)*, the Agency reveals the state for all $\omega \in \Omega$. In a *Partial Disclosure Equilibrium (PDE)*, the Agency withholds information for a non-empty set of states, so that $\varnothing \neq M \subset \Omega$. Finally, we distinguish a substantively important limiting case of partial disclosure, which we will refer to as a *Non-Disclosure Equilibrium (NDE)*. In an NDE, the disclosure set is the singleton $M = \{p_0^P\}$. This outcome is substantively equivalent to complete non-disclosure from the Policymaker's perspective, as her posterior belief upon observing non-disclosure remains her prior.

As in Section 4, we focus attention on belief-stability of equilibria (see Definition 1). Recall that an equilibrium is belief-stable if the belief-updating dynamic is locally convergent, meaning small errors in beliefs are self-correcting rather than self-amplifying. The equilibrium is belief-unstable otherwise.

This section proceeds by first establishing that all equilibria in the model follow a highly structured pattern: their disclosure sets are nested, and they alternate in terms of belief-stability. This alternating property implies that knowing the belief-stability of a single equilibrium is sufficient to determine the belief-stability of all others. The analysis then provides the specific conditions for the existence and belief-stability of the FDE and NDE before characterizing the conditions for the existence of PDE. The section concludes with comparative statics, examining how the outcomes in belief-stable equilibria change in response to model parameters.

Let the Policymaker's best-response function, $\hat{x}(x)$, be defined as the optimal policy given the belief that the Agency's disclosure rule, $M(x,\alpha,i)$, is based on a non-disclosure policy x. Formally,

$$\hat{x}(x) := \arg\max_p \int_{\omega \in N(x,\alpha,i)} u_P(p;\omega)\, dF(\omega), \tag{3}$$

where $N(x,\alpha,i) = \Omega \setminus M(x,\alpha,i)$ is the set of states the Agency would not disclose if the anticipated non-disclosure policy were x.

Every equilibrium non-disclosure policy $x^* \in X^*$ is a fixed point of this function, $x^* = \hat{x}(x^*)$. Let $X^*(\alpha,i)$ be the set of all equilibrium policies implemented by the Policymaker upon non-disclosure, x^*, for a given preference profile (α,i). The following proposition establishes that these equilibria exhibit a highly structured pattern.

Proposition 5.

[6] We assume here that the Policymaker is unbiased and show that all the results hold for biased Policymaker in the appendix.

1. *Let x_j^* and x_k^* be any two distinct equilibrium non-disclosure policies in $X^*(\alpha, i)$. The corresponding disclosure intervals are nested according to the Agency's utility for the non-disclosure policy. Formally,*

$$u_A(x_j^*; \omega, \alpha, i) \leq u_A(x_k^*; \omega, \alpha, i) \; \forall \, \omega \in \Omega \Leftrightarrow M(x_k^*, \alpha, i) \subseteq M(x_j^*, \alpha, i).$$

2. *Let the set X^* be ordered such that if $j < k$ then $x_j^* < x_k^*$, then the belief stability of equilibria must alternate along this ordering.*

Proof. See Appendix. □

A direct corollary of part (1) of Proposition 5 is that all equilibria can be strictly ranked by the Policymaker's ex-ante welfare. An equilibrium sustained by a non-disclosure policy more favorable to the Agency features a smaller disclosure set, increasing the Policymaker's residual uncertainty. This inverse relationship is starkest in the limiting cases. The NDE, if it exists, is most preferred by the Agency, as it reveals the least information. The FDE, when exists, maximizes the Policymaker's ex-ante welfare and minimizing that of the Agency.

Part (2) of Proposition 5 shows that for any given preference profile, belief-stable and belief-unstable equilibria must alternate along any ordered set of equilibrium policies absent disclosure. This property means we do not need to analyze the belief-stability of every equilibrium. Instead, if we can determine the belief-stability of a single one – for instance an FDE – we can infer belief-stability of all others. We formalize it in the following corollary.

Corollary 2. *For any given preference profile, knowing the belief-stability of one equilibrium is sufficient to determine belief-stability of all equilibria.*

The following proposition establishes conditions under which an FDE can be sustained and, critically, when it is belief-stable, providing necessary anchor for belief-stability analysis of all other equilibria.

Proposition 6.

1. (a) *If $u_A(p^P(\overline{\Omega}); \overline{\Omega}, \alpha, i) > u_A(p^P(\underline{\Omega}); \overline{\Omega}, \alpha, i)$, then for all conditions on primitives such that there exists a FDE s.t. $x^* = p^P(\underline{\Omega})$, that equilibrium is belief-stable;*

   (b) *If $u_A(p^P(\overline{\Omega}); \overline{\Omega}, \alpha, i) = u_A(p^P(\underline{\Omega}); \overline{\Omega}, \alpha, i)$, then for all conditions on primitives, such that there exists a FDE s.t. $x^* = p^P(\underline{\Omega})$, that equilibrium is belief-unstable;*

   (c) *If $u_A(p^P(\overline{\Omega}); \overline{\Omega}, \alpha, i) < u_A(p^P(\underline{\Omega}); \overline{\Omega}, \alpha, i)$, there does not exist a FDE such that $x^* = p^P(\underline{\Omega})$.*[7]

2. *(Seidmann and Winter, 1997) There exists an FDE if the Agency's utility, $u_A(\cdot)$, satisfies single-crossing.*

Proof. See Appendix. □

[7] Results symmetric to (1).a-(1).c hold for the FDE at $x^* = p^P(\overline{\Omega})$.

Part (1) of Proposition 6 addresses belief-stability of FDE, establishing that for an FDE sustained by a boundary policy to be stable, the Agency must *strictly* prefer disclosure at the opposite boundary. If the Agency is indifferent, the equilibrium is not robust to belief perturbations. At this point of indifference, an arbitrarily small perturbation to the Policymaker's belief causes the Agency's non-disclosure set to include states from neighborhoods of both $\underline{\Omega}$ and $\overline{\Omega}$. The Policymaker's best response, $\hat{x}(x)$, then, jumps discontinuously to a policy strictly greater than $p^P(\underline{\Omega})$, violating the condition for belief-stability.

It also provides a necessary condition for an FDE to be sustained. If the Agency has a strict incentive to conceal a boundary state, FDE is not incentive compatible, rendering belief-stability analysis obsolete. Finally, part (2), established by Seidmann and Winter (1997), provides a sufficient condition for the existence of an FDE.

Having established the belief-stability conditions for one potential anchor, we now turn to the other extreme. The following proposition characterizes conditions for the existence and belief-stability of the NDE.

Proposition 7.

1. *The NDE exists if and only if Agency's preference state-dependence α satisfies $\alpha \leq \alpha^{**}$ for a unique threshold $\alpha^{**} \in (0, \alpha^*]$ and the Agency's bias is $i = p_0^P$.*

2. *The NDE, if it exists, is belief-stable.*

Proof. See Appendix. □

Proposition 7 identifies the NDE as a belief-stable outcome for the case of no ex-ante preference conflict. We now characterize the conditions under which partial disclosure equilibria, the central focus of our analysis, can be sustained. The following proposition provides the necessary and sufficinet conditions for the existence of such equilibria, linking them to the degree of the Agency's preference state-dependence and the magnitude of its intrinsic bias.

Proposition 8.

1. *A PDE exists if and only if the Agency's state-dependence satisfies $\alpha \leq \alpha^*$ for a unique threshold $\alpha^* \in (0, 1)$. For any such α, the additive biases i that support this equilibrium form a non-empty, bounded interval $I^*(\alpha) \subset \Omega$.*

2. *The interval $I^*(\alpha)$ always contains the Policymaker's ex-ante optimum, $p_0^P \in I^*(\alpha)$.*

Proof. See Appendix. □

Proposition 8 provides the necessary and sufficient conditions for the existence of partial disclosure equilibrium. It confirms that partial disclosure is a robust possibility sustainable when the Agency's preferences are not excessively state-dependent (low α) and its intrinsic bias is moderate ($i \in I^*(\alpha)$).

Note that Seidmann and Winter's condition – single-crossing – is a restriction on preferences that is not required for our analysis of partial disclosure. With this in mind, observe that the parameter spaces supporting PDE and FDE are not mutually exclusive: if the

Agency's utility, $u_A(\cdot)$, satisfies single-crossing, then for any parameter profile (α, i) that supports a partial disclosure equilibrium ($\alpha \leq \alpha^*$ and $i \in I^*(\alpha)$), the game admits multiple equilibria.

The following propositions show the belief-stability criterion is equivalent to a specific comparative static property of the equilibrium non-disclosure policy and disclosure intervals with respect to the Agency's bias and state-dependence.

Proposition 9. *The equilibrium is belief-stable if and only if the equilibrium policy following non-disclosure, x^*, is*

1. *decreasing in the Agency's bias, i; and*
2. *decreasing in the Agency's preference state-dependence, α, when $x^* > i$, and increasing in α otherwise.*

Proof. See Appendix. □

Proposition 10. *The equilibrium is belief-stable if and only if the disclosure interval, $M(x^*, \alpha, i)$, is*

1. *expanding in the Agency's bias, i, when $x^* \leq i$, and contracting in i otherwise; and*
2. *expanding in the Agency's preference state-dependence, α.*

Proof. See Appendix. □

In canonical disclosure models (e.g., Milgrom 1981; Milgrom and Roberts 1986), a sophisticated receiver anticipates the sender's incentive to conceal unfavorable information. This anticipation leads the receiver to adopt a skeptical posture, making a pessimistic inference from any non-disclosure, which in turn compels the sender to be more forthcoming. In our partial disclosure equilibria, the underlying sophisticated behavior – in particuar, the equilibrium policy moving in opposition to the Agency's additive bias ($dx^*/di \leq 0$) – is analogous. Anticipating the pull of the incentives of a more biased Agency, the Policymaker adjusts the policy absent disclosure away from the Agency's ideal point. Belief-stability thus serves to select equilibria that may be though to embody the Policymaker's rational distrust of the Agency, given those incentives.

While Proposition 9 fully characterizes how policy x^* behaves in belief-stable equilibria, the disclosure interval, $M(x^*, \alpha, i)$, is a function of both x^* and i. Consequently, a change in the Agency's bias exerts both a direct influence on the boundaries of the disclosure interval and an indirect influence mediated through the adjustment of the equilibrium policy x^*.

Proposition 10 characterizes the net effect of parameters on the information disclosed. The comparative statics with respect to the Agency's preference state-dependence, α, is straightforward. A higher α implies that the Agency's preferences are more closely aligned with the state-contingent goals of the Policymaker, reducing the conflict of interest. In any belief-stable equilibrium, this greater alignment unambiguously leads to a larger disclosure interval and, thus, to more information transmission.

The effect of the Agency's additive bias, i, is more complex. An increase in i can represent either a convergence of ex-ante preferences (when $i < p_0^P$) or a divergence (when $i \geq p_0^P$).

