# Persuasion via Weak Institutions*

Elliot Lipnowski† *Columbia University*

Doron Ravid *University of Chicago*

Denis Shishkin *UC San Diego*

July 20, 2021

**Abstract**

A sender commissions a study to persuade a receiver, but influences the report with some probability. We show that increasing this probability can benefit the receiver and can lead to a discontinuous drop in the sender's payoffs. We also examine a public-persuasion setting, where we observe the report's susceptibility to influence restricts the amount of information the sender can provide. To derive our results, we geometrically characterize the sender's highest equilibrium payoff, which is based on the concave envelope of her *capped* value function.

*Lipnowski and Ravid acknowledge support from the National Science Foundation (grant SES-1730168). We would like to thank Roland Bénabou, Ben Brooks, Eddie Dekel, Wouter Dessein, Jon Eguia, Emir Kamenica, Navin Kartik, Stephen Morris, Pietro Ortoleva, Wolfgang Pesendorfer, Carlo Prato, Marzena Rostek, Zichang Wang, Richard van Weelden, Leeat Yariv, and various audiences for useful suggestions. We would also like to thank Chris Baker, Sulagna Dasgupta, Takuma Habu, and Elena Istomina for excellent research assistance.

†E-mail: `e.lipnowski@columbia.edu`, `dravid@uchicago.edu`, `dshishkin@ucsd.edu`.

# 1 Introduction

Many institutions routinely collect and disseminate information. Although the collected information is instrumental to its consumers, the goal of dissemination is often to persuade. Persuading one's audience, however, requires the audience to believe what one says. In other words, the institution must be credible, meaning it must be capable of delivering both good and bad news. Delivering bad news might be difficult, however, because doing so requires the institution to withstand pressure from its superiors. The current paper studies how an institution's susceptibility to such pressures influences its persuasiveness and the quality of the information it provides.

For concreteness, consider a head of state who wants to sway a large multinational firm to invest as much as possible in her country's economy. The firm can either make a large investment, 2, a small investment, 1, or no investment, 0. Whereas the country's leader wants to maximize the firm's expected investment, the firm's net benefit from investing depends on the state of the local economy, which can be either *good* or *bad*. When the economy is *good*, the firm makes a profit of 1 from a large investment, and $\frac{3}{4}$ from a small investment. Investing in a *bad* economy results in losses, yielding the firm a payoff of $-1$ and $-\frac{1}{4}$ from a large and small investment, respectively. Not investing always generates a payoff of zero to the firm, regardless of the state. Therefore, the firm will make a large (no) investment whenever it assigns a probability of at least $3/4$ to the economy being *good* (*bad*). For intermediate beliefs, the firm makes a small investment. The firm and the policymaker share a prior belief of $\mathbb{P}(good) = 0.5$.

To persuade the firm to invest, the leader commissions a report by the country's central bank. By specifying the report's parameters—its data, methods, assumptions, focus, and so on—the leader controls the information it conveys to the firm. Formally, the commissioned report is a signal structure, $\xi(\cdot|good)$ and $\xi(\cdot|bad)$, specifying a distribution over messages that the firm observes conditional on the state if the report is conducted as announced. To execute the report as planned, however, the bank must withstand the leader's behind-the-scenes pressures; that is, the firm observes a message drawn from $\xi$ only if the bank is *credible*, which occurs with probability $\chi$. With complementary probability, the bank is *influenced*, meaning it reveals the state to the leader and releases a message of the leader's choice. Once the message is realized, the firm observes it and chooses how much to invest, without knowing whether the report is credible.

When the central bank is fully credible, $\chi = 1$, it is committed to the official report.

As such, the leader can communicate any information she chooses, and so this example falls within the framework of Kamenica and Gentzkow (2011). Using their results, one can deduce the policymaker optimally chooses a symmetric binary signal,

$$\xi_1^*(g|good) = 3/4 \qquad \xi_1^*(g|bad) = 1/4,$$
$$\xi_1^*(b|good) = 1/4 \qquad \xi_1^*(b|bad) = 3/4.$$

Under this signal structure, the firm is willing to invest 2 following a $g$ signal, and 1 following a $b$ signal. Ex ante, the two signals occur with equal probability, leading the firm to invest 3/2 on average.

If the central bank were weaker, its messages would be less persuasive, because the firm would no longer take them at face value. To illustrate, suppose $\chi = 2/3$ and that the leader commissioned the same report as under full credibility. In this case, the firm could not possibly make a large investment after seeing $g$: Otherwise, the leader would always send $g$ when influencing the report, which would make a small investment strictly better for the firm. Thus, when $\chi = 2/3$, the leader's full-commitment report is not sufficiently persuasive to increase the firm's involvement in the local economy beyond its no-information investment of 1.

The leader can, however, overcome the firm's skepticism by asking the bank to release more information. In fact, when $\chi = 2/3$, commissioning a fully revealing report that sends $g$ if and only if the economy is *good* is optimal for the leader. In the resulting equilibrium, the leader always sends $g$ when influencing the report, whereas the firm makes a large investment when seeing $g$ and invests nothing otherwise. The reason the firm finds it optimal to invest 2 upon seeing $g$ is that the bank's official report is so informative that a $g$ message results in the firm assigning a probability of 3/4 to *good* economy despite the leader's possible interference. Because the firm sees the $g$ message with probability 2/3, it invests 4/3 on average in the leader's economy.

Because a weaker central bank results in the leader commissioning a more informative report, the firm may benefit from a reduction in the bank's credibility. To illustrate, observe that when $\chi = 1$, the firm is no better off with the leader's report than it was without it: in either case, the firm expects a profit of $\frac{1}{4}$. By contrast, when $\chi = 2/3$, the firm strictly benefits from the leader's communications, making an expected profit of $\frac{1}{2}$ from investing 2 after seeing $g$, and not investing otherwise. On average, the firm's profit equals $\frac{1}{3}$. Thus, the leader responds to the central bank's weakness by commissioning a report whose informativeness more than compensates the firm for the central bank's increased susceptibility.

To understand examples such as the one above, we study a general model of strategic communication between a receiver (R, he) and a sender (S, she) who cares only about R's action. R's preferences over his actions depend on an unknown state, $\theta$. To learn about $\theta$, R relies on information provided by an institution under S's control. The game begins with S publicly announcing an official reporting protocol, which is an informative signal about the state. With probability $\chi$, S's institution is credible, delivering R a message drawn according to the originally announced protocol. With complementary probability, the report is influenced: S learns the state and chooses what message to send to R. Seeing the message (but not its origin), R takes an action. Thus, $\chi$ represents the credibility, or strength, of S's institution; that is, the institution's ability to resist interference by its superiors.

At the extremes, our framework specializes to two prominent models of information transmission. When $\chi = 1$, S never gets to influence the report, so our setting reduces to one in which S publicly commits to her communication protocol at the beginning of the game. In other words, under full credibility, our model is equivalent to Bayesian persuasion (Kamenica and Gentzkow, 2011). When $\chi = 0$, R knows S is choosing the report's message ex post. Because messages are costless, S's communication is cheap talk (Crawford and Sobel, 1982; Green and Stokey, 2007), meaning our no-credibility case corresponds to a cheap talk game with state-independent preferences (Chakraborty and Harbaugh, 2010; Lipnowski and Ravid, 2020).

The corner cases of our model lend themselves to geometric analysis, commonly used in the information-design literature (e.g., Kamenica and Gentzkow, 2011; Alonso and Câmara, 2016; Ely, 2017). Such an analysis characterizes S's utility in her favorite equilibrium by looking at her *value function*, which specifies the highest value S can obtain from R responding optimally to a given posterior belief. As Kamenica and Gentzkow (2011) showed, concavifying this function gives S's largest equilibrium payoff in the Bayesian persuasion model. More recently, Lipnowski and Ravid (2020) observed that, as long as S only cares about R's actions, *quasi*concavifying S's value function—that is, taking the function's quasiconcave envelope—delivers her highest equilibrium payoff under cheap talk.

Our Theorem 1 uses the above-mentioned geometric approach to characterize S's maximal equilibrium value in the intermediate credibility case, $\chi \in (0, 1)$. To do so, the theorem partitions S's equilibrium messages into two sets: messages S willingly sends when influencing the report (e.g., $g$ in the above example), and messages communicated only by the official report. We show that whereas one can characterize S's payoffs from influenced reporting via quasiconcavification as in Lipnowski and Ravid (2020), one cannot simply use concavification to obtain S's utility from messages sent exclusively by official reporting. The reason is that, in equilibrium, S's utility from these messages cannot surpass the payoff she obtains under compromised reporting, because if it did, S would have a profitable deviation. To account for this incentive constraint, one must *cap* S's value function before concavifying it.

Using Theorem 1, we explore how the use of weaker institutions affects persuasion. Proposition 1 identifies situations in which R does better with a less credible S. In particular, the proposition shows such productive mistrust can occur when S wants to reveal intermediate information under full credibility. In such circumstances, a less credible S may choose to commission a report that releases more news that is bad for her, so that R believes messages that are good for S. We see this case in the central-bank example above: when $\chi = 1$, the leader never fully reveals any state, whereas under $\chi = 2/3$, the leader must occasionally reveal that the economy is *bad* in order to ensure the firm invests 2 when seeing $g$.

Our next result, Proposition 2, shows that small decreases in credibility lead to large drops in the sender's value for all interesting finite instances of our model. More precisely, we show such a collapse occurs at some full-support prior and some credibility level if and only if S can benefit from persuasion. Such a collapse is present in the above example: Whenever $\chi < 2/3$, the leader cannot induce the firm to invest 2 even when she chooses to commission a fully revealing report. Thus, the best the leader can do when $\chi < 2/3$ is to get an investment of 1 for sure by communicating no information—a drop of 1/3 from the 4/3 average investment the leader obtains when $\chi$ is exactly 2/3.

One can also construct examples in which S's value collapses at full credibility. For example, suppose the firm can make a gigantic investment of 10 in the leader's economy but finds doing so optimal if and only if it is *certain* the economy is *good*. Under full credibility, the leader can obtain a payoff of 5 by revealing the state, and have the firm invest nothing when the economy is *bad* and 10 when the economy is *good*. By contrast, when $\chi < 1$, none of the leader's messages can get the firm to make a gigantic investment: if it could, the leader would always send such a message when influencing the report in a *bad* economy; thus, that message could not possibly convince the firm that the economy is *good* for sure. As such, when $\chi < 1$, the leader cannot hope to obtain an investment larger than 2. Thus, even a tiny imperfection in the central bank's credibility causes a significant drop in the leader's payoff.

One may suspect the non-robustness of the full-credibility solution in the above modified example is rather special. Our third proposition confirms this suspicion. In particular, it shows S's value can collapse at full credibility if and only if R does not give S the benefit of the doubt; that is, to obtain her best feasible payoff, S must persuade R that some state is impossible. This property is clearly satisfied by the above modified example: The firm is willing to make a gigantic investment only if it assigns a zero probability to the economy being *bad*. Moreover, this property is non-generic: for a fixed finite setting, the set of environments where R fails to give S the benefit of the doubt has measure zero. Thus, although S's value often collapses due to small decreases in credibility, such collapses rarely occur at full credibility.

Section 5 abandons our general analysis in favor of a specific instance of public persuasion. In this specification, S uses her weak institution to release a public report whose purpose is to sway a population of receivers to take a favorable binary action. For example, S may be a seller who markets her product by sending it to reviewers or a leader using state-owned media to vie for the support of her populace. Each receiver's utility from taking S's favorite action is additively separable in the unknown state and his idiosyncratic type, which follows a well-behaved single-peaked distribution. We show (Claim 1) it is S-optimal for the official report to take an upper-censorship form, characterized by a threshold below which states are fully separated. States above this threshold are pooled into a single message, which S always sends when influencing the report. The information revealed turns out to be identical to the experiment that S uses under full commitment with an upper bound on the informativeness of her signal. Thus, in this setting, partial credibility has the same impact as bounding the amount of information that S can release in equilibrium.

We also consider several extensions. For example, we show that letting S know whether her announcement is credible before choosing the official report does not alter the S-favorite equilibrium: every S-favorite equilibrium when S does not know her credibility type is outcome-equivalent to an S-favorite equilibrium when her credibility type is her private information, and vice-versa. We also show the results of the baseline model readily extend to the case in which credibility is correlated with the state. This extension allows us to assess the relative value of credibility in a different state in specific examples. We illustrate by showing that in our public persuasion example from section 5, S especially prefers her institution to be resistant to pressure in bad states. Specifically, concentrating the credibility of S's institution in low states uniformly increases S's payoffs across all type distributions.

We conclude by further enriching our model to allow S to design her institution at a cost. More precisely, we let S publicly choose the probability with which reporting is credible in each state. S's credibility choice is made in ignorance of the state and comes at a cost that is a continuous and increasing function of the institution's average credibility. We explain how to adjust our analysis to this setting, and observe that R may benefit from an increase in S's costs, echoing the productive-mistrust phenomenon of the fixed-credibility model. By contrast, an infinitesimal increase in S's costs never leads to a sizable decrease in S's value, suggesting collapses of trust are a byproduct of rigid institutional structures.

**Related Literature.** This paper contributes to the literature on strategic information transmission. To place our work, consider two extreme benchmarks: full credibility and no credibility. Our full-credibility case is the model used in the Bayesian persuasion literature (Aumann and Maschler, 1995; Kamenica and Gentzkow, 2011; Kamenica, 2019),[1] which studies sender-receiver games in which a sender commits to an information-transmission strategy. The no-credibility specialization of our model reduces to cheap talk (Crawford and Sobel, 1982; Green and Stokey, 2007). In particular, we build on Lipnowski and Ravid (2020), who use the belief-based approach to study cheap talk under state-independent sender preferences.

Two recent papers (Min, 2020; Fréchette, Lizzeri, and Perego, 2020) study closely related models. Fréchette, Lizzeri, and Perego (2020) test experimentally the connection between the informativeness of the sender's communication and her credibility in the binary-state, binary-action version of our model. Min (2020) looks at a generalization of our model in which the sender's preferences can be state dependent. He shows the sender weakly benefits from a higher commitment probability. Applying Blume, Board, and Kawamura's (2007) results on noisy communication, Min (2020) also shows allowing the sender to commit with positive, rather than zero, probability strictly helps both players in Crawford and Sobel's (1982) uniform-quadratic example.

Other thematically related work studies games of information transmission while varying the (exogenous or endogenous) limits to communication. Some such work focuses on games of direct communication, showing how some manner of commitment power can be sustained (for either a sender or a receiver) via lying costs (e.g., Kartik, 2009; Guo and Shmaya, 2021; Nguyen and Tan, 2021), repeated interactions (e.g., Best and Quigley, 2020; Mathevet, Pearce, and Stacchetti, 2019), verifiable information

[1] See also Aumann and Maschler (1966).

(e.g., Glazer and Rubinstein, 2006; Sher, 2011; Hart, Kremer, and Perry, 2017; Ben-Porath, Dekel, and Lipman, 2019), informational control (e.g., Ivanov, 2010; Luo and Rozenas, 2018), or mediation (e.g., Goltsman et al., 2009; Salamanca, 2021). Other work considers models in which a sender chooses an experiment ex ante, asking how persuasion can be shaped by exogenous experiment constraints (e.g., Ichihashi, 2019; Perez-Richet and Skreta, 2021) or by signaling motives (e.g., Perez-Richet, 2014; Hedlund, 2017; Alonso and Câmara, 2018).

More broadly, weak institutions often serve as a justification for examining mechanism design under limited commitment (e.g., Bester and Strausz, 2001; Skreta, 2006). We complement this literature by relaxing a principal's commitment power in the control of information rather than incentives.

# 2 A Weak Institution

We analyze a game with two players: a sender (S, she) and a receiver (R, he). Whereas both players' payoffs depend on R's action, $a \in A$, R's payoff also depends on an unknown state, $\theta \in \Theta$. Thus, S and R have objectives $u_S \colon A \to \mathbb{R}$ and $u_R \colon A \times \Theta \to \mathbb{R}$, respectively, and each aims to maximize expected payoffs.

The game begins with S commissioning a report, $\xi \colon \Theta \to \Delta M$, to be delivered by a research institution. The state then realizes, and R receives a message $m \in M$ (without observing $\theta$). Given any $\theta$, S is credible with probability $\chi$, meaning $m$ is drawn according to the official reporting protocol, $\xi(\cdot|\theta)$. With probability $1 - \chi$, S is not credible, in which case S decides which message to send after privately observing $\theta$. Only S learns her credibility type, and she learns it only after announcing the official reporting protocol.

We impose some technical restrictions on our model.[2] Both $A$ and $\Theta$ are compact metrizable spaces with at least two elements, and the objectives $u_R$ and $u_S$ are continuous. Occasionally, we discuss the **finite** model, which is the special case in which both $\Theta$ and $A$ are finite. The state, $\theta$, follows some prior distribution $\mu_0 \in \Delta\Theta$, which is known to both players. Finally, we assume $M$ is an uncountable compact metrizable space.[3]

[2]For a compact metrizable space, $Y$, we denote by $\Delta Y$ the set of all Borel probability measures over $Y$, endowed with the weak* topology.

[3]We impose the richness condition to isolate the restrictions on communication that arise from credibility concerns alone. Our characterization of sender-optimal equilibrium values (Theorem 1) and

We now define an equilibrium, which consists of four objects: S's official reporting protocol, $\xi\colon \Theta \to \Delta M$, executed whenever S is credible; the strategy that S employs when not committed, that is, S's influencing strategy, $\sigma\colon \Theta \to \Delta M$; R's strategy, $\alpha\colon M \to \Delta A$; and R's belief map, $\pi\colon M \to \Delta\Theta$, assigning a posterior belief to each message. A $\chi$-equilibrium is an official reporting policy announced by S, $\xi$, together with a perfect Bayesian equilibrium of the subgame following S's announcement. Formally, a **$\chi$-equilibrium** is a tuple $(\xi, \sigma, \alpha, \pi)$ of measurable maps such that it is consistent with Bayesian updating, and both R and S behave optimally; that is,

1. $\pi\colon M \to \Delta\Theta$ is derived from $\mu_0$ via Bayes' rule, given message policy

$$\chi\xi + (1-\chi)\,\sigma\colon \Theta \to \Delta M,$$

   whenever possible;

2. $\alpha(m)$ is supported on $\mathrm{argmax}_{a\in A} \int_\Theta u_R\,(a,\cdot)\ \mathrm{d}\pi(\cdot|m)$ for all $m \in M$;

3. $\sigma(\theta)$ is supported on $M^*_\alpha := \mathrm{argmax}_{m\in M} \int_A u_S(\cdot)\,\mathrm{d}\alpha(\cdot|m)$ for all $\theta \in \Theta$.

We view S as a principal capable of steering R toward her favorite $\chi$-equilibria. Because such equilibria automatically satisfy S's incentive constraints on the choice of $\xi$, we omit said constraints for the sake of brevity.

Discussing what happens in our game as one varies the prior and credibility of S's institution is useful. Thus, given a belief $\mu \in \Delta\Theta$ and $\chi \in [0,1]$, let $\mathbf{G}(\chi,\mu)$ denote the version of our game in which the prior is $\mu$ and S's credibility is $\chi$. An equilibrium of $\mathbf{G}(\chi,\mu)$ is a $\chi$-equilibrium when $\mu$ is the state's distribution.

We analyze our model via the belief-based approach, commonly used in the literature on strategic communication. Specifically, we use the ex-ante distribution over R's posterior beliefs, $p \in \Delta\Delta\Theta$, as a substitute for S's official reporting protocol, S's strategy, and the equilibrium belief system, $\pi$. Clearly, every $\xi$, $\sigma$, and $\pi$ generate some such distribution over R's posterior belief. By Bayes' rule, this posterior distribution averages to the prior, $\mu_0$. That is, $p \in \Delta\Delta\Theta$ satisfies $\int \mu\ \mathrm{d}p\,(\mu) = \mu_0$. We refer to any $p$ that averages back to the prior as an **information policy**. Thus, only information policies can originate from some $\xi$, $\sigma$, and $\pi$. The fundamental result underlying the belief-based approach is that every information policy can be generated by some $\sigma$ and

the propositions of section 4 hold if $|M| \geq |A|$, or if $\Theta$ is finite and $|M| \geq 2|\Theta| - 1$; see Proposition 4.

$\pi$.[4] Let $\mathcal{R}(\mu_0)$ denote the set of all information policies when the prior is $\mu_0$.

The belief-based approach allows us to focus on the game's outcomes. Formally, an **outcome** is a triplet, $(p, s_o, s_i) \in (\Delta\Delta\Theta) \times \mathbb{R} \times \mathbb{R}$, representing R's posterior distribution, $p$, S's payoff when credible, $s_o$, and S's payoff when influencing the report, $s_i$. An outcome is a $\chi$**-equilibrium outcome** if it corresponds to a $\chi$-equilibrium.[5] Observe that knowing a $\chi$-equilibrium's outcome is sufficient for recovering each player's expected payoff: given an outcome $(p, s_o, s_i)$, S earns a payoff of $\chi s_o + (1-\chi)s_i$, whereas R's expected utility is $\int \max_{a\in A} \int u_R(a,\cdot)\,\mathrm{d}\mu\,\mathrm{d}p(\mu)$.

# 3 Persuasion with Partial Credibility

In this section, we characterize S's maximal $\chi$-equilibrium payoff. We begin by reviewing existing results that cover the extreme versions of our model. We then proceed to use these results to prove our main theorem, which covers the case in which $\chi$ is intermediate.

## 3.1 The Extreme Cases

The existing results that characterize our model's edge cases are geometric and rely on S's value function. To describe this function, define S's **value correspondence**,[6]

$$\begin{aligned} V \colon \Delta\Theta &\rightrightarrows \mathbb{R} \\ \mu &\mapsto \mathrm{co}\; u_S\left(\operatorname*{argmax}_{a\in A} \int u_R(a,\cdot)\,\mathrm{d}\mu\right). \end{aligned}$$

In words, $V(\mu)$ is the set of payoffs that S can obtain when R behaves optimally given posterior belief $\mu$. S's **value function**,

$$v(\mu) \coloneqq \max V(\mu),$$

identifies S's highest continuation payoff from inducing this posterior.[7]

[4]For example, see Aumann and Maschler (1995), Benoît and Dubra (2011), or Kamenica and Gentzkow (2011).

[5]Definition 2 in the Appendix spells out the definition of a $\chi$-equilibrium outcome.

[6]The notation "co" refers to the convex hull.

[7]Note (appealing to Berge's theorem) $V$ is a Kakutani correspondence, that is, a nonempty-compact-convex-valued, upper hemicontinuous correspondence. Therefore, $v$ is a well-defined, upper semicontinuous function.

When $\chi = 1$, S's official announcement is binding, and so the game reduces to the Bayesian persuasion model of Kamenica and Gentzkow (2011). In this case, S is hampered by two constraints: Bayesian updating and R's incentives. As explained in section 2, R being Bayesian is tantamount to restricting R's belief distribution to the set of information policies, $p \in \mathcal{R}(\mu_0)$. R's incentives mean S's expected utility from inducing a belief $\mu$ must come from $V(\mu)$. Maximizing S's payoff belief by belief, and across all information policies gives her payoff under full credibility,

$$\max_{p \in \mathcal{R}(\mu_0)} \int v(\cdot)\, \mathrm{d}p.$$

It is well-known (e.g. Aumann and Maschler, 1966; Kamenica and Gentzkow, 2011) that the function mapping the prior to this optimal value admits a geometric characterization: It is the pointwise-lowest concave and upper semicontinuous function that is everywhere above $v$. This function, which we denote by $\hat{v}$, is known as $v$'s **concave envelope**.

When $\chi = 0$, R knows S is choosing $m$ after observing the state. Being costless, these messages are cheap talk (Crawford and Sobel, 1982; Green and Stokey, 2007) and thus need to satisfy S's incentive constraints. Our assumption that S's preferences are state independent simplifies these constraints considerably: S must be indifferent between all on-path messages. The reason is that if S's payoffs across two distinct messages differ, S never sends the lower-payoff message. Combining this indifference condition with the restrictions imposed by Bayesian updating and R-optimality yields the following characterization of the set of equilibrium outcomes (see Aumann and Hart, 2003; Lipnowski and Ravid, 2020): in addition to $p$ being an information policy, $p \in \mathcal{R}(\mu_0)$, S's payoff must be incentive compatible for R for all on-path posterior beliefs; that is, $s_i \in V(\mu)$ $p$-almost surely. Lipnowski and Ravid (2020) show the highest payoff S can attain subject to these constraints is given by $v$'s ***quasi*concave envelope**—that is, the pointwise lowest *quasi*concave function that is everywhere above $v$—which we denote by $\bar{v}$.

Panels (a) and (b) in Figure 3 (see subsection 3.2) illustrate $v$'s quasiconcave and concave envelopes, respectively. These envelopes describe S's ability to benefit from communication by connecting points on the graph of S's value correspondence. With full credibility, S can connect such points using any affine segment. When $\chi = 0$, S's incentive constraints mean her payoff coordinate must remain constant; that is, S can use only flat segments.

We now demonstrate the solution of our model's extreme cases in two binary-state specifications: the introduction's central bank example and a news-outlet example.

**Example 1.** Let us reformulate the introduction's central-bank example in more succinct notation. The economy's state can be either 1 *(*good*)* or 0 *(*bad*)*. Given the binary state, we abuse notation and identify each belief $\mu \in \Delta\Theta$ with the probability it assigns to state 1. The international firm (R) can make a large investment, $a = 2$, a small investment $a = 1$, or no investment, $a = 0$. The country's leader (S) wants to maximize R's expected investment, regardless of the state, $u_S(a) = a$, where R's returns from investing depends on the state via

$$u_R(a, \theta) = \theta a - 0.25a^2.$$

It is straightforward to verify this formulation leads to the same payoffs, and therefore the same R best-response correspondence as presented in the introduction. Thus, S's value correspondence and value function are given by

$$V(\mu) = \begin{cases} \{0\} & \text{if } \mu < 1/4, \\ [0,1] & \text{if } \mu = 1/4, \\ \{1\} & \text{if } \mu \in (1/4, 3/4), \\ [1,2] & \text{if } \mu = 3/4, \\ \{2\} & \text{if } \mu > 3/4, \end{cases} \qquad v(\mu) = \begin{cases} 0 & \text{if } \mu < 1/4, \\ 1 & \text{if } \mu \in [1/4, 3/4), \\ 2 & \text{if } \mu \geq 3/4. \end{cases} \tag{1}$$

Figure 1a graphs $v$ as a function of the probability R's belief assigns to state 1. Observe $v(\mu)$ increases with $\mu$.

Let us begin by solving the no-credibility case. When $\chi = 0$, S's maximal equilibrium value is given by the quasiconcave envelope of $v$ evaluated at the prior, $\bar{v}(\mu_0)$. Observe, however, that $v$ is itself quasiconcave by virtue of being monotone, and so $\bar{v} = v$. It follows that babbling is the best S can achieve via cheap talk.

Now suppose now S has full credibility. In this case, one can calculate S's highest equilibrium payoff via $v$'s concave envelope,

$$\hat{v}(\mu) = \begin{cases} 4\mu & \text{if } \mu \leq 1/4, \\ 1 + 2(\mu - 1/4) & \text{if } \mu \in [1/4, 3/4] \\ 2 & \text{if } \mu \geq 3/4. \end{cases} \tag{2}$$

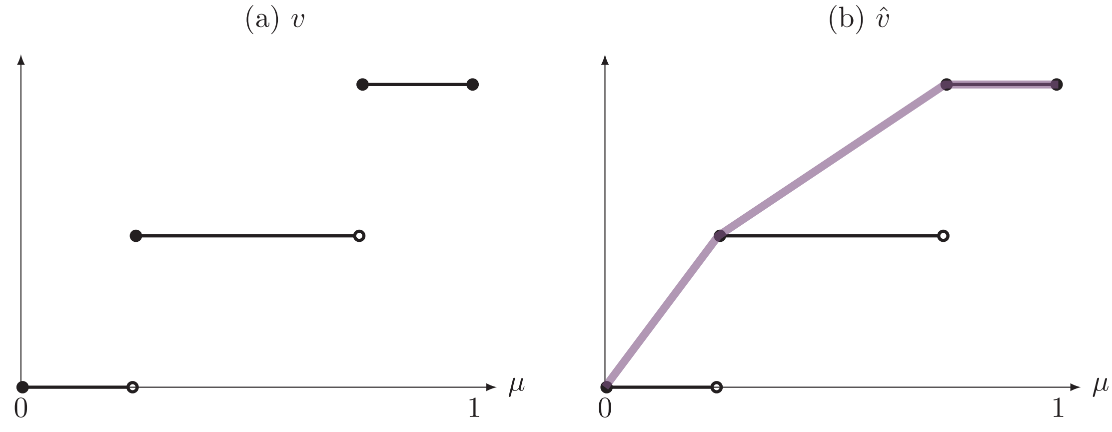

Figure 1: The value function and its concave envelope in Example 1.

For an illustration, see Figure 1. Evaluating this envelope at the prior, we get S's highest equilibrium payoff when $\chi = 1$ is $3/2$. As explained in the introduction, S can obtain this utility by commissioning a symmetric binary report that generates a signal realization that matches the state with probability $3/4$.

**Example 2.** Suppose the state is binary, $\Theta = \{0, 1\}$, and the prior, $\mu_0$, assigns state 1 a probability $\frac{1}{3}$. As in the previous example, identify each belief with the probability it assigns to $\theta = 1$. R chooses an action in $A = [0, 1]$ to minimize quadratic loss,

$$u_R(a, \theta) = -(a - \theta)^2.$$

S's preferences are given by

$$u_S(a) = (a - 1/3)^2 .$$

Observe R's unique best reply is to set his action equal to the expectation of the state, $a^*(\mu) = \mu$. Thus, S wants R's action to differ as much as possible from $1/3$, which is R's optimal action under the prior. One can interpret S as a news outlet whose goal is to influence R's beliefs as much as possible.

We now review how one would find S's favorite equilibrium in this example when $\chi$ is extreme. The first step is to find S's value function, which is given by

$$v(\mu) = u_S(a^*(\mu)) = (\mu - 1/3)^2.$$

Figure 2 depicts $v$ along with its concave ($\hat{v}$) and quasiconcave ($\bar{v}$) envelopes as functions of the probability $\mu$ R's belief assigns to state 1.

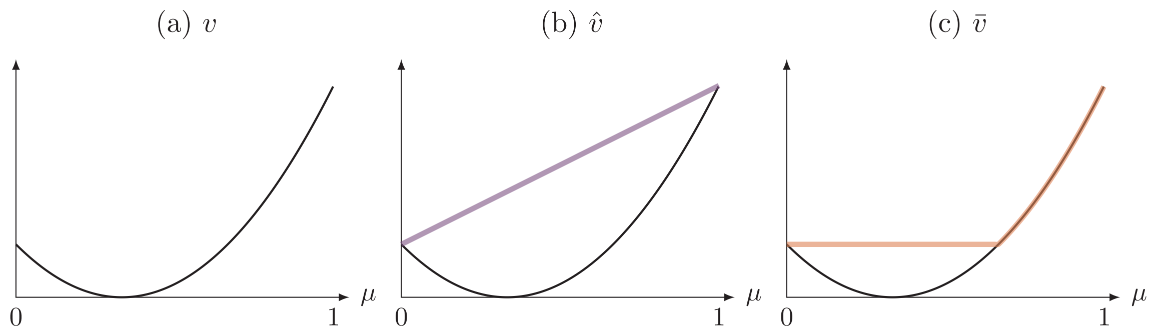


Figure 2: The value function and its quasiconcave and concave envelopes in Example 2.

Under $\chi = 1$, S's favorite equilibrium value is given by the concave envelope of $v$, evaluated at the prior. Since $v$ is convex, its concave envelope is given by the line connecting its value at the extreme beliefs, $\delta_0$ and $\delta_1$,

$$\hat{v}(\mu) = 1/9 + (1/3)\mu.$$

To attain this payoff, S uses a fully revealing official report. Because $\mu_0 = 1/3$ under the prior, S's full credibility payoff is 2/9.

Without any credibility, S's favorite equilibrium payoff is given by the quasiconcave envelope of $v$,

$$\bar{v}(\mu) = \max\{1/9, v(\mu)\} = \begin{cases} 1/9 & \text{if } \mu \leq 2/3, \\ (\mu - 1/3)^2 & \text{otherwise.} \end{cases}$$

S can attain the value $\bar{v}(\mu_0) = 1/9$ in a 0-equilibrium with only two on-path messages, $m_0$ and $m_1$, where S sends $m_1$ with probability 1 when the state is 1, and with probability 1/4 when the state is 0. Conditional on observing $m_0$, R knows the state is 0, and so chooses $a = 0$. If he observes $m_1$, R assigns a probability of 2/3 to the state being 1, and so chooses $a = 2/3$. Hence, S's expected payoff is 1/9, regardless of which message she sends.

## 3.2 The Intermediate Credibility Case

This section presents Theorem 1, which geometrically characterizes S's optimal $\chi$-equilibrium value for our general model. We begin by defining a class of $\chi$-equilibria that are sufficient for obtaining all of the game's $\chi$-equilibrium outcomes. We then explain how one can use this class of $\chi$-equilibria along with existing results on strategic communication to obtain an upper bound on S's maximal $\chi$-equilibrium payoff. Theorem 1 shows this bound is attainable, and thus delivers S's favorite $\chi$-equilibrium value.

In the Appendix, we show every $\chi$-equilibrium outcome is attainable via a restricted class of equilibria, termed $\chi$-nonical equilibria. A **$\chi$-nonical equilibrium** is a $\chi$-equilibrium, $(\xi, \sigma, \alpha, \pi)$, such that[8]

$$\sigma(\cdot|\theta) = \xi(\cdot|\theta, M_\alpha^*); \tag{3}$$

that is, given every state, S's strategy when influencing the report equals the official experiment's message distribution *conditional* on sending a message from $M_\alpha^*$. Informally, $\chi$-nonical equilibria are those in which an influenced institution censors any message S dislikes, repeatedly redrawing messages from the official report until it obtains a message that S approves.

Note the equilibria discussed in the introduction's example are all $\chi$-nonical. To see why, observe all the equilibria admit $g$ as S's unique payoff-maximizing message; that is, $M_\alpha^* = \{g\}$. Because influenced reporting only sends messages in $M_\alpha^*$, it follows that $\xi(g|\theta, M_\alpha^*) = \sigma(g|\theta) = 1$.

A useful property of $\chi$-nonical equilibria is that they remain equilibria when S has no credibility—as long as the prior is appropriately shifted. More specifically, if $(\xi, \sigma, \alpha, \pi)$ is a $\chi$-nonical equilibrium of $\mathbf{G}(\chi, \mu_0)$, it is also an equilibrium of $\mathbf{G}(0, \gamma)$, where $\gamma := \mathbb{E}[\pi|m \in M_\alpha^*]$ is the posterior R would have if he learned $m$ is in $M_\alpha^*$ but did not observe $m$ itself. To see why, observe first that because $(\xi, \sigma, \alpha, \pi)$ is an equilibrium of $\mathbf{G}(\chi, \mu_0)$, S's influencing strategy $\sigma(\theta)$ is supported on $M_\alpha^*$ for all $\theta$, and thus satisfies S's incentive constraints in $\mathbf{G}(0, \gamma)$. Similarly, because R's posterior belief, $\pi$, and resulting action distribution, $\alpha$, are taken from an equilibrium, $\alpha(m)$ must maximize R's utility given the posterior $\pi(m)$ for every message $m$. All that remains is to verify that $(\xi, \sigma, \alpha, \pi)$ is consistent with Bayesian updating when played in $\mathbf{G}(0, \gamma)$, a fact that follows from $(\xi, \sigma, \alpha, \pi)$ being $\chi$-nonical. [9]

[8]Formally, given $\theta \in \Theta$, define $\xi(\hat{M}|\theta, M_\alpha^*) := \xi(\hat{M} \cap M_\alpha^*|\theta)/\xi(M_\alpha^*|\theta)$ for every Borel $\hat{M} \subseteq M$ if $\xi(M_\alpha^*|\theta) > 0$, and define $\xi(\cdot|\theta, M_\alpha^*) := \sigma(\cdot|\theta)$ otherwise.

[9]For intuition, suppose the model is finite and that $M_\alpha^*$ is finite as well. In this case, every on-path $m \in M_\alpha^*$ and every $\theta \in \Theta$ satisfy

$$\pi(\theta|m) = \frac{\mu_0(\theta)[\chi\xi(M_\alpha^*|\theta) + 1 - \chi]\sigma(m|\theta)}{\sum_{\theta' \in \Theta} \mu_0(\theta')[\chi\xi(M_\alpha^*|\theta') + 1 - \chi]\sigma(m|\theta')} = \frac{\gamma(\theta)\sigma(m|\theta)}{\sum_{\theta'} \gamma(\theta')\sigma(m|\theta')},$$

We now proceed to obtain an upper bound on S's value from an arbitrary $\chi$-nonical equilibrium, $(\xi, \pi, \alpha, \sigma)$. Letting $k := \mathbb{P}\{m \notin M_\alpha^*\}$ be the probability that R observes an S-suboptimal message, this equilibrium gives S a utility of

$$\mathbb{E}[u_S] = k\mathbb{E}[u_S|m \notin M_\alpha^*] + (1-k)\mathbb{E}[u_S|m \in M_\alpha^*]. \tag{4}$$

We find that because $(\xi, \pi, \alpha, \sigma)$ is $\chi$-nonical, one can use quasiconcavification to bound $\mathbb{E}[u_S|m \in M_\alpha^*]$ from above. To do so, notice that all messages in $M_\alpha^*$ generate the same payoff for S. Moreover, this payoff equals S's utility when playing $(\xi, \pi, \alpha, \sigma)$ in $\mathbf{G}(0, \gamma)$, because then S's message comes from $\sigma$, which is supported on $M_\alpha^*$. Therefore, $\mathbb{E}[u_S|m \in M_\alpha^*]$ is a no-credibility equilibrium value under prior $\gamma$, and thus must be smaller than the largest such payoff, $\bar{v}(\gamma)$.