In standard applications with sufficient symmetry (such as in the case with uniform prior and quadratic objective functions we studied above), the Policymaker's ex-ante optimum, p_0^P, and the equilibrium non-disclosure policy, x^*, both lie on the same side of the Agency's bias.[8] According to Proposition 10, this implies that greater ex-ante preference misalignment will result in *more* information disclosure.

This finding stands in sharp contrast to the canonical predictions of cheap-talk models, where preference divergence typically undermines communication. The mechanism here is driven by the Policymaker's skepticism. As established in Proposition 9, x^* always shifts to oppose the Agency's bias. When x^* and p_0^P are on the same side of i, preference divergence makes the non-disclosure outcome more punitive, leading to greater disclosure.

This counterintuitive result would be reversed, however, if the equilibrium geometry is such that the non-disclosure policy, x^*, and the ex-ante optimum, p_0^P, lie on opposite sides of the Agency's bias i ($(x^* - i) \cdot (p_0^P - i) < 0$). This would require a strong skew in the prior distribution or utility functions to pull the Policymaker's conditional expectation, x^*, so far from the unconditional expectation, p_0^P, that it crosses the Agency's bias. When this occurs, an increase in preference divergence moves the skeptical policy x^* closer to the Agency's ideal point, making non-disclosure less punitive and so leading to less disclosure.

The existence of these two distinct equilibrium geometries, which may generate opposing comparative statics, makes it important to identify the underlying conditions that govern which type of outcome is possible. The following proposition provides such a condition, linking the geometric properties of equilibria for any bias i to the equilibrium structure at the specific point of no ex-ante bias, $i = p_0^P$.

Proposition 11.

1. *All partial disclosure equilibria satisfy $(x^* - i) \cdot (p_0^P - i) > 0$ for any $i \neq p_0^P$ if and only if at $i = p_0^P$, the unique belief-stable equilibrium is the Non-Disclosure Equilibrium.*

2. *There exist no belief-stable partial disclosure equilibria satisfying $(x^* - i) \cdot (p_0^P - i) < 0$ for any i if at $i = p_0^P$, the unique belief-stable equilibrium is the Non-Disclosure Equilibrium.*

Proof. See Appendix. □

Proposition 11 may be seen as providing a regularity condition that can serve as a diagnostic tool for characterizing the model's outcomes. The condition that the NDE is the unique belief-stable equilibrium at $i = p_0^P$ is met in standard symmetric settings, including the uniform-quadratic example analyzed earlier in the paper.

The proposition's second statement establishes that when this regularity condition holds, the geometric configuration where $(x^* - i) \cdot (p_0^P - i) < 0$ is ruled out for all belief-stable equilibria. This provides a clear scope for the model's comparative static predictions. Specifically, under this condition, any belief-stable equilibrium will feature the property that greater ex-ante preference divergence leads to more information disclosure.

[8] E.g., if the Agency's ex-ante preference is to the left of the Policymaker's, the non-disclosure set is typically concentrated to the left of p_0^P, and the resulting equilibrium non-disclosure policy lies to the right of the Agency's bias ($x^* > i$).

Conversely, if the condition from Proposition 11 fails, we cannot guarantee against equilibria with the $(x^* - i) \cdot (p_0^P - i) < 0$ geometry. It remains possible, however, that all belief-stable equilibria still satisfy $(x^* - i) \cdot (p_0^P - i) > 0$, in which case the comparative static of increased disclosure with greater preference divergence holds everywhere.

6 Robustness

In this section, we test the robustness of our core findings by relaxing two key assumptions: perfect message verifiability and the restriction to a binary disclosure choice. To maintain analytical tractability and isolate the impact of each modification, we conduct this analysis within the framework of the uniform prior and state-independent preferences model from Section 3. Finally, we assume, for tractability, that $i \geq 0$, the case of $i \leq 0$ is symmetric.

6.1 Partial Verifiability

Throughout the preceding analysis, all messages observed by the Policymaker are assumed to be perfectly verifiable. That is, whenever the Agency discloses a state, the Policymaker can infer with certainty that the message accurately reflects the Agency's observation. Formally, the Agency can send a message $m = \omega$ if and only if it observes state ω.

In this section, we relax this assumption by introducing partial verifiability, allowing the Agency to distort information. Specifically, the Agency may send a point message $m = \tilde{\omega} \in \Omega$ that differs from the true state ω. The Policymaker, in turn, has the ability to verify whether the received message accurately reflects the underlying state.

Verification occurs probabilistically: with probability $q \in [0, 1]$, the Policymaker receives a signal $s(m)$ that indicates whether the reported message is truthful. If the message is truthful ($m = \omega$), the Policymaker observes $s(m) =$ True; if the message is distorted ($m \neq \omega$), the Policymaker observes $s(m) =$ False. With probability $1 - q$, the verification mechanism is inconclusive, and the Policymaker receives no additional information, observing $s(m) = \varnothing$. When $q = 1$, all messages are perfectly verifiable, meaning that any distortion by the Agency is immediately detected by the Policymaker. Conversely, when $q = 0$, messages are never verifiable, and the Policymaker receives no information beyond the reported message itself.

Observe that in this game with partial verifiability, a full disclosure equilibrium cannot exist. Consider an Agency with $i \geq 0$ that observes the state realization $\omega = \underline{\Omega}$. The Agency will refrain from distorting its message only if the expected policy implemented following a distortion is at least as far from its ideal point as the policy implemented when the true state is disclosed. This condition implies that the Policymaker must implement at least $p = \underline{\Omega}$ whenever verification is inconclusive.

However, since the probability of verification is independent of whether the message corresponds to the true state or is strategically distorted: in FDE all types including $\omega \neq \underline{\Omega}$ choose to disclose their state by sending $m = \omega$. The Policymaker, thus, must incorporate this uncertainty into their policy decision. Following an inconclusive verification outcome, $s(m) = \varnothing$, the Policymaker's sequentially optimal policy must be closer to i than to $\underline{\Omega}$. This, in turn, implies that type $\omega = \underline{\Omega}$ strictly prefers to distort its message rather than disclose truthfully, contradicting the existence of a full disclosure equilibrium.

The following proposition characterizes an equilibrium of this game:

Proposition 12. *There is a belief-stable equilibrium such that*

$$p^* = \begin{cases} m, & \text{if } s(m) = T, \\ \frac{i(i-y^*)}{i-y^*-1}, & \text{if } s(m) = F, \\ z^*, & \text{if } s(m) = \varnothing \end{cases} ; \; m^*(\omega) = \begin{cases} \omega, \text{if } \omega \in [y^*, 2 \cdot i - y^*] \cap [-1, 1], \\ \tilde{\omega}, else, \end{cases}$$

where

$$y^* = \frac{i(1+q) - 1 + \sqrt{(1 - i(1+q))^2 - 4i^2 q}}{2}, \quad z^* = 0, \quad \tilde{\omega} \sim U\big[[y^*, 2 \cdot i - y^*] \cap [-1, 1]\big],$$

Proof. See Appendix. □

In this equilibrium, the Agency fully discloses its information when the realized state falls within the interval $[i - \sqrt{(y-i)^2}, i + \sqrt{(y-i)^2}] \cap [-1, 1]$. When the realized state lies outside this region, the Agency distorts its information, instead sending a message drawn from a uniform distribution over this interval, effectively mimicking the prior distribution.

Figure 6 illustrates the disclosure boundaries as a function of the Agency's ideal point i for different values of the verifiability parameter q.

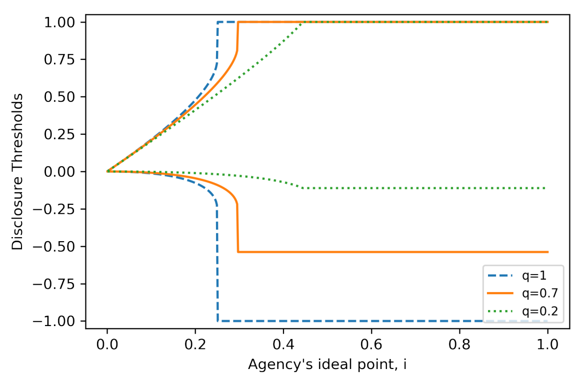

Figure 6: Disclosure intervals as a function of the Agency's ideal point i for different levels of partial verifiability q.

The following proposition characterizes the comparative statics of the disclosure intervals with respect to the Agency's ideal point i and the verifiability parameter q.

Proposition 13. *The disclosure interval within which the Agency fully reveals the state to the Policymaker,*

$$[y^*, 2 \cdot i - y^*],$$

such that $y^* = \frac{i \cdot (1+q) - 1 + \sqrt{(1 - i(1+q))^2 - 4 \cdot i^2 \cdot q}}{2}$ *exhibits the following properties:*

1. *it is increasing in i;*

2. *it is increasing in q.*

Proof. See Appendix. □

As the Agency's ideal point moves further from the prior mean, the range of disclosed states expands. Higher verifiability reduces the incentive to engage in misreporting, leading to a contraction in the set of states that remain undisclosed.

6.2 Agency's Vagueness

We demonstrate that an equilibrium outcome analogous to the belief-stable partial disclosure equilibrium (the "guarded" equilibrium from Proposition 1 can be sustained as a Sequential Equilibrium (SE) if we allow the Agency to strategically choose the precision of its verifiable message. We continue to work with the assumptions of the special case: $\Omega = [-1, 1]$, $f(\cdot)$ is the uniform density on Ω, and utility functions are $u_P(p) = -(\omega - p)^2$ and $u_A(p) = -(i - p)^2$.

Suppose that for any true state $\omega \in \Omega$, the Agency can select any message $T \subseteq \Omega$ such that $\omega \in T$. The message T is verifiable, meaning that the Policymaker learns with certainty that the true state ω is an element of T. The Agency's message space is thus a set $\{T \subseteq \Omega \mid \omega \in T\}$. A message strategy for the Agency is a function $m : \Omega \to \mathcal{P}(\Omega)$, where $\mathcal{P}(\Omega)$ is the power set of Ω, such that for every $\omega \in \Omega$, $\omega \in m(\omega)$.

If the Agency sends a singleton message $m(\omega) = \{\omega\}$, we say the state ω is fully disclosed. If $m(\omega) = T$ where $T = \{\omega\}$, the message is vague. Upon receiving a message T, the Policymaker updates her beliefs via Bayes' rule when applicable, conditional on her conjecture of the Agency's strategy $m(\cdot)$ and the prior $f(\cdot)$. If $T = \{\omega\}$, the Policymaker's posterior belief assigns probability one to state ω.

Let $\overline{\omega}(T) := \arg\max_{\omega \in T} |i - \omega|$.[9] Substantively, state $\overline{\omega}(T)$ is the one furthest away from the Agency's most preferred policy i. Next, denote the family of all off-path messages available to the Agency observing realization ω as $\mathcal{T}_{off}(\omega) := \{T : \omega \in T, T \neq \{\omega\}, T \neq N_G\}$.

The following proposition characterizes a sequential equilibrium supported by this system of beliefs:

Proposition 14. *The following strategy profile and system of beliefs can be supported in a Sequential Equilibrium:*

1. *Agency's Strategy $m^*(\omega)$:*

$$m^*(\omega) = \begin{cases} \{\omega\} & \text{if } \omega \in M_G, \\ N_G & \text{if } \omega \in N_G, \end{cases}$$

   where $M_G = \{\omega' \in \Omega \mid -(i - \omega')^2 \geq -(i - x_G)^2\}$, x_G is the policy equal to one implemented in the guarded partial disclosure equilibrium (Proposition 1), and $N_G = \Omega \setminus M_G$.

2. *Policymaker's Strategy $p^*(T)$:*

$$p^*(T) = \mathbb{E}[\omega \mid T; m^*(\cdot), \beta(\cdot|T)],$$

   where $\beta(\cdot|T)$ are Policymaker's beliefs given by

[9] If multiple such states exist, ties can be broken arbitrarily, e.g., by selecting the smallest such $\tilde{\omega}$.

3.

$$\beta(\omega|T, m(.)) = \begin{cases} 1, & T = \{\omega\}; \\ \mathbb{1}_{\omega \in N_G} \cdot \frac{f(\omega)}{\int_{N_G} f(\hat{\omega}) d\hat{\omega}}, & T = N_G; \\ \mathbb{1}_{\omega = \overline{\omega}(T)}, & T \in \mathcal{T}_{off}(\omega). \end{cases} \quad (4)$$

Proof. See Appendix. □

Note that system of Policymaker's beliefs described in the proposition is "skeptical", attributing any deviation to the type $\overline{\omega}(T)$ within the deviated message T that is least favorable to the Agency. The construction of such beliefs is crucial for deterring deviations to arbitrary vague messages not specified on the equilibrium path.

Proposition 14 demonstrates that the equilibrium outcome characterized in Proposition 1 – the unique belief-stable partial disclosure equilibrium – is robust to a significant enlargement of the Agency's message space. The equilibrium assessment (m^*, p^*, β) retains the same partition of types into disclosure and non-disclosure sets, inducing the same policy absent disclosure. Consequently, the sequential equilibrium of the model with vague communication retains the key features and comparative statics of the belief-stable equilibrium, emphasizing that the mere availability of arbitrarily precise messages does not inherently improve communication.

7 Discussion

7.1 Policymaker's Optimal Choice of Agency

Policymakers in many institutional settings often have discretion over the selection of agents or advisors from whom they receive policy advice. The biases of these agents significantly influence the quality and nature of the information transmitted.

The standard intuition, borne of the extensive cheap-talk signaling literature, is that information-revelation increases in the preference proximity between the sender and receiver. A related idea is the benchmark "ally principle" from the delegation literature, which posits that principals should delegate authority to agents with co-aligned preferences (Bendor and Meirowitz, 2004). A substantial body of literature has shown that this expectation might fail, and a principal might prefer an agent with preferences divergent from her own, when agents' information is endogenous. Che and Kartik (2009) argue that differences in preferences between advisors and decision-makers create incentives for advisors to acquire information, which can benefit decision-makers when the advisors' biases are moderate. Similarly, Krishna and Morgan (2001) demonstrate that the presence of another biased expert can enhance the decision-maker's ability to extract information from a biased advisor. Gailmard and Patty (2007) characterize scenarios where greater agent bias encourages the acquisition of expertise, as biased agents are more motivated to influence outcomes. Prendergast (2007) also shows that higher agent bias may motivate agents to exert more effort beyond what might be achieved via monetary incentives.

In contrast to the previous literature, we have shown that even when information is exogenous and readily available to the Agency, strategic policy implementation by the Policymaker is sometimes – depending on the nature of the equilibrium played by the Policymaker and the Agency – not enough to guarantee full disclosure. Our analysis suggests that, assuming that the Policymaker is not in a position to effect the selection of the full-disclosure equilibrium when partial disclosure equilibria are possible, she may be better off selecting an Agency with sufficiently divergent ex ante policy preference. Doing so can guarantee that the only equilibrium possible is full-disclosure, and helps increase disclosure in the belief-stable partial disclosure equilibria when they are present.

7.2 More Discussion TBA

Appendix

Lemma 1

Because information is verifiable, in case of disclosure, subsequent beliefs are independent of Agency strategy:

$$\Pr(\omega|m=\omega') = \begin{cases} 1 & \text{if } \omega = \omega' \\ 0 & \text{if } \omega \neq \omega'. \end{cases} \tag{5}$$

In contrast, $p(\omega|\varnothing)$ is determined by Bayes Rule and depends on the Agency's disclosure strategy $m(\omega)$:

$$p(\omega|m=\varnothing) = \frac{\mathbb{1}(m(\omega)=\varnothing)p(\omega)}{\int_{-1}^{1}\mathbb{1}(m(\omega'=\varnothing))p(\omega')d\omega'}, \tag{6}$$

where $\mathbb{1}$ is an indicator function. For the quadratic utility function, the Policymaker's expected utility is maximized at $p = \mathbb{E}[\omega|m]$, thus equation 1 follows.

Lemma 2

The Agency strictly prefers to disclose its information ω to the Policymaker when $p = \omega$ yields higher utility than $p = x$:

$$-(i-\omega)^2 > -(i-x)^2,$$

$$\implies \omega \in \begin{cases} (x, 2\cdot i - x) & \text{if } x \leq i \\ (2\cdot i - x, x) & \text{if } x \geq i. \end{cases}$$

Because the support of the distribution if ω is $[-1,1]$, the Agency discloses $\omega \in (x, 2\cdot i - x) \cap [-1,1]$ for $i > x$ and does not disclose $\omega \in [-1,1] \setminus [x, 2\cdot i - x]$. The Agency is indifferent for $\omega = x$ and $\omega = 2\cdot i - x$. The argument for $i \leq x$ case is symmetric.

Finally, note that because $x = E[\omega|m(\omega) = \varnothing]$, when $i \geq x$, $x < 0$.

Proposition 1

1. There exists a full-disclosure sequential equilibrium in this game. Suppose the Agency discloses all states it observes, and after non-disclosure ($m = \varnothing$), the Policymaker selects policy $x^* = -1$ if $i > 0$, and $x^* = +1$ if $i < 0$, and selects either $x^* = -1$ or $x^* = +1$ if $i = 0$. Then, given x^*, Agency prefers $p^* = \omega$ to $p^* = x^*$ $\forall\omega$.

To establish that this strategy profile is supported in a sequential equilibrium, it remains to show that beliefs supporting $p^*(\varnothing) = x^*$ as an optimal choice for the Policymaker are fully consistent with the strategy profile. A sequence of completely mixed disclosure strategies $\{\mu^k(\omega)\}_1^\infty$ can be constructed with

$$\mu^k(\omega) = \begin{cases} 1 - \varepsilon^k(\omega) & \text{if } \omega \in M(x^*, i) \\ \varepsilon^k(\omega) & \text{if } \omega \in [-1,1] \setminus M(x^*, i). \end{cases} \tag{7}$$

If $\{\varepsilon^k(\omega)\}_1^\infty \to 0$ for every ω, then $\{\mu^k(\omega)\}_1^\infty \to \mu^*(\omega)$ for every $\omega \in \Omega$. In particular, let $\varepsilon^k(\omega')$ converge faster than $\varepsilon^k(\omega'')$ for every ω', ω'' s.t. $\omega' > \omega''$; then $\{x^{*k}\}_{k=1}^\infty := \{\mathbb{E}[\omega|m = \varnothing; \mu^k]\}_{k=1}^\infty$ converges to -1. Likewise, if $\varepsilon^k(\omega')$ converges faster than $\varepsilon^k(\omega'')$ for every ω', ω'' s.t. $\omega' < \omega''$, it converges to +1. Thus, the full-disclosure equilibria are SE.

2. Suppose $i = 0$. If $x^* = 0$, the Agency is indifferent between disclosing and not disclosing $\omega = 0$, and strictly prefers not disclosing ($m = \varnothing$) for all $\omega \neq 0$. If the Agency does not disclose for all $\omega \neq 0$, then, regardless of $\mu(0)$, $x^* = \mathbb{E}[\omega|m = \varnothing, \mu^*] = 0$. It remains to show that the Policymaker's beliefs are fully consistent. Because beliefs following $m = \omega \ \forall\omega$ are determined by the verifiability of information, and because $m = \varnothing$ occurs with positive probability in equilibrium, there are no off-path-of-play information sets, and so beliefs are fully consistent with $\mu^*(\cdot)$.

3. Lemma 1 and Lemma 2 describe Policymaker's and Agency's equilibrium behavior. Absent disclosure, the Policymaker selects policy x^* such that

$$x^* = E[\omega|m = \varnothing; \mu^*].$$

For $i > 0$ and $[x^*, 2 \cdot i - x^*] \subset [-1, 1]$, x^* solves equation

$$\begin{aligned} x = E[\omega|m = \varnothing, \mu^*] &= \frac{\int_{-1}^{x} \omega \cdot f(\omega)d\omega + \int_{2\cdot i-x}^{+1} \omega \cdot f(\omega)d\omega}{F(\infty) - F(2\cdot i - x) + F(x) - F(-1)} \\ = -\frac{\int_{x}^{2\cdot i-x} \omega \cdot f(\omega)d\omega}{1 - F(2 \cdot i - x) + F(x)} &= -\frac{\int_{x}^{2\cdot i-x} \omega/2 \; d\omega}{1 - (2 \cdot i - x + 1)/2 + (x + 1)/2} \\ &= \frac{(i - x) \cdot i}{i - x - 1} \end{aligned}$$

Therefore, x^* solves

$$\begin{aligned} x &= \frac{(i - x) \cdot i}{i - x - 1} \\ &\Leftrightarrow \\ i \cdot x - x^2 - x &= i^2 - i \cdot x \\ &\Leftrightarrow \\ x^2 - x \cdot (-1 + 2 \cdot i) + i^2 &= 0 \end{aligned}$$

Thus, $x^* \in \{x_E, x_G\}$ where

$$x_E := \frac{2 \cdot i - 1 - \sqrt{1 - 4 \cdot i}}{2}$$

and

$$x_G := \frac{2 \cdot i - 1 + \sqrt{1 - 4 \cdot i}}{2}.$$

Hence, there can be two equilibria, one where the Policymaker chooses x_E absent disclosure and the other one where the Policymaker chooses x_G.

Note that, for $i > 0$, $x_E < x_G < 0$, and thus the disclosure intervals corresponding to these two equilibria are nested:

$$[x_G, \ 2 \cdot i - x_G] \subset [x_E, 2 \cdot i - x_E].$$

Because the condition $[x^*, 2\cdot i - x^*] \subset [-1, 1]$, which we imposed at the beginning, must be satisfied, it must be that $x^* \in [-1, 0]$, or, equivalently, $\frac{2\cdot i-1-\sqrt{1-4\cdot i}}{2} > -1$ and $\frac{2\cdot i-1+\sqrt{1-4\cdot i}}{2} < 0$. These inequalities are satisfied when $i \in [0, 1/4]$. When $i > 1/4$ neither guarded nor expansive equilibrium exists.

Because information is verifiable, and $m = \varnothing$ occurs on the path of play, beliefs are fully consistent and hence these equilibria are sequential equilibria.