Next, we explain how to use concavification to bound $\mathbb{E}[u_S|m \notin M_\alpha^*]$. Toward this goal, fix some $m$ outside $M_\alpha^*$, and let $\mu := \pi(m)$ be its induced posterior belief. Observe that S's payoff from sending $m$ must satisfy two constraints. First, R must be acting optimally given his posterior, meaning S's payoff must lie below $v(\mu)$. And second, S's payoff must be below the utility she gets from sending a message in $M_\alpha^*$; otherwise, S would prefer to send $m$ rather than $M_\alpha^*$ when influencing the report. Because S's utility from sending messages in $M_\alpha^*$ is lower than $\bar{v}(\gamma)$, it follows that S's payoff from $m$ must also be below $\bar{v}(\gamma)$. Thus, S's payoff from $m$ must be lower than

$$v_{\wedge\gamma}(\mu) := \bar{v}(\gamma) \wedge v(\mu).$$

It follows that $\mathbb{E}[u_S|m \notin M_\alpha^*]$ is smaller than the utility that S would obtain in a full-credibility game in which her value function is $v_{\wedge\gamma}$ and R's belief is $\beta := \mathbb{E}[\pi|m \notin M_\alpha^*]$; that is,[10]

$$\mathbb{E}[u_S|m \notin M_\alpha^*] \leq \hat{v}_{\wedge\gamma}(\beta),$$

where $\hat{v}_{\wedge\gamma}$ is $v_{\wedge\gamma}$'s concave envelope.

Observe that $\hat{v}_{\wedge\gamma}$ expresses only some of the restrictions imposed by partial credibility on S's ability to derive value from her official report. For an explanation, note $\hat{v}_{\wedge\gamma}$

---

where the first equality follows from $(\xi, \pi, \alpha, \sigma)$ being $\chi$-nonical, and the second equality from dividing both the numerator and the denominator by the probability R observes a message from $M_\alpha^*$ when $(\xi, \pi, \alpha, \sigma)$ is played in $\mathbf{G}(\chi, \mu_0)$.

[10] One can also verify this inequality directly: $\mathbb{E}[u_S(m)|m \notin M_\alpha^*] \leq \mathbb{E}[v_{\wedge\gamma} \circ \pi(m)|m \notin M_\alpha^*] \leq \mathbb{E}[\hat{v}_{\wedge\gamma} \circ \pi(m)|m \notin M_\alpha^*] \leq \hat{v}_{\wedge\gamma}(\beta)$, where the first and second inequalities following from $u_S(m) \leq v_{\wedge\gamma} \circ \pi(m) \leq \hat{v}_{\wedge\gamma} \circ \pi(m)$, and the last inequality following from Jensen.

is obtained by first capping S's value function at $\bar{v}(\gamma)$ and then concavifying (see Figure 3). Prima facie, each operation seems reasonable: the cap prevents S from sending messages that would lead to profitable deviations under influenced reporting, whereas concavification expresses the fact that S is committed to the announced experiment when reporting is credible. Still, each operation involves a relaxation of some incentive constraint. First, for an arbitrary $\chi$-equilibrium, S's payoff when influencing the report may be lower than $\bar{v}(\gamma)$, and so using $\bar{v}(\gamma)$ as a cap may not eliminate all of S's profitable deviations. And second, whereas concavification assumes all posteriors beliefs are feasible, inducing posteriors at which the cap binds may be impossible, because at such posteriors, R may be unwilling to give S a payoff as low as $\bar{v}(\gamma)$. As we discuss later, showing neither of these relaxations matters at S's favorite $\chi$-equilibrium lies at the crux of Theorem 1's proof.

(a) $\bar{v}$ (b) $\hat{v}$ (c) $\hat{v}_{\wedge\gamma}$

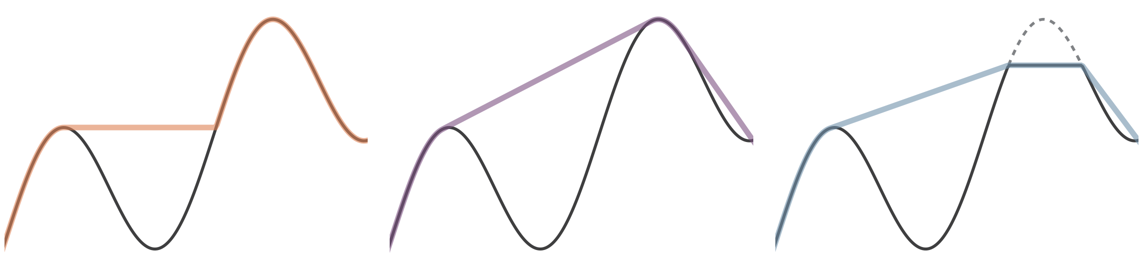

Figure 3: Quasiconcave envelope, concave envelope, and concave envelope with a cap.

So far, we have argued S's value from a fixed $\chi$-nonical equilibrium is lower than

$$v^*(\beta, \gamma, k) := k\hat{v}_{\wedge\gamma}(\beta) + (1-k)\bar{v}(\gamma).$$

Of course, the above bound holds only for the equilibrium that induced $(\beta, \gamma, k)$. To attain an upper bound across all equilibria, we maximize the above expression over all $(\beta, \gamma, k)$ satisfying two restrictions necessary for a $\chi$-nonical equilibrium. For the first restriction, observe that telling R whether $m$ is in $M^*_\alpha$ (without telling her $m$ itself) results in her having a posterior of $\beta$ with probability $k$, and a posterior of $\gamma$ with probability $1-k$. Bayesian updating therefore requires

$$k\beta + (1-k)\gamma = \mu_0; \tag{R-BP}$$

that is, R's average posterior must equal his prior. For the second restriction, take any event (i.e., Borel set) $\hat{\Theta} \subseteq \Theta$, and observe the probability this event occurs and R sees a message from $M_\alpha^*$ is $(1-k)\gamma(\hat{\Theta})$. Moreover, this probability should be at least as high as the probability that $\hat{\Theta}$ occurs and reporting is non-credible—that is, $(1-\chi)\mu_0(\hat{\Theta})$—because influenced reporting always sends a message from $M_\alpha^*$. In other words,

$$(1-k)\gamma \geq (1-\chi)\mu_0. \tag{$\chi$-BP}$$

Thus, we have obtained the following upper bound on S's maximal $\chi$-equilibrium value:

$$v_\chi^*(\mu_0) := \max_{\beta,\gamma\in\Delta\Theta,\ k\in[0,1]} \quad v^*(\beta,\gamma,k) \tag{5}$$

$$\text{s.t.} \quad (\text{R-BP}) \text{ and } (\chi\text{-BP}). \tag{6}$$

Applying the existing literature's results, one can easily show the above bound is attained when $\chi$ is extreme. For an explanation, suppose first S has no credibility, meaning $\chi = 0$. In this case, the only way to satisfy ($\chi$-BP) is to set $k = 0$ and $\gamma = \mu_0$. As such, $v_0^*(\mu_0) = \bar{v}(\mu_0)$, which is attained by S's favorite 0-equilibrium. Suppose now S has full credibility; that is, $\chi = 1$. Because both $\bar{v}$ and $\hat{v}_{\wedge\gamma}$ are bounded from above by $\hat{v}$ (see Figure 3), it follows that every $\beta$, $\gamma$, and $k$ that satisfy (R-BP) must also have

$$k\hat{v}_{\wedge\gamma}(\beta) + (1-k)\bar{v}(\gamma) \leq k\hat{v}(\beta) + (1-k)\hat{v}(\gamma) \leq \hat{v}(\mu_0),$$

where the last inequality follows from concavity of $\hat{v}$. Because $\hat{v}(\mu_0)$ is attainable in a 1-equilibrium, we therefore have $\hat{v}(\mu_0) \leq v_1^*(\mu_0) \leq \hat{v}(\mu_0)$; that is, S obtains $v_1^*(\mu_0)$ in her favorite 1-equilibrium.

Our main theorem shows the above bound is also tight when $\chi$ is intermediate.

**Theorem 1.** *A $\chi$-equilibrium exists in which S's value is $v_\chi^*(\mu_0)$. Moreover, any such $\chi$-equilibrium is S-optimal.*

When $v_\chi^*(\mu_0) = \bar{v}(\mu_0)$, one can attain the program's value using a cheap-talk equilibrium. Thus, suppose $v_\chi^*(\mu_0) > \bar{v}(\mu_0)$. In this case, the theorem's proof uses the model's extreme cases to transform any $(\beta, \gamma, k)$ that solves (5) into a $\chi$-nonical equilibrium. For a sketch, start by taking an S-favorite equilibrium of $\mathbf{G}(0,\gamma)$—say, $(\xi_\gamma, \sigma_\gamma, \alpha_\gamma, \pi_\gamma)$—that generates S a payoff of $\bar{v}(\gamma)$. Next, find an equilibrium $(\xi_\beta, \sigma_\beta, \alpha_\beta, \pi_\beta)$ of the game $\mathbf{G}(1,\beta)$ that gives S a utility of $\hat{v}_{\wedge\gamma}(\beta)$, and in which all of S's messages yield a value smaller than $\bar{v}(\gamma)$ (we defer explaining how to obtain such an equilibrium until the next paragraph). Standard measure-theoretic arguments ensure we can select these two equilibria in a way that partitions the set of on-path messages into two: the set $M_\gamma$ of messages sent by $\sigma_\gamma$ and the set $M_\beta$ of messages sent by $\xi_\beta$. One can then proceed by pasting these two equilibria into a single $\chi$-nonical equilibrium in which S's value is $v^*_\chi(\mu_0)$. When $M_\gamma$, $M_\beta$, and $\Theta$ are all finite, the construction goes as follows. For any message $m \in M_\beta \cup M_\gamma$, define R's strategy and beliefs according to the equilibrium that generates $m$ on path; that is, $(\alpha^*(m), \pi^*(m)) = (\alpha_\gamma(m), \pi_\gamma(m))$ for $m \in M_\gamma$, and $(\alpha^*(m), \pi^*(m)) = (\alpha_\beta(m), \pi_\beta(m))$ for $m \in M_\beta$. S's influencing strategy is defined via $\sigma^* := \sigma_\gamma$, whereas the official report is defined in two parts. First, for $m$ in $M_\gamma$ take

$$\xi^*(m|\theta) = \left[\frac{(1-k)\gamma(\theta)}{(1-\chi)\mu_0(\theta)} - 1\right]\left(\frac{1-\chi}{\chi}\right)\sigma_\gamma(m|\theta),$$

which is a well-defined probability due to (R-BP) and ($\chi$-BP). And second, set the probability the official report sends a message $m$ in $M_\beta$ to be

$$\xi^*(m|\theta) = \left[\frac{k\beta(\theta)}{\chi\mu_0(\theta)}\right]\xi_\beta(m|\theta).$$

Let us now verify $(\xi^*, \sigma^*, \alpha^*, \pi^*)$ is an equilibrium of the original game, $\mathbf{G}(\chi, \mu_0)$. Observe all on-path messages come from either $M_\beta$ or $M_\gamma$, and that R's beliefs and (mixed) action following any such message come from an equilibrium with a prior of $\beta$ and $\gamma$, respectively, and so R's strategy must be optimal given her beliefs. Optimality of S's behaviors follows from two facts. First, all messages in $M_\gamma$ lead $\alpha_\gamma$ to take an action yielding S a payoff of $\bar{v}(\gamma)$. And, second, all messages in $M_\beta$ result in $\alpha_\beta$ giving S a lower payoff. It follows the same is true for $\alpha^*$, and so S has no incentive to deviate when influencing the report. It remains to check whether R's beliefs obey Bayes' rule. To do so, observe our construction guarantees the probability the state is $\theta$ and R observes the message $m$ is $(1-k)\gamma(\theta)\sigma_\gamma(m|\theta)$ when $m$ is in $M_\gamma$, and $k\beta(\theta)\xi_\beta(m|\theta)$ when $m$ is in $M_\beta$. Therefore, Bayes' rule dictates R must assign $\theta$ a probability of

$$\frac{k\beta(\theta)\xi_\beta(m|\theta)}{\sum_{\theta' \in \Theta} k\beta(\theta')\xi_\beta(m|\theta')} = \pi_\beta(\theta|m) = \pi^*(\theta|m),$$

after seeing a message $m \in M_\beta$, and

$$\frac{(1-k)\sigma_\gamma(m|\theta)\gamma(\theta)}{\sum_{\theta' \in \Theta}(1-k)\sigma_\gamma(m|\theta')\gamma(\theta')} = \pi_\gamma(\theta|m) = \pi^*(\theta|m)$$

after $m \in M_\gamma$. In other words, $(\xi^*, \sigma^*, \alpha^*, \pi^*)$ is consistent with Bayesian updating.

We now return to finding an appropriate $(\xi_\beta, \sigma_\beta, \alpha_\beta, \pi_\beta)$. Our task is to define this profile so that it is an equilibrium of $\mathbf{G}(1, \beta)$ satisfying two criteria: First, it gives S a payoff of $\hat{v}_{\wedge\gamma}(\beta)$. And second, all of S's messages yield a payoff weakly under $\bar{v}(\gamma)$. Toward this goal, take $\tilde{\mathbf{G}}(1, \beta)$ to be the modified version of the game in which S's utility from high-payoff actions is reduced to $\bar{v}(\gamma)$; that is, replace $u_S$ with

$$u_{S\wedge\gamma}(a) := \min\{u_S(a), \bar{v}(\gamma))\}.$$

In this modified game, S's value function equals $v_{\wedge\gamma}$, and so $\hat{v}_{\wedge\gamma}(\beta)$ is S's highest equilibrium value in $\tilde{\mathbf{G}}(1, \beta)$. Moreover, S's payoff in this game differs from her payoff in the original game only when R takes an action $a$ for which $u_S(a) > \bar{v}(\gamma)$. Therefore, to obtain the desired strategy profile, it is enough to find an S-favorite equilibrium of $\tilde{\mathbf{G}}(1, \beta)$ with the property that all on-path messages deliver S a payoff below $\bar{v}(\gamma)$ in the *original* game, $\mathbf{G}(1, \beta)$. We find one can obtain such a profile by taking $(\xi_\beta, \sigma_\beta, \alpha_\beta, \pi_\beta)$ to be an S-favorite equilibrium of $\tilde{\mathbf{G}}(1, \beta)$ in which R chooses (mixed) actions that yield a $u_S$ of $\bar{v}(\gamma)$ or below when indifferent.[11] The intuition is similar to Kamenica and Gentzkow's (2011) judge example in that it uses the following property of the full-commitment solution: the concave envelope must be affine on the line connecting $\beta$ with $\pi_\beta(m)$ for almost every $m$. To use this property, notice $(\xi_\beta, \sigma_\beta, \alpha_\beta, \pi_\beta)$ results in $u_S$ being strictly above $\bar{v}(\gamma)$ after a message $m$ only if R is unwilling to take an action for which $u_S$ is below $\bar{v}(\gamma)$ when holding a belief of $\mu := \pi_\beta(m)$. A simple limit argument then implies no such action can be optimal for R at any nearby belief as well. Therefore, $\hat{v}_{\wedge\gamma}(\mu') = \hat{v}_{\wedge\gamma}(\mu) = \bar{v}(\gamma) > \hat{v}_{\wedge\gamma}(\beta)$ must hold for any belief $\mu'$ close to $\mu$ that lies on the line between $\mu$ and $\beta$. But this inequality can hold only if $\hat{v}_{\wedge\gamma}$ fails to be affine on this line, a contradiction. It follows that $(\xi_\beta, \sigma_\beta, \alpha_\beta, \pi_\beta)$ satisfies the desiderata.

We now demonstrate the theorem by applying it to the simple two-state examples from before.

**Example 1** (continued)**.** Consider again the reformulation of the introduction's example. Thus, $\Theta = \{0, 1\}$, $A = \{0, 1, 2\}$, $u_S(a) = a$, and $u_R(a, \theta) = \theta a - 0.25a^2$. Recall we identify each belief with the probability it assigns to $\theta = 1$. S's value function is given by equation (1), whereas (2) gives its concave envelope. Because the value function is

[11] Observe such a selection is consistent with S-favored tie-breaking in the *modified* game, because in that game, S cannot obtain a payoff higher than $\bar{v}(\gamma)$.

monotone, it is quasiconcave and so equal to its quasiconcave envelope.

By Theorem 1, we can find S's highest $\chi$-equilibrium payoff by solving the program (5). To obtain this solution, we first make a few useful observations regarding the program's constraints. Examining (R-BP) for this binary-state setting reveals it is equivalent to two conditions: First, either $\beta \leq \mu_0 \leq \gamma$ or $\gamma \leq \mu_0 \leq \beta$. And second, if $\beta \neq \gamma$, then $k = k_{\beta,\gamma}$, where

$$k_{\beta,\gamma} = \frac{\gamma - \mu_0}{\gamma - \beta}. \tag{7}$$

Moreover, because the state space is finite (indeed, binary), the measure-valued constraint ($\chi$-BP) can be evaluated state by state. Specifically, it is enough to check

$$(1-k)\gamma \;\; \geq \;\; (1-\chi)\mu_0 \tag{8}$$

$$(1-k)(1-\gamma) \;\; \geq \;\; (1-\chi)(1-\mu_0). \tag{9}$$

Let us make two observations assuming $\beta \leq \mu_0 < \gamma$ (which turns out to be the relevant case for this example). First, in this range, the weight $k_{\beta,\gamma}$ is increasing in $\beta$ and $\gamma$. Second, because $\mu_0 = 1 - \mu_0 = 1/2$, the right-hand sides of (8) and (9) is the same. It follows (9) implies (8) whenever $\gamma \geq \mu_0 = 1/2$.

We now establish that whenever the program (5) attains a value strictly greater than 1, it admits a solution $(\beta, \gamma, k)$ with $\gamma = \gamma^* := 3/4$. To see $\gamma \geq \gamma^*$ if $v^*_\chi(\mu_0) > 1$, observe that if $\gamma < \gamma^*$, then $v^*(\beta, \gamma, k) \leq k\bar{v}(\gamma) + (1-k)\bar{v}(\gamma) = \bar{v}(\gamma) = v(\gamma) \leq 1$ must hold for any $(\beta, k)$ such that $(\beta, \gamma, k)$ is feasible. Next, we argue one can modify any $(\beta, \gamma, k)$ with $\gamma > \gamma^*$ in a way that replaces $\gamma$ with $\gamma^*$ and does not decrease S's objective. For this purpose, let $k^* := k_{\beta,\gamma^*}$, and observe that $\beta \leq \mu_0 < \gamma^* < \gamma$ implies $(\beta, \gamma^*, k^*)$ satisfies (R-BP) and $k^* \leq k$. In addition,

$$\begin{aligned}
(1-k^*)(1-\gamma^*) \;\; &= \;\; (1-\mu_0) - k^*(1-\beta) \\
&\geq \;\; (1-\mu_0) - k(1-\beta) \\
&= \;\; (1-k)(1-\gamma) \\
&\geq \;\; (1-\chi)(1-\mu_0),
\end{aligned}$$

where the equalities hold because both $(\beta, \gamma, k)$ and $(\beta, \gamma^*, k^*)$ satisfy (R-BP), the first inequality holds because $k^* \leq k$, and the second inequality holds because $(\beta, \gamma, k)$ satisfies (9). Hence, $(\beta, \gamma^*, k^*)$ satisfies (9), and therefore ($\chi$-BP). Thus, $(\beta, \gamma^*, k^*)$ is feasible in program (5). Moreover, that $\bar{v}(\gamma^*) = \bar{v}(\gamma) > \hat{v}_{\wedge\gamma}(\beta)$ and $k^* \leq k$ implies $v^*(\beta,\gamma^*,k^*) \geq v^*(\beta,\gamma,k)$. Thus, because $(\beta,\gamma,k)$ is optimal, $(\beta,\gamma^*,k^*)$ must be optimal as well.

Collecting our observations to this point, we have that it is enough to choose $(\beta,\gamma,k)$ so that either

(a) $\beta=\gamma=\mu_0$ and $k=0$, which is always feasible, or

(b) $\beta\in[0,\mu_0]$, $\gamma=\gamma^*$, and $k=k_{\beta,\gamma}$, which is feasible if and only if (9) holds.

Moreover, we know that decreasing $\beta$ in case (b) always relaxes (9) by lowering $k_{\beta,\gamma}$. Thus, letting

$$\tilde{v}^*(\beta) := v^*(\beta,\gamma^*,k_{\beta,\gamma^*})$$

for $\beta\in[0,\mu_0]$, it follows S's optimal value is the higher of 1 (the payoff she attains in case (a)) and the maximal payoff attainable in case (b), which is given by $\sup_\beta \tilde{v}^*(\beta)$, where the supremum is taken over the interval of $\beta\in[0,\mu_0]$ that satisfy (9) when paired with $\gamma^*$ and $k_{\beta,\gamma^*}$.

To see which of the above two cases is optimal, we can use the formula for $\hat{v}$ derived in (2) to compute

$$\begin{aligned}
\tilde{v}^*(\beta) &= k_{\beta,\gamma^*}\hat{v}_{\wedge\gamma}(\beta)+(1-k_{\beta,\gamma^*})\bar{v}(\gamma^*)\\
&= \tfrac{\gamma^*-\mu_0}{\gamma^*-\beta}\hat{v}(\beta)+\tfrac{\mu_0-\beta}{\gamma^*-\beta}2\\
&= \begin{cases} 1+1/(3-4\beta) & \text{if } \beta\in[0,1/4],\\ 3/2 & \text{if } \beta\in[1/4,1/2].\end{cases}
\end{aligned}$$

Observe the above objective is increasing on $[0,1/4]$, constant on $[1/4,\mu_0]$, and globally strictly higher than $1=v(\mu_0)$. It follows that case (a) is optimal only when no $\beta\in[0,\mu_0]$ can satisfy (9) when paired with $\gamma^*$ and $k_{\beta,\gamma^*}$. In addition, whenever such a $\beta$ is feasible, it is optimal to set it to be the highest feasible value below $1/4$ from case (b).

Let us solve the program. First, if $\chi<2/3$, then even setting $\beta=0$ (which relaxes (9) as much as possible) violates (9), and so we cannot improve upon feasible solution $(\beta,\gamma,k)=(1/2,1/2,0)$, which yields value $v^*_\chi(\mu_0)=1$. Second, if $\chi\geq 3/4$, then setting $\beta=1/4$ satisfies (9) and maximizes $\tilde{v}^*(\beta)$, so that $(\beta,\gamma,k)=(1/4,3/4,1/2)$ is optimal—delivering the full-commitment value of $v^*_\chi(\mu_0)=3/2$. Finally, consider the case in which $2/3 \leq \chi < 3/4$. In this case, direct computation shows

$$\beta_\chi^* = \frac{3\chi - 2}{4\chi - 2} \in [0, 1/4]$$

uniquely satisfies (9) with equality, and so is the highest feasible value of $\beta$. Because $k_{\beta_\chi^*,\gamma_\chi^*} = 2\chi - 1$, it follows that $(\beta, \gamma, k) = (\beta_\chi^*, 3/4, 2\chi - 1)$ solves (5), giving S a utility of $v_\chi^*(\mu_0) = \tilde{v}_\chi^*(\beta_\chi^*) = 2\chi$.

To summarize, S's maximal equilibrium payoff is given by

$$v_\chi^*(1/2) = \begin{cases} 1 & \text{if } \chi < \frac{2}{3}, \\ 2\chi & \text{if } \chi \in \left[\frac{2}{3}, \frac{3}{4}\right], \\ 3/2 & \text{if } \chi \geq \frac{3}{4}. \end{cases}$$

The way S obtains the above value depends on $\chi$. When $\chi < 2/3$, it is best for S to leave R uninformed. When $\chi = 1$, S is best commissioning the report described in the introduction, $\xi_1$. To obtain her full-credibility payoff when $\chi \in [3/4, 1)$, S commissions a report that induces the same information about $\theta$ in equilibrium. Specifically, S uses an official experiment that sends either $m_1$ or $m_0$ according to

$$\xi_\chi^*(m_1|1) = 1 - \frac{1}{4\chi}, \qquad \xi_\chi^*(m_0|0) = \frac{3}{4\chi},$$

whereas influenced reporting sends $m_1$ regardless of the state.

To find S's optimal equilibrium when $\chi \in [2/3, 3/4)$, we begin by using the solution to (5) to find the distribution of R's posterior belief via a two-step procedure. First, one splits $\mu_0$ into $\beta$ with probability $k_{\beta,\gamma}$, and $\gamma$ with probability $1 - k_{\beta,\gamma}$. Second, one splits the first step's outcome across the nearest posteriors that include it in their convex hull and for which $v$ agrees with the appropriate envelope—$\hat{v}_{\wedge\gamma}$ for $\beta$, and $\bar{v}$ for $\gamma$. Examining Figure 4 reveals that splitting $\gamma_\chi^* = 3/4$ is not necessary, whereas $\beta_\chi^*$ is split across $1/4$ and $0$, with probability $4\beta$ and $1 - 4\beta$, respectively. Thus, S's optimal equilibrium induces three posteriors, $3/4$, $1/4$, and $0$, with probability $1 - k_{\beta_\chi^*,\gamma_\chi^*} = 2 - 2\chi$, $4\beta_\chi^* k_{\beta_\chi^*,\gamma_\chi^*} = 6\chi - 4$, and $k_{\beta_\chi^*,\gamma_\chi^*}(1 - 4\beta_\chi^*) = 3 - 4\chi$, respectively.[12] Given this posterior distribution, one can construct a $\chi$-nonical equilibrium in which influenced reporting sends only the messages that split $\gamma$, whereas the official report sends all on-path messages. Letting $m_\mu$ be the message inducing posterior $\mu \in \{0, 1/4, 3/4\}$, this

[12] If $\chi = 2/3$, posterior $1/4$ has zero probability.

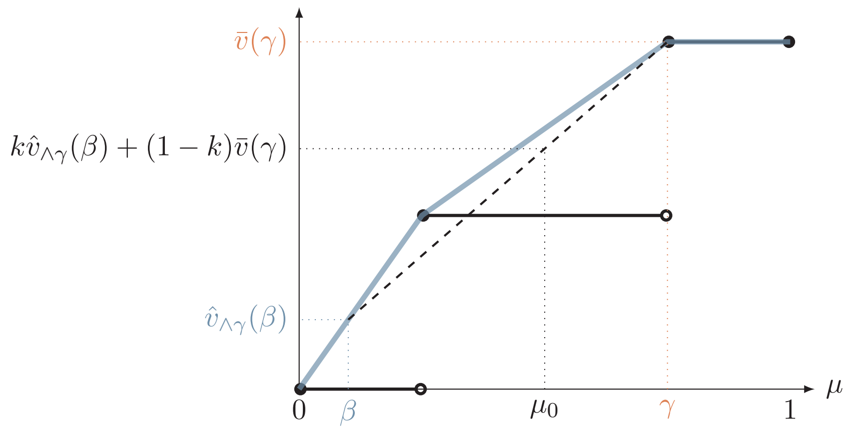


Figure 4: Calculating S value for feasible $\beta$ and $\gamma$ for Example 1.

equilibrium has $\sigma$ sending $m_{3/4}$ regardless of the state, and the official report being equal to

$$\xi^*_\chi(m_0|0) = \frac{6}{\chi} - 8, \qquad \xi^*_\chi(m_{\frac{1}{4}}|0) = \frac{6}{5\chi}\left[6\chi - 4\right],$$
$$\xi^*_\chi(m_0|1) = 0, \qquad \xi^*_\chi(m_{\frac{1}{4}}|1) = \frac{4}{5\chi}\left[6\chi - 4\right].$$

**Example 2** (continued)**.** Recall the state is binary, $\Theta = \{0, 1\}$, and that $\mu_0\{1\} = 1/3$. R chooses an action in $A = [0, 1]$ to minimize quadratic loss, $u_R(a, \theta) = -(a - \theta)^2$, whereas S wants R's action to be as far away from $1/3$ as possible, $u_S(a) = (a - 1/3)^2$. Identifying each belief with the probability it assigns state 1, $a^*(\mu) = \mu$ is R's unique best response, and S's value function is given by

$$v(\mu) = u_S(a^*(\mu)) = (\mu - 1/3)^2 = 1/9 + (\mu - 2/3)\mu.$$

Let us find S's favorite $\chi$-equilibrium when $\chi$ is intermediate. By Theorem 1, S's value in this case is given by the program (5). A straightforward computation reveals

$$\hat{v}_{\wedge\gamma}(\mu) = \begin{cases} \frac{1}{9} & \text{if } \gamma \leq \frac{2}{3}, \\ \frac{1}{9} + \left(\gamma - \frac{2}{3}\right)\gamma & \text{if } \mu \geq \gamma > \frac{2}{3}, \\ \frac{1}{9} + \left(\gamma - \frac{2}{3}\right)\mu & \text{if } \gamma \geq \max\{\mu, \frac{2}{3}\}. \end{cases}$$

Figure 5 plots $\hat{v}_{\wedge\gamma}$ for different $\gamma$, putting the probability R's belief assigns to state 1 on the horizontal axis. As can be seen in the figure, $\hat{v}_{\wedge\gamma}(\mu)$ increases in an affine way for $\mu \in [0, \gamma]$ and remains flat at $\bar{v}(\gamma)$ thereafter.

We now argue one can focus on the case in which $\beta \leq \mu_0 \leq \gamma$ and $\beta \neq \gamma$. For this purpose, recall (R-BP) can be satisfied if and only if either $\beta \leq \mu_0 \leq \gamma$, or $\gamma \leq \mu_0 \leq \beta$. Notice in the latter case $\hat{v}_{\wedge\gamma}(\cdot)$ is constant and equals $1/9$ for all beliefs, which is also attainable by setting $\gamma = \beta = \mu_0$. Moreover, $(\mu_0, \mu_0, \chi)$ trivially satisfies ($\chi$-BP). Therefore, the program (5) attains a value if and only if this value is feasible for some $\beta \leq \mu_0 \leq \gamma$, and so one can restrict attention to this case. Moreover, the value attained by setting $\beta = \gamma = \mu_0$ can also be attained by setting $(\beta, \gamma, k) = (0, \mu_0, 0)$, so we may further take $\beta \neq \gamma$ without loss of optimality.

Next, given $\beta \leq \mu_0 \leq \gamma$ and $\beta \neq \gamma$, we show we can always set $\beta = 0$ at the optimum. For an explanation, recall (R-BP) can be satisfied only if $k = k_{\beta,\gamma}$ (see equation (7) and the surrounding discussion). Because $\beta \leq \mu_0 \leq \gamma$, $k_{\beta,\gamma}$ is decreasing with $\beta$, and so lowering $\beta$ relaxes ($\chi$-BP). Therefore, replacing $\beta$ with $\beta = 0$ does not hurt feasibility. Moreover, such a replacement has no impact on $v^*$'s value, because $\hat{v}_{\wedge\gamma}$ is affine on $[0, \gamma]$ (see Figure 5). Hence, if $(\beta, \gamma, k_{\beta,\gamma})$ solve (5), so do $(0, \gamma, k_{0,\gamma})$.

Thus, we have reduced the task of solving (5) to finding the $\gamma \in [\mu_0, 1]$ that maximizes $v^*(0, \gamma, k_{0,\gamma})$, subject to ($\chi$-BP). A simple calculation reveals $v^*(0, \gamma, k_{0,\gamma}) = \hat{v}_{\wedge\gamma}(\mu_0)$, and so increases with $\gamma$ (see Figure 5). Therefore, to solve the program, we need to choose $\gamma$ to equal the highest value ($\chi$-BP) allows. To find this value, we substitute $k_{0,\gamma} = \gamma/\mu_0$ into ($\chi$-BP) to obtain two inequalities,

$$\mu_0 \geq (1-\chi)\mu_0 \tag{10}$$

$$\frac{\mu_0}{\gamma}(1-\gamma) \geq (1-\chi)(1-\mu_0). \tag{11}$$

Notice (10) holds for all $\chi$, and so the only binding constraint is (11). Rearranging this constraint and substituting $\mu_0 = 1/3$ shows it is equivalent to

$$\gamma \leq 1/(3-2\chi) \in [1/3, 1].$$

Saturating this constraint, we obtain S's utility in her favorite equilibrium is

$$v^*_\chi(\mu_0) = \begin{cases} 1/9 & \text{if } \chi \leq 3/4, \\ 2\chi/[3(9-6\chi)] & \text{otherwise.} \end{cases}$$

To obtain this value, S commissions a binary-message official report. When $\chi \leq$

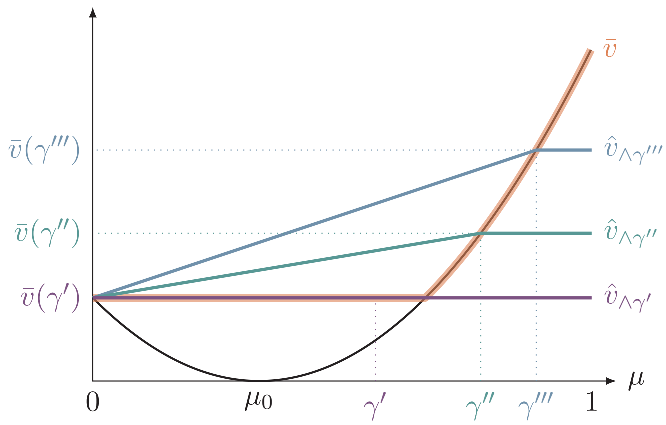


Figure 5: The concave envelope ($\hat{v}_{\wedge\gamma}$) of the capped value function ($v_{\wedge\gamma}$) for various values of $\gamma$ for Example 2.

3/4, this report equals S's strategy in her best 0-equilibrium, which also equals S's influencing strategy. When $\chi > 3/4$, the official report is fully revealing, and S pretends the state is 1 whenever the report is not credible.

# 4 Varying Credibility

This section uses Theorem 1 to conduct general comparative statics in the model's finite version. First, we study how a decrease in S's credibility affects R's value. In particular, we provide sufficient conditions for R to benefit from a less credible S. Second, we show that small reductions in S's credibility can often lead to a large drop in S's payoffs. Finally, we note these drops rarely occur at full credibility. In other words, the full credibility value is usually robust to small imperfections in S's commitment power.

## 4.1 Productive Mistrust

We now study how a decrease in S's credibility affects R's value and the informativeness of S's equilibrium communication. In general, the less credible the sender, the smaller the set of equilibrium information policies.[13] However, that the set of equilibrium policies shrinks does not mean less information is transmitted in S's preferred equilibrium. Our introductory example is a case in point, showing that lowering S's

[13]See Lemma 1 in the appendix. In particular, as credibility decreases, the lemma's condition 3 becomes more restrictive.

credibility can result in a more informative equilibrium (à la Blackwell, 1953). Moreover, in that example, this additional information is used by R, who obtains a strictly higher value when S's credibility is lower. In what follows, we refer to this phenomenon as productive mistrust and provide sufficient conditions for it to occur.

Our key sufficient condition involves S's optimal information policy under full credibility. Given prior $\mu$, an information policy $p \in \mathcal{R}(\mu)$ is a **show-or-best** (SOB) policy if it is supported on $\{\delta_\theta\}_{\theta\in\Theta} \cup \operatorname{argmax}_{\mu' \in \Delta[\text{supp}(\mu)]} v(\mu')$. In words, $p$ is an SOB policy if it either shows the state to R or brings R to a posterior that attains S's best feasible value. Say S is a **two-faced SOB** if, for every binary-support prior $\mu \in \Delta\Theta$, every $p \in \mathcal{R}(\mu)$ is outperformed by an SOB policy $p' \in \mathcal{R}(\mu)$; that is, $\int_{\Delta\Theta} v \, \mathrm{d}p \leq \int_{\Delta\Theta} v \, \mathrm{d}p'$. Figure 6 depicts an example in which S is a two-faced SOB. Note productive mistrust cannot occur in this example: One can show that if S's favorite equilibrium policy changes as credibility declines, no information must become S-optimal.[14] As such, R need not benefit from a less credible S.

Finally, say a model is **generic** if R is (i) not indifferent between any two actions at any degenerate belief, and (ii) not indifferent between any three actions at any binary-support belief.[15]

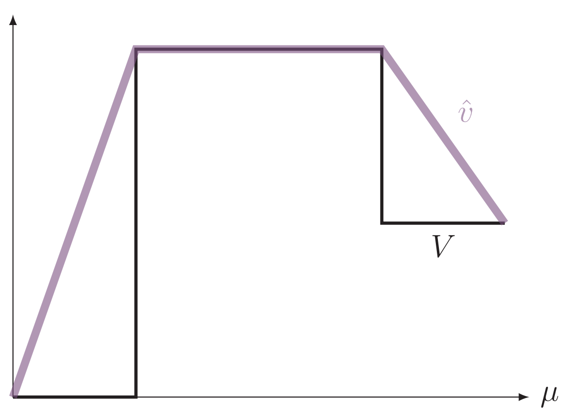


Figure 6: Sender is a two-faced SOB

Proposition 1 below shows that in generic finite settings, S not being a two-faced SOB is sufficient for productive mistrust to occur for some full-support priors at some

[14]For an explanation, observe the claim is obvious for priors that allow S to attain her first-best under no information. For other priors, a feasible $(k, \beta, \gamma)$ exists that improves on S's no-information payoff if and only if a feasible $(k, \beta, \gamma)$ exists giving S her full-credibility payoff.

[15]Given a fixed finite $A$ and $\Theta$, genericity holds for (Lebesgue) almost every $u_R \in \mathbb{R}^{A\times\Theta}$. In particular, it holds if $u_R(a,\theta) \neq u_R(a',\theta)$ for all distinct $a, a' \in A$ and all $\theta \in \Theta$, and $\frac{u_R(a_1,\theta_1)-u_R(a_2,\theta_1)}{u_R(a_1,\theta_2)-u_R(a_2,\theta_2)} \neq \frac{u_R(a_2,\theta_1)-u_R(a_3,\theta_1)}{u_R(a_2,\theta_2)-u_R(a_3,\theta_2)}$ for all distinct $a_1, a_2, a_3 \in A$ and all distinct $\theta_1, \theta_2 \in \Theta$.

credibility levels. Intuitively, S being an SOB means a highly credible S has no bad information to hide: under full credibility, S's bad messages are maximally informative, subject to keeping R's posterior fixed following S's good messages. S not being an SOB at some prior means her bad messages optimally hide some instrumental information. By reducing S's credibility just enough to make the full-credibility solution infeasible, one can push her to reveal some of that information to R. In other words, S commits to potentially revealing more extreme bad information in order to preserve the credibility of her good messages. Proposition 1 below formalizes this intuition.