4. Now suppose that $i < 0$ and $[2i - x^*, x^*] \subset [-1, 1]$. In equilibrium, x^* solves

$$x = E[\omega|m = \varnothing, \mu^*] = \frac{\int_{-1}^{2i-x^8x} \omega \cdot f(\omega)d\omega + \int_{x}^{+1} \omega \cdot f(\omega)d\omega}{F(1) - F(x^*) + F(2i - x^*) - F(-1)} = \frac{i(i - x^*)}{1 + i - x^*},$$

which yields $x^* \in \{x_E, x_G\}$, where

$$x_E := \frac{1}{2}(2i + 1 + \sqrt{1 + 4i})$$

and

$$x_G := \frac{1}{2}(2i + -\sqrt{1 + 4i}).$$

Note that $[2i - x_G, x_G] \subset [2i - x_E, x_E]$, i.e. more information is disclosed in the expansive equilibrium than in the guarded equilibrium, and $0 \leq x_G \leq x_E$, implying $i \in [-\frac{1}{4}, 0]$. Because $m = \varnothing$ occurs on the path of play, beliefs are fully consistent with the strategy profile.

Proposition 2

1. From Proposition 1, x_F is independent of i except at $i = 0$, where there is a jump discontinuity (downward).

2. x_G is defined in parts 3 and 4 of PropositIon 1. Differentiating each wrt i, we have

$$\frac{dx_G}{di} = 1 - \frac{1}{\sqrt{1 - 4 \cdot i}} \leq 0 \text{ for } i \in [0, 1/4), \tag{8}$$

and

$$\frac{dx_G}{di} = 1 - \frac{1}{\sqrt{1 + 4 \cdot i}} \leq 0 \text{ for } i \in [-1/4, 0),$$

hence x_G is decreasing in $i \in [-1/4, 1/4)$.

3. x_E is defined in parts 3 and 4 of Proposition 1. Differentiating each wrt i, we have

$$\frac{dx_E}{di} = 1 + \frac{1}{\sqrt{1 - 4 \cdot i}} \geq 0, \text{ for } i \in [0, 1/4), \tag{9}$$

and

$$\frac{dx_E}{di} = 1 + \frac{1}{\sqrt{1 + 4 \cdot i}} \geq 0, \text{ for } i \in [-1/4, 0).$$

Note that x_E for $i \leq 0$ evaluated at $i = 0$ is 1, whereas x_E for $i \geq 0$ evaluated at $i = 0$ is -1. Thus, x_E has a downward discontinuity at $i = 0$.

Proposition 3

1. From Proposition 1, the disclosure interval in the full disclosure equilibrium does not depend on i. Consider the case $i \geq 0$. From Lemma 2, the upper bounds of the disclosure intervals depend on i both directly and indirectly via x, and the lower bounds depend on i via x.

Differentiating the upper bound,

$$\frac{d(2 \cdot i - x^*)}{di} = \underbrace{2}_{direct\ effect} \underbrace{-\frac{dx^*}{di}}_{indirect\ effect}.$$

From (9), in the expansive equilibrium, the indirect effect of increasing i dominates the direct effect, lowering the upper bound of the disclosure interval.

$$\frac{d(2 \cdot i - x_E)}{di} = 2 - 1 - \frac{1}{\sqrt{1 - 4 \cdot i}} \leq 0.$$

From Proposition 2, the lower bound of the disclosure interval is increasing; thus the Agency discloses less in the expansive equilibrium as i increases.

3. In the guarded equilibrium, both the direct and indirect effects of increasing i are aligned, resulting in an expansion of the disclosure interval. From Lemma 2 and (8), the derivative of the upper bound of the disclosure interval in the guarded equilibrium is

$$\frac{d(2 \cdot i - x_E)}{di} = 2 - 1 + \frac{1}{\sqrt{1 - 4 \cdot i}} \geq 0.$$

This positive derivative indicates that as i increases, the upper threshold $2i - x_G$ increases. From Proposition 2, the lower threshold x_G decreases with i, and so the disclosure interval expands as i increases in the guarded equilibrium.

Proposition 4

Let us first consider the guarded SE. Policymaker chooses $x_G = (2 \cdot i - 1 + \sqrt{1 - 4 \cdot i})/2$ absent disclosure. Note that Agency's best response to Policymaker's selection of policy $x_0 \in [x_G - \varepsilon, x_G + \varepsilon]$ where $\varepsilon = \sqrt{1 - 4 \cdot i}$ will be to disclose states ω such that $\omega \in [x_0, 2 \cdot i - x_0]$. Importantly, Policymaker's best response to this disclosure strategy is to select policy x_1 such that

$$x_1 = -\frac{\int_{x_0}^{2 \cdot i - x_0} \omega \cdot f(\omega) d\omega}{1 - (2 \cdot i - x_0 + 1)/2 + (x_0 + 1)/2} = \frac{(i - x_0) \cdot i}{i - x_0 - 1}.$$

Note that $x_0 \leq x_1$ when

$$x_0 \leq \frac{(i - x_0) \cdot i}{i - x_0 - 1} \Leftrightarrow x_0 \in [\frac{2 \cdot i - 1 - \sqrt{1 - 4 \cdot i}}{2}, \frac{2 \cdot i - 1 + \sqrt{1 - 4 \cdot i}}{2}]. \tag{10}$$

and x_0 exceeds x_1 otherwise. Therefore, when $\varepsilon = \sqrt{1-4\cdot i}$ and $x_0 \in [x_G - \varepsilon, x_G]$ policy $x_0 \leq x_1$. Further, when $x_0 \in [x_G - \varepsilon, x_G]$ policy x_1 cannot exceeds x_G. Assume instead that x_1 exceeds x_G :

$$x_1 = \frac{(i - x_0)\cdot i}{i - x_0 - 1} > x_G = \frac{2\cdot i - 1 + \sqrt{1-4\cdot i}}{2}$$
$$x_0 > \frac{2\cdot i - 1 + \sqrt{1-4\cdot i}}{2},$$

which contradicts expression (10). Therefore, when $x_0 \in [x_G - \varepsilon, x_G]$, $|x_G - x_0| < |x_G - x_1|$. By similar logic, when $x_0 \in [x_G, x_G + \varepsilon]$, $|x_G - x_0| < |x_G - x_1|$. Therefore,

$$\forall x_0 : x_0 \in [x_G - \varepsilon, x_G + \varepsilon], \varepsilon = \sqrt{1-4\cdot i}, |x_G - x_0| < |x_G - x_1|, \tag{11}$$

and the guarded equilibrium is belief stable.

The expansive equilibrium, in contrast, is belief-unstable. Note that for any policy x_0 chosen absent disclosure from the interval $[x_E, x_E + \varepsilon]$, where $\varepsilon = \sqrt{1-4\cdot i}$, the Agency responses by disclosing states $\omega \in [x_0, 2\cdot i - x_0]$. The Policymaker's response to this is

$$x_1 = -\frac{\int_{x_0}^{2\cdot i - x_0} \omega \cdot f(\omega) d\omega}{1 - (2\cdot i - x_0 + 1)/2 + (x_0 + 1)/2} = \frac{(i - x_0)\cdot i}{i - x_0 - 1}.$$

By statement (11), $x_1 > x_0$. Now, let us consider $x_0 \in [-1, x_E]$. Because

$$x_1 = \frac{(i - x_0)\cdot i}{i - x_0 - 1}$$

and $x_0 \leq x_1$ if and only if $x_0 \in [\frac{2\cdot i - 1 - \sqrt{1-4\cdot i}}{2}, \frac{2\cdot i - 1 + \sqrt{1-4\cdot i}}{2}]$, when $x_0 < x_E$ policy $x_1 < x_0$. Therefore, guarded equilibrium is never belief-stable.

Finally, when equilibrium is fully revealing and $i > 0$, the Agency's best response to any policy $x_0 \in [-1, x_E]$ is $x_1 = \frac{(i - x_0)\cdot i}{i - x_0 - 1}$ s.t. $x_1 \leq x_0$. Therefore, fully revealing equilibrium will be belief stable. When $i = 0$, policy x_E converges to -1 and fully revealing equilibrium is belief-unstable.

Proposition 5

1. Since the Agency's utility function u_A is continuous in all its arguments, the boundaries of the set $N(x, \alpha, i)$ are continuous functions of the hypothetical policy $x \in (\underline{\Omega}, \overline{\Omega})$. Given ω is drawn from a continuous distribution F, the integral defining $N(\cdot)$ is a continuous function of its (continuously moving) boundaries. Therefore, $\hat{x}(x)$ is continuous for all $x \in (\underline{\Omega}, \overline{\Omega})$.

Let the set of all equilibrium non-disclosure policies, including any boundary FDEs (X^*) be strictly ordered. An equilibrium x_j^* is belief-stable if, in a neighborhood of x_j^*, the graph of $\hat{x}(x)$ crosses the 45-degree line from above. It is belief-unstable if it crosses from below. Given continuity of $\hat{x}(x)$ for all $x \in (\underline{\Omega}, \overline{\Omega})$, the elements of X^* must alternate in their belief-stability properties.

2. Let (x_j^*, M_j^*) and (x_k^*, M_k^*) constitute two distinct partial disclosure equilibria, where $M_j^* = M(x_j^*, \alpha, i)$ and $M_k^* = M(x_k^*, \alpha, i)$.

Assume that $u_A(x_j^*;\omega_0,\alpha,i) \leq u_A(x_k^*;\omega_0,\alpha,i)$ but $M_k^* \not\subseteq M_j^*$. Then there exists ω_0 s.t., $\omega_0 \in \{M_k^* \backslash M_j^*\}$. This implies: $u_A(p^P(\omega_0);\omega_0,\alpha,i) \geq u_A(x_k^*;\omega_0,\alpha,i)$ and $u_A(p^P(\omega_0);\omega_0,\alpha,i) < u_A(x_j^*;\omega_0,\alpha,i)$. Thus, $u_A(x_j^*;\omega_0,\alpha,i) > u_A(p^P(\omega_0);\omega_0,\alpha,i) \geq u_A(x_k^*;\omega_0,\alpha,i)$. This simplifies to $u_A(x_j^*;\omega_0,\alpha,i) > u_A(x_k^*;\omega_0,\alpha,i)$, presenting a contradiction.

The proof for the converse implication follows a symmetric argument and is omitted.

Proposition 6

1. We analyze the belief-stability of a FDE characterized by $x^* = p^P(\underline{\Omega})$; the analysis for $x^* = p^P(\overline{\Omega})$ is symmetric.

An FDE is belief-stable if small perturbations to the Policymaker's beliefs about the off-path policy do not lead to divergent best responses. Formally, it is belief-stable iff

$$\lim_{\varepsilon \to 0^+} \frac{\hat{x}(p^P(\underline{\Omega}) + \varepsilon) - \hat{x}(p^P(\underline{\Omega}))}{\varepsilon} \leq 1. \tag{12}$$

When $\hat{x}(x)$ is right-differentiable at $p^P(\underline{\Omega})$, this can condition simplifies to

$$\lim_{x \to p^P(\underline{\Omega})^+} \frac{d\hat{x}}{dx} \leq 1.$$

The differentiability of $\hat{x}(x)$ at $p^P(\underline{\Omega})$ is not guaranteed. If the non-disclosure set, $N(x,\alpha,i)$, is disconnected for x in a right-neighborhood of $p^P(\underline{\Omega})$, $\hat{x}(x)$ will be discontinuous, $\lim_{x \to p^P(\underline{\Omega})^+} \hat{x}(x) \neq p^P(\underline{\Omega})$. Such discontinuity implies not belief-stability. We, thus, focus on the case where $\hat{x}(x)$ is right-differentiable, which requires that $\exists \delta > 0$: the non-disclosure interval $\forall x \in [p^P(\underline{\Omega}), p^P(\underline{\Omega}) + \delta)$, is a single interval $N(x) = [\underline{\Omega}, \omega_b(x)]$, where $\omega_b(x)$ is such that $u_A(p^P(\omega_b(x)); \omega_b(x), \alpha, i) = u_A(x; \omega_b(x), \alpha, i)$.

Define

$$K(x,y;\alpha,i) := \int_\Omega \frac{\partial}{\partial p} u_P(p;\omega)\bigg|_{p=y} dF(\omega | \omega \in N(x,\alpha,i)). \tag{13}$$

The best response $\hat{x}(x)$ is implicitly defined by $K(x,\hat{x}(x);\alpha,i) = 0$. Applying the Implicit Function Theorem yields

$$\frac{d\hat{x}}{dx} = -\frac{\partial K(x,y;\alpha,i)/\partial x}{\partial K(x,y;\alpha,i)/\partial y}\bigg|_{y=\hat{x}(x)}. \tag{14}$$

We evaluate Equation 14 as $x \to p^P(\underline{\Omega})^+$. The strict concavity of u_P ensures that the denominator, $\partial K/\partial y$, is always strictly negative.

$$\lim_{x \to p^P(\underline{\Omega})^+} \frac{\partial K}{\partial y}\bigg|_{y=\hat{x}(x)} = \frac{\partial^2 u_P(p;\omega)}{\partial p^2}\bigg|_{p=p^P(\underline{\Omega}),\, \omega=\underline{\Omega}} < 0. \tag{15}$$

Next, we evaluate the numerator. Define

$$\begin{aligned} A(x,y,\alpha,i) &:= \int_{\omega \in N(x,\alpha,i)} \frac{\partial}{\partial p} u_P(p;\omega)|_{p=y} f(\omega) d\omega \\ B(x,\alpha,i) &:= \int_{\omega \in N(x,\alpha,i)} f(\omega) d\omega. \end{aligned} \tag{16}$$

The equilibrium condition $K = 0$ implies $A(x = x^*, y = \hat{x}(x)) = 0$. Therefore, the numerator of Equation 14 simplifies to

$$\frac{\partial K(x, y; \alpha, i)}{\partial x}\Big|_{y=\hat{x}(x)} = \frac{dA/dx \cdot B - dB/dx \cdot A}{B^2}\Bigg|_{y=\hat{x}(x)} = \frac{dA/dx}{B}\Bigg|_{y=\hat{x}(x)}. \tag{17}$$

As $x \to p^P(\underline{\Omega})$, both numerator and denumerator converge to zero. We apply L'Hôpital's Rule to determine $\lim_{x \to p^P(\underline{\Omega})^+} \frac{\partial K(x,y;\alpha,i)}{\partial x}\Big|_{y=\hat{x}(x)}$.

$$\lim_{x \to p^P(\underline{\Omega})^+} \frac{dB}{dx} = \lim_{x \to p^P(\underline{\Omega})^+} f(\omega_b(x)) \cdot \frac{d\omega_b(x)}{dx}. \tag{18}$$

We analyze components of $\frac{dA}{dx}$ separately, denoting

$$\begin{aligned} \lim_{x \to p^P(\underline{\Omega})^+} \frac{dA}{dx} &= \lim_{x \to p^P(\underline{\Omega})^+} \underbrace{\frac{\partial u_P}{\partial p}\Bigg| p = \hat{x}(x), \omega = \omega_b(x) \cdot f(\omega_b(x)) \cdot \frac{d\omega_b}{dx}}_{\text{Term 1}} \\ &+ \lim_{x \to p^P(\underline{\Omega})^+} \underbrace{\frac{d\hat{x}}{dx} \cdot \int_{\underline{\Omega}}^{\omega_b(x)} \frac{\partial^2 u_P}{\partial p^2}\Bigg| p = \hat{x}(x) f(\omega) d\omega}_{\text{Term 2}}. \end{aligned} \tag{19}$$

Given $\partial u_p/\partial p$ approaches 0 as x approaches $p^P(\underline{\Omega})$,

$$\begin{aligned} \lim_{x \to p^P(\underline{\Omega})^+} \frac{d}{dx}(\text{Term 1}) &= \lim_{x \to p^P(\underline{\Omega})^+} \frac{d}{dx}\left(\frac{\partial u_P}{\partial p}\Bigg|_{p=\hat{x}(x), \omega=\omega_b(x)} \cdot f(\omega_b(x)) \cdot \frac{d\omega_b}{dx}\right) \\ &= \left(\frac{\partial^2 u_P}{\partial p \partial \omega} \cdot \frac{d\omega_b}{dx} + \frac{\partial^2 u_P}{\partial p^2} \cdot \frac{d\hat{x}}{dx}\right)\Bigg|_{p^P(\underline{\Omega}), \underline{\Omega}} \cdot f(\underline{\Omega}) \cdot \frac{d\omega_b}{dx}\Bigg|_{p^P(\underline{\Omega})}. \end{aligned} \tag{20}$$

$$\begin{aligned} \lim_{x \to p^P(\underline{\Omega})^+} \frac{d}{dx}(\text{Term 2}) &= \lim_{x \to p^P(\underline{\Omega})^+} \left(\frac{d}{dx}\left(\int_{\underline{\Omega}}^{\omega_b(x)} \frac{\partial^2 u_P}{\partial p^2}\Bigg|_{p=\hat{x}(x)} \cdot f(\omega) d\omega\right)\right) \cdot \frac{d\hat{x}}{dx} \\ &+ \lim_{x \to p^P(\underline{\Omega})^+} \left(\int_{\underline{\Omega}}^{\omega_b(x)} \frac{\partial^2 u_P}{\partial p^2}\Bigg|_{p=\hat{x}(x)} \cdot f(\omega) d\omega\right) \cdot \frac{d^2\hat{x}}{dx^2} \\ &= \lim_{x \to p^P(\underline{\Omega})^+} \frac{\partial^2 u_P}{\partial p^2}\Bigg|_{\hat{x}(x), \omega_b} \cdot f(\omega_b) \cdot \frac{d\omega_b}{dx} \cdot \frac{d\hat{x}}{dx} + 0 \\ &= \frac{\partial^2 u_P}{\partial p^2}\Bigg|_{p^P(\underline{\Omega}), \underline{\Omega}} \cdot f(\underline{\Omega}) \cdot \frac{d\omega_b}{dx}\Bigg|_{p^P(\underline{\Omega})} \cdot \frac{d\hat{x}}{dx}\Bigg|_{p^P(\underline{\Omega})}. \end{aligned} \tag{21}$$

Therefore,

$$\begin{aligned} \lim_{x \to p^P(\underline{\Omega})^+} \frac{\partial K}{\partial x}\Bigg|_{y=\hat{x}(x)} &= \frac{\left(\frac{\partial^2 u_P}{\partial p \partial \omega} \cdot \frac{d\omega_b}{dx} + 2 \cdot \frac{\partial^2 u_P}{\partial p^2} \cdot \frac{d\hat{x}}{dx}\right)\Big|_{p^P(\underline{\Omega}, \underline{\Omega}} \cdot f(\underline{\Omega}) \cdot \frac{d\omega_b}{dx}\Big|_{p^P(\underline{\Omega})}}{f(\underline{\Omega}) \cdot \frac{d\omega_b}{dx}\Big|_{p^P(\underline{\Omega}}} \\ &= \frac{\partial^2 u_P}{\partial p \partial \omega}\Bigg|_{p^P(\underline{\Omega}), \underline{\Omega}} \cdot \frac{d\omega_b}{dx}\Bigg|_{p^P(\underline{\Omega})} + 2 \cdot \frac{\partial^2 u_P}{\partial p^2}\Bigg|_{p^P(\underline{\Omega}), \underline{\Omega}} \cdot \frac{d\hat{x}}{dx}\Bigg|_{p^P(\underline{\Omega})}. \end{aligned} \tag{22}$$

Substituting (22) into the implicit function formula gives an equation for the slope $\frac{d\hat{x}}{dx}\big|_{p^P(\underline{\Omega})}$:

$$\begin{aligned} \frac{d\hat{x}}{dx}\Big|_{p^P(\underline{\Omega})} &= -\frac{\left(\frac{\partial^2 u_P}{\partial p \partial \omega}\cdot\frac{d\omega_b}{dx} + 2\cdot\frac{\partial^2 u_P}{\partial p^2}\cdot\frac{d\hat{x}}{dx}\right)\big|_{p^P(\underline{\Omega}),\underline{\Omega}}}{\frac{\partial^2 u_P}{\partial p^2}\big|_{p^P(\underline{\Omega}),\underline{\Omega}}} \\ &= \left(-\frac{\frac{\partial^2 u_P}{\partial p\partial\omega}\cdot\frac{d\omega_b}{dx}}{\frac{\partial^2 u_P}{\partial p^2}} - 2\cdot\frac{d\hat{x}}{dx}\right)\Big|_{p^P(\underline{\Omega}),\underline{\Omega}}. \end{aligned} \tag{23}$$

Solving for the derivative yields the definitive expression for the slope of the best-response function at the boundary:

$$\frac{d\hat{x}}{dx}\Big|_{p^P(\underline{\Omega})} = -\frac{1}{3}\frac{\frac{\partial^2 u_P}{\partial p\partial\omega}}{\frac{\partial^2 u_P}{\partial p^2}}\cdot\frac{d\omega_b}{dx}\Big|_{p^P(\underline{\Omega}),\underline{\Omega}},$$

where $\omega_b(x)$ is defined by the Agency's indifference $u_A(p = x; \omega = \omega_b, \alpha, i) = u_A(p = p^P(\omega_b); \omega = \omega_b, \alpha, i)$. Then

$$\begin{aligned} \frac{d\omega_b(x)}{dx} &= -\frac{\partial(u_A(x;\omega_b,\alpha,i) - u_A(p^P(\omega_b);\omega_b,\alpha,i))/\partial x}{\partial(u_A(x;\omega_b,\alpha,i) - u_A(p^P(\omega_b);\omega_b,\alpha,i)/\partial\omega_b} \\ &= -\frac{\partial u_A/\partial p|_{p=x,\omega=\omega_b}}{\partial u_A/\partial\omega|_{p=x,\omega=\omega_b} - (\partial u_A/\partial p\cdot\frac{dp^P(\omega_b)}{d\omega_b} + \partial u_A/\partial\omega)|_{p=p^P(\omega_b),\omega=\omega_b}} \end{aligned} \tag{24}$$

As x approaches $p^P(\underline{\Omega})$, ω_b approaches $\underline{\Omega}$, thus

$$\frac{d\omega_b(x)}{dx}\Big|_{p^P(\underline{\Omega})} = -\frac{\partial u_A/\partial p|_{p=p^P(\omega_b),\omega=\omega_b}}{-\partial u_A/\partial p|_{p=p^P(\omega_b),\omega=\omega_b}\cdot\frac{dp^P(\underline{\Omega})}{d\omega}} = \frac{1}{\frac{dp^P(\underline{\Omega})}{d\omega}} \tag{25}$$

Given $p^P(\omega) : \frac{\partial u_P(p;\omega)}{\partial p} = 0$, we have $\frac{d}{d\omega}\frac{\partial u_P(p;\omega)}{\partial p} = \frac{\partial^2 u_P(p;\omega)}{\partial p^2}\cdot\frac{dp^P(\omega)}{d\omega} + \frac{\partial^2 u_P(p;\omega)}{\partial\omega\partial p}\cdot\frac{d\omega}{d\omega} = 0$. Therefore,

$$\frac{d\hat{x}}{dx}\Big|_{p^P(\underline{\Omega})} = 1/3 \leq 1 \tag{26}$$

and FDE is belief-stable at $x^* = p^P(\underline{\Omega})$ if $\hat{x}(\cdot)$ is differentiable.