**Proposition 1.** *Consider a finite and generic model in which S is not a two-faced SOB. Then, a full-support prior and credibility levels $\chi' < \chi$ exist such that every S-optimal $\chi'$-equilibrium is strictly better for R than every sender-optimal $\chi$-equilibrium.*[16]

The proposition builds on the binary-state case, extending to the general case via a continuity argument. We now sketch the binary-state argument. To follow the argument, consulting Figure 7, which depicts the relevant objects for the central bank example, is useful. Because the model is generic, $\bar{v}$ has a non-degenerate interval of maximizers (which correspond to beliefs in $[3/4, 1]$ in the figure). Fixing a prior near this interval but toward the nearest kink, we then find the lowest $\chi \in [0, 1]$ at which S still obtains her perfect credibility value. In the central-bank example, one can use any prior in $(1/4, 3/4)$. If we choose $\mu_0 = 1/2$, we take $\chi$ to be $3/4$, which is the lowest credibility level that delivers S's full-commitment payoff. At this $\chi$, S's favorite equilibrium information policy $p$ is unique and is supported on the beliefs $(\beta, \gamma)$ that solve Theorem 1's program (see $\gamma = 3/4$ and $\beta = 1/4$ in the figure). These beliefs are interior, and $\hat{v}$ has a kink at $\beta$. Although $\gamma$ remains optimal in Theorem 1's program for any additional small reduction in credibility, ($\chi$-BP) means one must replace $\beta$ with a new belief $\beta'$ that is further from the prior. Relying on the set of beliefs being one-dimensional, we show this new solution results in an information policy $p'$ that is strictly more informative than $p$. Intuitively, one can attain $p'$ from $p$ using two consecutive splittings, each of which involves an increase in informativeness: First, $\beta$ is split between $\gamma$ and $\beta'$, and then $\beta'$ is split between $\beta$ and another posterior (0 in the figure). This posterior lies even further from the prior than $\beta'$ does, and gives S a strictly lower continuation value than $\beta$. Hence, the additional information $p'$ provides to R over $p$ is instrumental, strictly increasing R's utility.

[16] Moreover, when $|\Theta| = 2$, every sender-optimal $\chi'$-equilibrium is more Blackwell-informative than every sender-optimal $\chi$-equilibrium.

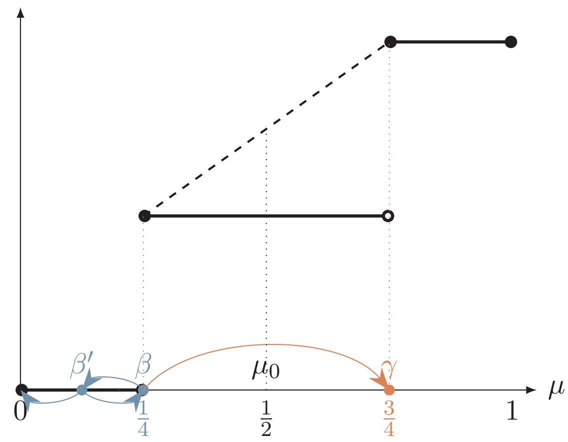


Figure 7: Productive mistrust in Example 1.

## 4.2 Collapse of Trust

Theorem 1 immediately implies lowering S's credibility can only decrease her value.[17] Below, we show this decrease is often discontinuous. In other words, small decreases in S's credibility often result in a large drop in S's benefits from communication.

**Proposition 2.** *In a finite model, the following are equivalent:*

*(i) A collapse of trust never occurs:*

$$\lim_{\chi' \nearrow \chi} v^*_{\chi'}(\mu_0) = v^*_\chi(\mu_0)$$

*for every $\chi \in [0,1]$ and every full-support prior $\mu_0$.*

*(ii) Commitment is of no value: $v_1^* = v_0^*$.*

*(iii) No conflict occurs: $v(\delta_\theta) = \max v(\Delta\Theta)$ for every $\theta \in \Theta$.*

Let us sketch Proposition 2's proof. To this end, notice two of the proposition's three implications are immediate. First, whenever no conflict occurs, S can reveal the state in an incentive-compatible way while obtaining her first-best payoff (given R's incentives), meaning commitment is of no value; that is, (iii) implies (ii). Second, because S's highest equilibrium value increases with her credibility, commitment having

[17]In Appendix A.1.3 we show credibility *increases* have a continuous payoff effect: a sufficiently small increase in S's credibility never results in a large gain in S's benefits from communication. Thus, S's value is an upper semicontinuous function of $\chi$. Proposition 2 implies lower semicontinuity is frequently violated.

no value means S's best equilibrium value is constant (and, a fortiori, continuous) in the credibility level; that is, (ii) implies (i).

To show (i) implies (iii), we show that any failure of (iii) implies the failure of (i). To do so, we fix a full-support prior $\mu_0$ at which $\bar{v}$ is minimized. Because conflict occurs, $\bar{v}$ is nonconstant and thus takes values strictly greater than $\bar{v}(\mu_0)$. By Theorem 1, one has that $v^*_\chi(\mu_0) > \bar{v}(\mu_0)$ if and only if a feasible triplet $(\beta, \gamma, k)$ with $k < 1$ exists such that $\bar{v}(\gamma) > \bar{v}(\mu_0)$. Using upper semicontinuity of $\bar{v}$, we show such a triplet is feasible for credibility $\chi$ if and only if $\chi$ is weakly greater than some strictly positive $\chi^*$. We thus have

$$v^*_{\chi^*}(\mu_0) \geq k\bar{v}(\mu_0) + (1-k)\bar{v}(\gamma) > \bar{v}(\mu_0) = \max_{\chi \in [0,\chi^*)} v^*_\chi(\mu_0),$$

where the first inequality follows from $\mu_0$ minimizing $\bar{v}$; that is, a collapse of trust occurs.

## 4.3 Robustness of the Commitment Case

Given the large and growing literature on optimal persuasion with commitment, one may wonder whether the commitment solution is robust to small decreases in S's credibility. Proposition 3 shows the answer is almost always.

**Proposition 3.** *In a finite model, the following are equivalent:*

*(i) The full-commitment value is robust:* $\lim_{\chi \nearrow 1} v^*_\chi(\mu_0) = v^*_1(\mu_0)$ *for every full-support* $\mu_0$.

*(ii) S gets the benefit of the doubt: Every* $\theta \in \Theta$ *is in the support of some member of* $\operatorname{argmax}_{\mu \in \Delta\Theta} v(\mu)$.

Thus, the proposition shows S's full-credibility value is robust if and only if S can persuade R to take her favorite action without ruling out any states. A sufficient condition for the latter is that R is willing to take S's preferred undominated action at some full-support belief, a property that holds generically whenever the model is finite.[18] Thus, although small decreases in credibility often lead to a collapse in S's value, these collapses rarely occur at $\chi = 1$.

The argument behind Proposition 3 establishes a four-way equivalence between

[18]More precisely, Proposition 3 implies S's full-credibility value is robust whenever an S-best action among those not strictly dominated for R is a best reply for some full-support belief. It follows from Lemma 1 in Lipnowski, Ravid, and Shishkin (2021) that this property holds for Lebesgue-almost every preference specification for fixed finite $A$ and $\Theta$.

(a) S getting the benefit of the doubt,

(b) $\bar{v}$ being maximized by a full-support prior $\gamma$,

(c) a full-support $\gamma$ existing such that $\hat{v}_{\wedge\gamma}$ and $\hat{v}$ agree over all full-support prior(s), and

(d) robustness to limited credibility.

To see (a) implies (b), notice that whenever S gets the benefit of the doubt, one can find a full-support prior in the convex hull of the beliefs in which R is willing to give S her first-best action. Splitting this prior across those beliefs gives an information policy that delivers S her highest feasible payoff for all posteriors, meaning S can attain this payoff using cheap talk. For the converse direction, one can use the fact that $\max \bar{v}(\Delta\Theta) = \max v(\Delta\Theta)$. Specifically, this fact implies $\bar{v}$ is maximized at a full-support prior $\gamma$ if and only if one can split $\gamma$ in a way that attains $v$'s maximal value at all posteriors, because $\bar{v}$ gives S's highest cheap-talk payoff for every prior. S getting the benefit of the doubt then follows from $\gamma$ having full support.

For the equivalence of (b) and (c), note that in finite models, $\hat{v}$ and $\hat{v}_{\wedge\gamma}$ are both continuous. Therefore, the two functions agree over all full-support priors if and only if they are equal, which is equivalent to the cap on $v_{\wedge\gamma}$ being non-binding; that is, $\gamma$ maximizes $\bar{v}$.

To see why (c) is equivalent to (d), fix some full-support $\mu_0$, and consider two questions about Theorem 1's program. First, which beliefs can serve as $\gamma$ for $\chi < 1$ large enough? Second, how do the optimal $(k, \beta)$ for a given $\gamma$ change as $\chi$ goes to 1? For the first question, the answer is that $\gamma$ is feasible for some $\chi < 1$ if and only if $\gamma$ has full support.[19] For the second question, one can show it is always optimal to choose $(k, \beta)$ so as to make ($\chi$-BP) bind while still satisfying (R-BP).[20] Direct computation reveals that as $\chi$ goes to 1, every such $(k, \beta)$ must converge to $(1, \mu_0)$. Combined,

[19]Suppose the model is finite. It is easy to see that every full-support $\gamma$ admits some $\beta$ and $k < 1$ that make (R-BP) hold. Moreover, ($\chi$-BP) is also satisfied at $(\beta, \gamma, k)$ for all sufficiently high $\chi$, because ($\chi$-BP)'s right-hand side converges to zero as $\chi \to 1$. Conversely, observe that if $\gamma(\theta) = 0$, ($\chi$-BP) is violated at $\theta$ for all $\chi < 1$, because $\mu_0$ has full support.

[20]To see why, for any feasible $(k, \beta, \gamma)$, a $(k', \beta')$ exists such that $(k', \beta', \gamma)$ is feasible, ($\chi$-BP) binds, and $k' \geq k$. By (R-BP), $\beta' = \frac{k}{k'}\beta + \left(1 - \frac{k}{k'}\right)\gamma$. Because $\hat{v}_{\wedge\gamma}$ is concave and $\hat{v}_{\wedge\gamma}(\gamma) = \bar{v}(\gamma)$,

$$\begin{aligned} k'\hat{v}_{\wedge\gamma}(\beta') + (1-k')\bar{v}(\gamma) &= k'\hat{v}_{\wedge\gamma}\left(\tfrac{k}{k'}\beta + \left(1 - \tfrac{k}{k'}\right)\gamma\right) + (1-k')\,\bar{v}(\gamma) \\ &\geq k\hat{v}_{\wedge\gamma}(\beta) + (k'-k)\,\hat{v}_{\wedge\gamma}(\gamma) + (1-k')\,\bar{v}(\gamma) = k\hat{v}_{\wedge\gamma}(\beta) + (1-k)\,\bar{v}(\gamma). \end{aligned}$$

one obtains that as $\chi$ increases, S's optimal value converges to $\max_{\gamma \in \text{int}(\Delta\Theta)} \hat{v}_{\wedge\gamma}(\mu_0)$. Thus, S's value is robust to limited credibility if and only if some full-support $\gamma$ exists for which $\hat{v}_{\wedge\gamma} = \hat{v}$ for all full-support priors; that is, (c) is equivalent to (d). The proposition follows.

# 5 Persuading the Public

This section uses our results to analyze a situation in which a single sender is interested in persuading a population of receivers to take a favorable action. For example, S could be a government of a small open economy trying to encourage foreigners to invest in the local market, a seller advertising to entice consumers to buy her product, or a leader vying for the support of her populace. To persuade the receivers, S commissions a weak institution (e.g., a central bank, product reviewer company, or state-owned media outlet) to issue a public report. In this section, we analyze the S-optimal report under partial credibility.

We modify our model as follows. The report of S's institution is now publicly revealed to a unit mass of receivers. After observing the institution's report, receivers simultaneously take a binary action. Each receiver $i$ cares only about his own action, $a_i \in A = \{0, 1\}$. Receiver $i$'s payoff from $a_i$ is given by $a_i(\theta - \omega_i)$, where $\theta \in \Theta = [0, 1]$ is the unknown state, distributed according to an atomless, full-support prior $\mu_0$, and $\omega_i \in \mathbb{R}$ is receiver $i$'s type. The mass of receivers whose type is below $\omega$ is given by $H(\omega)$, an absolutely continuous cumulative distribution function whose density $h$ is continuous, strictly quasiconcave, and strictly positive on $(0, 1)$. S's objective is to maximize the proportion of receivers taking action 1.

An equilibrium of the modified game is tuple, $(\xi, \sigma, \alpha, \pi)$, where $\xi\colon \Theta \to \Delta M$, $\sigma\colon \Theta \to \Delta M$, and $\pi\colon M \to \Delta\Theta$ respectively represent S's official report, S's strategy when not committed, and the public's belief mapping, as in the original game. We let $\alpha\colon M \to [0, 1]$ represent the proportion of receivers taking action 1 conditional on the realized message. Observe action 1 is optimal for receiver $i$ if and only if $\omega_i \leq E\mu$, where $\mu \in \Delta\Theta$ is the publicly held posterior about $\theta$, and $E$ maps beliefs to their associated expectations.[21] As such, given a posterior $\mu$, the proportion of receivers taking action 1 is given by $H(E\mu)$. Thus, a **$\chi$-equilibrium** is a tuple $(\xi, \sigma, \alpha, \pi)$ in which $\pi$ is derived from $\mu_0$ via Bayes' rule, $\alpha(\cdot) = H(E\pi(\cdot))$, and $\sigma(\theta)$ is supported on

[21]That is, $E\mu \coloneqq \int \theta \, \mathrm{d}\mu(\theta)$ for all $\mu \in \Delta\Theta$.

$\arg\max_{m \in M} \alpha(m)$ for all $\theta$.

Theorem 1 applies readily to the current setting. Because $H(E\mu)$ is the proportion of the population taking action 1 given posterior $\mu \in \Delta\Theta$, S's continuation payoff from a public message inducing $\mu$ is $v(\mu) := H(E\mu)$. Taking $v$ to be S's value function, we can directly apply Theorem 1 to the current game.

Next, we use Theorem 1 to find S's optimal $\chi$-equilibrium. We begin with the extreme credibility levels. Suppose first S has no credibility; that is, $\chi = 0$. In this case, S's optimal value is given by the quasiconcave envelope of S's value function evaluated at the prior, $\bar{v}(\mu_0)$. Because an increasing transformation of an affine function is quasiconcave, $v = H \circ E = \bar{v}$. Hence, with no credibility, S cannot benefit from communication.

Suppose now that S has full credibility; that is, $\chi = 1$. In this case, S's maximal $\chi$-equilibrium value equals $v$'s maximal expected value across all information policies, $p \in \mathcal{R}(\mu_0)$. Notice that a given information policy $p$ yields an expected value of $\int H(\cdot)\,\mathrm{d}\mu$, where $\mu = p \circ E^{-1} \in \Delta\Theta$ is the distribution of the population's posterior mean. As such, maximizing S's value across all information policies is the same as maximizing the expectation of $H(\cdot)$ across all posterior mean distributions produced by some information policy. Such posterior mean distributions are characterized via the notion of mean-preserving spreads.[22] Formally, we say $\mu \in \Delta\Theta$ is a **mean-preserving spread** of $\tilde{\mu} \in \Delta\Theta$, denoted by $\mu \succeq \tilde{\mu}$, if

$$\int_0^{\hat{\theta}} \mu[0,\theta]\,\mathrm{d}\theta \geq \int_0^{\hat{\theta}} \tilde{\mu}[0,\theta]\,\mathrm{d}\theta,\ \forall \hat{\theta} \in [0,1],\ \text{with equality at } \hat{\theta} = 1. \tag{MPS}$$

As noted by the literature, in this setting, mean-preserving spreads are synonymous with more information. To be more precise, say that an information policy $p$ is more informative than another information policy $p'$ if every decision-maker prefers $p$ to $p'$. Then, $p$ being more informative than $p'$ implies the posterior-mean distribution induced by $p$ is a mean-preserving-spread of the one induced by $p'$. Moreover, if $\mu \succeq \mu'$, one can find two information policies, $p$, $p'$, with the property that $p$ is more informative than $p'$, and such that $p$ and $p'$ induce $\mu$ and $\mu'$, respectively. Because $\mu_0$ is the posterior-mean distribution induced by full information, it follows $\mu$ can arise as the posterior mean distribution of some information policy if and only if $\mu_0 \succeq \mu$. Therefore, S's

[22] See Blackwell and Girshick (1979) and Rothschild and Stiglitz (1970).

value under full credibility is given by

$$\hat{v}(\mu_0) = \max_{\mu \in \Delta\Theta:\ \mu \preceq \mu_0} \int H(\cdot)\,\mathrm{d}\mu.$$

The solution to the above program is dictated by the shape of the CDF $H$. Because the CDF's density, $h$, is strictly quasiconcave, $H$ is a convex-concave function over $[0,1]$. Said differently, some $\omega^* \in [0,1]$ exists such that $H$ is strictly convex on $[0,\omega^*]$ and strictly concave on $[\omega^*,1]$. As noted by Kolotilin (2018) and Dworczak and Martini (2019), when $H$ is convex-concave, the above program can be solved via $\theta^*$ upper censorship, which we now formally define.[23] Under full credibility, $\theta^*$ upper censorship arises whenever S's official report reveals (pools) all states below (above) $\theta^*$. Given such an official reporting protocol, it is optimal for S to say the state is above $\theta^*$ whenever she influences the report. Thus, we say $(\xi,\sigma)$ is a **$\theta^*$-upper-censorship pair** if every $\theta \in \Theta$ has $\sigma(\cdot|\theta) = \delta_1$ and[24]

$$\xi(\cdot|\theta) = \begin{cases} \delta_\theta & \text{if } \theta \in [0,\theta^*), \\ \delta_1 & \text{if } \theta \in [\theta^*,1]. \end{cases}$$

We refer to the resulting distribution of posterior means as a $\theta^*$ upper censorship of $\mu_0$. One can describe this distribution formally by introducing some notation. For a bounded measurable $f:\Theta \to \mathbb{R}_+$ and $\mu \in \Delta\Theta$, define the measure $f\mu$ on $\Theta$ via $f\mu(\hat{\Theta}) \coloneqq \int_{\hat{\Theta}} f\,\mathrm{d}\mu$ for each Borel $\hat{\Theta} \subseteq \Theta$. Then, the $\theta^*$ **upper censorship** of $\mu_0$ is given by

$$\mathbf{1}_{[0,\theta^*)}\mu_0 + \mu_0[\theta^*,1]\delta_{\mathbb{E}_{\mu_0}[\theta|\theta\geq\theta^*]}.$$

Observe this distribution coincides with $\mu_0$ below the cutoff. Above the cutoff, the distribution has a single atom at the mean of $\mu_0$'s conditional on $\theta$ being above $\theta^*$. At the optimum, $\theta^*$ is chosen so that the atom lies in the concave region of $H$.

We find upper-censorship pairs are also optimal when credibility is partial, although the reasoning is more delicate. One complication is that not every upper-censorship pair induces a $\chi$-equilibrium. The reason is that under partial credibility, the posterior

[23]That upper censorship solves the full-credibility problem has been discussed by the aforementioned papers under slightly different assumptions. Still, we provide an elementary proof in the appendix for completeness.

[24]Observe our description $\theta^*$-upper-censorship pair breaks from the common convention used in the literature (e.g., Kolotilin, 2018; Dworczak and Martini, 2019) of identifying messages with their induced posterior mean. In particular, R's posterior mean conditional on $m=1$ need not equal 1.

mean following message 1 can be strictly below the posterior mean induced by other messages, thereby violating S's incentive constraints. To avoid such a violation, the mean induced by message 1 must be above the upper-censorship cutoff, $\theta^*$, which holds if and only if[25]

$$\int(\theta-\theta^*)(1-\mathbf{1}_{[0,\theta^*)}\chi)\mathrm{d}\mu_0(\theta)\geq 0. \tag{$\theta^*$-IC}$$

Observe that with intermediate credibility, the left-hand side of ($\theta^*$-IC) is continuous and strictly decreasing in $\theta^*$, strictly positive for $\theta^*=0$, and strictly negative for $\theta^*=1$.[26] As such, ($\theta^*$-IC) holds whenever $\theta^*$ is below the unique upper-censorship cutoff at which it holds with equality, a cutoff that we denote by $\bar{\theta}_\chi$.

Another complication arising from partial credibility is that a $\theta^*$-upper-censorship pair does not typically yield an upper censorship of $\mu_0$ as its posterior mean distribution. Instead, every $\theta^*$-upper-censorship pair with $\theta^*\leq\bar{\theta}_\chi$ turns out to yield a $\theta^*$ upper censorship of

$$\bar{\mu}_\chi:=\mathbf{1}_{[0,\bar{\theta}_\chi)}\chi\mu_0+\left(1-\chi\mu_0[0,\bar{\theta}_\chi)\right)\delta_{\bar{\theta}_\chi},$$

which is the posterior mean distribution induced by the $\bar{\theta}_\chi$-upper-censorship pair.[27]

Claim 1 below shows that upper censorship always yields an S-optimal $\chi$-equilibrium. Moreover, to find the optimal censorship cutoff, one can solve the full-credibility problem with the modified prior $\bar{\mu}_\chi$. As such, in this setting, partial credibility can be seen as bounding the amount of information S can provide in equilibrium.

**Claim 1.** *Some $\theta^*\in[0,\bar{\theta}_\chi]$ exists such that the $\theta^*$ upper censorship of $\bar{\mu}_\chi$, denoted by $\mu_{\chi,\theta^*}$, satisfies*

$$v_\chi^*(\mu_0)=\hat{v}(\bar{\mu}_\chi)=\int H(\cdot)\,\mathrm{d}\mu_{\chi,\theta^*}.$$

*Moreover, the corresponding $\theta^*$-upper-censorship pair is an S-optimal $\chi$-equilibrium that induces $\mu_{\chi,\theta^*}$ as its posterior mean distribution.*

The intuition for why $v_\chi^*(\mu_0)\geq\hat{v}(\bar{\mu}_\chi)$ is straightforward. Recall $\hat{v}(\bar{\mu}_\chi)=\int H\,\mathrm{d}\mu_{\chi,\theta^*}$, where $\mu_{\chi,\theta^*}$ is a $\theta^*$ upper censorship of $\bar{\mu}_\chi$ for some $\theta^*\in[0,1]$. Because $\bar{\mu}_\chi$'s support is in $[0,\bar{\theta}_\chi]$, any $\theta$ upper censorship of $\bar{\mu}_\chi$ for a $\theta$ above $\bar{\theta}_\chi$ is just $\bar{\mu}_\chi$ itself. Thus,

---

[25]To see this equivalence, note R's posterior mean conditional on seeing message 1 from a $\theta^*$-upper-censorship pair equals $\frac{\int\theta\left[\mathbf{1}_{[\theta^*,1]}\chi+1-\chi\right]\mathrm{d}\mu_0}{\int[\mathbf{1}_{[\theta^*,1]}\chi+1-\chi]\mathrm{d}\mu_0}=\frac{\int\theta[1-\mathbf{1}_{[0,\theta^*)}\chi]\mathrm{d}\mu_0}{\int[1-\mathbf{1}_{[0,\theta^*)}\chi]\mathrm{d}\mu_0}$, which is larger than $\theta^*$ only if ($\theta^*$-IC) holds.

[26]Recall $\mu_0$ is assumed to be an atomless, full-support distribution over $[0,1]$.

[27]That $\bar{\mu}_\chi$ is induced by a $\bar{\theta}_\chi$-upper-censorship pair follows from the observation that this pair leads R to have a posterior mean of $\bar{\theta}_\chi$ after seeing $m=1$.

assuming $\theta^*$ is in $[0, \bar{\theta}_\chi]$ is without loss. Given such a $\theta^*$, one can induce the posterior mean distribution $\mu_{\chi,\theta^*}$ in a $\chi$-equilibrium (with the original prior $\mu_0$) using a $\theta^*$-upper-censorship pair. As such, S's maximal $\chi$-equilibrium value is at least as high as the value generated by $\mu_{\chi,\theta^*}$; that is, $v^*_\chi(\mu_0) \geq \int H \, \mathrm{d}\mu_{\chi,\theta^*} = \hat{v}(\bar{\mu}_\chi)$.

To show $v^*_\chi(\mu_0) \leq \hat{v}(\bar{\mu}_\chi)$, we use Theorem 1 to find an S-favorite $\chi$-nonical equilibrium that communicates less information than some $\theta^*$-upper-censorship pair for some $\theta^* \leq \bar{\theta}_\chi$. Because more informative policies induce more spread-out distributions, it follows that this $\chi$-nonical equilibrium induces a posterior mean distribution that is a mean-preserving contraction of the distribution induced by the said $\theta^*$-upper-censorship pair, which in turn induces a mean-preserving contraction of $\bar{\mu}_\chi$. Because the mean-preserving-spread relation is transitive, this S-favorite $\chi$-nonical equilibrium must induce a posterior mean distribution that is feasible under full commitment when the prior is $\bar{\mu}_\chi$, and so $v^*_\chi(\mu_0) \leq v^*_1(\bar{\mu}_\chi) = \hat{v}(\bar{\mu}_\chi)$, as required.

# 6 Extensions and Discussion

## 6.1 Robustness and Equilibrium Selection

Proposition 3 shows S receiving the benefit of the doubt is necessary and sufficient for robustness of S's full-commitment value to small decreases in credibility, assuming S can coordinate R toward the S-favorite $\chi$-equilibrium. One might argue, however, that imperfect credibility compromises S's status as a principal and, through it, her ability to choose which equilibrium is played. In Appendix B.2, we ask when S's commitment value is **strongly robust**, that is, when S's almost-perfect credibility payoff equals the best she can do with full commitment, regardless of which equilibrium is selected under imperfect credibility. Under a finiteness condition, we show such robustness holds if and only if S's coordination ability is immaterial when credibility is perfect. Said differently, S's best commitment value is strongly robust if and only if S has only one equilibrium payoff when $\chi = 1$. Lipnowski, Ravid, and Shishkin (2021) provide necessary and sufficient conditions for the latter condition to hold, and show these conditions hold for Lebesgue-almost all instances of the finite model. Thus, S's full-commitment value typically remains robust even if she cannot dictate which equilibrium the players are using.

## 6.2 Signaling Credibility

Our baseline model assumes S announces her official report before knowing whether the announcement is credible. In practice, S may be privy to institutional features that affect her chances of influencing the report before she commissions it. To understand such situations, Appendix B.3 considers a modified model in which S learns her credibility type before announcing the official reporting protocol.

By letting S commission a different official report based on her credibility, the modified model allows S to signal whether she can influence the report's message. However, such signaling turns out to have no impact on S's attainable payoffs. More precisely, every interim S-payoff profile (i.e., every pair specifying S's payoffs conditional on each credibility type) is attainable in a pooling equilibrium in which both credibility types choose the same official experiment. It follows that pooling equilibria are without loss as far as S payoffs are concerned.

Appendix B.3 also shows an S-payoff profile is attainable in a pooling equilibrium if and only if it is attainable in a $\chi$-equilibrium. That every pooling equilibrium payoff profile is attainable in a $\chi$-equilibrium follows from definition: a pooling equilibrium of the modified game requires the same conditions as a $\chi$-equilibrium, except S must also be willing to announce the equilibrium experiment conditional on her credibility type. For the converse direction, we show every $\chi$-equilibrium can be implemented as a pooling equilibrium of the signaling game by appropriately constructing R's behavior off path.

To summarize, the Appendix establishes a three-way equivalence between S's payoffs in all equilibria of the signaling game, all pooling equilibria of the signaling game, and $\chi$-equilibria of the original game. It follows that informing S of her ability to influence the report before its announcement has no impact on S's achievable payoffs.

## 6.3 State-Dependent Credibility

Throughout the paper, we assumed S's credibility is independent of the state of the world. However, in many applications, it is natural for S's credibility to be correlated with the state. For example, an autocrat may be more likely to influence the media in a rich economy with abundant resources than in a country where resources are scarce (e.g., Egorov, Guriev, and Sonin, 2009). To capture such correlation, suppose the probability of S's official report when the state is $\theta$ is given by $\boldsymbol{\chi}(\theta)$. Whereas most of the paper assumed the **credibility function**, $\boldsymbol{\chi} : \Theta \to [0, 1]$ is constant, here we discuss the **state-dependent credibility** (SDC) model, which imposes no restrictions on $\boldsymbol{\chi}$.[28] We proceed informally, relegating the mathematical details to Appendix A.

Theorem 1 generalizes to the SDC model with a minor modification. To present this modification, let us recall some convenient notation: given a bounded and measurable $f\colon \Theta \to \mathbb{R}$ and $\mu \in \Delta\Theta$, let $f\mu$ denote the measure on $\Theta$ given by $f\mu(\hat{\Theta}) \coloneqq \int_{\hat{\Theta}} f \, \mathrm{d}\mu$. Appendix A shows an S-optimal $\boldsymbol{\chi}$-equilibrium exists and yields S a utility of[29]

$$
\begin{aligned}
v^*_{\boldsymbol{\chi}}(\mu_0) = \max_{\beta,\gamma\in\Delta\Theta,\ k\in[0,1]} \quad & k\hat{v}_{\wedge\gamma}(\beta) + (1-k)\bar{v}(\gamma) && (12)\\
\text{s.t.} \quad & k\beta + (1-k)\gamma = \mu_0, \\
& (1-k)\gamma \geq (\mathbf{1}-\boldsymbol{\chi})\mu_0. && (\boldsymbol{\chi}\text{-BP})
\end{aligned}
$$

With the above characterization in hand, the propositions of section 4 extend to the SDC model in a straightforward manner; see the Appendix for precise statements.

The SDC model allows us to analyze the value of credibility at different states for specific examples. To illustrate, consider the public-persuasion example of section 5. Recall that under state-independent credibility—that is, $\boldsymbol{\chi}(\theta) = \chi$ for all $\theta$—S uses a $\theta^*$-upper-censorship pair in her favorite equilibrium, where $\theta^*$ is chosen to solve the full-credibility problem with a modified prior. The same result remains true for the SDC model. More precisely, let $\bar{\theta}_{\boldsymbol{\chi}}$ be the highest $\theta^*$ such that $\theta^*$-upper censorship is a $\boldsymbol{\chi}$-equilibrium; that is, $\bar{\theta}_{\boldsymbol{\chi}}$ is the unique solution to

$$\int (\theta - \bar{\theta}_{\boldsymbol{\chi}})(1 - \mathbf{1}_{[0,\bar{\theta}_{\boldsymbol{\chi}})}) \, \mathrm{d}\mu_0(\theta) = 0.$$

In addition, take $\bar{\mu}_{\boldsymbol{\chi}}$ to be the posterior-mean distribution that arises in the $\boldsymbol{\chi}$-equilibrium in which S uses the $\bar{\theta}_{\boldsymbol{\chi}}$-upper-censorship pair,

$$\bar{\mu}_{\boldsymbol{\chi}} = \mathbf{1}_{[0,\bar{\theta}_{\boldsymbol{\chi}})}\boldsymbol{\chi}\mu_0 + (1 - \boldsymbol{\chi}\mu_0[0,\bar{\theta}_{\boldsymbol{\chi}}))\delta_{\bar{\theta}_{\boldsymbol{\chi}}}.$$

Then, Claim 1 generalizes as is but with $\bar{\theta}_{\boldsymbol{\chi}}$ and $\bar{\mu}_{\boldsymbol{\chi}}$, respectively, replacing $\bar{\theta}_{\chi}$ and $\bar{\mu}_{\chi}$. In other words, S's maximal $\boldsymbol{\chi}$-equilibrium value is given by

$$v^*_{\boldsymbol{\chi}}(\mu_0) = \max_{\mu\in\Delta\Theta:\mu\preceq\bar{\mu}_{\boldsymbol{\chi}}} \int H(\cdot) \, \mathrm{d}\mu.$$

[28] As with all functions in this paper, we require $\boldsymbol{\chi}$ to be Borel measurable.

[29] Let $\mathbf{1}$ and $\mathbf{0}$ denote constant functions taking value 1 and 0, respectively, when the domain is not ambiguous.

Armed with this generalization, we can see S is less constrained by her credibility whenever $\bar{\mu}_{\boldsymbol{\chi}}$ increases in a mean-preserving-spread sense. Said differently, whenever $\bar{\mu}_{\boldsymbol{\chi}} \succeq \bar{\mu}_{\tilde{\boldsymbol{\chi}}}$, S prefers $\boldsymbol{\chi}$ to $\tilde{\boldsymbol{\chi}}$ regardless of the population's type distribution. Using a constructive argument, one can show the converse is also true: S prefers $\boldsymbol{\chi}$ to $\tilde{\boldsymbol{\chi}}$ for all population-type distributions *only if* $\bar{\mu}_{\tilde{\boldsymbol{\chi}}}$ is a mean-preserving spread of $\bar{\mu}_{\boldsymbol{\chi}}$. We present this result in Claim 2 below.

**Claim 2.** *In the state-dependent-credibility version of the public-persuasion setting, the following are equivalent:*

*(i) S prefers with* $\boldsymbol{\chi}$ *over* $\tilde{\boldsymbol{\chi}}$ *for all type distributions.*[30]

*(ii) The distribution* $\bar{\mu}_{\boldsymbol{\chi}}$ *is a mean-preserving spread of* $\bar{\mu}_{\tilde{\boldsymbol{\chi}}}$.

*(iii) For all* $\theta \in [0, \bar{\theta}_{\tilde{\boldsymbol{\chi}}}]$: $\int_0^{\hat{\theta}} \int_0^{\theta} (\boldsymbol{\chi} - \tilde{\boldsymbol{\chi}})\, \mathrm{d}\mu_0\, \mathrm{d}\theta \geq 0$.

The economic intuition behind the claim is that credibility is most valuable when the conflict between S's ex-ante and ex-post incentives is large. This intuition is expressed most clearly in the claim's part (iii), which says S prefers credibility to be concentrated in low states. Broadly speaking, low states are those that S benefits from revealing ex ante but would like to hide ex post. The more credibility S has in those states, the less S's ex-post incentives interfere with his ex-ante payoffs, and so the higher S's value is.

## 6.4 Investing in Credibility

Our analysis so far has taken the credibility of S's institutions to be exogenously given. Whereas this assumption seems reasonable for understanding short-run behavior, in the long run, S may have the ability redesign her institutions. To accommodate such situations, this section extends the state-dependent-credibility model of section 6.3 to endogenize S's credibility function $\boldsymbol{\chi}$. Specifically, suppose S can choose any measurable $\boldsymbol{\chi}\colon \Theta \to [0,1]$ at a cost of $c\left(\int \boldsymbol{\chi}\, \mathrm{d}\mu_0\right)$ prior to the persuasion game, where $c\colon [0,1] \to \mathbb{R}_+$ is continuous and strictly increasing.[31] Then, S chooses $\boldsymbol{\chi}$ to solve

$$v_c^{**}(\mu_0) = \max_{\boldsymbol{\chi}} \left[ v_{\boldsymbol{\chi}}^*(\mu_0) - c\left(\int \boldsymbol{\chi}\, \mathrm{d}\mu_0\right)\right].$$

[30] That is, $v_{\boldsymbol{\chi}}^*(\mu_0) \geq v_{\tilde{\boldsymbol{\chi}}}^*(\mu_0)$ for all $H$ admitting a continuous, quasiconcave density.

[31] The substantive results reported below would remain if we assigned each $\boldsymbol{\chi}$ a cost of $C(\boldsymbol{\chi})$ for some $||\cdot||_\infty$-continuous $C$ with the property that $C(\boldsymbol{\chi}) > C(\boldsymbol{\chi}')$ whenever $\boldsymbol{\chi} \geq \boldsymbol{\chi}'$ and $\boldsymbol{\chi}$ is not $\mu_0$-almost surely equal to $\boldsymbol{\chi}'$.

Clearly, S never invests in greater credibility than is necessary to induce her equilibrium information. As such, S always chooses $(\boldsymbol{\chi}, k, \beta, \gamma)$ so that ($\boldsymbol{\chi}$-BP) holds with equality. Combining this observation with (R-BP) yields

$$\int \boldsymbol{\chi} \, \mathrm{d}\mu_0 = k\beta(\Theta) = k.$$

S's problem therefore reduces to

$$\begin{aligned} v_c^{**}(\mu_0) = \max_{\beta,\gamma \in \Delta\Theta,\ k \in [0,1]} \quad & k\hat{v}_{\wedge\gamma}(\beta) + (1-k)\bar{v}(\gamma) - c(k) \\ \text{s.t.} \quad & k\beta + (1-k)\gamma = \mu_0. \end{aligned}$$

We now discuss how our results change when credibility is endogenized as above. We begin by revisiting productive mistrust. Similar to R's ability to benefit from a decrease in exogenous credibility, R can also benefit from an increase in S's credibility costs. Recall our introductory example and suppose the cost function is given by $c(k) = \frac{\lambda}{2}k^2$ for some $\lambda > 0$. For any $\lambda \in [2, 3)$, one can verify S's optimal investment choice is unique and leads to the following equilibrium information policy:

$$p_\lambda^* = \left[1 - \left(\frac{6}{\lambda} - 2\right)\right] \left(\frac{1}{3}\delta_0 + \frac{2}{3}\delta_{\frac{3}{4}}\right) + \left[\frac{6}{\lambda} - 2\right] \left(\frac{1}{2}\delta_{\frac{1}{4}} + \frac{1}{2}\delta_{\frac{3}{4}}\right).$$

Direct computation reveals R's corresponding equilibrium payoff is $\frac{1}{4} - \frac{1}{4\lambda}$, which increases with $\lambda$.[32]

Whereas reducing $\chi$ in our main model often leads to a discontinuous drop in S's payoff (Proposition 2), a uniformly small increase in $c$ cannot. The reason is that the set of feasible $(\beta, \gamma, k)$ in Theorem 1's program is independent of the cost, and the cost enters S's objective separably. Therefore,

$$|v_c^{**}(\mu_0) - v_{\tilde{c}}^{**}(\mu_0)| \leq ||c - \tilde{c}||_\infty.$$

Thus, in the endogenous-credibility model, small cost changes have small effects on S's value.

We now note that making credibility endogenous leads to simple institutional structures in the public-persuasion setting of section 5. To describe these structures, say $\boldsymbol{\chi}$ is a **cutoff credibility function** if a $\theta^* \in [0, 1]$ exists for which $\boldsymbol{\chi}(\theta) = 1$ whenever

[32] One can also verify that $p_\lambda^*$ is Blackwell-increasing in $\lambda$.

$\theta < \theta^*$, and $\boldsymbol{\chi}(\theta) = 0$ otherwise; that is, $\boldsymbol{\chi} = \mathbf{1}_{[0,\theta_*)}$. The following proposition shows it is optimal to choose cutoff credibility, that is, to invest in perfect credibility in low states and none in high states.

**Claim 3.** *An optimal credibility choice exists in the endogenous credibility version of the public-persuasion setting. Moreover, any optimal choice (along with S-optimal equilibrium) in this setting is a cutoff credibility function and entails full revelation by the official reporting protocol.*

The claim is based on three observations. The first observation is that S prefers to concentrate her credibility in low states, which implies S can improve upon any non-cutoff $\boldsymbol{\chi}$ with a less expensive cutoff credibility function $\mathbf{1}_{[0,\theta^*)}$. The second observation is that it is S-optimal to use a $\tilde{\theta}$-upper-censorship pair for some $\tilde{\theta}$. Thus, S's official report reveals states below $\tilde{\theta}$, whereas states above $\tilde{\theta}$ are pooled into a single high message that S sends for sure when influencing the report. The third observation is that S never invests in extraneous credibility. It follows that setting $\theta^*$ above $\tilde{\theta}$ is suboptimal, because a $\tilde{\theta}$-upper-censorship pair treats states above $\tilde{\theta}$ in the same way under both influenced and official reporting. All that remains is to note that when $\boldsymbol{\chi} = \mathbf{1}_{[0,\theta^*)}$, the $\tilde{\theta}$-upper-censorship pair's official report always reveals the states on path.