Finally, assume, contrary to the proposition, that an FDE exists and that both of the following conditions hold

$$u_A(p^P(\overline{\Omega}); \underline{\Omega}, \alpha, i) > u_A(p^P(\underline{\Omega}); \underline{\Omega}, \alpha, i), \tag{27}$$

$$u_A(p^P(\underline{\Omega}); \overline{\Omega}, \alpha, i) > u_A(p^P(\overline{\Omega}); \overline{\Omega}, \alpha, i). \tag{28}$$

In an FDE, the Agency discloses the state ω for all $\omega \in \Omega$. Since the Agency's utility u_A is strictly concave in the policy, for any interior policy absent disclosure $x \in \big(p^P(\underline{\Omega}), p^P(\overline{\Omega})\big)$, the set of states ω for which the Agency's ideal point $p^A(p^P(\omega), \alpha, i)$ is closer to x than to $p^P(\omega)$ would be a non-empty set. Therefore, any policy x^* that sustains an FDE must be $x^* \in \{p^P(\underline{\Omega}), p^P(\overline{\Omega})\}$.

Suppose $x^* = p^P(\underline{\Omega})$. Note that the Agency observing the state $\overline{\Omega}$ prefers to conceal the state given Inequality 28. The case of $x^* = p^P(\overline{\Omega})$ is symmetric, thus no FDE can exist.

2. The proof of sufficiency for the single-crossing condition is provided by Seidmann and Winter (1997) and is omitted.

Proposition 7

1. For the NDE to be an equilibrium, the Agency must prefer the non-disclosure policy p_0^P to disclosing the true state ω for all $\omega \in \Omega$. The incentive compatibility condition is

$$u_A(p_0^P; \omega, \alpha, i) \geq u_A(p^P(\omega); \omega, \alpha, i) \quad \forall \omega \in \Omega.$$

Given that $p^P(\omega) = \omega$, this simplifies to

$$u_A(p_0^P; \omega, \alpha, i) \geq u_A(\omega; \omega, \alpha, i) \quad \forall \omega \in \Omega.$$

2. For an interior equilibrium, the belief-stability condition $\frac{d\hat{x}}{dx}|_{x=x^*} \leq 1$ is equivalent to the condition derived from the Implicit Function Theorem

$$\left.\frac{\partial K(x, y; \alpha, i = p_0^P)}{\partial x}\right|_{x=p_0^P,\, y=p_0^P} + \left.\frac{\partial K(x, y; \alpha, i = p_0^P)}{\partial y}\right|_{x=p_0^P,\, y=p_0^P} \leq 0.$$

We evaluate the two terms at the NDE, where $x = p_0^P, i = p_0^P$. First, consider the partial derivative with respect to y. Given the strict concavity of of $u_P(\cdot)$ in p, this term is strictly negative.

$$\left.\frac{\partial K}{\partial y}\right|_{x=p_0^P, y=p_0^P} = \mathbb{E}\left[\left.\frac{\partial^2 u_P(p;\omega)}{\partial p^2}\right|_{p=p_0^P} \middle| \omega \in N(p_0^P)\right] < 0.$$

The core of the proof is to show that $\frac{\partial A}{\partial x}$ is zero at this specific equilibrium. By the Leibniz Integral Rule

$$\left.\frac{\partial A}{\partial x}\right|_{x=p_0^P} = \pm \frac{\partial}{\partial p} u_P(p; \omega = p_0^P)|_{p=p_0^P} \cdot f(p_0^P) \cdot \frac{\partial}{\partial x} p_0^P. \tag{29}$$

By the model's definition, the Policymaker's ideal policy is $p^P(\omega) = \omega$. This implies that the Policymaker's marginal utility is zero whenever the policy matches the state. Therefore, $\frac{\partial A}{\partial x}|_{x=p_0^P} = 0$ and the Non-Disclosure Equilibrium, if it exists, is belief-stable.

Proposition 8

The proof proceeds in two main steps. First, we establish the existence and uniqueness of the threshold α^*. Second, for any $\alpha \leq \alpha^*$, we establish the existence of a bounded interval $I^*(\alpha)$ containing p_0^P.

A partial disclosure equilibrium exists if there is a non-disclosure policy $x \in \Omega$ and a non-empty, proper subset of states $M \subset \Omega$ such that: (i) $u_A(p^P(\omega); \omega, \alpha, i) \geq u_A(x; \omega, \alpha, i)$ $\forall \omega \in M$, $u_A(p^P(\omega); \omega, \alpha, i) \leq u_A(x; \omega, \alpha, i)$ $\forall \omega \notin M$ and (ii) $x = \arg\max_p \mathbb{E}[u_P(p; \omega) | \omega \notin M]$. The Agency's decision to disclose depends on the sign of the net gain from full disclosure of state ω, defined as $\Delta(x, \omega; \alpha, i) := u_A(p^P(\omega); \omega, \alpha, i) - u_A(x; \omega, \alpha, i)$. A partial disclosure equilibrium is possible only if there exists an x such that the set $M = \{\omega \in \Omega \mid \Delta(x, \omega; \alpha, i) \geq 0\}$ is nonempty and is not equal to Ω.

We first show that M expands monotonically in α. Note that because u_A is strictly concave, the Agency discloses a fixed state ω over a fixed induced policy x when $p^P(\omega)$ is close to p^A than x. As α increases, the Agency's ideal point $p^A(p^P(\omega), \alpha, i)$ moves closer to

$p^P(\omega)$. Thus, if a partial disclosure fails to exist for some (α_0, i), it must also fail to exist for all (α, i) such that $\alpha > \alpha_0$. Similarly, if a partial disclosure exists for some (α_0, i), it must also exist for all (α, i) such that $\alpha \leq \alpha_0$.

Now consider the limit cases. At $\alpha = 1$, the Agency's ideal point is $p^A = p^P(\omega)$. For any $x \neq p^P(\omega)$, $\Delta_A(x, \omega; 1, i) > 0$. The Agency strictly prefers to disclose every state rather than have any policy $x \neq p^P(\omega)$ implemented. The partial disclosure is not sustainable. Conversely, at $\alpha = 0$ there exists i for which the set of non-disclosure is non-empty. In particular, if $i = p_0^P$, a partial disclosure equilibrium can be sustained by the non-disclosure policy $x = p_0^P$. Since a partial disclosure equilibrium exists for $\alpha = 0$ (for some i) but not for $\alpha = 1$ (for all i), and M is monotonically expanding in α, there must exist a unique threshold $\alpha^* \in (0, 1)$ such that a partial disclosure equilibrium exists for some i if and only if $\alpha \leq \alpha^*$.

Fix any $\alpha \in [0, \alpha^*]$. By definition of α^*, there exist at least one i that supports a partial disclosure equilibrium. Let $I^*(\alpha)$ be the set of all such i. We must show that this non-empty (by definition) interval is bounded and $p_0^P \in I^*(\alpha)$. Given Agency's utility function is concave, for every α there exist unique ideal points $\underline{I}(\alpha)$ and $\overline{I}(\alpha)$ such that an Agency with ideal point $\underline{I}(\alpha)$ (respectively, $\overline{I}(\alpha)$) is indifferent between inducing the policy p_0^P (through non-disclosure) and inducing policy $p^P(\underline{\Omega})$ (respectively, $p^P(\overline{\Omega})$) through disclosure of the boundary state. That is, $u_A(p_0^P; \underline{\Omega}, \alpha, \underline{I}(\alpha)) = u_A(p^P(\underline{\Omega}); \underline{\Omega}, \alpha, \underline{I}(\alpha))$ and $u_A(p_0^P; \overline{\Omega}, 0, \overline{I}) = u_A(p^P(\overline{\Omega}); \overline{\Omega}, 0, \overline{I})$.

These critical ideal points define an interval $I^*(\alpha) \subseteq [\underline{I}(\alpha), \overline{I}(\alpha)] \subset \Omega$. For any Agency's ideal point $i \in I^*(\alpha)$, a partial disclosure equilibrium can be sustained. Finally, while policy p_0^P is not necessarily the optimal policy absent disclosure, if the Agency cannot conceal boundary states even when induced policy absent disclosure is p_0^P, it never conceals states in equilibrium.

Proposition 9

1. Consider an equilibrium characterized by a policy x^* absent disclosure. By the implicit function theorem and given equation 13,

$$\frac{\partial x^*}{\partial i} = -\frac{\partial K(x, y; \alpha, i)/\partial i|_{x=x^*, y=x^*}}{\partial K(x, y; \alpha, i)/\partial x|_{x=x^*, y=x^*}}. \tag{30}$$

We first determine the sign of the numerator, $\partial K(x, y; \alpha, i)/\partial i|_{x=x^*, y=x^*}$.

$$\begin{aligned}\partial K(x, y; \alpha, i)/\partial i|_{x=x^*, y=x^*} =& \frac{\partial}{\partial i} \int_\Omega \frac{\partial}{\partial p} u_P(p; \omega)\bigg|_{p=x^*} dF(\omega|\omega \in N(x^*, \alpha, i)) = \\ & \frac{\partial}{\partial i} \frac{\int_{\omega \in N(x^*, \alpha, i)} \frac{\partial}{\partial p} u_P(p; \omega)|_{p=x^*} dF(\omega)}{\int_{\omega \in N(x^*, \alpha, i)} dF(\omega)}\end{aligned} \tag{31}$$

Given Equations 16

$$\frac{\partial}{\partial i} K(x, y; \alpha, i)\bigg|_{x=x^*, y=x^*} = \frac{\partial}{\partial i} \frac{A}{B} = \frac{\partial A/\partial i \cdot B - \partial B/\partial i \cdot A}{B^2}. \tag{32}$$

Because $A(x = x^*, y = \hat{x}(x)) = 0$, the sign of $\partial K(x, y; \alpha, i)/\partial i$ at equilibrium is determined by the sign of $\partial A/\partial i$. By Leibniz Integral Rule, when $M(x^*, \alpha, i) = [x^*, \overline{M}(x^*, \alpha, i)]$

$$\partial A/\partial i = -\frac{\partial}{\partial p} u_P(p; \omega = \overline{M}(x^*, \alpha, i))|_{p=x^*} \cdot f(\overline{M}(x^*, \alpha, i)) \cdot \frac{\partial}{\partial i}\overline{M}(x^*, \alpha, i) < 0, \tag{33}$$

where $\frac{\partial}{\partial p} u_P(x^*, \overline{M}(x^*, \alpha, i)) > 0$ follows $u_P(.)$ concavity; and given concavity of $u_A(\cdot)$ holds, an increase in the Agency's ideal point i shifts its indifference points outwards, so $\partial \overline{M}/\partial i > 0$. Alternatively, if the disclosure interval is $M(x^*, \alpha, i) = [\underline{M}(x^*, \alpha, i), x^*]$,

$$\partial A/\partial i = \frac{\partial}{\partial p} u_P(p; \omega = \underline{M}(x^*, \alpha, i))|_{p=x^*} \cdot f(\underline{M}(x^*, \alpha, i)) \cdot \frac{\partial}{\partial i}\underline{M}(x^*, \alpha, i) < 0, \tag{34}$$

where $\frac{\partial}{\partial p} u_P(x^*, \underline{M}(x^*, \alpha, i)) < 0$ follows $u_P(.)$ concavity and $\frac{\partial}{\partial i}\underline{M}(x^*, \alpha, i) > 0$ given Agency's objective function satisfies concavity. Therefore

$$\partial K(x, y; \alpha, i)/\partial i|_{x=x^*, y=x^*} < 0 \tag{35}$$

and

$$\text{sign}\frac{\partial x^*}{\partial i} = \text{sign}\left.\frac{\partial K(x, y; \alpha, i)}{\partial x}\right|_{x=x^*, y=x^*}. \tag{36}$$

By the Implicit Function Theorem

$$\frac{d\hat{x}(x)}{dx} = -\frac{\left.\frac{\partial K(x,y;\alpha,i)}{\partial x}\right|_{y=\hat{x}(x)}}{\left.\frac{\partial K(x,y;\alpha,i)}{\partial y}\right|_{y=\hat{x}(x)}}. \tag{37}$$

Given $\hat{x}(x)$ is optimal given no disclosure, $\left.\frac{\partial K(x,y;\alpha,i)}{\partial y}\right|_{y=\hat{x}(x)} < 0$. Therefore,

$$\begin{cases} \frac{d\hat{x}(x)}{dx} \leq 1, & \frac{\partial K(x,y=\hat{x}(x);\alpha,i)}{\partial x} + \frac{\partial K(x,y=\hat{x}(x);\alpha,i)}{\partial y} \leq 0, \\ \frac{d\hat{x}(x)}{dx} > 1, & \frac{\partial K(x,y=\hat{x}(x);\alpha,i)}{\partial x} + \frac{\partial K(x,y=\hat{x}(x);\alpha,i)}{\partial y} > 0. \end{cases} \tag{38}$$

Finally, by the chain rule

$$\partial K(y = x^*, x = x^*; \alpha, i)/\partial x^* = \left.\frac{\partial K(x, y; \alpha, i)}{\partial y}\right|_{y=x^*, x=x^*} + \left.\frac{\partial K(x, y; \alpha, i)}{\partial x}\right|_{y=x^*, x=x^*}. \tag{39}$$

Combining equations 39, 38, and 36, belief-stable equilibria exhibit $\partial x^*/\partial i \leq 0$ (non-increasing x^* as i increases), while belief-unstable equilibria exhibit $\partial x^*/\partial i \geq 0$.

2. Consider function $K(\cdot)$ defined at 13. By the implicit function theorem

$$\frac{\partial x^*}{\partial \alpha} = -\frac{\partial K(x, y; \alpha, i)/\partial \alpha|_{x=x^*, y=x^*}}{\partial K(x, y; \alpha, i)/\partial x|_{x=x^*, y=x^*}}. \tag{40}$$

Following logic equivalent to that in part 1 of Proposition 9, the sign of $\frac{\partial x^*}{\partial \alpha}$ in belief-stable equilibria is determined by the sign of $\partial A/\partial\alpha$. If the disclosure interval is $M(x^*, \alpha, i) = [x^*, \overline{M}(x^*, \alpha, i)]$,

$$\partial A/\partial\alpha = -\frac{\partial}{\partial p}u_P(p; \omega = \overline{M}(x^*, \alpha, i)) \mid_{p=x^*} \cdot f(\overline{M}(x^*, \alpha, i)) \cdot \frac{\partial}{\partial\alpha}\overline{M}(x^*, \alpha, i) < 0, \tag{41}$$

where $\frac{\partial}{\partial\alpha}\overline{M}(x^*, \alpha, i) > 0$ follows monotonicity argument from Proposition 8. If $M(x^*, \alpha, i) = [\underline{M}(x^*, \alpha, i), x^*]$,

$$\partial A/\partial\alpha = \frac{\partial}{\partial p}u_P(p; \omega = \underline{M}(x^*, \alpha, i)) \mid_{p=x^*} \cdot f(\underline{M}(x^*, \alpha, i)) \cdot \frac{\partial}{\partial\alpha}\underline{M}(x^*, \alpha, i) > 0, \tag{42}$$

where $\frac{\partial}{\partial\alpha}\underline{M}(x^*, \alpha, i) < 0$ follows monotonicity argument from Proposition 8.

Proposition 10

1. Consider the case where the equilibrium non-disclosure policy satisfies $x^* \le i$; the argument for $x^* > i$ is symmetric. In this case, the disclosure interval is $M(i) = [x^*, \overline{M}(x^*, \alpha, i)]$, where the upper boundary $\overline{M}$ is defined by the Agency's indifference.

From Proposition 9, in any belief-stable partial disclosure equilibrium, the equilibrium non-disclosure policy x^* is a strictly decreasing function of the Agency's bias i. Therefore, $\frac{dx^*}{di} < 0$. The lower boundary strictly decreases.

The upper boundary $\overline{M}$ is a function of both the equilibrium policy x^* and the bias i. Its total derivative with respect to i is given by the chain rule:

$$\frac{d\overline{M}}{di} = \underbrace{\frac{\partial\overline{M}}{\partial i}}_{\text{Direct Effect}} + \underbrace{\frac{\partial\overline{M}}{\partial x^*} \cdot \frac{dx^*}{di}}_{\text{Indirect Effect}}.$$

Holding x^* constant, an increase in bias i shifts the Agency's ideal point further from x^*, making it more willing to disclose states far from x^*. Therefore, $\frac{\partial\overline{M}}{\partial i} > 0$. Holding i constant, a decrease in the non-disclosure policy x^* makes non-disclosure a worse outside option for the Agency, strengthening its incentive to disclose. Thus, $\frac{\partial\overline{M}}{\partial x^*} < 0$.

The direct and indirect effects are mutually reinforcing, causing the upper boundary to strictly increase: $\frac{d\overline{M}}{di} > 0$. Since the lower boundary x^* strictly decreases and the upper boundary $\overline{M}$ strictly increases with i, it follows that for any $i_2 > i_1$, we have $M(i_1) \subset M(i_2)$. The disclosure interval is strictly expanding in i.

2. Consider the disclosure interval $M = [x^*, \overline{M}(x^*(\alpha), \alpha, i)]$. From Proposition 9, the equilibrium policy x^* is a strictly decreasing function of state-dependence α. Thus, $\frac{dx^*}{d\alpha} < 0$.

The total derivative of the upper boundary with respect to α is:

$$\frac{d\overline{M}}{d\alpha} = \frac{\partial\overline{M}}{\partial\alpha} + \frac{\partial\overline{M}}{\partial x^*} \cdot \frac{dx^*}{d\alpha},$$

where $\frac{\partial\overline{M}}{\partial\alpha} > 0$ and $\frac{\partial\overline{M}}{\partial x^*} < 0$. The total derivative is positive.

Since the lower boundary strictly decreases and the upper boundary strictly increases with α, the disclosure interval M is strictly expanding in the Agency's preference state-dependence.

Proposition 11

The proof relies on a key property of the Policymaker's best-response function, $\hat{x}(x;i)$. For any fixed interior policy x, the function $\hat{x}(x;i)$ is strictly decreasing in the Agency's bias, i.

Assume, without loss of generality, that at $i = p_0^P$, there exists an equilibrium $x^* > p_0^P$. Now, consider a small increase in the Agency's bias to $i' = p_0^P + \epsilon$ for some small $\varepsilon > 0$. For ε small enough and given continuity of $\hat{x}(\cdot)$ there will be an equilibrium $x^* > i$.

Assume that at $i = p_0^P$, the unique fixed point of $\hat{x}(x)$ is $x = p_0^P$. Because the NDE is stable, this implies that for all $x > p_0^P$, $\hat{x}(x; i = p_0^P) < x$, and for all $x < p_0^P$, $\hat{x}(x; i = p_0^P) > x$. Consider any $i > p_0^P$. The best-response function for this new bias, let's call it $\hat{x}_i(x)$, must lie at or below the original function for $i = p_0^P$, i.e., $\hat{x}_i(x) \leq \hat{x}_{p_0^P}(x)$ for all x. This rules out the existence of any new equilibrium at a policy greater than p_0^P. Thus, any belief-stable equilibrium x^* for an agent with bias $i > p_0^P$ must satisfy $x^* \leq p_0^P$.

Proposition 12

When the Policymaker receives message m and signal $s(m) = T$, the Policymaker implements policy equal to the message observed. Next, signal $s(m) = \varnothing$ and signal $s(m) = F$ both should produce state-independent policies in equilibrium. This implies that Agency's distortion must replicate prior distribution on the disclosure interval. Denote policy the Policymaker implements following signal $s(m) = \varnothing$ as z and policy the Policymaker implements following signal $s(m) = F$ as x.

Next, consider an Agency with an ideal point i. Note that if an Agency with state's realization $\tilde{\omega}$ prefers to disclose its information to the Policymaker instead of distorting it, any Agency with state $\omega : |\omega - i| < |\tilde{\omega} - i|$ will disclose its state instead of concealing it. In this case, disclosing state produces policy $q \cdot \omega + (1 - q) \cdot z$ while distorting it leads to policy $q \cdot x + (1-q) \cdot z$. Thus, there exists a threshold y such that the Agency discloses states $\omega \in [y, 2 \cdot i - y] \cap [-1, 1]$ and distorts states otherwise.