## 6.5 Simple Communication

Whereas our main theorem simplifies the task of finding S's optimal $\chi$-equilibrium, it says nothing about the complexity of S's communication with R. The following proposition sheds light on this issue by providing restrictions on the total number of messages required for implementing S's favorite equilibrium.

**Proposition 4.** *Some S-optimal $\chi$-equilibrium exists with no more than $|A|$ distinct messages sent on path and no more than $2|\Theta| - 1$ distinct messages sent on path if $\Theta$ is finite.*[33]

The proposition's bounds are known for the benchmark cases of full credibility and no credibility. Kamenica and Gentzkow (2011) point out how, following the revelation principle (Myerson, 1986), any official reporting protocol can be converted into an outcome-equivalent direct-recommendation mechanism in which S recommends an

[33]The corollary's bounds also hold when credibility is state-dependent, that is, for $\boldsymbol{\chi}$-equilibria.

action to R and R always obeys. In the no-credibility case, analogously converting an equilibrium into one with direct recommendations would not yield an appropriate bound on the number of on-path messages, because S-optimal equilibrium sometimes requires R to mix between different best responses to maintain S incentives. Nevertheless, Lipnowski and Ravid (2020) establish a revelation principle in the style of Bester and Strausz (2001): any equilibrium S payoff of the no-credibility game can be attained with every on-path message being a *pure* action recommendation. Even though R may mix in response to a given pure action recommendation, the recommendation is always one of his best responses—in fact an S-preferred best response—to the belief it induces. Together, these two revelation principles show some S-optimal equilibrium entails no more than $|A|$ on-path messages if credibility is extreme.

Next, supposing $\Theta$ is finite, Kamenica and Gentzkow (2011) appeal to Carathéodory's theorem to show an S-optimal equilibrium exists in which the set of on-path beliefs is affinely independent. Because every affinely independent subset of $\Delta\Theta$ has cardinality no greater than $|\Theta|$, it follows that some S-optimal equilibrium entails no more than $|\Theta|$ on-path messages in the full-credibility game. A nearly identical appeal to Carathéodory's theorem shows the same result holds for the no-credibility case as well. Because $|\Theta| \leq 2|\Theta| - 1$, it follows that the message bounds in the statement of the proposition hold when credibility is extreme.

A naive application of the results for extreme $\chi$ to the case in which $\chi$ is intermediate delivers message bounds that are higher than those of Proposition 4. For an explanation, recall Theorem 1 transforms a $(\beta, \gamma, k)$ that solves the program (5) into an S-favorite $\chi$-equilibrium by pasting together an S-optimal equilibrium of $\mathbf{G}(0, \gamma)$ with an appropriately chosen S-optimal equilibrium of $\mathbf{G}(1, \beta)$. Hence, the maximal number of messages required for S's favorite $\chi$-equilibrium must be lower than the sum of the number used for the two component games. Thus, if one were to appeal to the above-mentioned results without further argument, one would obtain message bounds of $2|A|$ and (if $\Theta$ is finite) $2|\Theta|$, whereas Proposition 4 says these bounds can be reduced to $|A|$ and $2|\Theta| - 1$, respectively.

To obtain tighter bounds on the required number of messages, we show one can always choose $(\beta, \gamma, k)$ so that the full-credibility game with prior $\beta$, $\mathbf{G}(1, \beta)$, admits an S-favorite equilibrium $(\xi_\beta, \sigma_\beta, \alpha_\beta, \pi_\beta)$ with the property that every on-path message gives S an expected payoff *strictly* below $\bar{v}(\gamma)$.[34] Using this equilibrium, one can obtain the

[34]See Lemma 17. This lemma can also be useful in solving Theorem 1's program. For example, in the context of the central-bank example, the lemma directly implies setting $\beta \leq \frac{1}{4}$ is optimal whenever

above-mentioned message bounds as follows. For the action-based bound, start by using Kamenica and Gentzkow's (2011) revelation principle to implement $(\xi_\beta, \sigma_\beta, \alpha_\beta, \pi_\beta)$ using action recommendations, and pick any S-optimal direct-recommendation equilibrium of $\mathbf{G}(0, \gamma)$ as constructed by Lipnowski and Ravid (2020). Observe that the former equilibrium uses only actions with an S-payoff strictly below $\bar{v}(\gamma)$, whereas only actions with higher payoff are recommended by the latter. Therefore, the two equilibria use disjoint action sets for their recommendations, and so one can paste the two equilibria together without using more than $|A|$ messages.

To tighten the state-based message bound, suppose $\Theta$ is finite and observe it is enough to show some S-optimal equilibrium of $\mathbf{G}(0, \gamma)$ generates no more than $|\Theta| - 1$ posterior beliefs for R (and so requires no more than $|\Theta| - 1$ messages). Assume otherwise for a contradiction. Then, $\gamma$ is interior in the set of beliefs at which $\bar{v} \geq \bar{v}(\gamma)$, and so we can move $\gamma$ closer to the prior, while lowering $k$ to preserve (R-BP), and maintain (or raise) $\bar{v}(\gamma)$. Owing to the special structure of $(\beta, \gamma, k)$, we know $\hat{v}_{\wedge\gamma}(\beta) < \bar{v}(\gamma)$, so that this adjustment raises S's expected payoff. Moreover, because this adjustment makes $(1 - k)\gamma(\theta) = \mu_0(\theta) - k\beta(\theta)$ larger for every state $\theta$, it relaxes ($\boldsymbol{\chi}$-BP) and thus contradicts the optimality of $(\beta, \gamma, k)$. It follows that the S-optimal equilibria of $\mathbf{G}(0, \gamma)$ and $\mathbf{G}(1, \beta)$ require only $|\Theta| - 1$ and $|\Theta|$ messages, respectively. Pasting these equilibria together, one obtains an S-optimal $\boldsymbol{\chi}$-equilibrium with only $2|\Theta| - 1$ messages.

# References


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

———. 2018. "On the value of persuasion by experts." *Journal of Economic Theory* 174:103–123.

Aumann, Robert J and Sergiu Hart. 2003. "Long Cheap Talk." *Econometrica* 71 (6):1619–1660.

Aumann, Robert J and M Maschler. 1966. "Game theoretic aspects of gradual disarmament." *Report of the US Arms Control and Disarmament Agency* 80:1–55.


S can outperform her no-credibility payoff.

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

Ben-Porath, Elchanan, Eddie Dekel, and Barton L Lipman. 2019. "Mechanisms with evidence: Commitment and robustness." *Econometrica* 87 (2):529–566.

Benoît, Jean-Pierre and Juan Dubra. 2011. "Apparent Overconfidence." *Econometrica* 79 (5):1591–1625.

Best, James and Daniel Quigley. 2020. "Persuasion for the Long Run." Working paper.

Bester, Helmut and Roland Strausz. 2001. "Contracting with Imperfect Commitment and the Revelation Principle: The Single Agent Case." *Econometrica* 69 (4):1077–1098.

Blackwell, David. 1953. "Equivalent comparisons of experiments." *The Annals of Mathematical Statistics* 24 (2):265–272.

Blackwell, David A and Meyer A Girshick. 1979. *Theory of games and statistical decisions*. Courier Corporation.

Blume, Andreas, Oliver J Board, and Kohei Kawamura. 2007. "Noisy talk." *Theoretical Economics* 2 (4):395–440.

Chakraborty, Archishman and Rick Harbaugh. 2010. "Persuasion by cheap talk." *American Economic Review* 100 (5):2361–2382.

Crawford, Vincent P and Joel Sobel. 1982. "Strategic information transmission." *Econometrica* 50 (6):1431–1451.

de Clippel, Geoffroy. 2008. "An axiomatization of the inner core using appropriate reduced games." *Journal of Mathematical Economics* 44 (3-4):316–323.

Dworczak, Piotr and Giorgio Martini. 2019. "The simple economics of optimal persuasion." *Journal of Political Economy* 127 (5):1993–2048.

Egorov, Georgy, Sergei Guriev, and Konstantin Sonin. 2009. "Why Resource-Poor Dictators Allow Freer Media: A Theory and Evidence from Panel Data." *American Political Science Review* 103 (4):645–668.

Ely, Jeffrey C. 2017. "Beeps." *American Economic Review* 107 (1):31–53.

Fréchette, Guillaume, Alessandro Lizzeri, and Jacopo Perego. 2020. "Rules and Commitment in Communication: An Experimental Analysis." Working paper.

Gentzkow, Matthew and Emir Kamenica. 2016. "A Rothschild-Stiglitz approach to Bayesian persuasion." *American Economic Review* 106 (5):597–601.

Glazer, Jacob and Ariel Rubinstein. 2006. "A study in the pragmatics of persuasion: a game theoretical approach." *Theoretical Economics* 4 (1):395–410.

Goltsman, Maria, Johannes Hörner, Gregory Pavlov, and Francesco Squintani. 2009. "Mediation, arbitration and negotiation." *Journal of Economic Theory* 144 (4):1397–1420.

Green, Jerry R and Nancy L Stokey. 2007. "A two-person game of information transmission." *Journal of Economic Theory* 135 (1):90–104.

Guo, Yingni and Eran Shmaya. 2021. "Costly miscalibration." *Theoretical Economics* 16 (2):477–506.

Hart, Sergiu, Ilan Kremer, and Motty Perry. 2017. "Evidence games: Truth and commitment." *American Economic Review* 107 (3):690–713.

Hedlund, Jonas. 2017. "Bayesian persuasion by a privately informed sender." *Journal of Economic Theory* 167:229–268.

Ichihashi, Shota. 2019. "Limiting Sender's information in Bayesian persuasion." *Games and Economic Behavior* 117:276–288.

Ivanov, Maxim. 2010. "Informational control and organizational design." *Journal of Economic Theory* 145 (2):721–751.

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

Kamenica, Emir and Matthew Gentzkow. 2011. "Bayesian persuasion." *American Economic Review* 101 (6):2590–2615.

Kartik, Navin. 2009. "Strategic Communication with Lying Costs." *Review of Economic Studies* 76 (4):1359–1395.

Kolotilin, Anton. 2018. “Optimal information disclosure: A linear programming approach.” *Theoretical Economics* 13 (2):607–635.

Lipnowski, Elliot and Doron Ravid. 2020. “Cheap Talk With Transparent Motives.” *Econometrica* 88 (4):1631–1660. URL `https://ideas.repec.org/a/wly/emetrp/v88y2020i4p1631-1660.html`.

Lipnowski, Elliot, Doron Ravid, and Denis Shishkin. 2021. “Perfect Bayesian Persuasion.” *Working paper* .

Luo, Zhaotian and Arturas Rozenas. 2018. “Strategies of election rigging: Trade-offs, determinants, and consequences.” *Quarterly Journal of Political Science* 13 (1):1–28.

Mathevet, Laurent, David Pearce, and Ennio Stacchetti. 2019. “Reputation and Information Design.” *Working paper* .

Min, Daehong. 2020. “Bayesian Persuasion under Partial Commitment.” Working paper.

Myerson, Roger B. 1986. “Multistage games with communication.” *Econometrica: Journal of the Econometric Society* :323–358.

Nguyen, Anh and Yong Teck Tan. 2021. “Bayesian Persuasion with Costly Messages.” Working paper.

Perez-Richet, Eduardo. 2014. “Interim bayesian persuasion: First steps.” *American Economic Review* 104 (5):469–474.

Perez-Richet, Eduardo and Vasiliki Skreta. 2021. “Test Design under Falsification.” Working paper.

Rockafellar, Ralph Tyrell. 2015. *Convex analysis.* Princeton university press.

Rothschild, Michael and Joseph E Stiglitz. 1970. “Increasing risk: I. A definition.” *Journal of Economic Theory* 2 (3):225–243.

Salamanca, Andrés. 2021. “The value of mediated communication.” *Journal of Economic Theory* 192:105191.

Sher, Itai. 2011. “Credibility and determinism in a game of persuasion.” *Games and Economic Behavior* 71 (2):409–419.

Skreta, Vasiliki. 2006. “Sequentially Optimal Mechanisms.” *Review of Economic Studies* 73 (4):1085–1111.

Srivastava, Sashi Mohan. 2008. *A course on Borel sets*, vol. 180. Springer Science & Business Media.

# Online Appendix

# A Main Results

## A.1 Toward the Proof of Theorem 1

Throughout this subsection, we work with the more general model with state-dependent credibility $\boldsymbol{\chi}$ laid out in subsection 6.3. In order to make the relationship even more transparent, we adopt the following notational convention. For a compact metrizable space $Y$, a probability measure $\mu \in \Delta Y$, and a function $f\colon Y \to \mathbb{R}$ that is bounded and measurable, let $f(\mu) \coloneqq \int_Y f \, \mathrm{d}\mu \in \mathbb{R}$ denote the average value of $f$. In particular, for any credibility function $\boldsymbol{\chi}$, the scalar $\boldsymbol{\chi}(\mu_0)$ is simply the total probability that the report is not subject to influence. While accommodating this more general model entails some notational cost, all conceptual content of the proof is identical in the special case of constant credibility, the generalization requiring no additional arguments. We therefore encourage the reader to read the entire proof with the special case that the function $\boldsymbol{\chi}$ is a constant $\chi$ in mind.

We now provide a brief overview of the proof. The proof begins by showing an equivalence between the set of $\boldsymbol{\chi}$-equilibrium outcomes, the set of $\boldsymbol{\chi}$-nonical equilibrium outcomes, and the existence of a particular decomposition of the equilibrium information policy. This decomposition makes it easy to see the program (12) is a relaxation of the program that maximizes S's value across all $\chi$-equilibrium outcomes. In particular, the program (12) enables S to induce posteriors that would generate too high a continuation payoff for S. The proof's next part establishes this constraint is non-binding at the optimum. We then conclude by explicitly writing the program that finds S's favorite equilibrium outcome and showing its value is identical to that of (12).

### A.1.1 Characterization of All Equilibrium Outcomes

In this section, we characterize the full range of $\boldsymbol{\chi}$-equilibrium outcomes, which we define below. In short, a $\boldsymbol{\chi}$-equilibrium outcome consists of a description of the information R receives in equilibrium (which is jointly constructed by the official reporting protocol and an influencing S's messaging strategy), an expected payoff that S gets conditional on the official reporting protocol being used, and an expected payoff that S gets conditional on having the opportunity to influence.

To present unified proofs including for the case of $\boldsymbol{\chi} = \mathbf{1}$ and $\boldsymbol{\chi} = \mathbf{0}$, we adopt the notational convention that $\frac{0}{0} = 1$ wherever it appears.

In line with the main text, define the following straightforward generalization of a A $\chi$-nonical equilibrium to the case of state-dependent credibility.

**Definition 1.** *A* $\boldsymbol{\chi}$***-nonical equilibrium*** *is a* $\boldsymbol{\chi}$*-equilibrium* $(\xi, \sigma, \alpha, \pi)$ *such that every Borel* $\hat{M} \subseteq M_\alpha^*$ *has* $\xi(\hat{M}|\cdot) = \xi(M_\alpha^*|\cdot)\ \sigma(\hat{M}|\cdot)$.

The above definition imposes further structure on a $\boldsymbol{\chi}$-equilibrium. The requirement pertains to the set $M_\alpha^*$ of highest-payoff messages for S, which are necessarily the only messages an influencing S chooses. The condition says the conditional distribution messages in $M_\alpha^*$ is identical for the official experiment and for an influencing sender's choices, in any state for which the official report sometimes sends messages in $M_\alpha^*$. Informally, the condition says all differences in how the official and influenced report communicate are through whether or not they send a message in $M_\alpha^*$ in a given state.

**Definition 2.** *Say* $(p, s_o, s_i) \in \Delta\Delta\Theta \times \mathbb{R} \times \mathbb{R}$ *is a* $\boldsymbol{\chi}$***-equilibrium outcome*** *if some* $\boldsymbol{\chi}$*-equilibrium* $(\xi, \sigma, \alpha, \pi)$ *exists whose induced receiver belief distribution, official-report sender payoff, and influenced-report sender payoff are* $(p, s_o, s_i)$*, that is,*

$$
\begin{aligned}
p &= \left(\int_\Theta \Big[\boldsymbol{\chi}\xi + (\mathbf{1} - \boldsymbol{\chi})\sigma\Big]\, \mathrm{d}\mu_0\right) \circ \pi^{-1} \\
s_o &= \int_\Theta \tfrac{\boldsymbol{\chi}}{\boldsymbol{\chi}(\mu_0)} \int_M u_S(\alpha(m))\, \mathrm{d}\xi(m|\cdot)\, \mathrm{d}\mu_0 \\
s_i &= \int_\Theta \tfrac{\mathbf{1}-\boldsymbol{\chi}}{1-\boldsymbol{\chi}(\mu_0)} \int_M u_S(\alpha(m))\, \mathrm{d}\sigma(m|\cdot)\, \mathrm{d}\mu_0.
\end{aligned}
$$

*If, further,* $(\xi, \sigma, \alpha, \pi)$ *is a* $\boldsymbol{\chi}$*-nonical equilibrium, then we say* $(p, s_o, s_i)$ *is a* $\boldsymbol{\chi}$***-nonical equilibrium outcome****.*

The following lemma adopts a belief-based approach, directly characterizing $\boldsymbol{\chi}$-equilibrium outcomes of our game.

**Lemma 1.** *For* $(p, s_o, s_i) \in \Delta\Delta\Theta \times \mathbb{R} \times \mathbb{R}$*, the following are equivalent:*

1. $(p, s_o, s_i)$ *is a* $\boldsymbol{\chi}$*-equilibrium outcome;*

2. $(p, s_o, s_i)$ *is a* $\boldsymbol{\chi}$*-nonical equilibrium outcome;*

3. *Some* $k \in [0, 1]$*,* $g$*,* $b \in \Delta\Delta\Theta$ *exist such that*

*(i)* $kb + (1-k)g = p \in \mathcal{R}(\mu_0)$;

*(ii)* $(1-k)\int_{\Delta\Theta} \mu \, \mathrm{d}g(\mu) \geq (\mathbf{1} - \boldsymbol{\chi})\mu_0$;

*(iii)* $g\{\mu \in \Delta\Theta : s_i \in V(\mu)\} = b\{\mu \in \Delta\Theta : \min V(\mu) \leq s_i\} = 1$;

*(iv)* $s_i - s_o \in \frac{k}{\boldsymbol{\chi}(\mu_0)} \left[ s_i - \int_{\mathrm{supp}(b)} s_i \wedge V \, \mathrm{d}b \right]$.[35]

The first two parts of the lemma are self explanatory. The third part says that the information policy $p$ can be decomposed into two separate random posteriors, $b$ and $g$, satisfying three conditions. Condition (ii) says the barycenter of $g$ satisfies ($\chi$-BP). Condition (iii) says R is willing to give S a continuation payoff equal to $s_i$ after all posteriors induced by $g$, and a lower continuation payoff for any posterior induced by $b$. And condition (iv) says there is a way to select R's best response to posteriors in $b$ so that no posterior generates a payoff above $s_i$ and so that S's average payoff conditional on her report coming from the official protocol adds up to $s_o$.

We now give an overview of Lemma 1. Obviously, 2 implies 1. Therefore, the proof proceeds by completing a cycle, showing that 1 implies 3, and that 3 implies 2. To show 1 implies 3, we take an equilibrium and partition the set of on-path messages into two subsets: the set of "good" messages for S to send (i.e., those that give S the highest possible expected payoff out of any possible message), and the complementary "bad" messages. Following this decomposition, one can obtain $g$ and $b$ by looking at the distribution of R's posterior beliefs conditional on the message being in the "good" or "bad" set, respectively. Letting $k$ be the probability S sends a "bad" message, one obtains condition (i) from the usual Bayesian reasoning. Condition (ii) then follows from similar reasoning as explained in the main text, whereas conditions (iii) and (iv) follow from S's incentive constraints. To prove 3 implies 2, we use the decomposition provided by 3 to construct a $\boldsymbol{\chi}$-nonical equilibrium.

*Proof.* We show that 1 implies 3 and 3 implies 2, noting 2 obviously implies 1.

Let us first show 1 implies 3. To that end, suppose $(\xi, \sigma, \alpha, \pi)$ is a $\boldsymbol{\chi}$-equilibrium

[35] Here, $s_i \wedge V : \Delta\Theta \rightrightarrows \mathbb{R}$ is the correspondence with $s_i \wedge V(\mu) = (-\infty, s_i] \cap V(\mu)$; it is a Kakutani correspondence (because $V$ is) on the restricted domain $\{\min V \leq s_i\} \supseteq \mathrm{supp}(b)$. The integral is the (Aumann) integral of a correspondence:

$$\int_{\mathrm{supp}(b)} s_i \wedge V \, \mathrm{d}b = \left\{ \int_{\mathrm{supp}(b)} \phi \, \mathrm{d}b \colon \phi \text{ is a measurable selector of } s_i \wedge V|_{\mathrm{supp}(b)} \right\}.$$

resulting in outcome $(p, s_o, s_i)$. Let

$$G \coloneqq \int_\Theta \sigma \,\mathrm{d}\left[\tfrac{\mathbf{1}-\boldsymbol{\chi}}{1-\boldsymbol{\chi}(\mu_0)}\mu_0\right] \text{ and } P \coloneqq \int_\Theta [\boldsymbol{\chi}\xi + (\mathbf{1}-\boldsymbol{\chi})\sigma]\,\mathrm{d}\mu_0 \in \Delta M$$

denote the probability measures over messages induced by non-committed behavior and by average sender behavior, respectively. Let $k \coloneqq 1 - P(M_\alpha^*)$ denote the ex-ante probability that a suboptimal message is sent. Sender incentive compatibility (which implies that $\sigma(M_\alpha^*|\cdot) = \mathbf{1}$) tells us that $k \in [0, \boldsymbol{\chi}(\mu_0)]$. Let $B \coloneqq \frac{1}{k}[P - (1-k)G]$ if $k > 0$; and let $B \coloneqq \int_\Theta \xi \,\mathrm{d}\mu_0$ otherwise. As barycenters of probability measures over $M$, the measures $G, P$ are in $\Delta M$. Measure $B$ on $M$ therefore has total measure 1. Therefore, $B \in \Delta M$ so long as $B$ is a positive measure, that is, $P \geq (1-k)G$. To see this measure inequality, notice

$$(1-k)G = P(M_\alpha^*) \int_\Theta \sigma \,\mathrm{d}\left[\tfrac{\mathbf{1}-\boldsymbol{\chi}}{1-\boldsymbol{\chi}(\mu_0)}\mu_0\right] \leq \int_\Theta \sigma \,\mathrm{d}\left[(\mathbf{1}-\boldsymbol{\chi})\mu_0\right] \leq P,$$

where the first inequality follows from sender incentives (implying influenced reporting only sends messages in $M_\alpha^*$). Now, define the induced belief distributions by these two distributions over messages, $g \coloneqq G \circ \pi^{-1}$ and $b \coloneqq B \circ \pi^{-1}$. By construction, $kb + (1-k)g = P \circ \pi^{-1} = p \in \mathcal{R}(\mu_0)$, that is, the first condition holds. Moreover, the second condition holds:

$$(1-k) \int_{\Delta\Theta} \mu \,\mathrm{d}g(\mu) = \int_M \pi \,\mathrm{d}[(1-k)G] = \int_{M_\alpha^*} \pi \,\mathrm{d}P \geq (\mathbf{1}-\boldsymbol{\chi})\mu_0,$$

where the inequality follows from the Bayesian property of $\pi$, together with the fact that $\sigma$ almost surely sends a message from $M_\alpha^*$ on the path of play. Next, observe that for any $m \in M$, sender incentive compatibility tells us that $u_S(\alpha(m)) \leq s_i$, and receiver incentive compatibility implies $\alpha(m) \in V(\pi(m))$. It follows directly that $g\{V \ni s_i\} = b\{\min V \leq s_i\} = 1$, that is, the third condition holds. Toward the fourth and final condition, let us view $\pi, \alpha$ as random variables on the probability space $\langle M, P \rangle$. Defining the conditional expectation $\phi_0 \coloneqq \mathbb{E}_B[u_S(\alpha)|\pi] : M \to \mathbb{R}$, the Doob-Dynkin lemma delivers a measurable function $\phi : \Delta\Theta \to \mathbb{R}$ such that $\phi \circ \pi =_{B-\text{a.e.}} \phi_0$. As $u_S(\alpha(m)) \in s_i \wedge V(m)$ for every $m \in M$, and the correspondence $s_i \wedge V$ is compact- and convex-valued, it must be that $\phi_0 \in_{B-\text{a.e.}} s_i \wedge V(\pi)$. Therefore, $\phi \in_{b-\text{a.e.}} s_i \wedge V$. Modifying $\phi$ on a $b$-null set, we may assume without loss that $\phi$ is a measurable selector of $s_i \wedge V$. Observe now that

$$\int_{\mathrm{supp}(b)} \phi \, \mathrm{d}b = \int_M \phi_0 \, \mathrm{d}B = \int_M \mathbb{E}_B[u_S(\alpha)|\pi] \, \mathrm{d}B = \int_M u_S \circ \alpha \, \mathrm{d}B.$$

Therefore, since $G(M^*_\alpha) = 1$,

$$\begin{aligned} s_o &= \int_M u_S \circ \alpha \, \mathrm{d}\tfrac{P-[1-\boldsymbol{\chi}(\mu_0)]G}{\boldsymbol{\chi}(\mu_0)} = \int_M u_S \circ \alpha \, \mathrm{d}\tfrac{kB+(1-k)G-[1-\boldsymbol{\chi}(\mu_0)]G}{\boldsymbol{\chi}(\mu_0)} \\ &= \int_M u_S \circ \alpha \, \mathrm{d}\left[\left(1-\tfrac{k}{\boldsymbol{\chi}(\mu_0)}\right)G + \tfrac{k}{\boldsymbol{\chi}(\mu_0)}B\right] = \left(1-\tfrac{k}{\boldsymbol{\chi}(\mu_0)}\right)s_i + \tfrac{k}{\boldsymbol{\chi}(\mu_0)}\int_{\mathrm{supp}(b)} \phi \, \mathrm{d}b, \end{aligned}$$

as required.

Now, we show 3 implies 2. As $M$ is an uncountable Polish space, the Borel isomorphism theorem (Theorem 3.3.13 Srivastava, 2008) says $M$ is isomorphic (as a measurable space) to $\{i,o\}\times\Delta\Theta$. We can therefore assume without loss that $M = \{i,o\}\times\Delta\Theta$.

Suppose $k \in [0,1]$, $g, b \in \Delta\Delta\Theta$ satisfy the four listed conditions so that 3 holds; and let $\phi$ be a measurable selector of $s_i \wedge V|_{\mathrm{supp}(b)}$ with $s_o = \left(1-\frac{k}{\boldsymbol{\chi}(\mu_0)}\right)s_i + \frac{k}{\boldsymbol{\chi}(\mu_0)}\int_{\mathrm{supp}(b)} \phi \, \mathrm{d}b$, which the fourth condition assures us exists.

We will construct a $\boldsymbol{\chi}$-nonical equilibrium from these objects that induces outcome $(p, s_o, s_i)$.

Let us proceed in two cases. First, consider the case that $s_o = s_i$. In this case, the fourth condition implies $b\{\phi = s_i\} = 1$, so that $p \in \mathcal{R}(\mu_0)$ has $p\{V \ni s_i\} = 1$. Hence, ($V$ being upper hemicontinuous) Lipnowski and Ravid (2020, Lemma 1) delivers an equilibrium $(\sigma, \alpha, \pi)$ of the pure cheap talk game generating receiver information distribution $p$ and sender payoff $s_i$. It follows immediately that $(\sigma, \sigma, \alpha, \pi)$ is a $\boldsymbol{\chi}$-nonical equilibrium that induces outcome $(p, s_i, s_i)$.

Henceforth, we focus on the remaining case that $s_o < s_i$. Without loss of generality, we may further assume $b\{\phi < s_i\} = 1$.[36] Define $\beta := \int_{\Delta\Theta} \mu \, \mathrm{d}b(\mu)$ and $\gamma := \int_{\Delta\Theta} \mu \, \mathrm{d}g(\mu)$. Let measurable $\eta_g : \Theta \to \Delta[\mathrm{supp}(g)] \subseteq \Delta\Delta\Theta$ and $\eta_b : \Theta \to \Delta[\mathrm{supp}(b)] \subseteq \Delta\Delta\Theta$ be signals that induce belief distribution $g$ for prior $\gamma$ and belief distribution $b$ for prior $\beta$, respectively, such that for each such signal the induced posterior belief is to equal

[36] Indeed, one could replace $k$ with $\tilde{k} := kb\{\phi < s_i\} > 0$, replace $b$ with $\tilde{b} := \frac{k}{\tilde{k}} b\left((\cdot) \cap \{\phi < s_i\}\right)$, and replace $g$ with $\tilde{g} := \frac{1}{1-\tilde{k}}(p - \tilde{k}\tilde{b})$.

the message itself. That is, for every Borel $\hat{\Theta} \subseteq \Theta$ and $\hat{D} \subseteq \Delta\Theta$,

$$\int_{\hat{\Theta}} \eta_b(\hat{D}|\cdot)\,\mathrm{d}\beta = \int_{\hat{D}} \mu(\hat{\Theta})\,\mathrm{d}b(\mu) \text{ and } \int_{\hat{\Theta}} \eta_g(\hat{D}|\cdot)\,\mathrm{d}\gamma = \int_{\hat{D}} \mu(\hat{\Theta})\,\mathrm{d}g(\mu).$$

Take some Radon-Nikodym derivative $\frac{\mathrm{d}\beta}{\mathrm{d}\mu_0} : \Theta \to \mathbb{R}_+$; changing it on a $\mu_0$-null set, we may assume that $\mathbf{0} \leq \frac{k}{\boldsymbol{\chi}}\frac{\mathrm{d}\beta}{\mathrm{d}\mu_0} \leq \mathbf{1}$ since $(1-k)\gamma \geq (\mathbf{1}-\boldsymbol{\chi})\mu_0$. With the above ingredients in hand, we can define the sender's influenced strategy and reporting protocol

$$\begin{aligned}
\sigma &\coloneqq \delta_i \otimes \eta_g : \Theta \to \Delta M,\\
\xi &\coloneqq \left(\mathbf{1} - \frac{k}{\boldsymbol{\chi}}\frac{\mathrm{d}\beta}{\mathrm{d}\mu_0}\right)\delta_i \otimes \eta_g + \frac{k}{\boldsymbol{\chi}}\frac{\mathrm{d}\beta}{\mathrm{d}\mu_0}\delta_o \otimes \eta_b : \Theta \to \Delta M.
\end{aligned}$$

Because $M_i \coloneqq \{i\} \times \Delta\Theta$ obviously has $\sigma(M_i|\cdot) = \mathbf{1}$ and $\xi(\hat{M}_i|\cdot) = \xi(M_i|\cdot)\ \sigma(\hat{M}_i|\cdot)$ for every Borel $\hat{M}_i \subseteq M_i$, it will follow that a $\boldsymbol{\chi}$-equilibrium with sender play described by $(\sigma, \xi)$ is in fact a $\boldsymbol{\chi}$-nonical equilibrium, as long as the receiver strategy $\alpha$ satisfies $M^*_\alpha \supseteq M_i$. To finish constructing such a $\boldsymbol{\chi}$-equilibrium, we define the receiver strategy and belief map for our proposed equilibrium as follows. Intuitively, an on-path message $(i, \mu)$ will lead to belief $\mu$ and a receiver best response that delivers payoff $s_i$ to the sender; an on-path message $(o, \mu)$ will lead to belief $\mu$ and receiver best response that delivers a potentially lower payoff to the sender, calibrated to give the target average payoff; and off-path messages are interpreted as equivalent to some on-path message so as not to introduce new incentive constraints. Formally, fix some $\hat{\mu} \in \mathrm{supp}(b)$, which will serve as a default beliefs and incentive-compatible receiver response for any off-path messages. We can then define a receiver belief map as

$$\begin{aligned}
\pi : M &\to \Delta\Theta\\
m &\mapsto \begin{cases} \mu & : \ m = (i, \mu) \text{ for } \mu \in \mathrm{supp}(g), \text{ or } m = (o, \mu) \text{ for } \mu \in \mathrm{supp}(b)\\ \hat{\mu} & : \ \text{otherwise.} \end{cases}
\end{aligned}$$

Finally, by Lipnowski and Ravid (2020, Lemma 2), some measurable $\alpha_b, \alpha_g : \Delta\Theta \to \Delta A$ exist such that:[37]

- $\alpha_b(\mu), \alpha_g(\mu) \in \mathrm{argmax}_{\tilde{\alpha} \in \Delta A}\, u_R(\tilde{\alpha}, \mu)\ \forall \mu \in \Delta\Theta$;

[37]The cited lemma delivers $\alpha_b|_{\mathrm{supp}(b)}, \alpha_g|_{\mathrm{supp}(g)}$. Then, as $\mathrm{supp}(p) \subseteq \mathrm{supp}(b) \cup \mathrm{supp}(g)$, we can extend both functions to the rest of their domains by making them agree on $\mathrm{supp}(p) \setminus [\mathrm{supp}(b) \cap \mathrm{supp}(g)]$.

- $u_S(\alpha_b(\mu)) = \phi(\mu)\ \forall \mu \in \text{supp}(b)$, and $u_S(\alpha_g(\mu)) = s_i\ \forall \mu \in \text{supp}(g)$.

From these selectors, we can define a receiver strategy as

$$\begin{aligned} \alpha : M &\to \Delta A \\ m &\mapsto \begin{cases} \alpha_b(\mu) & : \ m = (o, \mu) \text{ for some } \mu \in \text{supp}(b) \\ \alpha_g(\mu) & : \ m = (i, \mu) \text{ for some } \mu \in \text{supp}(g) \\ \alpha_b(\hat{\mu}) & : \ \text{otherwise.} \end{cases} \end{aligned}$$

We want to show that the tuple $(\xi, \sigma, \alpha, \pi)$ is a $\boldsymbol{\chi}$-equilibrium (hence, a $\boldsymbol{\chi}$-nonical equilibrium) resulting in outcome $(p, s_o, s_i)$. It is immediate from the construction of $(\sigma, \alpha, \pi)$ that sender incentive compatibility and receiver incentive compatibility hold, and that the expected sender payoff is $s_i$ given influenced reporting. It remains to verify that the induced receiver belief distribution is $p$, that the Bayesian property is satisfied, and that the expected sender payoff from the official report is $s_o$. We verify these features below, via a tedious computation.

Recall $\boldsymbol{\chi}\xi : \Theta \to \Delta M$ is defined as the pointwise product, i.e. for every $\theta \in \Theta$ and Borel $\hat{M} \subseteq M$, we have $(\boldsymbol{\chi}\xi)(\hat{M}|\theta) = \boldsymbol{\chi}(\theta)\xi(\hat{M}|\theta)$; and similarly for $(\mathbf{1} - \boldsymbol{\chi})\sigma$. To see that the Bayesian property holds, observe that every Borel $D \subseteq \Delta\Theta$ satisfies

$$\begin{aligned} [(\mathbf{1} - \boldsymbol{\chi})\sigma + \boldsymbol{\chi}\xi](\{o\} \times D|\cdot) &= k \tfrac{\mathrm{d}\beta}{\mathrm{d}\mu_0} \eta_b(D|\cdot) \\ [(\mathbf{1} - \boldsymbol{\chi})\sigma + \boldsymbol{\chi}\xi](\{i\} \times D|\cdot) &= \left[(\mathbf{1} - \boldsymbol{\chi}) + \boldsymbol{\chi}\left(\mathbf{1} - \tfrac{k}{\boldsymbol{\chi}} \tfrac{\mathrm{d}\beta}{\mathrm{d}\mu_0}\right)\right] \eta_g(D|\cdot) \\ &= \left(\mathbf{1} - k \tfrac{\mathrm{d}\beta}{\mathrm{d}\mu_0}\right) \eta_g(D|\cdot). \end{aligned}$$

Now, take any Borel $\hat{M} \subseteq M$ and $\hat{\Theta} \subseteq \Theta$, and let $D_z := \left\{\mu \in \Delta\Theta : \ (z, \mu) \in \hat{M}\right\}$ for $z \in \{i, o\}$. Observe that

$$\begin{aligned} &\int_\Theta \int_{\hat{M}} \pi(\hat{\Theta}|m) \, \mathrm{d}[(1 - \boldsymbol{\chi})\sigma + \boldsymbol{\chi}\xi](m|\cdot) \, \mathrm{d}\mu_0 \\ = &\int_\Theta \left(\int_{\{o\} \times D_o} + \int_{\{i\} \times D_i}\right) \pi(\hat{\Theta}|m) \, \mathrm{d}[(1 - \boldsymbol{\chi})\sigma + \boldsymbol{\chi}\xi](m|\cdot) \, \mathrm{d}\mu_0 \\ = &\int_\Theta \left[k \tfrac{\mathrm{d}\beta}{\mathrm{d}\mu_0} \int_{D_o} \mu(\hat{\Theta}) \, \mathrm{d}\eta_b(\mu|\cdot) + \left(1 - k \tfrac{\mathrm{d}\beta}{\mathrm{d}\mu_0}\right) \int_{D_i} \mu(\hat{\Theta}) \, \mathrm{d}\eta_g(\mu|\cdot)\right] \mathrm{d}\mu_0 \\ = k &\int_\Theta \int_{D_o} \mu(\hat{\Theta}) \, \mathrm{d}\eta_b(\mu|\cdot) \, \mathrm{d}\beta + \int_\Theta \int_{D_i} \mu(\hat{\Theta}) \, \mathrm{d}\eta_g(\mu|\cdot) \, \mathrm{d}[\mu_0 - k\beta] \end{aligned}$$

$$
\begin{aligned}
&= k\int_{\Theta}\int_{D_o}\mu(\hat{\Theta})\,\mathrm{d}\eta_b(\mu|\cdot)\,\mathrm{d}\beta + (1-k)\int_{\Theta}\int_{D_i}\mu(\hat{\Theta})\,\mathrm{d}\eta_g(\mu|\cdot)\,\mathrm{d}\gamma \\
&= k\int_{D_o}\int_{\Theta}\mu(\hat{\Theta})\,\mathrm{d}\mu(\theta)\,\mathrm{d}b(\mu) + (1-k)\int_{D_i}\int_{\Theta}\mu(\hat{\Theta})\,\mathrm{d}\mu(\theta)\,\mathrm{d}g(\mu) \\
&= k\int_{D_o}\mu(\hat{\Theta})\,\mathrm{d}b(\mu) + (1-k)\int_{D_i}\mu(\hat{\Theta})\,\mathrm{d}g(\mu).
\end{aligned}
$$

Let us see that the above computation implies both that $(\xi,\sigma,\pi)$ satisfies the Bayesian property (making $(\xi,\sigma,\alpha,\pi)$ a $\boldsymbol{\chi}$-equilibrium) and that its induced belief distribution is $p$. First, observe that

$$
\begin{aligned}
&\int_{\Theta}\int_{\hat{M}}\pi(\hat{\Theta}|m)\,\mathrm{d}[(1-\boldsymbol{\chi})\sigma+\boldsymbol{\chi}\xi](m|\cdot)\,\mathrm{d}\mu_0 \\
&= k\int_{D_o}\mu(\hat{\Theta})\,\mathrm{d}b(\mu) + (1-k)\int_{D_i}\mu(\hat{\Theta})\,\mathrm{d}g(\mu) \\
&= k\int_{\hat{\Theta}}\eta_b(D_o|\cdot)\,\mathrm{d}\beta + (1-k)\int_{\hat{\Theta}}\eta_g(D_i|\cdot)\,\mathrm{d}\gamma \\
&= \int_{\hat{\Theta}}\eta_b(D_o|\cdot)\,\mathrm{d}[k\beta] + \int_{\hat{\Theta}}\eta_g(D_i|\cdot)\,\mathrm{d}[\mu_0-k\beta] \\
&= \int_{\hat{\Theta}}\left[k\tfrac{\mathrm{d}\beta}{\mathrm{d}\mu_0}\eta_b(D_o|\cdot) + \left(\mathbf{1}-k\tfrac{\mathrm{d}\beta}{\mathrm{d}\mu_0}\right)\eta_g(D_i|\cdot)\right]\,\mathrm{d}\mu_0 \\
&= \int_{\hat{\Theta}}[(\mathbf{1}-\boldsymbol{\chi})\sigma+\boldsymbol{\chi}\xi](\hat{M}|\cdot)\,\mathrm{d}\mu_0,
\end{aligned}
$$

verifying the Bayesian property. Second, for any Borel $D\subseteq\Delta\Theta$, we can specialize to the case of $D_o=D_i=D$ and $\hat{\Theta}=\Theta$, showing the equilibrium probability of the receiver posterior belief belonging to $D$ is exactly

$$
\int_{\Theta}[(\mathbf{1}-\boldsymbol{\chi})\sigma+\boldsymbol{\chi}\xi](\{i,o\}\times D|\cdot)\,\mathrm{d}\mu_0 = k\int_D\mathbf{1}\,\mathrm{d}b + (1-k)\int_D\mathbf{1}\,\mathrm{d}g = p(D).
$$

Finally, the expected sender payoff conditional on reporting not being influenced is given by:

$$
\begin{aligned}
&\int_\Theta \int_M u_S\left(\alpha(m)\right)\,\mathrm{d}\xi(m|\cdot)\,\mathrm{d}\left[\frac{\boldsymbol{\chi}}{\boldsymbol{\chi}(\mu_0)}\mu_0\right] \\
&= \int_\Theta \left[\left(1-\frac{k}{\boldsymbol{\chi}}\frac{\mathrm{d}\beta}{\mathrm{d}\mu_0}\right)\int_{\Delta\Theta} u_S\left(\alpha(i,\mu)\right)\,\mathrm{d}\eta_g(\mu|\cdot) + \frac{k}{\boldsymbol{\chi}}\frac{\mathrm{d}\beta}{\mathrm{d}\mu_0}\int_{\Delta\Theta} u_S\left(\alpha(o,\mu)\right)\,\mathrm{d}\eta_b(\mu|\cdot)\right]\,\mathrm{d}\left[\frac{\boldsymbol{\chi}}{\boldsymbol{\chi}(\mu_0)}\mu_0\right] \\
&= \int_\Theta \left[\left(\mathbf{1}-\frac{k}{\boldsymbol{\chi}}\frac{\mathrm{d}\beta}{\mathrm{d}\mu_0}\right)\int_{\Delta\Theta} s_i\,\mathrm{d}\eta_g(\mu|\cdot) + \frac{k}{\boldsymbol{\chi}}\frac{\mathrm{d}\beta}{\mathrm{d}\mu_0}\int_{\mathrm{supp}(b)} \phi(\mu)\,\mathrm{d}\eta_b(\mu|\cdot)\right]\,\mathrm{d}\left[\frac{\boldsymbol{\chi}}{\boldsymbol{\chi}(\mu_0)}\mu_0\right] \\
&= s_i + \frac{k}{\boldsymbol{\chi}(\mu_0)}\int_\Theta \left[-s_i + \int_{\mathrm{supp}(b)} \phi(\mu)\,\mathrm{d}\eta_b(\mu|\theta)\right]\,\mathrm{d}\beta(\theta) \\
&= \left[1-\frac{k}{\boldsymbol{\chi}(\mu_0)}\right]s_i + \frac{k}{\boldsymbol{\chi}(\mu_0)}\int_{\Delta\Theta}\int_\Theta \phi(\mu)\,\mathrm{d}\mu(\theta)\,\mathrm{d}b(\mu) \\
&= \frac{(1-k)-[1-\boldsymbol{\chi}(\mu_0)]}{\boldsymbol{\chi}(\mu_0)}s_i + \frac{k}{\boldsymbol{\chi}(\mu_0)}\int_{\mathrm{supp}(b)} \phi\,\mathrm{d}b \\
&= s_o,
\end{aligned}
$$

as required.

□

### A.1.2 Proof of Theorem 1

We begin with a simple technical lemma on the geometry of concave envelopes and the belief distributions that attain them.

**Lemma 2.** *If $f : \Delta\Theta \to \mathbb{R}$ is upper semicontinuous, $\hat{f}$ is $f$'s concave envelope, $\beta \in \Delta\Theta$, and $b \in \mathcal{R}(\beta)$ has $\int f\,\mathrm{d}b = \hat{f}(\beta)$, then $b\{\mu \in \Delta\Theta : \hat{f}|_{\mathrm{co}\{\beta,\mu\}} \text{ affine}\} = 1$.*

*Proof.* First, observe that every concave, non-affine function $\varphi : [0,1] \to \mathbb{R}$ has $\varphi(z) > z\varphi(1) + (1-z)\varphi(0)$ for every $z \in (0,1)$. Hence, it suffices to show $\hat{f}\left(\frac{1}{2}\beta + \frac{1}{2}\mu\right) = \frac{1}{2}\hat{f}(\beta) + \frac{1}{2}\hat{f}(\mu)$ a.s.-$b(\mu)$. Equivalently, because concavity of $\hat{f}$ implies $\frac{1}{2}\hat{f}(\beta) + \frac{1}{2}\hat{f}(\mu) - \hat{f}\left(\frac{1}{2}\beta + \frac{1}{2}\mu\right) \le 0$ for every $\mu \in \Delta\Theta$, we need only show $\int\left[\frac{1}{2}\hat{f}(\beta) + \frac{1}{2}\hat{f}(\mu)\right]\mathrm{d}b(\mu)$ and $\int f\left(\frac{1}{2}\beta + \frac{1}{2}\mu\right)\mathrm{d}b(\mu)$ coincide. To show this identity, observe that (as $\hat{f}$ is concave, upper semicontinuous, and everywhere above $f$)

$$
\begin{aligned}
\hat{f}(\beta) = \int\left[\tfrac{1}{2}\hat{f}(\beta) + \tfrac{1}{2}f\right]\mathrm{d}b \le \quad & \int\left[\tfrac{1}{2}\hat{f}(\beta) + \tfrac{1}{2}\hat{f}\right]\mathrm{d}b \\
\le \quad & \int \hat{f}\left(\tfrac{1}{2}\beta + \tfrac{1}{2}\mu\right)\mathrm{d}b(\mu) \;\; \le \hat{f}\left(\int\left[\tfrac{1}{2}\beta + \tfrac{1}{2}\mu\right]\mathrm{d}b(\mu)\right) = \hat{f}(\beta).
\end{aligned}
$$

Hence, all of the above expressions are equal, delivering the lemma. □

Before proceeding to the proof of Theorem 1, we prove a useful lemma about the theorem's auxiliary program. In short, the lemma shows that a relaxation built into this program—that S can be held to payoff $\bar{v}(\gamma)$ even at beliefs at which every R best response gives S a higher payoff—is payoff-irrelevant at an optimum.

**Lemma 3.** *If $(\beta, \gamma, k)$ solve program (12) and have $\hat{v}_{\wedge\gamma}(\beta) < \bar{v}(\gamma)$, and $b \in \mathcal{R}(\beta)$ has $\int v_{\wedge\gamma}\, \mathrm{d}b = \hat{v}_{\wedge\gamma}(\beta)$, then $v_{\wedge\gamma}(\mu) \in V(\mu)$ for every $\mu \in \operatorname{supp}(b)$. In particular, $b\{\min V \leq \bar{v}(\gamma)\} = 1$.*

*Proof.* Given the definition of $v_{\wedge\gamma}$, and given that $V$ is nonempty-compact-convex-valued, it suffices to show $w(\mu) \leq \bar{v}(\gamma)$ for $\mu \in \operatorname{supp}(b)$, where $w := \min V$. Then, because $V$ is upper hemicontinuous, it suffices to show $b\{w \leq \bar{v}(\gamma)\} = 1$. To that end, define $D := \left\{\mu \in \Delta\Theta : \ \hat{v}_{\wedge\gamma}|_{\operatorname{co}\{\beta,\mu\}} \text{ affine}\right\}$. Applying Lemma 2 to $v_{\wedge\gamma}$ implies $b(D) = 1$, so the lemma will follow if we can show $w|_D \leq \bar{v}(\gamma)$.

Let us establish that every $\mu \in D$ has $w(\mu) \leq \bar{v}(\gamma)$. The result is obvious if $v(\mu) < \bar{v}(\gamma)$, so let us focus on the case that $v(\mu) \geq \bar{v}(\gamma)$. For such $\mu$, note that every proper convex combination $\mu'$ of $\beta$ and $\mu$ has $v(\mu') < \bar{v}(\gamma)$—for otherwise $\hat{v}_{\wedge\gamma}(\beta) < \hat{v}_{\wedge\gamma}(\mu') = \hat{v}_{\wedge\gamma}(\mu)$, violating the definition of $D \ni \mu$. It follows that $\mu$ is in the closure of $\{v \leq \bar{v}(\gamma)\} \subseteq \{w \leq \bar{v}(\gamma)\}$. Lower semicontinuity of $w$ then implies $w(\mu) \leq \bar{v}(\gamma)$. $\square$

We now prove our main theorem. In fact, we prove its generalization to the setting of section 6.3: An S-optimal $\boldsymbol{\chi}$-equilibrium exists, giving S payoff $v^*_{\boldsymbol{\chi}}(\mu_0)$.

*Proof.* By Lemma 1, the supremum sender value over all $\boldsymbol{\chi}$-equilibrium outcomes is

$$
\begin{aligned}
\tilde{v}^*_{\boldsymbol{\chi}}(\mu_0) := \sup_{b,g \in \Delta\Delta\Theta,\ k \in [0,1],\ s_o, s_i \in \mathbb{R}} & \left\{\boldsymbol{\chi}(\mu_0) s_o + [1 - \boldsymbol{\chi}(\mu_0)] s_i\right\} \\
\text{s.t.} \qquad & kb + (1-k)g \in \mathcal{R}(\mu_0),\ (1-k)\int_{\Delta\Theta} \mu \, \mathrm{d}g(\mu) \geq (\mathbf{1} - \boldsymbol{\chi})\mu_0, \\
& g\{V \ni s_i\} = b\{\min V \leq s_i\} = 1, \\
& s_o \in \left(1 - \tfrac{k}{\boldsymbol{\chi}(\mu_0)}\right) s_i + \tfrac{k}{\boldsymbol{\chi}(\mu_0)} \int_{\operatorname{supp}(b)} s_i \wedge V \, \mathrm{d}b.
\end{aligned}
$$

Given any feasible $(b, g, k, s_o, s_i)$ in the above program, replacing the associated measurable selector of $s_i \wedge V|_{\operatorname{supp}(b)}$ with the weakly higher function $s_i \wedge v|_{\operatorname{supp}(b)}$, and raising $s_o$ to $\left(1 - \frac{k}{\boldsymbol{\chi}(\mu_0)}\right) s_i + \frac{k}{\boldsymbol{\chi}(\mu_0)} \int_{\operatorname{supp}(b)} s_i \wedge v \, \mathrm{d}b$, weakly raises the objective and preserve all constraints. Therefore,

$$\begin{aligned}
\tilde{v}^*_{\boldsymbol{\chi}}(\mu_0) = \sup_{b,g\in\Delta\Delta\Theta,\ k\in[0,1],\ s_i\in\mathbb{R}} & \left\{(1-k)s_i + k\int_{\mathrm{supp}(b)} s_i \wedge v \,\mathrm{d}b\right\} \\
\text{s.t.} \qquad & kb + (1-k)g \in \mathcal{R}(\mu_0),\ (1-k)\int_{\Delta\Theta} \mu \,\mathrm{d}g(\mu) \geq (\mathbf{1}-\boldsymbol{\chi})\mu_0, \\
& g\{V \ni s_i\} = b\{\min V \leq s_i\} = 1.
\end{aligned}$$

Given any feasible $(b, g, k, s_i)$ in the latter program, replacing $(g, s_i)$ with any $(g^*, s_i^*)$ such that $\int_{\Delta\Theta} \mu \,\mathrm{d}g^*(\mu) = \int_{\Delta\Theta} \mu \,\mathrm{d}g(\mu)$, $g^*\{V \ni s_i^*\} = 1$, and $s_i^* \geq s_i$ will preserve all constraints and weakly raise the objective. Moreover, Lipnowski and Ravid (2020, Lemma 1 and Theorem 2) tell us that any $\gamma \in \Delta\Theta$ has $\max_{g\in\mathcal{R}(\gamma),s_i\in\mathbb{R}:\ g\{V\ni s_i\}=1} s_i = \bar{v}(\gamma)$.[38] Therefore,

$$\begin{aligned}
\tilde{v}^*_{\boldsymbol{\chi}}(\mu_0) = \sup_{\beta,\gamma\in\Delta\Theta,\ k\in[0,1],\ b\in\mathcal{R}(\beta)} & \left\{(1-k)\bar{v}(\gamma) + k\int_{\Delta\Theta} v_{\wedge\gamma} \,\mathrm{d}b\right\} \\
\text{s.t.} \qquad & k\beta + (1-k)\gamma = \mu_0,\ (1-k)\gamma \geq (\mathbf{1}-\boldsymbol{\chi})\mu_0, \\
& b\{\min V \leq \bar{v}(\gamma)\} = 1.
\end{aligned}$$

Trivially, the program (12) that defines $v^*_{\boldsymbol{\chi}}(\mu_0)$ is a relaxation of the above program; that is, for every feasible $(\beta, \gamma, k, b)$ for the above program, $(\beta, \gamma, k)$ is feasible in (12) and generates a weakly higher objective there; that is, $\tilde{v}^*_{\boldsymbol{\chi}}(\mu_0) \leq v^*_{\boldsymbol{\chi}}(\mu_0)$. We now prove the opposite inequality also holds, thereby completing the theorem's proof. Notice the program (12) has an upper semicontinuous objective and compact constraint set, and so admits some solution $(\beta, \gamma, k)$. We now argue some $(\tilde{\beta}, \tilde{\gamma}, \tilde{k}, b)$ exists that is feasible for the above program, and such that

$$(1-\tilde{k})\bar{v}(\tilde{\gamma}) + \tilde{k}\int v_{\wedge\tilde{\gamma}} \,\mathrm{d}b \geq k\hat{v}_{\wedge\gamma}(\beta) + (1-k)\bar{v}(\gamma),$$

and so $\tilde{v}^*_{\boldsymbol{\chi}}(\mu_0) \geq v^*_{\boldsymbol{\chi}}(\mu_0)$. If $\hat{v}_{\wedge\gamma}(\beta) < \bar{v}(\gamma)$, then Lemma 3 delivers $b$ such that $(\beta, \gamma, k, b)$ is as desired. Otherwise, $\hat{v}_{\wedge\gamma}(\beta) = \bar{v}(\gamma)$, and so quasiconcavity of $\bar{v}$ implies $\bar{v}(\mu_0) \geq k\hat{v}_{\wedge\gamma}(\beta) + (1-k)\bar{v}(\gamma)$—meaning $(\mu_0, \mu_0, 0, \delta_{\mu_0})$ is as desired. The theorem follows. $\square$

[38] Note that, $g\{V \ni s_i\} = 1$ implies $s_i \in \bigcap_{\mu\in\mathrm{supp}(g)} V(\mu)$ because $V$ is upper hemicontinuous.

### A.1.3 Consequences of Lemma 1 and Theorem 1

In this subsection we record some properties of the $\boldsymbol{\chi}$-equilibrium payoff set and S's favorite $\chi$-equilibrium payoff. We use these properties in the subsequent analysis.

**Corollary 1.** *The set of $\chi$-equilibrium outcomes $(p, s_o, s_i)$ at prior $\mu_0$ is a compact-valued, upper hemicontinuous correspondence of $(\mu_0, \chi)$ on $\Delta\Theta \times [0,1]$.*

*Proof.* Let $Y_G$ be the graph of $V$ and $Y_B$ be the graph of $[\min V, \max u_S(A)]$, both compact because $V$ is a Kakutani correspondence.

Let $X$ be the set of all $(\mu_0, p, g, b, \chi, k, s_o, s_i) \in (\Delta\Theta) \times (\Delta\Delta\Theta)^3 \times [0,1]^2 \times [\text{co } u_S(A)]^2$ such that:

- $kb + (1-k)g = p$;
- $(1-\chi)\int_{\Delta\Theta} \mu \, \mathrm{d}g(\mu) + \chi \int_{\Delta\Theta} \mu \, \mathrm{d}b(\mu) = \mu_0$;
- $(1-k)\int_{\Delta\Theta} \mu \, \mathrm{d}g(\mu) \geq (1-\chi)\mu_0$;
- $g \otimes \delta_{s_i} \in \Delta(Y_G)$ and $b \otimes \delta_{s_i} \in \Delta(Y_B)$;
- $k \int_{\Delta\Theta} \min V \, \mathrm{d}b \leq (k-\chi)\, s_i + \chi s_o \leq k \int_{\Delta\Theta} s_i \wedge v \, \mathrm{d}b$.

As an intersection of compact sets, $X$ is itself compact. By Lemma 1, the equilibrium outcome correspondence has a graph which is a projection of $X$, and so is itself compact. Therefore, it is compact-valued and upper hemicontinuous.

□

**Corollary 2.** *For any $\mu_0 \in \Delta\Theta$, the map*

$$\{\boldsymbol{\chi} : \Theta \to [0,1] : \ \boldsymbol{\chi} \textit{ measurable}\} \to \mathbb{R}$$
$$\boldsymbol{\chi} \mapsto v^*_{\boldsymbol{\chi}}(\mu_0)$$

*is weakly increasing.*

*Proof.* This result follows immediately from Theorem 1 (the general version, with state-dependent credibility, proven above) because increasing credibility weakly expands the constraint set. □

**Corollary 3.** *For any $\mu_0 \in \Delta\Theta$, the map*

$$\begin{aligned} [0,1] &\to \mathbb{R} \\ \chi &\mapsto v^*_\chi(\mu_0) \end{aligned}$$

*is weakly increasing and right-continuous.*

*Proof.* That it is weakly increasing is a specialization of Corollary 2. That it is upper semicontinuous (and so, since nondecreasing, it is right-continuous) follows directly from Corollary 1. □

**Corollary 4.** *For any $\chi \in [0,1]$, the map $v^*_\chi : \Delta\Theta \to \mathbb{R}$ is upper semicontinuous.*

*Proof.* This result is immediate from Corollary 1. □

## A.2 Productive Mistrust: Proof of Proposition 1

In this section, we prove Proposition 1 as stated in the main text. Whereas this proposition is stated for state-independent credibility, it immediately implies the following result for the case in which credibility is allowed to depend on the state:

**Corollary 5.** *Consider a finite and generic model in which S is not a two-faced SOB. Then, a full support prior and state-dependent credibility levels $\boldsymbol{\chi}' < \boldsymbol{\chi}$ exist such that every S-optimal $\boldsymbol{\chi}'$ equilibrium is strictly better for R than every S-optimal $\boldsymbol{\chi}$-equilibrium.*

As explained in the main text, one can divide the proof of Proposition 1 into two parts. The first part proves the proposition for the case in which $\Theta$ is binary. The second part uses a continuity argument to extend the binary-state result to any finite-state environment.

### A.2.1 Productive Mistrust with Binary States

We first verify our sufficient conditions for productive mistrust to occur in the binary-state world in the lemma below. In addition to being a special case of the proposition, it will also be an important lemma for proving the more general result.

To this end, it is useful to introduce a more detailed language for our key SOB condition. Given a prior $\mu \in \Delta\Theta$, say S is **an SOB at** $\mu$ if every $p \in \mathcal{R}(\mu)$ is outperformed by an SOB policy $p' \in \mathcal{R}(\mu)$.

**Lemma 4.** *Suppose $|\Theta| = 2$, the model is finite and generic, and a full-support belief $\mu \in \Delta\Theta$ exists such that the sender is not an SOB at $\mu$. Then, a full-support prior $\mu_0$ and credibility levels $\chi' < \chi$ exist such that every S-optimal $\chi'$-equilibrium is both strictly better for R and more Blackwell-informative than every S-optimal $\chi$-equilibrium.*

*Moreover, some full-support belief $\mu_+$ exists such that any solution $(\beta, \gamma, k)$ to the program in Theorem 1 at prior $\mu_0$ and credibility level in $\{\chi, \chi'\}$ has $\gamma = \mu_+$.*

*Proof.* First, notice that the genericity assumption delivers full-support $\mu'$ such that $V(\mu') = \{\max v(\Delta\Theta)\}$.

Name our binary-state space $\{0, 1\}$ and identify $\Delta\Theta = [0, 1]$ in the obvious way. The function $v : [0, 1] \to \mathbb{R}$ is upper semicontinuous and piecewise constant, which implies that its concave envelope $v_1^*$ is piecewise affine. That is, some $n \in \mathbb{N}$ and $\{\mu^i\}_{i=0}^n$ exist such that $0 = \mu^0 \le \cdots \le \mu^n = 1$ and $v_1^*|_{[\mu^{i-1},\mu^i]}$ is affine for every $i \in \{1, \ldots, n\}$. Taking $n$ to be minimal, we can assume that $\mu^0 < \cdots < \mu^n$ and the slope of $v_1^*|_{[\mu^{i-1},\mu^i]}$ is strictly decreasing in $i$. Therefore, some $i_0, i_1 \in \{0, \ldots, n\}$ exist such that $i_1 \in \{i_0, i_0 + 1\}$ and $\operatorname{argmax}_{\tilde\mu\in[0,1]} v_1^*(\tilde\mu) = [\mu^{i_0}, \mu^{i_1}]$. That the sender is not an SOB at $\mu$ implies that $i_0 > 1$ or $i_1 < n - 1$. Without loss of generality, say $i_0 > 1$. Now let $\mu_- \coloneqq \mu^{i_0-1}$ and $\mu_+ \coloneqq \mu^{i_0}$.

We now find a $\mu_0 \in (\mu_-, \mu_+)$ such that $\bar v|_{[\mu_0,\mu_+)}$ is constant and lies strictly below $v_1^*|_{[\mu_0,\mu_+)}$. To do so, recall the model is finite, and so $\bar v$ has finite range and is piecewise constant. It follows some $\epsilon > 0$ exists such that $\bar v$ is constant on $(\mu_+ - \epsilon, \mu_+)$. Since $v_1^*\colon [0, 1] \to \mathbb{R}$ is concave and upper semicontinuous, it is in fact continuous, and so admits an $\tilde\epsilon \in (0, \mu_+)$ such that every $\tilde\mu \in (\mu_+ - \tilde\epsilon, \mu_+)$ has

$$v_1^*(\tilde\mu) > \max\left[\bar v([0, 1]) \setminus \{\max \bar v([0, 1])\}\right] \ge \bar v(\tilde\mu),$$

where the last inequality follows from $\bar v|_{[0,\mu_+)} \le v_1^*|_{[0,\mu_+)} < v_1^*(\mu_+)$. Thus, the desired properties are satisfied by any $\mu_0 \in (\max\{\mu_-, \mu_+ - \epsilon, \mu_+ - \tilde\epsilon\}, \mu_+)$. Let $\mu_0$ be one such belief.

To summarize, the beliefs $\mu_-, \mu_0, \mu_+ \in [0, 1]$ are such that: $0 < \mu_- < \mu_0 < \mu_+$; $\hat v_{\wedge\mu_+} = \hat v = v_1^*$ is affine on $[\mu_-, \mu_+]$ and on no larger interval; $\hat v_{\wedge\mu_+}$ is strictly increasing on $[0, \mu_+]$; $v_0^* = \bar v$ is constant on $[\mu_0, \mu_+)$.

Let $\chi \in [0, 1]$ be the smallest credibility level such that $v_\chi^*(\mu_0) = v_1^*(\mu_0)$, which exists by Corollary 3. That $v_0^*(\mu_0) < v_1^*(\mu_0)$ implies $\chi > 0$. Notice $\mu_+$ has full support, because $0 \le \mu_- < \mu_+ \le \mu' < 1$. It follows that $\chi < 1$. Consider now the following claim.

Claim: *Given $\chi' \in [0, \chi]$, suppose that*

$$(\beta', \gamma', k') \in \text{argmax}_{(\beta,\gamma,k)\in[0,1]^3} \left\{ k\hat{v}_{\wedge\gamma}(\beta) + (1-k)\bar{v}(\gamma) \right\} \tag{13}$$
$$s.t. \qquad k\beta + (1-k)\gamma = \mu_0, \ (1-k)(\gamma, 1-\gamma) \geq (1-\chi')(\mu_0, 1-\mu_0),$$

*and the objective attains a value strictly higher than $\bar{v}(\mu_0)$. Then:*

- $\gamma' = \mu_+$ *and* $\beta' \leq \mu_-$.
- *If $b' \in \mathcal{R}(\beta')$ and $g' \in \mathcal{R}(\gamma')$ are such that $p' = k'b' + (1-k')g'$ is the information policy of an S-optimal $\chi'$-equilibrium, then $b'[0, \mu_-] = g'\{\mu_+\} = 1$.*

We now prove the claim.

Suppose first $\gamma' > \mu_+$ for a contradiction, and let $k'' > 0$ be the unique solution to $k''\beta' + (1-k'')\mu_+ = \mu_0$. Observe $k'' < k'$, and so

$$\begin{aligned}(1-k'')(\mu_+, 1-\mu_+) &= (\mu_0, 1-\mu_0) - k''(\beta', 1-\beta') \\ &\geq (\mu_0, 1-\mu_0) - k'(\beta', 1-\beta') \\ &= (1-k')(\gamma', 1-\gamma') \geq (1-\chi')(\mu_0, 1-\mu_0).\end{aligned}$$

Because

$$k''\hat{v}_{\wedge\mu_+}(\beta') + (1-k'')\bar{v}(\mu_+) \geq k''\hat{v}_{\wedge\gamma'}(\beta') + (1-k'')\bar{v}(\gamma') > k'\hat{v}_{\wedge\gamma'}(\beta') + (1-k')\bar{v}(\gamma'),$$

$(\beta', \mu_+, k'')$ is a feasible solution that would strictly outperform $(\beta', \gamma', k')$, contradicting optimality of $(\beta', \gamma', k')$. It follows $\gamma' \leq \mu_+$.

Next, notice that $\bar{v}$—as a weakly quasiconcave function which is nondecreasing and nonconstant over $[\mu_0, \mu_+]$—is nondecreasing over $[0, \mu_+]$. Moreover, $\lim_{\mu\nearrow\mu_+} \bar{v}(\mu) = \bar{v}(\mu_0) < \bar{v}(\mu_+)$. Therefore, if $\gamma' < \mu_+$, it would follow that $k'\hat{v}_{\wedge\gamma'}(\beta') + (1-k')\bar{v}(\gamma') \leq \bar{v}(\gamma') \leq \bar{v}(\mu_0)$. Given the hypothesis that $(\beta', \gamma', k')$ strictly outperforms $\bar{v}(\mu_0)$, it follows that $\gamma' = \mu_+$. A direct implication is that

$$(\beta', k') \in \text{argmax}_{(\beta,k)\in[0,1]^2} \left\{ k\hat{v}_{\wedge\mu_+}(\beta) + (1-k)\max v[0, \mu_+] \right\}$$
$$\text{s.t.} \qquad k\beta + (1-k)\mu_+ = \mu_0, \ (1-k)(1-\mu_+) \geq (1-\chi')(1-\mu_0).$$

Let us now see why we cannot have $\beta' \in (\mu_-, \mu_0)$. As $\hat{v}_{\wedge\mu_+}$ is affine on $[\mu_+, \mu_-]$, replacing such $(k', \beta')$ with $(k, \mu_-)$ which satisfies $k\mu_- + (1-k)\mu_+ = \mu_0$ necessarily has $(1-k)(\mu_+, 1-\mu_+) \gg (1-\chi')(\mu_0, 1-\mu_0)$. This would contradict minimality of $\chi$. Therefore, $\beta' \leq \mu_-$.

We now prove the second bullet. First, every $\mu < \mu_+$ satisfies $v(\mu) \leq v_1^*(\mu) < v_1^*(\mu_+) = v(\mu_+)$. This property implies that $\delta_{\mu_+}$ is the unique $g \in \mathcal{R}(\mu_+)$ with $\inf v(\operatorname{supp} g) \geq v(\mu_+)$. Therefore, $g' = \delta_{\mu_+}$. Second, the measure $b' \in \mathcal{R}(\beta')$ can be expressed as $b' = (1-\lambda)b_L + \lambda b_R$ for $b_L \in \Delta[0, \mu_-]$, $b_R \in \Delta(\mu_-, 1]$, and $\lambda \in [0, 1)$. Notice that $(\mu_-, v(\mu_-))$ is an extreme point of the subgraph of $v_1^*$, and therefore an extreme point of the subgraph of $\hat{v}_{\wedge\mu_+}$. Taking the unique $\hat{\lambda} \in [0, \lambda]$ such that $\hat{b} \coloneqq (1-\hat{\lambda})b_L + \hat{\lambda}\delta_{\mu_-} \in \mathcal{R}(\beta')$, it follows that $\int_{[0,1]} \hat{v}_{\wedge\mu_+} \, \mathrm{d}\hat{b} \geq \int_{[0,1]} \hat{v}_{\wedge\mu_+} \, \mathrm{d}b'$, strictly so if $\hat{\lambda} < \lambda$. But $\hat{\lambda} < \lambda$ necessarily if $\lambda > 0$, since $\int_{[0,1]} \mu \, \mathrm{d}\beta_R(\mu) > \mu_-$. Optimality of $b'$ then implies that $\lambda = 0$, i.e. $b'[0, \mu_-] = 1$. This observation completes the proof of the claim.

With the claim in hand, we can now prove the lemma. The claim implies that, for credibility level $\chi$, any solution $(\beta^*, \gamma^*, k^*)$ of the program (13) is such that $\gamma^* = \mu_+$, $k^* = \frac{\mu_+ - \mu_0}{\mu_+ - \beta^*}$, and $\beta^*$ solves

$$\max_{\beta \in [0, \mu_-]} \left\{ \frac{\mu_+ - \mu_0}{\mu_+ - \beta} \hat{v}_{\wedge\mu_+}(\beta) + \frac{\mu_0 - \beta}{\mu_+ - \beta} \bar{v}(\mu_+) \right\}.$$

Note that because $\bar{v}(\mu_+) = v(\mu_+) = \hat{v}_{\wedge\mu_+}(\mu_+)$, any $\beta \in [0, \mu_-]$ has

$$\frac{\mu_+ - \mu_0}{\mu_+ - \beta} \hat{v}_{\wedge\mu_+}(\beta) + \frac{\mu_0 - \beta}{\mu_+ - \beta} \hat{v}_{\wedge\mu_+}(\mu_+) \leq \hat{v}_{\wedge\mu_+} \left( \frac{\mu_+ - \mu_0}{\mu_+ - \beta} \beta + \frac{\mu_0 - \beta}{\mu_+ - \beta} \mu_+ \right) = \hat{v}_{\wedge\mu_+}(\mu_0)$$

by concavity of $\hat{v}_{\wedge\mu_+}$. Moreover, the inequality is strict for $\beta < \mu_-$ but holds with equality for $\beta = \mu_-$, since $\hat{v}_{\wedge\mu_+}$ is affine on $[\mu_-, \mu_+]$ and on no larger interval. Hence, the unique solution to (13) is $(\mu_-, \mu_+, k^*)$, where $k^*\mu_- + (1-k^*)\mu_+ = \mu_0$. Moreover, the minimality property defining $\chi$ implies that $(1-k^*)(1-\mu_+) = (1-\chi)(1-\mu_0)$.

Given $\chi' < \chi$ sufficiently close to $\chi$, one can verify directly that $(\beta', \mu_+, k')$ is feasible, where

$$k' \coloneqq 1 - \tfrac{1-\chi'}{1-\chi}(1-k^*) \text{ and } \beta' \coloneqq \tfrac{1}{k'} \left[ \mu_0 - (1-k')\mu_+ \right].$$

As $\hat{v}_{\wedge\mu_+}$ is a continuous function, it follows that $v_{\chi'}^*(\mu_0) \nearrow v_\chi^*(\mu_0)$ as $\chi' \nearrow \chi$. In particular, $v_{\chi'}^*(\mu_0) > v_0^*(\mu_0)$ for $\chi' < \chi$ sufficiently close to $\chi$. Fix such a $\chi'$.

Let $p'$ be any S-optimal $\chi'$-equilibrium information policy. Appealing to the claim, some $b' \in \mathcal{R}(\beta') \cap \Delta[0, \mu_-]$ exists such that $p' \in \text{co}\{b', \delta_{\mu_+}\}$. Therefore, $p'$ is weakly more Blackwell-informative than $p^*$. Finally, as $(1-k^*)(1-\mu_+) = (1-\chi)(1-\mu_0)$ and $\chi' < \chi$, feasibility of $p'$ tells us that $p' \neq p^*$. Therefore (the Blackwell order being antisymmetric), $p'$ is strictly more informative than $p^*$.

All that remains is to show that the receiver's optimal payoff is strictly higher given $p'$ than given $p^*$. To that end, fix sender-preferred receiver best responses $a_-$ and $a_+$ to $\mu_-$ and $\mu_+$, respectively. As the receiver's optimal value given $p^*$ is attainable using only actions $\{a_-, a_+\}$, and the same value is feasible given only information $p'$ and using only actions $\{a_-, a_+\}$, it suffices to show that there are beliefs in the support of $p'$ to which neither of $\{a_-, a_+\}$ is a receiver best response. But, every $\mu \in [0, \mu_-)$ satisfies

$$v(\mu) \leq \bar{v}(\mu) < \bar{v}(\mu_-) = \min\{\bar{v}(\mu_-), \bar{v}(\mu_+)\};$$

that is, $\max u_S \left(\text{argmax}_{a \in A}\, u_R(a, \mu)\right) < \min\{u_S(a_-), u_S(a_+)\}$. The result follows. □

#### A.2.2 Productive Mistrust with Many States: Proof of Proposition 1

Given Lemma 4, we need only prove the proposition for the case of $|\Theta| > 2$, which we do below. The proof intuition is as follows. Using the binary-state logic, one can always obtain a binary-support prior $\mu_0^\infty$ and credibility levels $\chi' < \chi$ such that R strictly prefers every S-optimal $\chi'$-equilibrium to every S-optimal $\chi$-equilibrium. We then find an interior direction through which to approach $\mu_0^\infty$, while keeping S's optimal equilibrium value under both credibility levels continuous. Genericity ensures that such a direction exists despite $\bar{v}$ being discontinuous. The continuity in S's value from the identified direction then ensures upper hemicontinuity of S's optimal equilibrium policy set; that is, the limit of every sequence of S-optimal equilibrium policies from said direction must also be optimal under $\mu_0^\infty$. Now, if the proposition were false, one could construct a convergent sequence of S-optimal equilibrium policies from said direction for each credibility level, $\{p_n^\chi, p_n^{\chi'}\}_{n \geq 0}$, such that R would weakly prefer $p_n^\chi$ to $p_n^{\chi'}$. As R's payoffs are continuous, R being weakly better off under $\chi$ than under $\chi'$ along the sequences would imply the same at the sequences' limits. Notice, though, that such limits must be S-optimal for the prior $\mu_0^\infty$ by the choice of direction, meaning that productive mistrust fails at $\mu_0^\infty$; that is, we have a contradiction. Below, we proceed with the formal proof.

*Proof.* Suppose some prior with binary support $\Theta_2 = \{\theta_1, \theta_2\}$ exists at which S is not

an SOB. Let $\bar{s} := \max v\,(\Delta\Theta_2)$, and define the R value function $v_R \colon \Delta\Delta\Theta \to \mathbb{R}$ via $v_R(p) := \int_{\Delta\Theta} \max_{a\in A} u_R(a,\mu)\,\mathrm{d}p(\mu)$. Lemma 4 delivers some $\mu_0^\infty \in \Delta\Theta$ with support $\Theta_2$ and credibility levels $\chi'' < \chi'$ such that every S-optimal $\chi''$-equilibrium is strictly better for R than every S-optimal $\chi'$-equilibrium. Consider the following claim.
Claim: Some sequence $\{\mu_0^n\}$ of full-support priors exists that converges to $\mu_0^\infty$ with

$$\liminf_{n\to\infty} v_\chi^*(\mu_0^n) \geq v_\chi^*(\mu_0^\infty) \text{ for } \chi \in \{\chi', \chi''\}.$$

Before proving the claim, let us argue that it implies the proposition. Given the claim, assume for contradiction that: for every $n \in \mathbb{N}$, prior $\mu_0^n$ admits some S-optimal $\chi'$-equilibrium and $\chi''$-equilibrium, $\Psi'_n = (p'_n, s'_{in}, s'_{on})$ and $\Psi''_n = (p''_n, s''_{i\,n}, s''_{on})$, respectively, such that $v_R(p'_n) \geq v_R(p''_n)$. Dropping to a subsequence if necessary, we may assume by compactness that $(\Psi'_n)_n$ and $(\Psi''_n)_n$ converge (in $\Delta\Delta\Theta \times [\text{co } u_S(A)]^2$) to some $\Psi' = (p', s'_i, s'_o)$ and $\Psi'' = (p'', s''_i, s''_o)$ respectively. By Corollary 1, for every credibility level $\chi$, the set of $\chi$-equilibria is an upper hemicontinuous correspondence of the prior. Therefore, $\Psi'$ and $\Psi''$ are $\chi'$- and $\chi''$-equilibria, respectively, at prior $\mu_0^\infty$. Continuity of $v_R$ (by Berge's theorem) then implies that $v_R(p') \geq v_R(p'')$. Finally, by the claim, it must be that $\Psi'$ and $\Psi''$ are S-optimal $\chi'$- and $\chi''$-equilibria, respectively, contradicting the definition of $\mu_0^\infty$. Therefore, some $n \in \mathbb{N}$ exists such that the full-support prior $\mu_0^n$ is as required for the proposition.