It immediately follows from the previous paragraph that when the Policymaker observes message m and signal $s(m) = F$, she implements policy $x^* = E[\omega|\omega \notin [y, 2 \cdot i - y]] = \frac{i \cdot (i-y)}{-1+i-y}$ when $2 \cdot i - y < 1$ and $x^* = \frac{y-1}{2}$ otherwise.

If the Policymaker observes $s(m) = \varnothing$, she implements equilibrium policy

$$z^* = E[\omega|s(m) = \varnothing] = E[Pr[\omega \in [y, 2 \cdot i - y]] \cdot m + Pr[\omega \notin [y, 2 \cdot i - y]] \cdot x] = 0. \quad (43)$$

Because both Agency that discloses information and the Agency that does not have equal probability to generate not-informative message, regardless of y, observing $s(m) = \varnothing$ conveys no information beyond prior about state's realization.

In any equilibrium, the following holds

$$y = x^* \cdot q + z^* \cdot (1 - q) = \frac{i \cdot (i - y) \cdot q}{-1 + i - y}.$$

We only focus on belief stable-equilibria, thus,

$$y^* = \frac{i \cdot (1 + q) - 1 + \sqrt{(1 - i(1 + q))^2 - 4 \cdot i^2 \cdot q}}{2}.$$

Proposition 13

1.
$$\frac{\partial y^*}{\partial i} = \frac{(1+q)\cdot\sqrt{-4\cdot i^2\cdot q + (1-i\cdot(1-q))^2} - (1-i\cdot(1-q)^2+q)}{2\cdot\sqrt{-4\cdot i^2\cdot q+(1-i\cdot(1+q))^2}}$$

Consider the numerator we denote as

$$N_i := (1+q)\cdot\sqrt{-4\cdot i^2\cdot q+(1-i\cdot(1-q))^2} - (1-i\cdot(1-q)^2+q).$$

Note that numerator is always negative when $q\in[0,1]$, $i\in[0,1]$. We will show that

$$(1+q)\cdot\sqrt{-4\cdot i^2\cdot q+(1-i\cdot(1+q))^2} < (1-i\cdot(1+q))\cdot(1+q)+4\cdot i\cdot q$$

Squaring both sides, we need to show

$$(1+q)^2((1-i\cdot(1+q))^2-4\cdot i^2\cdot q) < ((1-i\cdot(1+q))\cdot(1+q)+4\cdot i\cdot q)^2$$

which simplifies to

$$i\cdot(1-q)^2 < 2+2\cdot q$$

Since $i\in[0,1]$ and $q\in[0,1]$, we have $i\cdot(1-q)^2 \le (1-q)^2 \le 1$ and $2+2\cdot q \ge 2$. Since $1<2$, the inequality $i\cdot(1-q)^2 < 2+2\cdot q$ holds. Therefore, the numerator $N<0$, and hence $\frac{\partial y^*}{\partial i}<0$.

Because $\frac{\partial y^*}{\partial i}<0$, the lower disclosure boundary decreases in i while the upper disclosure boundary increases in i.

2.
$$\frac{\partial y^*}{\partial q} = i\cdot\frac{\sqrt{-4\cdot i^2\cdot q+(1-i\cdot(1-q))^2}-(1+i\cdot(1-q))}{2\cdot\sqrt{-4\cdot i^2\cdot q+(1-i\cdot(1+q))^2}}$$

Consider the numerator denoted as

$$N_q := \sqrt{-4\cdot i^2\cdot q+(1-i\cdot(1-q))^2}-(1+i\cdot(1-q)).$$

We aim to show that $\sqrt{-4\cdot i^2\cdot q+(1-i\cdot(1-q))^2} \le (1+i\cdot(1-q))$. Since both sides are non-negative, it is sufficient to show that their squares satisfy the inequality:

$$-4\cdot i^2\cdot q+(1-i\cdot(1-q))^2 \le (1+i\cdot(1-q))^2$$

Rearranging terms yields:

$$(1-i\cdot(1-q))^2-(1+i\cdot(1-q))^2 \le 4\cdot i^2\cdot q$$

Factoring the difference of squares, we have:

$$[(1-i\cdot(1-q))-(1+i\cdot(1-q))]\cdot[(1-i\cdot(1-q))+(1+i\cdot(1-q))] \le 4\cdot i^2\cdot q$$

$$[-2\cdot i\cdot(1-q)]\cdot[2] \le 4\cdot i^2\cdot q$$

$$-4\cdot i\cdot(1-q) \le 4\cdot i^2\cdot q$$

For $i \geq 0$, dividing both sides by $4i$ (when $i > 0$) or observing directly (when $i = 0$), we require:

$$-(1-q) \leq i \cdot q$$

$$q - 1 \leq i \cdot q$$

$$q(1-i) \leq 1$$

Since $0 \leq q \leq 1$ and $0 \leq i \leq 1$, it follows that $0 \leq 1 - i \leq 1$, and thus $0 \leq q(1-i) \leq 1$. The inequality $q(1-i) \leq 1$ always holds. Therefore, $N_q \leq 0$.

Because of that, $y_q^{*\prime} \leq 0$ and disclosure interval expands as q increases.

Proposition 14

Given the beliefs $\beta(\cdot|T, m(.))$, the Policymaker's policy choice $p^*(T)$ is optimal by definition for on-path messages. For an off-path message T_{off}, the belief is concentrated on a single state $\overline{\omega}(T_{off}) = \arg\max_{\tilde{\omega} \in T_{off}} |i - \tilde{\omega}|$, so $p^*(T_{off}) = \overline{\omega}(T_{off})$ is optimal.

If $\omega \in M_G$: By sending $m^*(\omega) = \{\omega\}$, the Agency's utility is $-(i-\omega)^2$. If it deviates to send N_G, its utility is $-(i - x_G)^2$. By definition of x_G, $-(i-\omega)^2 \geq -(i-x_G)^2$. If it deviates to some T_{off} (where $\omega \in T_{off}$), the policy will be $\overline{\omega}(T_{off})$. The utility is $-(i - \overline{\omega}(T_{off}))^2$. Since $\overline{\omega}(T_{off})$ is the state in T_{off} furthest from i, and $\omega \in T_{off}$, it must be that $|i - \overline{\omega}(T_{off})| \geq |i - \omega|$. Thus, $-(i - \overline{\omega}(T_{off}))^2 \leq -(i-\omega)^2$. So, no profitable deviation to T_{off}.

If $\omega \in N_G$: By sending $m^*(\omega) = N_G$, the Agency's utility is $-(i - x_G)^2$. If it deviates to send $\{\omega\}$, its utility is $-(i-\omega)^2$. By definition of x_G, $-(i - x_G)^2 > -(i-\omega)^2$. Deviations to T_{off} are deterred as above, as $-(i - \overline{\omega}(T_{off}))^2 \leq -(i-\omega)^2 < -(i-x_G)^2$.

Finally, to show that proposed beliefs are consistent, we construct a sequence of strictly mixed strategy profiles $(m^n(.), p^n(.))$ that converges to $(m^*(.), p^*(.))$ and corresponding sequence of Bayesian beliefs β^n that converges to β. Denote the family of all off-path messages available for the Agency observing ω for which ω is the furthest from the Agency's ideal point i as $\overline{\mathcal{T}}_{off}(\omega) := \{T : \omega = \overline{\omega}(T), \omega \in T, T \neq \{\omega\}, T \neq N_G\}$.

We construct $(m^n(.), p^n(.))$ as follows. Let $P^n(T|\omega)$ denote the probability type ω sends a message T. Suppose the Agency observing realization ω sends its equilibrium message $m^*(\omega)$ with probability $P^n(m^*(\omega)|\omega) = 1 - 1/n - 1/n^2$. With total probability $1/n$, the Agency sends an off-path message T_{off} selected uniformly from the set $\overline{\mathcal{T}}_{off}(\omega)$. With the remaining total probability $1/n^2$, the Agency sends an off-path message T_{off} selected uniformly from the set $\mathcal{T}_{off}(\omega) \backslash \overline{\mathcal{T}}_{off}(\omega)$.

As $n \to \infty$, for an off-path T_{off}, Bayes' rule for $\beta^n(\tilde{\omega}|T_{off}, m(.))$ requires

$$\beta^n(\tilde{\omega}|T_{off}, m(.)) = \frac{P^n(T_{off}|\tilde{\omega}) f(\tilde{\omega}|\tilde{\omega} \in T_{off})}{\int_{\omega \in T_{off}} P^n(T_{off}|\omega) dF(\omega|\omega \in T_{off})}.$$

If $\tilde{\omega} = \overline{\omega}(T_{off})$, the numerator term $P^n(T_{off}|\tilde{\omega})$ is $O(1/n)$. For any other $\omega' \in T_{off}$, $P_n(T_{off}|\omega')$ is $O(1/n^2)$. Thus, in the limit, the probability mass concentrates on $\overline{\omega}(T_{off})$. This ensures that $\lim_{n\to\infty} \beta^n(\cdot|T_{off}, m(.)) = \beta(\cdot|T_{off}, m(.))$ as specified. The on-path beliefs are similarly consistent.

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.

Aybas, Yunus, C. and Steven Callander. 2023. "Efficient Cheap Talk in Complex Environments." *Working paper* .

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

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

Callander, Steven. 2008. "A theory of policy expertise." *Quarterly Journal of Political Science* 3(2):123–140.

Callander, Steven, Nicolas S Lambert and Niko Matouschek. 2021. "The power of referential advice." *Journal of Political Economy* 129(11):3073–3140.

Che, Yeon-Koo and Navin Kartik. 2009. "Opinions as incentives." *Journal of Political Economy* 117(5):815–860.

Crawford, Vincent P and Joel Sobel. 1982. "Strategic information transmission." *Econometrica* pp. 1431–1451.

Denisenko, Anna, Catherine Hafer and Dimitri Landa. 2024. "Competence and Advice.".

Dye, Ronald A. 1985. "Disclosure of nonproprietary information." *Journal of accounting research* pp. 123–145.

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

Gailmard, Sean and John W Patty. 2007. "Slackers and zealots: Civil service, policy discretion, and bureaucratic expertise." *American Journal of Political Science* 51(4):873–889.

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.

Grossman, Sanford J. 1981. "The informational role of warranties and private disclosure about product quality." *The Journal of law and Economics* 24(3):461–483.

Jung, Woon-Oh and Young K Kwon. 1988. "Disclosure when the market is unsure of information endowment of managers." *Journal of Accounting research* pp. 146–153.

Kaufman, Herbert. 1981. "Fear of bureaucracy: A raging pandemic." *Public Administration Review* pp. 1–9.

Krishna, Vijay and John Morgan. 2001. "A model of expertise." *The Quarterly Journal of Economics* 116(2):747–775.

McCarty, Nolan. 2004. "The appointments dilemma." *American Journal of Political Science* 48(3):413–428.

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.

Prendergast, Canice. 2007. "The motivation and bias of bureaucrats." *American Economic Review* 97(1):180–196.

Seidmann, Daniel J and Eyal Winter. 1997. "Strategic information transmission with verifiable messages." *Econometrica: Journal of the Econometric Society* pp. 163–169.

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.