So all that remains is to prove the claim. To do this, we construct the desired sequence.

First, Lemma 4 delivers some $\gamma^\infty \in \Delta\Theta$ with support $\Theta_2$ such that $\bar{v}(\gamma^\infty) = \bar{s}$ and, for $\chi \in \{\chi', \chi''\}$, any solution $(\beta, \gamma, k)$ to the program in Theorem 1 at prior $\mu_0^\infty$ and credibility level $\chi$ has $\gamma = \gamma^\infty$.

Let us now show that there exists a closed convex set $D \subseteq \Delta\Theta$ which contains $\gamma^\infty$, has nonempty interior, and satisfies $\bar{v}|_D = \bar{s}$. Notice, first, that the genericity assumption delivers $\mu'$ with support $\Theta_2$ such that $V(\mu') = \{\bar{s}\}$. Then, for any $n \in \mathbb{N}$, let $B_n \subseteq \Delta\Theta$ be the closed ball (say with respect to the Euclidean metric) of radius $\frac{1}{n}$ around $\mu'$, and let $D_n := \text{co}\,[\{\gamma^\infty\} \cup B_n]$. As $v|_{\Delta\Theta_2} \leq \bar{s}$ and $\bar{v} = \max_{p\in\mathcal{R}(\cdot)} \inf v(\text{supp}(p))$ (see Lipnowski and Ravid 2020, Theorem 2), it follows $\bar{v}|_{\Delta\Theta_2} \leq \bar{s}$ as well. As $V$ is upper hemicontinuous, the hypothesis on $\mu'$ ensures that $\bar{v}|_{B_n} \geq v|_{B_n} = \bar{s}$ for sufficiently large $n \in \mathbb{N}$; quasiconcavity then tells us $\bar{v}|_{D_n} \geq \bar{s}$. Assume now, for a contradiction, that every $n \in \mathbb{N}$ has $\bar{v}|_{D_n} \not\leq \bar{s}$. That is, each $n \in \mathbb{N}$ admits some $\lambda_n \in [0,1]$ and $\mu'_n \in B_n$ such that $\bar{v}\left((1-\lambda_n)\gamma^\infty + \lambda_n \mu'_n\right) > \bar{s}$. In this case, each $n \in \mathbb{N}$ has $\bar{v}\left((1-\lambda_n)\gamma^\infty + \lambda_n \mu'_n\right) \geq \hat{s} := \min\left[\bar{v}(\Delta\Theta) \cap (\bar{s}, \infty)\right]$ (observe $\hat{s}$ is well-defined because $|\bar{v}(\Delta\Theta)| < \infty$ due to the model being finite). Dropping to a subsequence, we get a strictly increasing sequence $(n_\ell)_{\ell=1}^\infty$ of natural numbers such that (since $[0,1]$ is compact) $\lambda_{n_\ell} \xrightarrow{\ell\to\infty} \lambda \in [0,1]$ and $\bar{v}\left((1-\lambda_{n_\ell})\gamma^\infty + \lambda_{n_\ell}\mu'_{n_\ell}\right) \geq \hat{s}$ for every $\ell \in \mathbb{N}$. As $\bar{v}$ is upper semicontinuous, this would imply that $\bar{v}\left((1-\lambda)\gamma^\infty + \lambda\mu'\right) \geq \hat{s} > \bar{s}$, contradicting the definition of $\bar{s}$ and $\mu'$. Therefore, some $D \in \{D_{n_\ell}\}_{\ell=1}^\infty$ is as desired.

In what follows, let $\gamma_1 \in D$ be some interior element with full support. Then, for each $n \in \mathbb{N}$, define $\mu_0^n := \frac{n-1}{n}\mu_0^\infty + \frac{1}{n}\gamma_1$. We show that the sequence $(\mu_0^n)_{n=1}^\infty$—a sequence of full-support priors converging to $\mu_0^\infty$—is as desired. To that end, fix $\chi \in \{\chi', \chi''\}$ and some $(\beta, k) \in \Delta\Theta \times [0,1]$ such that $(\beta, \gamma^\infty, k)$ solves the program in Theorem 1 at prior $\mu_0^\infty$. Then, for any $n \in \mathbb{N}$, let:

$$
\begin{aligned}
\epsilon_n &:= \tfrac{1}{n-(n-1)k} \in (0,1], \\
\gamma_n &:= (1-\epsilon_n)\gamma^\infty + \epsilon_n\gamma_1 \in D, \\
k_n &:= \tfrac{n-1}{n}k \in [0,k).
\end{aligned}
$$

Given these definitions,

$$
\begin{aligned}
(1-k_n)\gamma_n &= \tfrac{1}{n}\left[n-(n-1)k\right]\gamma_n \\
&= \tfrac{1}{n}\left\{\left[n-(n-1)k-1\right]\gamma^\infty + \gamma_1\right\} \\
&= \tfrac{n-1}{n}(1-k)\gamma^\infty + \tfrac{1}{n}\gamma_1 \\
&\geq \tfrac{n-1}{n}(1-\chi)\mu_0^\infty + \tfrac{1}{n}\gamma_1 \geq (1-\chi)\mu_0^n,
\end{aligned}
$$

and

$$
\begin{aligned}
k_n\beta + (1-k_n)\gamma_n &= \tfrac{n-1}{n}k\beta + \tfrac{n-1}{n}(1-k)\gamma^\infty + \tfrac{1}{n}\gamma_1 \\
&= \tfrac{n-1}{n}\mu_0^\infty + \tfrac{1}{n}\gamma_1 = \mu_0^n.
\end{aligned}
$$

Therefore, $(\beta, \gamma_n, k_n)$ is $\chi$-feasible at prior $\mu_0^n$. As a result,

$$
\begin{aligned}
v_\chi^*(\mu_0^n) &\geq k_n\hat{v}_{\wedge\gamma_n}(\beta) + (1-k_n)\bar{v}(\gamma_n) \\
&= k_n\hat{v}_{\wedge\gamma}(\beta) + (1-k_n)\bar{v}(\gamma) \text{ (since } \bar{v}(\gamma_n) = u) \\
&\xrightarrow{n\to\infty} k\hat{v}_{\wedge\gamma}(\beta) + (1-k)\bar{v}(\gamma) = v_\chi^*(\mu_0^\infty).
\end{aligned}
$$

This proves the claim, and so too the proposition. $\square$

## A.3 Collapse of Trust: Proof of Proposition 2

*Proof.* Let us establish a four-way equivalence between the three conditions in the proposition's statement and the following state-dependent-credibility analogue of condition (i):

$(i)'$ Every $\boldsymbol{\chi} \in [0,1]^\Theta$ and full-support prior $\mu_0$ have $\lim_{\boldsymbol{\chi}' \nearrow \boldsymbol{\chi}} v^*_{\boldsymbol{\chi}'}(\mu_0) = v^*_{\boldsymbol{\chi}}(\mu_0)$, where convergence of $\boldsymbol{\chi}'(\cdot) \to \boldsymbol{\chi}(\cdot)$ is in the Euclidean topology on $\mathbb{R}^\Theta$.

Three of four implications are easy given Corollary 2. First, $(i)'$ trivially implies $(i)$. Second ($(iii)$ implies $(ii)$), if there is no conflict, then Lipnowski and Ravid (2020, Lemma 1) tells us that there is a 0-equilibrium with full information that generates sender value $\max v(\Delta\Theta) \geq v_1^*$; in particular, $v_0^* = v_1^*$. Third ($(ii)$ implies $(i)'$), if $v_0^* = v_1^*$, then Corollary 2 implies $v^*_{\boldsymbol{\chi}}$ is constant in $\boldsymbol{\chi}$, ruling out a collapse of trust (even under state-dependent credibility). Below we show that any conflict whatsoever implies a collapse of trust, that is, a failure of $(iii)$ implies a failure of $(i)$.

Suppose there is conflict; that is, $\min_{\theta\in\Theta} v(\delta_\theta) < \max v(\Delta\Theta)$ or, equivalently, $\min_{\theta\in\Theta} \bar{v}(\delta_\theta) < \max \bar{v}(\Delta\Theta)$. Taking a positive affine transformation of $u_S$, we may assume without loss that $\min \bar{v}(\Delta\Theta) = 0$ and (since $\bar{v}(\Delta\Theta) \subseteq u_S(A)$ is finite) $\min[\bar{v}(\Delta\Theta) \setminus \{0\}] = 1$. The set $D \coloneqq \arg\min_{\mu\in\Delta\Theta} \bar{v}(\mu) = \bar{v}^{-1}(-\infty, 1)$ is then open and nonempty. We can then consider some full-support prior $\mu_0 \in D$. For any scalar $\hat{\chi} \in [0,1]$, let

$$\Gamma(\hat{\chi}) \coloneqq \{(\beta,\gamma,k) \in \Delta\Theta \times (\Delta\Theta \setminus D) \times [0,1] \colon k\beta + (1-k)\gamma = \mu_0,\ (1-k)\gamma \geq (1-\hat{\chi})\mu_0\},$$

and $K(\hat{\chi})$ be its projection onto its last coordinate. As the correspondence $\Gamma$ is upper hemicontinuous and increasing (with respect to set containment), $K$ inherits the same properties. Next, notice that $K(1) \ni 1$ (as $\bar{v}$ is nonconstant by the hypothesis that a conflict exists, so that $\Delta\Theta \neq D$) and $K(0) = \varnothing$ (as $\mu_0 \in D$). Therefore, $\chi \coloneqq \min\{\hat{\chi} \in [0,1] :\ K(\hat{\chi}) \neq \varnothing\}$ exists and belongs to $(0,1]$.

Given any scalar $\chi' \in [0,\chi)$, it must be that $K(\chi') = \varnothing$. That is, if $\beta,\gamma \in \Delta\Theta$ and $k \in [0,1]$ with $k\beta + (1-k)\gamma = \mu_0$ and $(1-k)\gamma \geq (1-\chi')\mu_0$, then $\gamma \in D$. By Theorem 1, then, $v^*_{\chi'}(\mu_0) = \bar{v}(\mu_0) = 0$. There is, however, some $k \in K(\chi)$. By Theorem 1 and the definition of $\Gamma$, there is therefore a $\chi$-equilibrium generating ex-ante sender payoff of at least $k \cdot 0 + (1-k)\cdot 1 = (1-k) \geq (1-\chi)$. If $\chi < 1$, a collapse of trust occurs at credibility level $\chi$.

The only remaining case is the case that $\chi = 1$. In this case, some $\epsilon \in (0,1)$ and

$\mu \in \Delta\Theta \setminus D$ exist such that $\epsilon\mu \leq \mu_0$. Then

$$v^*_\chi(\mu_0) \geq \epsilon\bar{v}(\mu) + (1-\epsilon)\bar{v}\left(\tfrac{\mu_0-\epsilon\mu}{1-\epsilon}\right) \geq \epsilon.$$

So again, a collapse of trust occurs at credibility level $\chi$. $\square$

## A.4 Robustness: Proof of Proposition 3

Before proving the proposition, let us briefly observe that the proposition as stated is equivalent to the analogous statement for state-dependent credibility. Indeed, given Corollary 2, any prior $\mu_0$ and state-dependent credibility $\boldsymbol{\chi}$ has $v^*_\chi(\mu_0) \leq v^*_{\boldsymbol{\chi}}(\mu_0) \leq v^*_1(\mu_0)$ for $\chi = \min_{\theta\in\Theta} \boldsymbol{\chi}(\theta) \in [0,1]$. It follows immediately that $\lim_{\chi\nearrow 1} v^*_\chi(\mu_0) = v^*_1(\mu_0)$ if and only if $\lim_{\boldsymbol{\chi}\nearrow\mathbf{1}} v^*_{\boldsymbol{\chi}}(\mu_0) = v^*_{\mathbf{1}}(\mu_0)$, where convergence of $\boldsymbol{\chi} \to \mathbf{1}$ is in the Euclidean topology on $\mathbb{R}^\Theta$. That is, the stronger property of robustness of the commitment value to small *state-dependent* departures from perfect credibility is equivalent to that stated in the proposition.

We now proceed to proving the proposition for the case of state-independent credibility.

*Proof.* By Lipnowski and Ravid (2020, Lemma 1 and Theorem 2), S gets the benefit of the doubt (i.e. every $\theta \in \Theta$ is in the support of some member of $\operatorname{argmax}_{\mu\in\Delta\Theta} v(\mu)$) if and only if there is some full-support $\gamma \in \Delta\Theta$ such that $\bar{v}(\gamma) = \max v(\Delta\Theta)$.

First, given a full-support prior $\mu_0$, suppose $\gamma \in \Delta\Theta$ is full-support with $\bar{v}(\gamma) = \max v(\Delta\Theta)$. It follows immediately that $\hat{v}_{\wedge\gamma} = \hat{v} = v^*_1$. Let $r_0 := \min_{\theta\in\Theta} \frac{\mu_0\{\theta\}}{\gamma\{\theta\}} \in (0,\infty)$

and $r_1 \coloneqq \max_{\theta\in\Theta} \frac{\mu_0\{\theta\}}{\gamma\{\theta\}} \in [r_0, \infty)$. Then Theorem 1 tells us that, for $\chi \in \left[\frac{r_1 - r_0}{r_1}, 1\right)$:

$$
\begin{aligned}
v^*_\chi(\mu_0) \geq & \sup_{\beta\in\Delta\Theta,\ k\in[0,1]} \left\{ k v^*_1(\beta) + (1-k)v(\gamma) \right\} \\
& \quad \text{s.t.} \qquad k\beta + (1-k)\gamma = \mu_0,\ (1-k)\gamma \geq (1-\chi)\mu_0 \\
= & \sup_{k\in[0,1]} \left\{ k v^*_1 \left( \tfrac{\mu_0 - (1-k)\gamma}{k} \right) + (1-k)v(\gamma) \right\} \\
& \quad \text{s.t.} \quad (1-\chi)\mu_0 \leq (1-k)\gamma \leq \mu_0 \\
\geq & \sup_{k\in[0,1]} \left\{ k v^*_1 \left( \tfrac{\mu_0 - (1-k)\gamma}{k} \right) + (1-k)v(\gamma) \right\} \\
& \quad \text{s.t.} \quad (1-\chi)r_1 \leq (1-k) \leq r_0 \\
\geq & \sup_{k\in[0,1]} \left\{ k v^*_1 \left( \tfrac{\mu_0 - (1-k)\gamma}{k} \right) + (1-k)v(\gamma) \right\} \\
& \quad \text{s.t.} \quad (1-\chi)r_1 = (1-k) \\
= & \left[1 - (1-\chi)r_1\right] v^*_1 \left( \tfrac{\mu_0 - (1-\chi)r_1\gamma}{1-(1-\chi)r_1} \right) + (1-\chi)r_1 v(\gamma).
\end{aligned}
$$

But notice that $v^*_1$, being a concave function on a finite-dimensional space, is continuous on the interior of its domain. Therefore, $v^*_1 \left( \frac{\mu_0 - (1-\chi)r_1\gamma}{1-(1-\chi)r_1} \right) \to v^*_1(\mu_0)$ as $\chi \to 1$, implying $\liminf_{\chi\nearrow 1} v^*_\chi(\mu_0) \geq v^*_1(\mu_0)$. Finally, monotonicity of $\chi \mapsto v^*_\chi(\mu_0)$ implies $v^*_\chi(\mu_0) \to v^*_1(\mu_0)$ as $\chi \to 1$. That is, persuasion is robust to limited commitment.

Conversely, suppose that S does not get the benefit of the doubt (which of course implies $v$ is non-constant). Taking an affine transformation of $u_S$, we may assume without loss that $\max v(\Delta\Theta) = 1$ and (since $v(\Delta\Theta) \subseteq u_S(A)$ is finite) $\max[\bar{v}(\Delta\Theta) \setminus \{1\}] = 0$. Fix any full-support prior $\mu_0$, and consider any credibility level $\chi \in [0, 1)$. For any $\beta, \gamma \in \Delta\Theta,\ k \in [0, 1]$ with $k\beta + (1-k)\gamma = \mu_0$ and $(1-k)\gamma \geq (1-\chi)\mu_0$, that S does not get the benefit of the doubt implies (say by Lipnowski and Ravid (2020, Theorem 1)) that $\bar{v}(\gamma) \leq 0$, and therefore that $k\hat{v}_{\wedge\gamma}(\beta) + (1-k)v(\gamma) \leq 0$. Theorem 1 then implies that $v^*_\chi(\mu_0) \leq 0$.

Fix some full-support $\mu_1 \in \Delta\Theta$ and some $\gamma \in \Delta\Theta$ with $v(\gamma) = 1$. For any $\epsilon \in (0, 1)$, the prior $\mu_\epsilon \coloneqq (1-\epsilon)\gamma + \epsilon\mu_1$ has full support and satisfies

$$v^*_1(\mu_\epsilon) \geq (1-\epsilon)v(\gamma) + \epsilon v(\mu_1) \geq (1-\epsilon) + \epsilon \cdot \min v(\Delta\Theta).$$

For sufficiently small $\epsilon$, then, $v^*_1(\mu_\epsilon) > 0$. Persuasion is therefore not robust to limited commitment at prior $\mu_\epsilon$. $\square$

## A.5 Persuading the Public: Proofs from section 5

In this section, our approach is similar to Appendix A.1: At a minimal notational cost, we prove all results for a more general model of state-dependent credibility (section 6.3). The corresponding definitions are generalized in a straightforward way: The cutoff $\bar{\theta}_{\boldsymbol{\chi}}$ is the unique root of $\theta \mapsto \int (\theta - \bar{\theta}_{\boldsymbol{\chi}})(1 - \mathbf{1}_{[0,\bar{\theta}_{\boldsymbol{\chi}})}\boldsymbol{\chi}(\theta))\, \mathrm{d}\mu_0(\theta)$, and the posterior mean distribution induced by the $\bar{\theta}_{\boldsymbol{\chi}}$-upper-censorship pair is $\bar{\mu}_{\boldsymbol{\chi}} \coloneqq \mathbf{1}_{[0,\bar{\theta}_{\boldsymbol{\chi}})}\boldsymbol{\chi}\mu_0 + \big(1 - \chi\mu_0[0,\bar{\theta}_{\boldsymbol{\chi}})\big)\, \delta_{\bar{\theta}_{\boldsymbol{\chi}}}$.

### A.5.1 Mathematical preliminaries

In this subsection, we document some notations and basic properties that are useful for the present case of $\Theta = [0,1]$, with the sender's value depending only on the receiver's posterior expectation of the state. This environment is studied by Gentzkow and Kamenica (2016) and others. Throughout the subsection, let $\theta_0 \coloneqq E\mu_0$ be the prior mean; let

$$\mathcal{I} \coloneqq \{I : \mathbb{R}_+ \to \mathbb{R}_+ :\ I \text{ convex},\ I(0) = 0,\ I|_{[1,\infty)} \text{ affine}\};$$

let $I'$ denote the right-hand-side derivative of $I$ for any $I \in \mathcal{I}$; and let

$$\mathcal{I}(I) \coloneqq \{\hat{I} \in \mathcal{I} :\ I'(1) = \hat{I}'(1),\ I(1) = \hat{I}(1),\ \hat{I} \le I\}$$

for any $I \in \mathcal{I}$.

As Fact 1 below clarifies, members of $\mathcal{I}$ are in natural bijective correspondence with measures over $[0,1]$, where a member of $\mathcal{I}$ is derived from a measure by taking the running integral of its cumulative distribution function. Moreover, working with members of $\mathcal{I}$, rather than directly in terms of the measures they represent, is particularly convenient for reasoning about information. Indeed, if $I \in \mathcal{I}$ represents a prior distribution over the state, then Fact 2 implies members of the (geometrically interpretable) set $\mathcal{I}(I) \subseteq \mathcal{I}$ correspond exactly to distributions over posterior expectations that can be implemented by some Blackwell experiment.

**Fact 1.** *Let $\mathcal{M}$ be the set of finite positive Borel measures on $\Theta$.*

1. *For any $\eta \in \mathcal{M}$, the function $I_\eta : \mathbb{R}_+ \to \mathbb{R}_+$ given by $\bar{\theta} \mapsto \int_0^{\bar{\theta}} \eta[0,\theta]\,\mathrm{d}\theta$ is a member of $\mathcal{I}$.*

2. *For any $I \in \mathcal{I}$, the function $I'$ is the CDF of some $\eta \in \mathcal{M}$ such that $I_\eta = I$.*

3. *Any $\eta \in \mathcal{M}$ has total mass $I'_\eta(1)$ and, if $\eta \in \Delta\Theta$, has barycenter $E\eta = 1 - I_\eta(1)$.*

The proof of the above fact is immediate, invoking the fundamental theorem of calculus for the second point and integration by parts for the third.

**Fact 2.** *Given $\mu, \hat{\mu} \in \Delta\Theta$, the following are equivalent:*

1. *$\hat{\mu} = p \circ E^{-1}$ for some $p \in \mathcal{R}(\mu)$.*

2. *$\mu$ is a mean-preserving spread of $\hat{\mu}$.*

3. *$I_{\hat{\mu}} \in \mathcal{I}(I_\mu)$.*

That the last two points are equivalent is immediate from the definition of a mean-preserving spread. Equivalence between these conditions and the first is as described in Gentzkow and Kamenica (2016). To apply their results, given $\mu \in \Delta\Theta$, notice that:

- A convex function $I : [0,1] \to \mathbb{R}$ with $I(\theta) \leq I_\mu(\theta)$ and $I(\theta) \geq (\theta - E\mu)_+$ for every $\theta \in [0,1]$ extends (by letting it take slope 1 on $[1,\infty)$) to a member of $\mathcal{I}(I_\mu)$.

- Every element $I \in \mathcal{I}(I_\mu)$ has, for each $\theta \in [0,1]$,

$$I(\theta) - (\theta - E\mu) = \int_\theta^1 [1 - I'(\tilde{\theta})]\,\mathrm{d}\tilde{\theta} \geq 0,$$

  so that $I(\theta) \geq (\theta - E\mu)_+ = \max\{I_\mu(1) - I'_\mu(1)(1-\theta),\ 0\}$.

### A.5.2 Characterizing S-optimal equilibrium

This section begins by constructing a class of perturbations to a given element of $\mathcal{I}$. Using these perturbations, we provide an elementary proof that upper-censorship is an optimal persuasion policy in the one-dimensional mean-measurable model when S's objective is convex-concave. We then establish uniqueness of $\bar{\theta}_{\boldsymbol{\chi}}$ satisfying (the state-dependent version of) ($\theta^*$-IC) and characterize $I_{\bar{\mu}_{\boldsymbol{\chi}}}$. We conclude by proving (the state-dependent version of) Claim 1.

**Lemma 5.** *Suppose* $\bar{I} \in \mathcal{I}$, $I \in \mathcal{I}(\bar{I})$, *and* $\omega \in [0, 1]$. *Then some* $\theta^* \in [0, \omega]$, $\theta^{**} \in [\omega, 1]$, *and* $I^* \in \mathcal{I}(\bar{I})$ *exist such that:*

- $I^* = \bar{I}$ *on* $[0, \theta^*]$, $I^*$ *is affine on* $[\theta^*, \theta^{**}]$, *and* $I^{*\prime}(\theta) = 1$ *on* $[\theta^{**}, \infty)$*;*
- $I^* - I$ *is nonnegative on* $[0, \omega]$ *and nonpositive on* $[\omega, \infty)$*.*

The proof of the lemma is constructive. While tedious to formally verify that the construction is as desired, it is intuitive to picture. We illustrate in Figure 8. Given the curves $I$ and $\bar{I}$, we wish to construct the curve $I^* \in \mathcal{I}(\bar{I})$. In order to ensure that $I^*$ has the required level and slope at $\theta = 1$, we will construct it to lie above the tangent line $\theta \mapsto \theta - \theta_0$ of $\bar{I}$ at 1. Now, consider positively sloped lines through the point $(\omega, I(\omega))$. Convexity of $\bar{I}$ ensures that some such line lies everywhere below the graph of $\bar{I}$, whence continuity delivers such a line of shallowest slope. This line is necessarily tangent to $\bar{I}$ somewhere to the left of $\omega$: this point will be our $\theta^*$. The same line intersects the tangent line $\theta \mapsto \theta - \theta_0$ to the right of $\omega$: this will be our $\theta^{**}$. Finally, we construct $I^*$ to coincide with upper bound function $\bar{I}$ to the left of $\theta^*$, the $\theta^*$ tangent line on $[\theta^*, \theta^{**}]$, and the 1 tangent line $\theta \mapsto \theta - \theta_0$ to the right of $\theta^{**}$.

*Proof.* Let $\Lambda := \{\lambda \in [0, I'(\omega)] : \ I(\omega) - \lambda(\omega - \theta) \leq \bar{I}(\theta) \text{ for all } \theta \in [0, \omega]\}$. The set $\Lambda$ is closed because $\bar{I}$ is continuous, and it contains $I'(\omega)$ because $I$ is convex and below $\bar{I}$. So let $\lambda := \min \Lambda$. Minimality of $\lambda$, together with the fact that the continuous function $\theta \mapsto I(\omega) - \lambda(\omega - \theta) - \bar{I}(\theta) : [0, \omega] \to \mathbb{R}$ attains a maximum, implies some $\theta^* \in [0, \omega]$ has $I(\omega) - \lambda(\omega - \theta^*) = \bar{I}(\theta^*)$. We can then construct the function

$$
\begin{aligned}
I^* : \mathbb{R}_+ &\to \mathbb{R}_+ \\
\theta &\mapsto \begin{cases} \bar{I}(\theta) & : \ 0 \leq \theta \leq \theta^* \\ I(\omega) - \lambda(\omega - \theta) & : \ \theta^* \leq \theta \leq \omega \\ \max\{I(\omega) + \lambda(\theta - \omega), \ I(1) - I'(1)(1 - \theta)\} & : \ \omega \leq \theta. \end{cases}
\end{aligned}
$$

Note, that $I$ is convex implies $I(\omega) + \lambda(\omega - \omega) \geq I(1) - (1 - \omega) I'(1)$, which in particular ensures that $I(\omega)$ is well-defined. That $I$ is convex and $\lambda \leq I'(\omega)$ implies $I(\omega) + \lambda(1 - \omega) \leq I(1) - I'(1)(1 - 1)$. So some $\theta^{**} \in [\omega, 1]$ has $I^*(\theta) = I(\omega) + \lambda(\theta - \omega)$ for $\theta \in [\omega, \theta^{**}]$ and equal to $I(1) - I'(1)(1 - \theta)$ for $\theta \in [\theta^{**}, \infty)$, verifying the first bullet. Moreover, because $\lambda \geq 0$ and $I(1) - I'(1)(1 - \theta) \leq I(\theta) \leq \bar{I}(\theta)$ for each $\theta \in [0, 1]$, it follows by construction that $I^* - I$ is nonpositive on $[\omega, 1]$ and nonnegative on $[0, \omega]$.

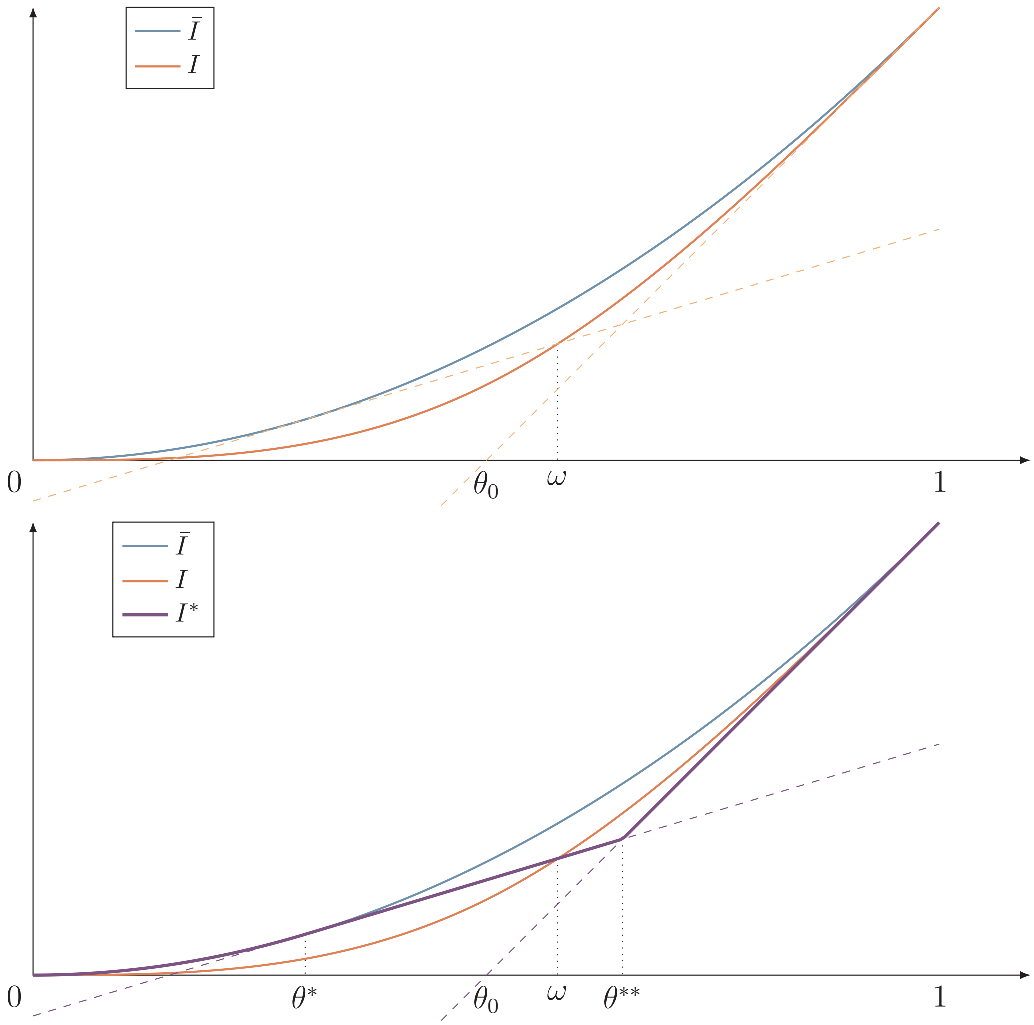


Figure 8: Construction of $(\theta^*, \theta^{**}, I^*)$ in Lemma 5

All that remains is to show that $I^* \in \mathcal{I}(\bar{I})$. Letting $\underline{I} : \mathbb{R}_+ \to \mathbb{R}_+$ be given by $\underline{I}(\theta) \coloneqq \max\{\bar{I}(1) - \bar{I}'(1)(1-\theta),\ 0\}$, we need to check that $\underline{I} \leq I^* \leq \bar{I}$ and $I^*$ is convex.

On $[0, \theta^*]$, we have $I^* = \bar{I} \geq \underline{I}$. On $[\theta^*, \omega]$, we have shown that $I^* \geq I \geq \underline{I}$, and we know $I^* \leq \bar{I}$ by the definition of $\lambda$. On $[\omega, \infty)$, we have shown that $I^* \leq I \leq \bar{I}$, and we have $I^* \geq \underline{I}$ by definition. So $\underline{I} \leq I^* \leq \bar{I}$ globally.

Finally, we verify convexity. Because the two affine functions coincide at $\theta^{**} \geq \theta^*$, we know that $I^*(\theta) = \max\{I(\omega) + \lambda(\theta - \omega),\ I(1) - (1-\theta)I'(1)\}$ for $\theta \in [\theta^*, \infty)$. A maximum of two affine functions, $I^*|_{[\theta^*,\infty)}$ is convex. Moreover, $I^*|_{[0,\theta^*]}$ is convex. Globally convexity then follows if $I^*$ is subdifferentiable at $\theta^*$. But $\lambda$ is a subdifferential of $\bar{I} \geq I^*$ at $\theta^*$, and the two functions coincide at $\theta^*$. It is therefore a subdifferential for $I^*$ at the same, as required. $\square$

**Lemma 6.** *Suppose $\tilde{H} : \Theta \to \mathbb{R}$ has $\tilde{H}(\cdot) = \tilde{H}(0) + \int_0^{(\cdot)} \tilde{h}(\theta)\,\mathrm{d}\theta$ for some $\tilde{h}$ of bounded variation. Then, for any $I, \hat{I} \in \mathcal{I}$ such that $I(1) - \hat{I}(1) = I'(1) - \hat{I}'(1) = 0$, we have*

$$\left[\tilde{H}(0)\hat{I}'(0) + \int_0^1 \tilde{H}\,\mathrm{d}\hat{I}'\right] - \left[\tilde{H}(0)I'(0) + \int_0^1 \tilde{H}\,\mathrm{d}I'\right] = \int_0^1 (\hat{I} - I)\,\mathrm{d}\tilde{h}.$$

*In particular, if $\tilde{H}$ is piecewise convex-or-concave; $\tilde{h}$ is a one-sided derivative of $\tilde{H}$; and $I, \hat{I} \in \mathcal{I}(\bar{I})$ for some $\bar{I} \in \mathcal{I}$ with $\bar{I}'(0) = 0$; then*

$$\int_0^1 \tilde{H}\,\mathrm{d}\hat{I}' - \int_0^1 \tilde{H}\,\mathrm{d}I' = \int_0^1 (\hat{I} - I)\,\mathrm{d}\tilde{h}.$$

*Proof.* Given the first formula, the last sentence follows from the fact that a piecewise monotone function on $[0,1]$ has bounded variation, the definition of $\mathcal{I}(\bar{I})$, and the observation that $I'(0) = \hat{I}'(0) = 0$ if $0 \leq I, \hat{I} \leq \bar{I}$ and $\bar{I}'(0) = 0$. To see the first formula holds, observe integration by parts yields

$$\begin{aligned}
&\left[\tilde{H}(0)\hat{I}'(0) + \int_0^1 \tilde{H}\,\mathrm{d}\hat{I}'\right] - \left[\tilde{H}(0)I'(0) + \int_0^1 \tilde{H}\,\mathrm{d}I'\right] \\
&= \tilde{H}(0)(\hat{I} - I)'(0) + \int_0^1 \tilde{H}\,\mathrm{d}(\hat{I} - I)' \\
&= \tilde{H}(0)(\hat{I} - I)'(0) + \left[(\hat{I} - I)'\tilde{H}\right]_0^1 - \int_0^1 (\hat{I} - I)'\,\mathrm{d}\tilde{H} \\
&= -\int_0^1 (\hat{I} - I)'(\theta)\tilde{h}(\theta)\,\mathrm{d}\theta \\
&= -\left[(\hat{I} - I)\tilde{h}\right]_0^1 + \int_0^1 (\hat{I} - I)\,\mathrm{d}\tilde{h} \\
&= \int_0^1 (\hat{I} - I)\,\mathrm{d}\tilde{h}.
\end{aligned}$$

□

We now complete our elementary proof that upper censorship is an optimal persuasion rule for convex-concave objectives. Recall, for $\theta^* \in [0,1]$ and $\mu \in \Delta\Theta$, a $\theta^*$ upper censorship of $\mu$ is

$$\mathbf{1}_{[0,\theta^*)}\mu + \mu[\theta^*, 1]\delta_{\mathbb{E}_\mu[\theta \mid \theta \geq \theta^*]} \in \Delta\Theta$$

if $\mu[\theta^*, 1] > 0$, and simply $\mu$ if $\mu[\theta^*, 1] = 0$.

**Lemma 7.** *Suppose $\tilde{H} : \Theta \to \mathbb{R}$ is continuous, and $\omega \in [0,1]$ is such that $\tilde{H}$ is* *(strictly) convex on $[0,\omega]$ and (strictly) concave on $[\omega,1]$. Then, if $\bar{\mu}\in\Delta\Theta$ has no atoms $<\max\operatorname{supp}(\bar{\mu})$, some (every) solution to $\max_{\mu\in\Delta\Theta:\ \mu\preceq\bar{\mu}}\int\tilde{H}\,\mathrm{d}\mu$ is a $\theta^*$ upper censorship of $\bar{\mu}$ for some $\theta^*\in[0,\omega]$. Moreover, if $\bar{\mu}(\theta^*,1]>0$, then $\mathbb{E}_{\bar{\mu}}[\theta\mid\theta\geq\theta^*]\geq\omega$; hence, this $\theta^*$ upper censorship puts probability $1$ on $[0,\theta^*]\cup[\omega,1]$.*

*Proof.* Let $\mu$ be a solution to the given program. Taking $\bar{I}:=I_{\bar{\mu}}$ and $I:=I_\mu$, note that the conditions of Lemma 5 are satisfied. Let $I^*\in\mathcal{I}$, $\theta^*\in[0,\omega]$, and $\theta^{**}\in[\omega,1]$ be as delivered by Lemma 5 and $\mu^*\in\Delta\Theta$ be such that $I^*=I_{\mu^*}$; recall this $\mu^*\preceq\bar{\mu}$. Then, by Lemma 6 (letting $\tilde{h}:=\tilde{H}'$),

$$\int_0^1\tilde{H}\,\mathrm{d}\mu^*-\int_0^1\tilde{H}\,\mathrm{d}\mu=\int_0^\omega(I^*-I)\,\mathrm{d}\tilde{h}+\int_\omega^1(I-I^*)\,\mathrm{d}(-\tilde{h}).$$

As $\tilde{h}$ is (strictly) increasing on $[0,\omega)$ and (strictly) decreasing on $[\omega,1]$, it follows from the definition of $I^*$ that $\int_0^1\tilde{H}\,\mathrm{d}\mu^*\geq\int_0^1\tilde{H}\,\mathrm{d}\mu$, (strictly so, given continuity of $I^*-I$, unless $I=I^*$). Optimality of $\mu$ then tells us that $\mu^*$ is optimal (and equal to $\mu$).

Now, let us establish that $\mu^*$ is a $\theta^*$ upper censorship of $\bar{\mu}$. First, that $I^{*\prime}|_{[\theta^{**},\infty)}=1$ implies $\mu^*(\theta^{**},1]=0$; and that $I^*|_{[\theta^*,\theta^{**}]}$ is affine implies $\mu^*(\theta^*,\theta^{**})=0$. Hence, $[\theta^*,1]\cap\operatorname{supp}(\mu^*)\subseteq\{\theta^*,\theta^{**}\}$. Further, because $\bar{\mu}\{\theta^*\}=0$ if $\bar{\mu}(\theta^*,1]>0$ (by hypothesis), we have $|[\theta^*,1]\cap\operatorname{supp}(\mu^*)|\leq 1$. Also by construction, $\mu^*[0,\theta]=\bar{\mu}[0,\theta]$ for every $\theta\in[0,\theta^*)$. As these properties—which are clearly also satisfied by a $\theta^*$ upper censorship of $\bar{\mu}$—characterize a unique distribution of any given mean, $\mu^*$ is a $\theta^*$ upper censorship of $\bar{\mu}$.

Finally, the "moreover" point follows from $\theta^{**}\geq\omega$, as guaranteed by Lemma 5, and the fact that $[\theta^*,1]\cap\operatorname{supp}(\mu^*)\subseteq\{\theta^*,\theta^{**}\}$. $\square$

Recall now our notational convention from section 6.3: Given a bounded and measurable $f\colon\Theta\to\mathbb{R}$ and $\mu\in\Delta\Theta$, let $f\mu$ denote the measure on $\Theta$ given by $f\mu(\hat{\Theta}):=\int_{\hat{\Theta}}f\,\mathrm{d}\mu$.

**Lemma 8.** *A unique $\bar{\theta}_{\boldsymbol{\chi}}\in[0,1]$ exists such that*[39] $I_{\boldsymbol{\chi}\mu_0}(\theta^*)=\int_0^{\theta^*}\boldsymbol{\chi}\mu_0[0,\theta]\,\mathrm{d}\theta$ *is*

$$\begin{cases}>\theta^*-\theta_0 & \textit{for } \theta^*\in[0,\bar{\theta}_{\boldsymbol{\chi}})\\ =\theta^*-\theta_0 & \textit{for } \theta^*=\bar{\theta}_{\boldsymbol{\chi}}\\ <\theta^*-\theta_0 & \textit{for } \theta^*\in(\bar{\theta}_{\boldsymbol{\chi}},1].\end{cases}$$

[39]Integration by parts shows that this definition of $\bar{\theta}_{\boldsymbol{\chi}}$ is equivalent to that in Equation $\theta^*$-IC.

*Moreover,* $\bar{\theta}_{\boldsymbol{\chi}} \geq \theta_0$ *and, if credibility is imperfect,* $\bar{\theta}_{\boldsymbol{\chi}} < 1$.

*Proof.* Let $\varphi(\theta^*) := (\theta^* - \theta_0) - \int_0^{\theta^*} \boldsymbol{\chi}\mu_0[0,\theta]\,\mathrm{d}\theta = \int_0^{\theta^*} (1 - \boldsymbol{\chi}\mu_0[0,\theta])\,\mathrm{d}\theta - \theta_0$ for $\theta^* \in \Theta$. Clearly, $\varphi$ is continuous and strictly increasing. Next, observe that $\varphi(\theta_0) = -\int_0^{\theta_0} \boldsymbol{\chi}\mu_0[0,\theta]\,\mathrm{d}\theta \leq 0$, and

$$\varphi(1) = (1 - \theta_0) - \int_0^1 \boldsymbol{\chi}\mu_0[0,\theta]\,\mathrm{d}\theta = I_{\mu_0}(1) - I_{\boldsymbol{\chi}\mu_0}(1) = I_{(1-\boldsymbol{\chi})\mu_0}(1) \geq 0,$$

with the last inequality being strict unless credibility is perfect (that is, unless $\mu_0\{\boldsymbol{\chi} = \mathbf{1}\} = 1$). The result then follows from the intermediate value theorem. □

In what follows, recall the mean distribution $\bar{\mu}_{\boldsymbol{\chi}}$ as defined for state-independent credibility $\boldsymbol{\chi} = \chi$ in section 5, and defined analogously for the general case in section 6.3.

**Lemma 9.** *For any* $\theta \in [0,1]$*, we have*

$$I_{\bar{\mu}_{\boldsymbol{\chi}}}(\theta) = \max\{I_{\boldsymbol{\chi}\mu_0}(\theta),\ \theta - \theta_0\} = \begin{cases} I_{\boldsymbol{\chi}\mu_0}(\theta) & : \ \theta \leq \bar{\theta}_{\boldsymbol{\chi}} \\ \theta - \theta_0 & : \ \theta \geq \bar{\theta}_{\boldsymbol{\chi}}. \end{cases}$$

*Moreover,* $E\bar{\mu}_{\boldsymbol{\chi}} = \theta_0$.

*Proof.* That $I_{\bar{\mu}_{\boldsymbol{\chi}}}$ coincides with $I_{\boldsymbol{\chi}\mu_0}$ on $[0, \bar{\theta}_{\boldsymbol{\chi}}]$ and has derivative 1 on $(\bar{\theta}_{\boldsymbol{\chi}}, 1]$ follows directly from the definition of $\bar{\mu}_{\boldsymbol{\chi}}$. Noting that $I_{\boldsymbol{\chi}\mu_0}(\bar{\theta}_{\boldsymbol{\chi}}) = \bar{\theta}_{\boldsymbol{\chi}} - \theta_0$ by Lemma 8, it follows that $I_{\bar{\mu}_{\boldsymbol{\chi}}}(\theta) = \theta - \theta_0$ for $\theta \in [\bar{\theta}_{\boldsymbol{\chi}}, 1]$.

Next, recall that $I_{\boldsymbol{\chi}\mu_0}(\theta) - (\theta - \theta_0)$ is nonnegative for $\theta \in [0, \bar{\theta}_{\boldsymbol{\chi}}]$ and nonpositive for $\theta \in [\bar{\theta}_{\boldsymbol{\chi}}, 1]$ by Lemma 8. Consequently, $I_{\bar{\mu}_{\boldsymbol{\chi}}}(\theta) = \max\{I_{\boldsymbol{\chi}\mu_0}(\theta),\ \theta - \theta_0\}$ for every $\theta \in [0,1]$.

Finally, $E\bar{\mu}_{\boldsymbol{\chi}} = 1 - I_{\bar{\mu}_{\boldsymbol{\chi}}}(1) = \theta_0$. □

We now prove the following generalization of Claim 1 to the case of state-dependent credibility:

**Claim 1*.** *Some* $\theta^* \in [0, \bar{\theta}_{\boldsymbol{\chi}}]$ *exists such that the* $\theta^*$ *upper censorship of* $\bar{\mu}_{\boldsymbol{\chi}}$*, denoted by* $\mu_{\boldsymbol{\chi},\theta^*}$*, satisfies*

$$v^*_{\boldsymbol{\chi}}(\mu_0) = \hat{v}(\bar{\mu}_{\boldsymbol{\chi}}) = \int H(\cdot)\,\mathrm{d}\mu_{\boldsymbol{\chi},\theta^*}.$$

*Moreover, the corresponding* $\theta^*$*-upper-censorship pair is an S-optimal* $\boldsymbol{\chi}$*-equilibrium that induces* $\mu_{\boldsymbol{\chi},\theta^*}$ *as its posterior mean distribution.*

*Proof.* First, we show that $\hat{v}(\bar{\mu}_{\boldsymbol{\chi}}) = \max_{\theta^* \in [0,\bar{\theta}_{\boldsymbol{\chi}}]} \int H \, \mathrm{d}\mu_{\boldsymbol{\chi},\theta^*}$, and that the maximum on the RHS is attained. By Lemma 7, some $\theta^* \in [0, \omega]$ exists such that $\hat{v}(\bar{\mu}_{\boldsymbol{\chi}}) = \int H \, \mathrm{d}\mu_{\boldsymbol{\chi},\theta^*}$. As $\bar{\mu}_{\boldsymbol{\chi}}[0, \bar{\theta}_{\boldsymbol{\chi}}] = 1$, we have $\mu_{\boldsymbol{\chi},\theta} = \mu_{\boldsymbol{\chi},\bar{\theta}_{\boldsymbol{\chi}}}$ for every $\theta \in [\bar{\theta}_{\boldsymbol{\chi}}, 1]$; so we may without loss take $\theta^* \leq \bar{\theta}_{\boldsymbol{\chi}}$. Furthermore, since

$$\int H \, \mathrm{d}\mu_{\boldsymbol{\chi},\theta^*} = \hat{v}(\bar{\mu}_{\boldsymbol{\chi}}) = \max_{\mu \preceq \bar{\mu}_{\boldsymbol{\chi}}} \int H \, \mathrm{d}\mu \geq \int H \, \mathrm{d}\mu_{\boldsymbol{\chi},\theta}$$

for every $\theta \in [0, \bar{\theta}_{\boldsymbol{\chi}}]$, the maximum is attained.

Next, given $\theta^* \in [0, \bar{\theta}_{\boldsymbol{\chi}}]$, we exhibit an equilibrium in which S communicates via a $\theta^*$-upper-censorship pair, and observe that this induces S value $\int H \, \mathrm{d}\mu_{\boldsymbol{\chi},\theta^*}$—in particular showing $\int H \, \mathrm{d}\mu_{\boldsymbol{\chi},\theta^*} \leq v^*_{\boldsymbol{\chi}}(\mu_0)$. To that end, define the belief map $\pi : M \to \Delta\Theta$ via

$$\pi(m) = \begin{cases} \delta_m & : \ m \in [0, \theta^*) \\ \gamma & : \ \text{otherwise}, \end{cases}$$

where $\gamma \coloneqq \frac{[1-\boldsymbol{\chi}\mathbf{1}_{[0,\theta^*)}]\mu_0}{1-\boldsymbol{\chi}\mu_0[0,\theta^*)}$ (with $\gamma \coloneqq \delta_1$ if $\boldsymbol{\chi}\mu_0[0,\theta^*) = 1$)—that is, $\gamma$ is R's posterior belief upon hearing message 1 given a $\theta^*$-upper-censorship pair. Then let R behavior be given by $\alpha \coloneqq H \circ E \circ \pi$. The Bayesian property is now straightforward, and the R incentive condition holds by construction. To verify that this is a $\boldsymbol{\chi}$-equilibrium, then, we need only check that S behavior is optimal under influenced reporting. As the set of interim own-payoffs S can induce with some message is $\{H(\theta) : \ \theta \in [0, \theta^*) \text{ or } \theta = E\gamma\}$, and $H$ is strictly increasing on $[0, 1]$, it remains to see that $E\gamma \geq \theta^*$, that is, that Equation $\theta^*$-IC is satisfied. This property follows directly from $\theta^* \leq \bar{\theta}_{\boldsymbol{\chi}}$ given Lemma 8, delivering S incentive compatibility. To show this equilibrium generates the required payoff, it suffices to show that the induced distribution $\mu$ of posterior means is equal to $\mu_{\boldsymbol{\chi},\theta^*}$. For any $\theta \in [0, \theta^*)$, notice that

$$\mu[0, \theta) = \int_0^\theta \boldsymbol{\chi} \, \mathrm{d}\mu_0 = \bar{\mu}_{\boldsymbol{\chi}}[0, \theta) = \mu_{\boldsymbol{\chi},\theta^*}[0, \theta).$$

Moreover, $|[\theta^*, 1] \cap \operatorname{supp}(\mu)| = 1 = |[\theta^*, 1] \cap \operatorname{supp}(\mu_{\boldsymbol{\chi},\theta^*})|$. Equality then follows from equality of their means (Lemma 9).

Finally, we show that $v^*_{\boldsymbol{\chi}}(\mu_0) \leq \hat{v}(\bar{\mu}_{\boldsymbol{\chi}})$. To that end, let $(\beta, \gamma, k)$ solve the program of Theorem 1 – and, without loss, say $\beta = \mu_0$ if $k = 0$. Let $\omega \coloneqq \omega^* \wedge E\gamma$, and see that $H(E\gamma) \wedge H$ is continuous, convex on $[0, \omega]$, and concave on $[\omega, 1]$. Therefore, by

Lemma 7, there is some $\theta^* \in [0, \omega]$ such that the $\theta^*$ upper censorship of $\beta$ belongs to $\text{argmax}_{\hat{\beta} \preceq \beta} \int H(E\gamma) \wedge H \, \mathrm{d}\hat{\beta}$. Let $\lambda \coloneqq \beta[0, \theta^*) \in [0, 1]$, $\eta \coloneqq \frac{\mathbf{1}_{[\theta^*, 1]}\beta}{1-\lambda} \in \Delta\Theta$, $\hat{\gamma} \coloneqq \frac{(1-k)\gamma + (1-\lambda)k\eta}{1-\lambda k} \in \Delta\Theta$, and $\hat{\beta} \coloneqq \frac{\mathbf{1}_{[0,\theta^*)}\beta}{\lambda} \in \Delta\Theta$.[40] Two observations will enable us to bound S payoffs across all equilibria. First, as a monotone transformation of an affine functional, $v = H \circ E$ is quasiconcave, implying $\bar{v} = v$. Second, Lemma 7 tells us $E\eta \geq \omega$, so that $H(E\gamma) \wedge H$ is concave on $\text{co}\{E\gamma, E\eta\}$. Now, observe that

$$\begin{aligned}
v^*_{\boldsymbol{\chi}}(\mu_0) &= k\hat{v}_{\wedge\gamma}(\beta) + (1-k)\bar{v}(\gamma) \\
&= k \int H(E\gamma) \wedge H \, \mathrm{d}\left[\mathbf{1}_{[0,\theta^*)}\beta + (1-\lambda)\delta_{E\eta}\right] + (1-k)H(E\gamma) \\
&= k\left[\lambda \int H \, \mathrm{d}\hat{\beta} + (1-\lambda)H(E\gamma) \wedge H(E\eta)\right] + (1-k)H(E\gamma) \wedge H(E\gamma) \\
&\leq k\lambda \int H \, \mathrm{d}\hat{\beta} + (1-k\lambda)H(E\gamma) \wedge H(E\hat{\gamma}) \\
&\leq \int H \, \mathrm{d}\left[k\lambda\hat{\beta} + (1-\lambda k)\delta_{E\hat{\gamma}}\right] \\
&\leq \hat{v}\left(k\lambda\hat{\beta} + (1-\lambda k)\delta_{E\hat{\gamma}}\right).
\end{aligned}$$

Letting $\hat{\mu} \coloneqq k\lambda\hat{\beta} + (1-\lambda k)\delta_{E\hat{\gamma}}$, the payoff ranking (and so too the claim) will follow if we show that $\hat{\mu} \preceq \bar{\mu}_{\boldsymbol{\chi}}$. As (appealing to Lemma 9) $E\bar{\mu}_{\boldsymbol{\chi}} = \theta_0 = E\hat{\mu}$, it suffices to show that $I_{\hat{\mu}} \leq I_{\bar{\mu}_{\boldsymbol{\chi}}}$.

For $\theta \in [0, E\hat{\gamma})$, we have $\delta_{E\hat{\gamma}}[0, \theta] = 0$. Therefore, over the interval $[0, E\hat{\gamma}]$, we have

$$I_{\hat{\mu}} = I_{\lambda k\hat{\beta}} + (1-\lambda k)I_{\delta_{E\hat{\gamma}}} = I_{\lambda k\hat{\beta}} \leq I_{k\beta} = I_{\mu_0} - I_{(1-k)\gamma} \leq I_{\mu_0} - I_{(\mathbf{1}-\boldsymbol{\chi})\mu_0} = I_{\boldsymbol{\chi}\mu_0}.$$

Now, as $I_{\hat{\mu}}(1) = 1 - \theta_0$ and (since $E\hat{\gamma} \geq \theta_0$) we have $I'_{\hat{\mu}}|_{(E\hat{\gamma},1)} = 1$, we know $I_{\hat{\mu}}(\theta) = \theta - \theta_0$ for $\theta \in [E\hat{\gamma}, 1]$. In particular, we learn that $I_{\hat{\mu}}(\theta) \leq \max\{I_{\boldsymbol{\chi}\mu_0}(\theta), \theta - \theta_0\}$ for $\theta \in [0, E\hat{\gamma}] \cup [E\hat{\gamma}, 1]$. Lemma 9 then tells us that $I_{\hat{\mu}} \leq I_{\bar{\mu}_{\boldsymbol{\chi}}}$. □

### A.5.3 Comparative Statics

Now, we prove Claim 2. In fact, because the proof applies without change, we prove a slightly stronger result, providing comparative statics results in the credibility function *and the prior*, holding the prior mean fixed. Specifically, given two pairs of parameters $\langle \mu_0, \boldsymbol{\chi} \rangle$ and $\langle \tilde{\mu}_0, \tilde{\boldsymbol{\chi}} \rangle$ such that $E\mu_0 = E\tilde{\mu}_0 = \theta_0$, we show that $v^*_{\boldsymbol{\chi}}(\mu_0) \geq v^*_{\tilde{\boldsymbol{\chi}}}(\tilde{\mu}_0)$ if and

[40] In case any of the described objects is defined by an expression with a zero denominator, we define it as follows: $\eta \coloneqq \delta_1$ if $\lambda = 1$, $\hat{\gamma} \coloneqq \delta_1$ if $\lambda k = 1$, and $\hat{\beta} \coloneqq \delta_0$ if $\lambda = 0$.

only if $\bar{\mu}_{\boldsymbol{\chi}} \succeq \bar{\tilde{\mu}}_{\tilde{\boldsymbol{\chi}}}$.

*Proof.* Appealing to Claim 1 and Lemma 6,

$$
\begin{aligned}
v^*_{\boldsymbol{\chi}}(\mu_0) - v^*_{\tilde{\boldsymbol{\chi}}}(\tilde{\mu}_0) &= \hat{v}(\bar{\mu}_{\boldsymbol{\chi}}) - \hat{v}(\bar{\tilde{\mu}}_{\tilde{\boldsymbol{\chi}}}) \\
&= \max_{I \in \mathcal{I}(I_{\bar{\mu}_{\boldsymbol{\chi}}})} \left[ H(0)I'(0) + \int_0^1 H \, \mathrm{d}I' \right] - \max_{\tilde{I} \in \mathcal{I}(I_{\bar{\tilde{\mu}}_{\tilde{\boldsymbol{\chi}}}})} \left[ H(0)\tilde{I}'(0) + \int_0^1 H \, \mathrm{d}\tilde{I}' \right] \\
&= \max_{I \in \mathcal{I}(I_{\bar{\mu}_{\boldsymbol{\chi}}})} \int_0^1 H \, \mathrm{d}I' - \max_{\tilde{I} \in \mathcal{I}(I_{\bar{\tilde{\mu}}_{\tilde{\boldsymbol{\chi}}}})} \int_0^1 H \, \mathrm{d}\tilde{I}' \\
&= \max_{I \in \mathcal{I}(I_{\bar{\mu}_{\boldsymbol{\chi}}})} \int_0^1 I \, \mathrm{d}h - \max_{\tilde{I} \in \mathcal{I}(I_{\bar{\tilde{\mu}}_{\tilde{\boldsymbol{\chi}}}})} \int_0^1 \tilde{I} \, \mathrm{d}h.
\end{aligned}
$$

Let $I_* \coloneqq I_{\bar{\mu}_{\boldsymbol{\chi}}}$ and $\tilde{I}_* \coloneqq I_{\bar{\tilde{\mu}}_{\tilde{\boldsymbol{\chi}}}}$. We now need to show that $\max_{I \in \mathcal{I}(I_*)} \int_0^1 I \, \mathrm{d}h \geq \max_{\tilde{I} \in \mathcal{I}(\tilde{I}_*)} \int_0^1 \tilde{I} \, \mathrm{d}h$ for every continuous, strictly quasiconcave $h : [0,1] \to \mathbb{R}$ if and only if $I_* \geq \tilde{I}_*$.

First, if $I_* \leq \tilde{I}_*$ then $\mathcal{I}(I_*) \subseteq \mathcal{I}(\tilde{I}_*)$, delivering the payoff ranking.

Conversely, suppose $I_* \not\leq \tilde{I}_*$. Then, elements of $\mathcal{I}$ being continuous, there are some $\theta_1, \theta_2 \in \Theta$ such that $\theta_1 < \theta_2$ and $I_* > \tilde{I}_*$ on $(\theta_1, \theta_2)$. If $h$ is increasing, then

$$
v^*_{\boldsymbol{\chi}}(\mu_0) - v^*_{\tilde{\boldsymbol{\chi}}}(\mu_0) = \int_0^1 I_* \, \mathrm{d}h - \int \tilde{I}_* \, \mathrm{d}h = \int_0^1 (I_* - \tilde{I}_*) \, \mathrm{d}h.
$$

As $(I_* - \tilde{I}_*)$ is strictly positive over $(\theta_1, \theta_2)$, globally bounded, and globally continuous, there is $\epsilon > 0$ small enough that $\left( \epsilon \int_0^{\theta_1} + \int_{\theta_1}^{\theta_2} + \epsilon \int_{\theta_2}^1 \right) \left[ I_*(\theta) - \tilde{I}_*(\theta) \right] \, \mathrm{d}\theta > 0$. It is then straightforward to construct a shock distribution whose continuous density $h$ satisfies $h'|_{(0,\theta_1) \cup (\theta_2, 1)} = \epsilon\zeta$ and $h'|_{(\theta_1, \theta_2)} = \zeta$ for some $\zeta > 0$. Such a shock distribution witnesses a failure of $v^*_{\boldsymbol{\chi}}(\mu_0) \geq v^*_{\tilde{\boldsymbol{\chi}}}(\tilde{\mu}_0)$.

□

# B Extensions: Proofs from section 6

This section contains the formal results reported in section 6, along with their proofs.

## B.1 On S's $\chi$-equilibrium Payoff Sets

We begin by providing results on the space of S payoffs which will be of use in two of the extensions that follow. First, we characterize the set of payoffs attainable in

$\chi$-equilibrium by an influencing S, in particular showing this payoff set is an interval. Then, we show the set of ex-ante S payoffs attainable in $\chi$-equilibrium is an interval as well. We note that these results also hold under the modified finite-message setting of Appendix B.2, via an appeal to Carathéodory's theorem.

Toward the proof, we first record a useful property of Kakutani correspondences.

**Fact 3.** *The range of a Kakutani correspondence from a nonempty, compact, convex space to $\mathbb{R}$ is a nonempty compact interval.*

*Proof.* Nonemptiness is trivial. Compactness of the range holds because the correspondence is upper hemicontinuous on a compact domain. Convexity follows from the intermediate value theorem for correspondences (e.g., Lemma 2 of de Clippel, 2008). □

Next, we will establish convexity and compactness of the sets of S's possible $\chi$-equilibrium ex-ante payoffs and payoffs from influencing. To do so, given any $\chi \in [0,1]$, we now provide a characterization of the set

$$S_i^\chi \coloneqq \{s_i \in \mathbb{R} \colon \ (p, s_o, s_i) \text{ is a } \chi\text{-equilibrium outcome for some } p, s_o\}.$$

**Lemma 10.** *Let $s_i \in \mathbb{R}$. Then $s_i \in S_i^\chi$ if and only if some $k \in [0,1], \gamma, \beta \in \Delta\Theta$ exist such that*

*(i)* $k\beta + (1-k)\gamma = \mu_0$,

*(ii)* $(1-k)\gamma \geqslant (1-\chi)\mu_0$,

*(iii)* $\max\{\underline{w}(\beta), \underline{w}(\gamma)\} \leq s_i \leq \bar{v}(\gamma)$.

*Moreover, the set $S_i^\chi$ is nonempty compact interval.*

*Proof.* By Lemma 1, $s_i \in S_i^\chi$ if and only if some $k \in [0,1], g, b \in \Delta\Delta\Theta$ exist such that

(i′) $kb + (1-k)g \in \mathcal{R}(\mu_0)$,

(ii′) $(1-k)\int \mu \, \mathrm{d}g(\mu) \geqslant (1-\chi)\mu_0$,

(iii′) $g\{V \ni s_i\} = b\{w \leq s_i\} = 1$.

Then the existence of $(k, g, b)$ satisfying (i′-iii′) immediately implies the existence of $(k, \gamma, \beta)$ satisfying (i-iii) by setting $\gamma \coloneqq \int \mu \, \mathrm{d}g(\mu), \beta \coloneqq \int \mu \, \mathrm{d}b(\mu)$. Conversely, let $(k, \gamma, \beta)$ satisfy (i-iii). By Lipnowski and Ravid's (2020) Theorem 2 and Corollary 3:

- Some $g \in \mathcal{R}(\gamma)$ exists with $g\{V \ni s_i\} = 1$ if and only if $s_i \in [\underline{w}(\gamma), \bar{v}(\gamma)]$,
- Some $b \in \mathcal{R}(\beta)$ exists with $b\{w \leq s_i\} = 1$ if and only if $s_i \geqslant \underline{w}(\beta)$.

Thus, we obtain the desired characterization.

Finally, to show the "moreover" part, rewrite the above characterization of $S_i^\chi$ as follows. Let $\mathcal{M}$ be the set of Borel measures on $\Theta$ and $\mathcal{G} \coloneqq \{\eta \in \mathcal{M} \colon (1-\chi)\mu_0 \leq \eta \leq \mu_0\}$, a compact convex subset. Define the functions

$$\begin{aligned}\tilde{v} \colon \mathcal{M} &\to \mathbb{R} \\ \eta &\mapsto \begin{cases} \bar{v}\left(\frac{\eta}{\eta(\Theta)}\right) & : \eta \neq 0 \\ \max \bar{v}(\Delta\Theta) & : \eta = 0 \end{cases}\end{aligned} \qquad \begin{aligned}\tilde{w} \colon \mathcal{M} &\to \mathbb{R} \\ \eta &\mapsto \begin{cases} \underline{w}\left(\frac{\eta}{\eta(\Theta)}\right) & : \eta \neq 0 \\ \min \underline{w}(\Delta\Theta) & : \eta = 0 \end{cases}\end{aligned}$$

$$\begin{aligned}\kappa \colon \mathcal{G} &\to \mathbb{R} \\ \eta &\mapsto \tilde{v}(\eta) - \max\{\tilde{w}(\eta), \tilde{w}(\mu_0 - \eta)\}.\end{aligned}$$

Then the above characterization implies that $s_i \in S_i^\chi$ if and only if some $\eta \in \mathcal{G}$ exists such that $s_i \in [\max\{\tilde{w}(\eta), \tilde{w}(\mu_0 - \eta), \tilde{v}(\eta)]$, because $(k, \gamma, \beta) \mapsto (1-k)\gamma$ is a surjection from the subset of $(k, \gamma, \beta) \in [0,1] \times \Delta\Theta^2$ satisfying (i-ii) to $\mathcal{G}$. But this means that $S_i^\chi = \tau(\mathcal{G}^*)$, where $\mathcal{G}^* \coloneqq \kappa^{-1}([0,\infty))$ and $\tau$ is a correspondence defined as

$$\begin{aligned}\tau \colon \mathcal{G}^* &\rightrightarrows \mathbb{R} \\ \eta &\mapsto [\max\{\tilde{w}(\eta), \tilde{w}(\mu_0 - \eta)\}, \tilde{v}(\eta)].\end{aligned}$$

We now proceed to show $S_i^\chi$ is a nonempty compact interval. First, observe that $\kappa$ is upper semicontinuous and quasiconcave—since both $\bar{v}$ and $-\underline{w}$ are and, therefore, so are $\tilde{v}$ and $-\tilde{w}$. Hence, the set $\kappa^{-1}([0,\infty)) = \mathcal{G}^*$ is compact and convex, and it is also nonempty because it contains $\mu_0$. Second, note that $\tau$ is a Kakutani correspondence since it is compact-convex-valued by definition; nonempty-valued by the definition of $\mathcal{G}^*$; and upper hemicontinuous by upper (resp. lower) semicontinuity of $\tilde{v}$ ($\tilde{w}$). Hence, the result follows from Fact 3. $\square$

Building on the previous two lemmas, the following result shows that the set of ex-ante S's $\chi$-equilibrium payoffs is convex.

**Lemma 11.** *The set $\{\chi s_o + (1-\chi)s_i \colon (p, s_o, s_i)$ is a $\chi$-equilibrium outcome$\}$ of ex-ante $\chi$-equilibrium payoffs is a nonempty compact interval.*

*Proof.* Define the correspondence

$$\begin{aligned}\varsigma\colon S_i^\chi &\rightrightarrows \mathbb{R}\\ s_i &\mapsto \{\chi s_o + (1-\chi)s_i\colon (p, s_o, s_i) \text{ is a } \chi\text{-equilibrium outcome}\}.\end{aligned}$$

We will show that $\varsigma$ is a Kakutani correspondence, which will give the desired result in light of Fact 3 and Lemma 10.

First, $\varsigma$ is nonempty-valued by the definition of $S_i^\chi$. Second, the graph of $\varsigma$ is compact as a continuous image of the compact space $X$ defined in the proof of Corollary 1. Therefore, $\varsigma$ is compact-valued and upper hemicontinuous.

Finally, we show that $\varsigma$ is convex-valued. Fix any $s_i \in S_i^\chi$, $s, s' \in \varsigma(s_i)$, $\lambda \in (0,1)$. By Lemma 1, there exist $k, k' \in [0,1], g, g', b, b' \in \Delta\Delta\Theta$ such that

$$\begin{aligned}
kb + (1-k)g &\in \mathcal{R}(\mu_0), & k'b' + (1-k')g' &\in \mathcal{R}(\mu_0),\\
(1-k)\int \mu \,\mathrm{d}g(\mu) &\geqslant (1-\chi)\mu_0, & (1-k')\int \mu \,\mathrm{d}g'(\mu) &\geqslant (1-\chi)\mu_0,\\
s \in (1-k)s_i + k\int_{\mathrm{supp}(b)} s_i \wedge V \,\mathrm{d}b, & & s' \in (1-k')s_i + k'\int_{\mathrm{supp}(b')} s_i \wedge V \,\mathrm{d}b'.
\end{aligned}$$

Let $s^* \coloneqq \lambda s + (1-\lambda)s'$, $k^* \coloneqq \lambda k + (1-\lambda)k'$, $g^* \coloneqq \lambda\frac{1-k}{1-k^*}g + (1-\lambda)\frac{1-k'}{1-k^*}g'$, and $b^* \coloneqq \lambda\frac{k}{k^*}b + (1-\lambda)\frac{k'}{k^*}b'$. Then, by Lemma 1, $(k^*, g^*, b^*)$ witness a $\chi$-equilibrium with expected payoff $s^*$ influencing payoff $s_i$. Thus, $\varsigma(s_i)$ is convex. $\square$

## B.2 Strong robustness: Proofs from section 6.1

This section provides the formal counterpart to section 6.1. We first introduce a notion of equilibrium that captures S's ex-ante incentives concerning her choice of official reporting protocol. We then characterize when, in the limit as credibility becomes perfect, *every* such equilibrium gives S a payoff approaching her full-credibility payoff $v_1^*(\mu_1)$.

Throughout Appendix B.2, we maintain the following assumption.

**Assumption 1.** *The sets $A$, $\Theta$, and $M$ are all finite, and $|M| \geq 2|\Theta|$.*

Thus, we specialize to the finite model, and further alter the model to impose that only finitely many messages are available to S. We make this finite-message assumption to simplify the analysis of players' behavior off of the equilibrium path. Observe this finite-message assumption does not alter the set of attainable $\chi$-equilibrium

$(s_o, s_i)$ pairs, relative to our original model with $M$ uncountable, given Lemma 1 and Carathéodory's theorem.

To formalize the relevant solution concept, let $\Xi$ denote the set of all official reporting protocols, i.e., maps $\xi\colon \Theta \to \Delta M$. Then let $\boldsymbol{\sigma}\colon \Theta \times \Xi \to \Delta M$, $\boldsymbol{\alpha}\colon M \times \Xi \to \Delta A$, and $\boldsymbol{\pi}\colon M \times \Xi \to \Delta\Theta$ denote S's influencing strategy, R's strategy, and R's belief map, respectively, that take into account the announced reporting protocol $\xi \in \Xi$. A **$\boldsymbol{\chi}$-PBE** is a tuple $(\xi, \boldsymbol{\sigma}, \boldsymbol{\alpha}, \boldsymbol{\pi})$ such that $(\xi', \boldsymbol{\sigma}(\cdot, \xi'), \boldsymbol{\alpha}(\cdot, \xi'), \boldsymbol{\pi}(\cdot, \xi'))$ is a $\boldsymbol{\chi}$-equilibrium for any $\xi' \in \Xi$, and $\xi \in \Xi$ maximizes S's ex-ante payoff

$$\int_\Theta \left( \int_M \left[ \int_A u_S(a) \, \mathrm{d}\boldsymbol{\alpha}(a|m,\xi) \right] \left[ \boldsymbol{\chi}(\theta) \, \mathrm{d}\xi(m|\theta) + (1 - \boldsymbol{\chi}(\theta)) \, \mathrm{d}\boldsymbol{\sigma}(m|\theta,\xi) \right] \right) \mathrm{d}\mu_0(\theta).$$

The following proposition, reported in section 6.1, characterizes when the full commitment value is robust to partial commitment under the worst equilibrium selection in a finitary setting. To state the result, define the **worst-value function**, $w(\mu) \coloneqq \min V(\mu)$, which identifies S's lowest continuation payoff from inducing belief $\mu$. Then, let $w^*_\chi(\mu_0)$ denote the **worst $\chi$-PBE S value**.[41]

**Proposition 5.** *The following are equivalent:*

1. *The full-commitment value is strongly robust to partial credibility:* $\lim_{\chi \nearrow 1} w^*_\chi(\mu_0) = v^*_1(\mu_0)$ *for every full-support prior* $\mu_0$.

2. *The full-commitment value is robust to equilibrium selection:* $w^*_1(\mu_0) = v^*_1(\mu_0)$ *for every full-support prior* $\mu_0$.

*Moreover, for almost every R objective, the full commitment value is strongly robust to partial credibility.*[42]

### B.2.1 Constructing off-path $\chi$-equilibria

In this subsection, we establish that every official reporting protocol that S chooses can be associated with appropriately "adversarial" continuation play. The possibility of such continuation play is useful for studying the range of S payoffs attainable in a $\chi$-PBE.

[41]Formally, $w^*_\chi(\mu_0)$ is the infimum over all $\chi$-PBE S values.

[42]That is, fixing finite $A$ and $\Theta$, for all but a Lebesgue-null (and nowhere dense) set of R objectives $u_R \in \mathbb{R}^{A\times\Theta}$ (and every S objective $u_S \in \mathbb{R}^A$), the full commitment value is strongly robust to partial credibility.

Toward constructing adversarial equilibria, we begin with a technical lemma showing any reporting protocol comprises part of a $\chi$-equilibrium in which R always chooses from a given restricted set of best responses.

**Lemma 12.** *If $\tilde{V} \subseteq V$ is a Kakutani correspondence and $\xi$ is any official reporting protocol, then some $\chi$-equilibrium $(\xi, \sigma, \alpha, \pi)$ exists such that $u_S(\alpha) \in \tilde{V}(\pi)$.*

*Proof.* Let $\Pi \coloneqq (\Delta\Theta)^M$ be the set of all R belief mappings and define correspondences

$$\begin{aligned}
\hat{S}\colon \Pi &\rightrightarrows \mathbb{R} \\
\pi &\mapsto \left[\max_{m\in M} \min \tilde{V}(\pi(m)), \max_{m\in M} \max \tilde{V}(\pi(m))\right], \\
\hat{M}\colon \Pi &\rightrightarrows M \\
\pi &\mapsto \left\{m \in M\colon\ \tilde{V}(\pi(m)) \cap \hat{S}(\pi) \neq \varnothing\right\}.
\end{aligned}$$

Observe that $\hat{S}$ is Kakutani, since $\tilde{V}$ is Kakutani and a finite maximum or minimum of upper or lower semicontinuous functions inherits the same semicontinuity. Therefore, $\hat{M}$ is nonempty-valued with closed graph. Now let $\Sigma \coloneqq (\Delta M)^\Theta$ and consider the correspondence mapping belief maps into S-IC influencing strategies (assuming R's strategy delivers S values from $\tilde{V}$)

$$\begin{aligned}
\hat{\Sigma}\colon\ \Pi &\rightrightarrows \Sigma \\
\pi &\mapsto \left\{\sigma \in \Sigma\colon\ \cup_{\theta\in\Theta} \operatorname{supp}(\sigma(\theta)) \subseteq \hat{M}(\pi)\right\},
\end{aligned}$$

and the correspondence mapping influencing strategies into consistent belief maps

$$\begin{aligned}
&\hat{\Pi}\colon\ \Sigma \rightrightarrows \Pi, \\
&\quad \sigma \mapsto \Big\{\pi \in \Pi\colon\ \pi(\theta|m) \int_\Theta \left[\chi\, \mathrm{d}\xi(m|\cdot) + (1-\chi)\sigma(m|\cdot)\right] \mathrm{d}\mu_0 \\
&\qquad\qquad = \left[\chi(\theta)\xi(m|\theta) + (1-\chi(\theta))\sigma(m|\theta)\right]\mu_0(\theta), \forall\theta \in \Theta, m \in M\Big\}.
\end{aligned}$$

It then follows that $\hat{\Sigma}$ and $\hat{\Pi}$ are both Kakutani. Therefore, the Kakutani fixed point theorem delivers some $\sigma \in \Sigma$ and $\pi \in \Pi$ such that $\sigma \in \hat{\Sigma}(\pi)$ and $\pi \in \hat{\Pi}(\sigma)$. Now, take any $s_i \in \hat{S}(\pi)$ and let $D \coloneqq \pi(M)$. Note that $s_i \wedge \tilde{V}|_D$ is nonempty-valued and so admits a selector $\phi\colon D \to \mathbb{R}$. Therefore, some $\tilde{\alpha}\colon D \to \Delta A$ exists such that $u_S(\tilde{\alpha}(m)) = \phi(m)$. Next, define $\alpha \coloneqq \tilde{\alpha} \circ \pi\colon M \to \Delta(A)$. It is then easy to verify that $(\xi, \sigma, \alpha, \pi)$ is a

$\chi$-equilibrium in the model with the state space $\Theta$ and the message space $M$. $\square$

To construct S-adversarial continuation play, it will be convenient to define the function

$$\begin{aligned}\hat{w}\colon \Delta\Theta &\to \mathbb{R}\\ \mu &\mapsto \sup_{p\in\mathcal{R}(\mu)} \int w \,\mathrm{d}p,\end{aligned}$$

and the payoff $\underline{s}_1(\mu_0) \coloneqq \hat{w}(\mu_0)$; when no confusion arises, we will omit the dependence on the prior and simply write $\underline{s}_1$. The following lemmas show (with finite states) that $\underline{s}_1$ is an upper bound on the S value in the worst $\chi$-equilibrium, an upper bound on the worst $\chi$-PBE value, and the minimal 1-PBE value.

**Lemma 13.** *Every official reporting protocol $\xi$ admits some $\chi$-equilibrium $(\xi, \sigma, \alpha, \pi)$ with ex-ante S payoff weakly below $\underline{s}_1$.*

*Proof.* Without loss, we can focus on the case that $\mu_0$ is of full support. Indeed, if we construct a $\chi$-equilibrium as desired (for official reporting protocol $\xi|_{\Theta_0}$) in the restricted model with state space $\Theta_0 \coloneqq \mathrm{supp}(\mu_0)$, then this equilibrium can be extended to a $\chi$-equilibrium in the true model, by fixing any $\theta_0 \in \Theta_0$ and extending $\sigma$ to $\Theta$ via $\sigma(\theta) \coloneqq \sigma(\theta_0)$ for $\theta \in \Theta \setminus \Theta_0$.

Note the lemma follows directly from Lemma 12 if we can find a Kakutani subcorrespondence $\tilde{V} \subseteq V$ such that $\mathrm{cav}[\max \tilde{V}](\mu_0) \leq \underline{s}_1$. We now show one can set $\tilde{V} \coloneqq [w, z]$ where $z$ is the upper semicontinuous envelope of $w$, given by

$$\begin{aligned}z\colon \Delta\Theta &\to \mathbb{R}\\ \mu &\mapsto \limsup_{\mu'\to\mu} w(\mu).\end{aligned}$$

First, $\tilde{V}$ is a Kakutani subcorrespondence of $V$ since $z$ is upper semicontinuous and lies above the lower semicontinuous function $w$. All that remains, then, is to show that $\underline{s}_1 \geq \mathrm{cav}[\max \tilde{V}](\mu_0) = \hat{z}(\mu_0)$. To do so, let us establish the stronger claim that $\hat{z}|_D = \hat{w}|_D$, where $D \subseteq \Delta\Theta$ is the set of full-support beliefs.

Define

$$\begin{aligned}\tilde{z}\colon \Delta\Theta &\to \mathbb{R}\\ \mu &\mapsto \limsup_{\mu'\to\mu} \hat{w}(\mu).\end{aligned}$$

It follows from concavity of $\hat{w}$ that $\tilde{z}$ is concave too. Hence, because $\tilde{z} \geq z$ and $\tilde{z}$ is upper semicontinuous by construction, it follows that $\tilde{z} \geq \hat{z}$.[43] Moreover, Theorem 10.4 from Rockafellar (2015) implies the concave function $\hat{w}|_D$ is continuous. Hence, $\tilde{z}|_D = \hat{w}|_D$ by the definition of $\tilde{z}$. That $\tilde{z} \geq \hat{z} \geq \hat{w}$ then implies $\hat{z}|_D = \hat{w}|_D$. $\square$

### B.2.2 On the range of PBE payoffs

Here, we provide a sufficient condition for a payoff to be compatible with $\chi$-PBE for an arbitrary $\chi$, as well as an exact characterization for the case of perfect credibility.

**Lemma 14.** *If $\Theta$ and $M$ are finite, and $s \in [\underline{s}_1 \wedge v^*_\chi(\mu_0),\ v^*_\chi(\mu_0)]$, then some $\chi$-PBE exists with ex-ante S payoff $s$.*

*Proof.* First, we argue that a $\chi$-equilibrium exists with ex-ante S payoff $s$. To that end, observe that Lemma 1 implies $(\delta_{\mu_0}, w(\mu_0), w(\mu_0))$ is a $\chi$-equilibrium outcome (as witnessed by $k = \chi$ and $g = b = \delta_{\mu_0}$). But then, as Theorem 1 says $v^*_\chi(\mu_0)$ is the highest $\chi$-equilibrium S payoff, it follows from Lemma 11 that every payoff in $[w(\mu_0), v^*_\chi(\mu_0)]$ is a $\chi$-equilibrium S payoff. Thus, $s$ is a $\chi$-equilibrium S payoff because $w(\mu_0) = \int w \,\mathrm{d}\delta_{\mu_0} \leq \underline{s}_1 \leq s$. So let $(\xi^*, \sigma^*, \alpha^*, \pi^*)$ be some $\chi$-equilibrium generating S payoff $s$.

Finally, construct $\boldsymbol{\sigma}^*$, $\boldsymbol{\alpha}^*$, and $\boldsymbol{\pi}^*$ as follows. First, let $\boldsymbol{\sigma}^*(\xi^*) := \sigma^*, \boldsymbol{\alpha}^*(\xi^*) := \alpha^*, \boldsymbol{\pi}^*(\xi^*) := \pi^*$. Second, for each $\xi \neq \xi^*$ find a $\chi$-equilibrium $(\xi, \boldsymbol{\sigma}^*(\xi), \boldsymbol{\alpha}^*(\xi), \boldsymbol{\pi}^*(\xi))$ with ex-ante S value of at most $\underline{s}_1$, which exists by Lemma 13. Since $s \geq \underline{s}_1$, $(\xi^*, \boldsymbol{\sigma}^*, \boldsymbol{\alpha}^*, \boldsymbol{\pi}^*)$ is a $\chi$-PBE as desired. $\square$

Finally, we characterize the set of all 1-PBE S payoffs.

**Lemma 15.** *The set of all 1-PBE S values is given by $[\underline{s}_1, \bar{s}_1]$, where*

$$\underline{s}_1 := \sup_{p \in \mathcal{R}(\mu_0)} \int w \,\mathrm{d}p, \quad \bar{s}_1 := v^*_1(\mu_0) = \max_{p \in \mathcal{R}(\mu_0)} \int v \,\mathrm{d}p.$$

*Proof.* Given a payoff $s \in \mathbb{R}$, Proposition 1 of Lipnowski, Ravid, and Shishkin (2021) tells us $s \in [\underline{s}_1, \bar{s}_1]$ if and only if $\xi \colon \Theta \to \Delta M$, $\boldsymbol{\alpha} \colon M \times \Xi \to \Delta A$, and $\boldsymbol{\pi} \colon M \times \Xi \to \Delta\Theta$ exist such that $\boldsymbol{\alpha}(m, \xi')$ maximizes R's expected payoff given belief $\boldsymbol{\pi}(m, \xi')$ for every $m \in M$ and $\xi' \in \Xi$; $\boldsymbol{\pi}(\cdot, \xi')$ satisfies the Bayesian property given prior $\mu_0$ for every $\xi' \in \Xi$; choosing $\xi' = \xi$ maximizes S's expected payoff given prior $\mu_0$ from profile $(\xi', \alpha)$

[43] In fact, one can show $\tilde{z} = \hat{z}$, but this fact is immaterial to the present argument.

over all $\xi' \in \Xi$; and this maximal expected S payoff is equal to $s$. These conditions are obviously implied by $(\xi, \boldsymbol{\sigma}, \boldsymbol{\alpha}, \boldsymbol{\pi})$ being a 1-PBE for some $\boldsymbol{\sigma}$; and conversely, pairing them with any best response $\boldsymbol{\sigma}$ for an influencing S (which exists because $M$ is finite) yields a 1-PBE. The lemma follows. □

### B.2.3 Proof of Proposition 5

*Proof.* The "moreover" part follows directly from the equivalence, given Proposition 4 of Lipnowski, Ravid, and Shishkin (2021). We now proceed to prove the equivalence.[44]

Because $v_1^* \geq w_1^*$ by definition, it suffices to show that $\lim_{\chi \nearrow 1} w_\chi^*(\mu_0) = w_1^*(\mu_0)$ for every full-support $\mu_0 \in \Delta\Theta$. Fix any full-support prior $\mu_0$, and denote $\underline{s}_\chi := w_\chi^*(\mu_0)$ for $\chi \in [0, 1)$. By Lemma 14, $\underline{s}_\chi \leq \underline{s}_1$ for all $\chi \in [0, 1]$ and, therefore, $\limsup_{\chi \nearrow 1} \underline{s}_\chi \leq \underline{s}_1 = w_1^*(\mu_0)$ by Lemma 15. It therefore remains to show that $\liminf_{\chi \nearrow 1} \underline{s}_\chi \geq \underline{s}_1$, which we do below.

Take an arbitrary $\epsilon > 0$. By definition of $\underline{s}_1$, some $p \in \mathcal{R}(\mu_0)$ exists such that $\int w \, \mathrm{d}p > \underline{s}_1 - \epsilon$. Moreover, because $\Theta$ is finite, we may further assume $|\mathrm{supp}(p)| \leq |\Theta| \leq |M|$, in light of Carathéodory's theorem. For each $\mu \in \mathrm{supp}(p)$, let $N(\mu) \subseteq \Delta\Theta$ be some open neighborhood of $\mu$ on which $w > w(\mu) - \epsilon$, which exists because $w$ is lower semicontinuous. Making $\{N(\mu)\}_{\mu \in \mathrm{supp}(p)}$ smaller if necessary, we may assume without loss that these finitely many neighborhoods are pairwise disjoint. Because $\mathrm{supp}(p)$ is finite, some $\underline{\chi} \in (0, 1)$ exists that $\frac{1}{\underline{\chi} p(\mu) + (1 - \underline{\chi})} [(1 - \eta) p(\mu) \mu + \eta \Delta\Theta] \subseteq N(\mu)$ for each $\mu \in \mathrm{supp}(p)$, and so (since $\Delta\Theta$ is convex) the containment holds as well when we replace $\underline{\chi}$ with any $\chi \in (\underline{\chi}, 1)$.

Consider now, any $\chi \in (\underline{\chi}, 1)$, and fix some $(\xi, \boldsymbol{\sigma}, \boldsymbol{\alpha}, \boldsymbol{\pi})$ generating S payoff $s \in \mathbb{R}$. Let $\xi_p \in \Xi$ be some official reporting protocol that would, if credibility were perfect, generate belief distribution $p$ for R. Modifying $\xi_p$ if necessary, we may assume without loss that any two distinct messages from $M_p := \{m \in M : \int_\Theta \xi_p(m|\cdot) \, \mathrm{d}\mu_0 > 0\}$ would generate distinct beliefs.[45] Hence, every belief $\mu \in \mathrm{supp}(p)$ admits a unique $m_\mu \in M_p$ such that every $\theta \in \Theta$ has $\frac{\xi_p(m|\theta)\mu_0(\theta)}{\int_\Theta \xi_p(m|\cdot) \, \mathrm{d}\mu_0} = \mu(\theta)$. If S chooses official reporting protocol $\xi_p$ and sends message $m_\mu$ for some $\mu \in \mathrm{supp}(p)$, then the Bayesian property implies R's belief will be $\boldsymbol{\pi}(m_\mu, \xi_p) \in \frac{1}{\chi p(\mu) + (1 - \chi)} [(1 - \eta) p(\mu) \mu + \eta \Delta\Theta] \subseteq N(\mu)$, so that R

[44] Whereas the "moreover" part relies on $A$ being finite, the proof of the equivalence relies only on $\Theta$ and $M$ being finite.

[45] That is, we assume without loss that any two distinct $m, m' \in M_p$ have $\frac{\xi_p(m|\theta)\mu_0(\theta)}{\int_\Theta \xi_p(m|\cdot) \, \mathrm{d}\mu_0} \neq \frac{\xi_p(m'|\theta)\mu_0(\theta)}{\int_\Theta \xi_p(m'|\cdot) \, \mathrm{d}\mu_0}$ for some $\theta \in \Theta$.

rationality implies S has continuation value exceeding $w(\mu)-\epsilon$. But because S chooses $\xi \in \Xi$ optimally, and has the option to choose $\xi_p$, it must be that

$$\begin{aligned} s &\geq \int_\Theta \left( \int_M \left[ \int_A u_S(a)\, \mathrm{d}\boldsymbol{\alpha}(a|m,\xi_p) \right] \mathrm{d}\left[\chi \xi_p(m|\theta) + (1-\chi)\, \boldsymbol{\sigma}(m|\theta,\xi_p)\right] \right) \mathrm{d}\mu_0(\theta) \\ &\geq \chi \int_\Theta \int_M \left[ \int_A u_S(a)\, \mathrm{d}\boldsymbol{\alpha}(a|m,\xi_p) \right] \mathrm{d}\xi_p(m|\theta)\, \mathrm{d}\mu_0(\theta) + (1-\chi) \min w(\Delta\Theta) \\ &\geq \chi \int_{\Delta\Theta} (w-\epsilon)\, \mathrm{d}p + (1-\chi) \min w(\Delta\Theta) \\ &\geq \chi(\underline{s}_1 - \epsilon) + (1-\chi) \min w(\Delta\Theta) \end{aligned}$$

Because $s$ was the payoff from an arbitrary $\chi$-PBE, we learn that every $\chi \in (\underline{\chi}, 1)$ has $\underline{s}_\chi \geq \chi(\underline{s}_1 - \epsilon) + (1-\chi) \min w(\Delta\Theta)$, which converges to $\underline{s}_1 - \epsilon$ as $\chi$ converges to 1. Hence, $\liminf_{\chi \nearrow 1} \underline{s}_\chi \geq \underline{s}_1 - \epsilon$. But $\epsilon$ was itself an arbitrary positive constant, so that $\liminf_{\chi \nearrow 1} \underline{s}_\chi \geq \underline{s}_1$, as desired. □

## B.3 Signaling Credibility: Proofs from section 6.2

In this section we present the formal analysis of the modified game in which S can signal her credibility through the choice of the official reporting protocol.

We start by introducing the modified game and notation. At the beginning, S privately learns her credibility type $t \in T = \{o, i\}$, i.e., whether the message will be determined according to the official protocol ($t = o$), or it will be possible to influence it ($t = i$). Then the game proceeds exactly as in our main model.

We focus on perfect Bayesian equilibria in which R's off-path beliefs satisfy a standard *"no signaling what you don't know"* restriction. To capture this idea, we define the equilibrium as follows. Let $(\xi_o, \xi_i) \in \Xi^2$ denote S's signaling strategy.[46] Let $\tilde{\boldsymbol{\chi}} : \Xi \to [0,1]$ denote R's belief mapping from an announced official reporting protocol to S's posterior credibility. Then, a $\chi$ **signaling PBE ($\chi$-SPBE)** is a tuple $(\xi_o, \xi_i, \boldsymbol{\sigma}, \boldsymbol{\alpha}, \tilde{\boldsymbol{\chi}}, \boldsymbol{\pi})$ of measurable maps[47] such that

1. $\tilde{\boldsymbol{\chi}}$ is derived from $\chi$ via Bayes' rule, given signal $t \mapsto \xi_t$, whenever possible.[48]

[46] To simplify notation, here we focus on pure signaling strategies. An analogous result holds for mixed signaling strategies.

[47] We define maps $\boldsymbol{\sigma}, \boldsymbol{\alpha}, \boldsymbol{\pi}$ as in the definition of $\chi$-PBE in section 6.1, but additionally require the maps be measurable, where we view the space $\Xi$ of measurable maps $\Theta \to \Delta M$ as a measurable space in which every subset is measurable. To simplify notation, let $\boldsymbol{\sigma}_\xi := \boldsymbol{\sigma}(\cdot, \xi)$ and similarly for $\boldsymbol{\alpha}$, $\tilde{\boldsymbol{\chi}}$, and $\boldsymbol{\pi}$.

[48] For convenience, we identify with $\tilde{\boldsymbol{\chi}}_\xi$ the element of $\Delta T$ that assigns probability $\tilde{\boldsymbol{\chi}}_\xi$ to $\{o\}$

2. $(\xi, \boldsymbol{\sigma}_\xi, \boldsymbol{\alpha}_\xi, \boldsymbol{\pi}_\xi)$ is a $\tilde{\boldsymbol{\chi}}_\xi$-equilibrium (for prior $\mu_0$) for each $\xi \in \Xi$.

3. $\xi_t$ maximizes $\boldsymbol{s}_t(\cdot)$ over $\Xi$, for each $t \in \{o, i\}$, where

$$
\begin{aligned}
\boldsymbol{s_o} &\colon \Xi \to \mathbb{R} \\
&\quad \xi \mapsto \int_\Theta \int_M u_S(\boldsymbol{\alpha}_\xi(m)) \, \mathrm{d}\xi(m|\cdot) \, \mathrm{d}\mu_0, \\
\boldsymbol{s_i} &\colon \Xi \to \mathbb{R} \\
&\quad \xi \mapsto \int_\Theta \int_M u_S(\boldsymbol{\alpha}_\xi(m)) \, \mathrm{d}\boldsymbol{\sigma}_\xi(m|\cdot) \, \mathrm{d}\mu_0.
\end{aligned}
$$

We call $(\max_\Xi \boldsymbol{s_o}, \max_\Xi \boldsymbol{s_i}) = (\boldsymbol{s_o}(\xi_o), \boldsymbol{s_i}(\xi_i))$ the corresponding S **payoff vector**. A **pooling $\chi$-SPBE** is one in which $\xi_o = \xi_i$.

Note that the above definition is equivalent to perfect Bayesian equilibria in which R updates joint beliefs over $T \times \Theta$, satisfying a "no signaling what you don't know" refinement. Indeed, since the official protocol announcement cannot convey information about the state, the $T$-marginal $\tilde{\boldsymbol{\chi}}_\xi$ pins down the joint belief $\tilde{\boldsymbol{\chi}}_\xi \otimes \mu_0$. Then, given the form of R's incentive constraints after a message is received, it is enough to keep track of only the $\Theta$-marginal $\boldsymbol{\pi}_\xi$.

Recall, $\underline{w} : \Delta\Theta \to \mathbb{R}$ is the quasiconvex envelope of $w$—that is, the pointwise highest quasiconvex and lower semi-continuous function that is everywhere below $w$, or, equivalently, $-\underline{w} = \overline{-w}$. It follows directly from Lipnowski and Ravid (2020) that a sender-worst 0-equilibrium exists and delivers S payoff $\underline{w}(\mu_0)$.

The following proposition establishes the equivalence between $\chi$-equilibrium payoff vectors and $\chi$-SPBE payoff vectors for S.

**Proposition 6.** *Fixing $(s_o, s_i) \in \mathbb{R}^2$, the following are equivalent:*

(a) *$(s_o, s_i)$ is a $\chi$-SPBE S payoff vector;*

(b) *$(s_o, s_i)$ is a pooling $\chi$-SPBE S payoff vector;*

(c) *$(p, s_o, s_i)$ is a $\chi$-equilibrium outcome for some $p \in \mathcal{R}(\mu_0)$.*

*Proof.* First, (b) trivially implies (a).

Now, let us show (c) implies (b). To do so, consider some $\chi$-equilibrium $(\xi, \sigma, \alpha, \pi)$ generating outcome $(p, s_o, s_i)$. Observe that, for each $\xi' \in \Xi \setminus \{\xi\}$, some uncountable

Borel $M_{\xi'} \subset M$ exists such that $\int_\Theta \xi'(M_{\xi'}|\cdot)\,\mathrm{d}\mu_0 = 0$.[49] It then follows readily from Theorem 2 of Lipnowski and Ravid (2020) that some 0-equilibrium $(\xi', \sigma_{\xi'}, \alpha_{\xi'}, \pi_{\xi'})$ exists giving S payoff $\underline{w}(\mu_0)$ with messages restricted to $M_{\xi'}$, that is, with $\boldsymbol{\sigma}_{\xi'}(M_{\xi'}|\cdot) = \mathbf{1}$. We now proceed to construct a pooling $\chi$-SPBE. Define an influencing sender strategy $\boldsymbol{\sigma}$ and credibility belief function $\tilde{\boldsymbol{\chi}}$ by letting, for each $\xi' \in \Xi$,

$$(\boldsymbol{\sigma}_{\xi'},\ \tilde{\boldsymbol{\chi}}_{\xi'}) := \begin{cases} (\sigma,\ \chi) & :\ \xi' = \xi \\ (\sigma_{\xi'},\ 0) & :\ \xi' \neq \xi. \end{cases}$$

Next, fix some $\mu_* \in \operatorname{argmin}_{\Delta\Theta} w$ and some R best response $a_*$ to $\mu_*$ with $u_S(a_*) = w(\mu_*)$. Define a receiver strategy $\boldsymbol{\alpha}$ and belief map (concerning the state) $\boldsymbol{\pi}$ by letting, for each $\xi' \in \Xi$ and $m \in M$,

$$(\boldsymbol{\alpha}_{\xi'}(m),\ \boldsymbol{\pi}_{\xi'}(m)) := \begin{cases} (\alpha(m),\ \pi(m)) & :\ \xi' = \xi \\ (\alpha_{\xi'}(m),\ \pi_{\xi'}(m)) & :\ \xi' \neq \xi,\ m \notin M_{\xi'} \\ (\delta_{a_*},\ \mu_*) & :\ \xi' \neq \xi,\ m \in M_{\xi'}. \end{cases}$$

By construction, $(\xi, \xi, \boldsymbol{\sigma}, \boldsymbol{\alpha}, \tilde{\boldsymbol{\chi}}, \boldsymbol{\pi})$ satisfies conditions 1 and 2 of the definition of $\chi$-SPBE. Moreover, observe that, by Lemma 10, some $\gamma, \beta \in \Delta\Theta$ exist such that $s_i \geq \max\{\underline{w}(\beta), \underline{w}(\gamma)\}$ and $\mu_0 \in \operatorname{co}\{\gamma, \beta\}$. Hence, $s_i \geq \underline{w}(\mu_0)$ since $\underline{w}$ is quasiconvex. Therefore, condition 3 of the definition of a $\chi$-SPBE is satisfied because $\boldsymbol{s_i}(\xi) = s_i \geqslant \underline{w}(\mu_0) = \boldsymbol{s_i}(\xi')$ and $\boldsymbol{s_o}(\xi) = s_o \geqslant \min_{\Delta\Theta} w = \boldsymbol{s_o}(\xi')$ for all $\xi' \in \Xi \setminus \{\xi\}$. Therefore, $(\xi, \xi, \boldsymbol{\sigma}, \boldsymbol{\alpha}, \tilde{\boldsymbol{\chi}}, \boldsymbol{\pi})$ is a pooling $\chi$-SPBE with S's payoff vector $(s_o, s_i)$ as desired.

It remains to show that (a) implies (c). To that end, suppose $(s_o, s_i)$ is some $\chi$-SPBE payoff vector, as witnessed by $\chi$-SPBE $(\xi_o, \xi_i, \boldsymbol{\sigma}, \boldsymbol{\alpha}, \tilde{\boldsymbol{\chi}}, \boldsymbol{\pi})$ generating payoff vector $(s_o, s_i)$, and let the functions $\boldsymbol{s_o}, \boldsymbol{s_i}$ be as defined in the definition of a $\chi$-SPBE; recall $\boldsymbol{s_o}, \boldsymbol{s_i} \leq s_i$ and $\boldsymbol{s_i}(\xi_i) = s_i$. For any $\xi \in \Xi$ with $\tilde{\boldsymbol{\chi}}_\xi = 1$, that $\boldsymbol{s_i}(\xi) \leq s_i$ implies we can assume without loss (modifying $\boldsymbol{\alpha}_\xi(m)$ and $\boldsymbol{\pi}_\xi(m)$ for some $m \in M$ with $\int_\Theta \xi(m|\cdot)\,\mathrm{d}\mu_0 = 0$, and modifying $\boldsymbol{\sigma}_\xi$, similarly to in the proof of Lemma 15) that $\boldsymbol{s_i}(\xi) = s_i$. Therefore, $\boldsymbol{s_i}(\xi_i) = \boldsymbol{s_i}(\xi_o) = s_i$. For each $\xi \in \{\xi_o, \xi_i\}$, then, Lemma 1

[49] For any Borel probability measure $\eta$ on $[0,1]$, construct an uncountable Borel $\eta$-null $X \subseteq [0,1]$ as follows. First, express $\eta = \lambda\eta_d + (1-\lambda)\eta_c$ for some $\lambda \in [0,1]$ and $\eta_d, \eta_c \in \Delta[0,1]$ with $\eta_d$ discrete and $\eta_c$ atomless; define the co-countable set $\hat{X} := \{x \in [0,1] :\ \eta_d\{x\} = 0\}$. Let $F$ denote the (continuous) CDF of $\eta_c$. If $F$ is constant on some nondegenerate interval $I \subseteq [0,1]$, then $X := \hat{X} \cap I$ is as desired. Otherwise, $X := \hat{X} \cap F^{-1}(\mathcal{C})$ is as desired, where $\mathcal{C} \subset [0,1]$ is the Cantor set.

Finally, such $M_{\xi'}$ exists since $\int_\Theta \xi'\,\mathrm{d}\mu_0$ is a Borel probability measure on $M$, and the measurable space $M$ is isomorphic to $[0,1]$ by the Borel isomorphism theorem.

delivers $k_\xi \in [0,1]$ and $g_\xi, b_\xi \in \Delta\Delta\Theta$ satisfying

$$\begin{aligned}
&k_\xi b_\xi + (1-k_\xi)g_\xi \in \mathcal{R}(\mu_0),\\
&(1-k_\xi)\int \mu \, \mathrm{d}g_\xi(\mu) \geqslant (1-\tilde{\boldsymbol{\chi}}_\xi)\mu_0,\\
&g_\xi\{s_i \in V\} = b_\xi\{s_i \geq \min V\} = 1,\\
&s_i - \boldsymbol{s_o}(\xi) \in \tfrac{k_\xi}{\tilde{\boldsymbol{\chi}}_\xi}\left[s_i - \int s_i \wedge V \, \mathrm{d}b_\xi\right].
\end{aligned}$$

But then consider

$$\begin{aligned}
k &:= \chi k_{\xi_o} + (1-\chi)k_{\xi_i} \in [0,1),\\
b &:= \tfrac{\chi k_{\xi_o}}{k} b_{\xi_o} + \left(1 - \tfrac{\chi k_{\xi_o}}{k}\right) b_{\xi_i} \in \Delta\Delta\Theta,\\
g &:= \left(1 - \tfrac{(1-\chi)(1-k_{\xi_i})}{1-k}\right) g_{\xi_o} + \tfrac{(1-\chi)(1-k_{\xi_i})}{1-k} g_{\xi_i} \in \Delta\Delta\Theta.
\end{aligned}$$

Direct computations with $(k,g,b)$ then show, by Lemma 1, that $(kb+(1-k)g, s_o, s_i)$ is a $\chi$-equilibrium outcome.

□

## B.4 Investing in Credibility: Proofs from section 6.4

In this section, we prove Claim 3 concerning the public persuasion application with costly endogenous credibility. Toward the proof, we first establish the following lemma.

**Lemma 16.** *For any non-cutoff credibility choice (i.e. any $\boldsymbol{\chi}$ such that there is no $\theta^* \in [0,1]$ with $\boldsymbol{\chi} = \mathbf{1}_{[0,\theta^*)}$ $\mu_0$-a.s.), there is some cutoff credibility choice that yields $S$ a strictly higher best equilibrium payoff net of costs.*

*Proof.* Consider any credibility choice $\boldsymbol{\chi}$ not of the desired form. In particular, this implies that $\boldsymbol{\chi}$ is not $\mu_0$-a.s. equal to 1, so that $\boldsymbol{\chi}\mu_0(\Theta) < 1$.

As $\mu_0$ is atomless, there is some $\theta^* \in [0,1)$ such that $\mu_0[0,\theta^*) = \boldsymbol{\chi}\mu_0(\Theta)$. That $\mathbf{1}_{[0,\theta^*)}\mu_0 \neq \boldsymbol{\chi}\mu_0$ but the two have the same total measure implies that $\mathrm{supp}[(\mathbf{1}-\boldsymbol{\chi})\mu_0]$ intersects $[0,\theta^*)$. For each $\theta_* \in [0,\theta^*]$, define the function $\eta_{\theta_*} := I_{\mathbf{1}_{[0,\theta_*)}\mu_0} - I_{\boldsymbol{\chi}\mu_0} : \mathbb{R}_+ \to \mathbb{R}$. By construction, its right-hand-side derivative at any $\theta$ is given by $\eta'_{\theta_*}(\theta) = \int_0^\theta (\mathbf{1}_{[0,\theta_*)} - \boldsymbol{\chi}) \, \mathrm{d}\mu_0$. In particular, this implies (since $\boldsymbol{\chi}\mu_0$ strictly first-order stochastically dominates $\mathbf{1}_{[0,\theta^*)}$) that $\eta'_{\theta^*}$ is globally nonnegative, weakly quasiconcave with peak at $\theta^*$, and not globally zero. In particular, $\eta_{\theta^*}(0) = 0$ yields $\eta_{\theta^*} \geq 0$ and $\epsilon := \frac{1}{2}\eta_{\theta^*}(\theta^*) > 0$. Now, with the prior being atomless and $\eta_{\theta^*}$ continuous, there is some $\theta_* \in [0, \theta^*)$ close enough to $\theta^*$ to ensure that $\eta_{\theta_*}(\theta_*) \geq \epsilon$ and $\mu_0(\theta_*, \theta^*] \leq \epsilon$. Let $\eta := \eta_{\theta_*}$.

As $\eta'$ is weakly quasiconcave on $[0, 1]$ (with peak at $\theta_*$), we have $\inf \eta'[0, 1] = \min\{\eta'(0), \eta'(1)\} = \min\{0, \eta'(1)\}$. But

$$\eta'(1) = \int_0^{\theta_*} 1 \, \mathrm{d}\mu_0 - \int_0^1 \boldsymbol{\chi} \, \mathrm{d}\mu_0 = \mu_0[0, \theta_*] - \mu_0[0, \theta^*] \geq -\epsilon,$$

so that $\eta'|_{[0,1]} \geq -\epsilon$.

Let us now observe that $\eta$ is nonnegative over $[0, 1]$. First, any $\theta \in [0, \theta_*]$ has $\eta(\theta) = \eta_{\theta^*}(\theta) \geq 0$. Next, any $\theta \in [\theta_*, 1]$ has

$$\eta(\theta) = \eta(\theta_*) + \int_{\theta_*}^{\theta} \eta'(\tilde{\theta}) \, \mathrm{d}\tilde{\theta} \geq \epsilon + (1 - \theta_*)(-\epsilon) = \theta_* \epsilon > 0.$$

So $I_{\mathbf{1}_{[0,\theta_*)}\mu_0} \geq I_{\boldsymbol{\chi}\mu_0}$ globally. Lemma 9 then implies that $\bar{\mu}_{\mathbf{1}_{[0,\theta_*)}} \succeq \bar{\mu}_{\boldsymbol{\chi}}$. Finally, Claim 2 tells us that $v^*_{\mathbf{1}_{[0,\theta_*)}}(\mu_0) \geq v^*_{\boldsymbol{\chi}}(\mu_0)$. Meanwhile, the cost of credibility $\mathbf{1}_{[0,\theta_*)}$, is strictly below that of credibility $\boldsymbol{\chi}$. $\square$

Now, we prove Claim 3

*Proof.* Consider any credibility choice $\boldsymbol{\chi}$ and accompanying $\boldsymbol{\chi}$-equilibrium. Lemma 16 shows that $\boldsymbol{\chi}$ is a cutoff credibility choice with cutoff $\theta_* \in [0, 1]$, or can be replaced with one for a strict improvement to the objective. Our analysis of public persuasion says that the $\boldsymbol{\chi}$-equilibrium entails influenced $\theta^*$ upper censorship for some cutoff $\theta^* \in [0, 1]$, or can be replaced with it for a strict improvement to the objective. Our main-text observation on the endogenous credibility problem (that no gratuitous credibility should be purchased) tells us that $\theta_* \leq \theta^*$, or else $\theta_*$ can be lowered to $\theta^*$ for a strict gain to the objective. But then, since $\boldsymbol{\chi}|_{[\theta_*,1]} = 0$, it is purely a normalization to set $\theta^* = \theta_*$.

The above observations tell us that we may as well restrict to the case that there is some cutoff $\theta^* \in [0, 1]$ such that S invests in cutoff credibility choice with cutoff $\theta^*$, official reporting always reveals the state, and influenced reporting reveals itself but provides no further information.

Thus, S solves (where the argument for $H$ on the right is taken to be 1 when $\theta^* = 1$)

$$\max_{\theta^* \in [0,1]} \int_0^{\theta^*} H \,\mathrm{d}\mu_0 - c\left(\mu_0[0,\theta^*)\right) + H\left(\frac{\int_{\theta^*}^1 \theta \,\mathrm{d}\mu_0(\theta)}{\mu_0[\theta^*,1]}\right).$$

This program is continuous with compact domain, so that an optimum exists. $\square$

## B.5 Simple Communication: Proof from section 6.5

We begin with a lemma showing that the program (12) always admits a solution with additional structure. In particular, whenever S-optimal $\boldsymbol{\chi}$-equilibrium requires the official reporting protocol to differ from an influencing S's behavior, we can assume without loss that every message sent by official reporting is *strictly* suboptimal for an influencing S.

**Lemma 17.** *One of the following holds:*

1. *The triple $(\beta,\gamma,k) = (\mu_0,\mu_0,0)$ is an optimal solution to program (12);*
2. *Some optimal solution $(\beta,\gamma,k)$ to program (12) and $b \in \mathcal{R}(\beta)$ exist with $k > 0$, $\int v_{\wedge\gamma}\,\mathrm{d}b = \hat{v}_{\wedge\gamma}(\beta)$, and $b\{v < \bar{v}(\gamma)\} = 1$.*

*Proof.* As observed in (the SDC generalization of) Theorem 1, program (12) admits some solution $(\beta,\gamma,k)$. Further, some $b \in \mathcal{R}(\beta)$ exists with $\int v_{\wedge\gamma}\,\mathrm{d}b = \hat{v}_{\wedge\gamma}(\beta)$ because $\mathcal{R}(\beta)$ is compact and $b \mapsto \int v_{\wedge\gamma}\,\mathrm{d}b$ is upper semicontinuous. Letting $D := \{v \geq \bar{v}(\gamma)\} \subseteq \Delta\Theta$, we have nothing to show if $b(D) = 0$, so suppose $b(D) > 0$.

Now, let $k' := k[1-b(D)] \in [0,1)$; let $\gamma' := \frac{1}{1-k'}\left[(1-k)\gamma + k\int_D \mu\,\mathrm{d}b(\mu)\right] \in \Delta\Theta$; and let $\beta' := \frac{1}{1-b(D)}\int_{(\Delta\Theta)\setminus D} \mu\,\mathrm{d}b(\mu)$ if $b(D) < 1$, and $\beta' := \mu_0$ if $b(D) = 1$. Because $k'\beta' + (1-k')\gamma' = k\beta + (1-k)\gamma$ and $(1-k')\gamma' \geq (1-k)\gamma$ by construction, $(\beta',\gamma',k')$ is feasible in (12). In what follows, we show that $(\beta',\gamma',k')$ is an optimal solution to (12) with the desired features.

First, by construction, $\gamma'$ is in the closed convex hull of $\{\bar{v} \geq \bar{v}(\gamma)\}$. But $\{\bar{v} \geq \bar{v}(\gamma)\}$ is closed and convex because $\bar{v}$ is upper semicontinuous and quasiconcave, implying $\bar{v}(\gamma') \geq \bar{v}(\gamma)$. If $k' = 0$ (in which case $\beta' = \gamma' = \mu_0$ by construction), this ranking implies $\bar{v}(\gamma') \geq (1-k)\bar{v}(\gamma) + k\hat{v}_{\wedge\gamma}(\beta)$, so that $(\beta',\gamma',k')$ is optimal too, establishing the claim.

We now focus on the remaining case that $0 < k' < 1$. That $\bar{v}(\gamma') \geq \bar{v}(\gamma)$ implies $b' := \frac{1}{1-b(D)} b\left((\cdot) \cap D\right) \in \mathcal{R}(\beta')$ has $b'\{v < \bar{v}(\gamma')\} = 1$. Moreover,

$$\begin{aligned}
(1-k')\bar{v}(\gamma') + k'\hat{v}_{\wedge\gamma'}(\beta') &\geq (1-k')\bar{v}(\gamma') + k' \int v_{\wedge\gamma'} \, \mathrm{d}b' \\
&= [1-k+k\beta(D)]\, \bar{v}(\gamma') + k' \int v_{\wedge\gamma'} \, \mathrm{d}b' \\
&= (1-k)\bar{v}(\gamma') + k \int v_{\wedge\gamma'} \, \mathrm{d}b \\
&\geq (1-k)\bar{v}(\gamma) + k \int v_{\wedge\gamma} \, \mathrm{d}b.
\end{aligned}$$

Optimality of $(\beta, \gamma, k)$ in (12) then implies $(\beta', \gamma', k')$ is optimal too. Therefore, the inequalities in the above chain must hold with equality, whence the first line of the above chain yields $\hat{v}_{\wedge\gamma'}(\beta') = \int v_{\wedge\gamma'} \, \mathrm{d}b'$. Thus, $(\beta', \gamma', k')$ and $b'$ are as required. $\square$

Although our main purpose for the above lemma is to prove Proposition 4, it is worth noting that Lemma 17 can be useful in narrowing the search for a solution to Theorem 1's program. For example, in the context of the central bank example, the lemma immediately implies that (for any $\chi$ at which S can do strictly better than her no-credibility value) one optimally sets $\beta \leq \frac{1}{4}$.

We now proceed to prove the corollary.

*Proof of Proposition 4.* By Lemma 17, some optimal solution $(\beta, \gamma, k)$ to program (12) exists such that either (1) $(\beta, \gamma, k) = (\mu_0, \mu_0, 0)$ or (2) $k > 0$, and some $\tilde{b} \in \mathcal{R}(\beta)$ has $\int v_{\wedge\gamma} \, \mathrm{d}\tilde{b} = \hat{v}_{\wedge\gamma}(\beta)$ and $\tilde{b}\{v < \bar{v}(\gamma)\} = 1$. Let $s_i := \bar{v}(\gamma)$.

In case 1, we will observe that some $g \in \mathcal{R}(\mu_0)$ exists with $g\{V \ni s_i\} = 1$ and $|\mathrm{supp}(g)|$ is weakly below the given cardinality bound. In case 2, we will observe that some $b \in \mathcal{R}(\beta)$ and $g \in \mathcal{R}(\gamma)$ exist with $b\{v < s_i\} = g\{V \ni s_i\} = 1$, and $|\mathrm{supp}(b)| + |\mathrm{supp}(g)|$ is weakly below the given cardinality bound. In either case, the proof of Lemma 1 (applied with $b = g$ in case 1) yields an S-optimal equilibrium that respects the cardinality bound on on-path messages.

First, we prove the bound based on the number of actions. Letting $A_+ := \{a \in A : \ u_S(a) \geq s_i\}$, (the proof of) Proposition 2 from Lipnowski and Ravid (2020) delivers some $g \in \mathcal{R}(\gamma)$ such that $g\{V \ni s_i\} = 1$ and $|\mathrm{supp}(g)| \leq |A_+|$. In case 1, nothing remains to show, so let us now focus on case 2. As $b \in \mathcal{R}(\beta)$ is such that $\mathrm{argmax}_{a \in A} \int u_R(a, \cdot) \, \mathrm{d}\mu \subseteq A \backslash A_+$ a.s.-$b(\mu)$, (the proof of) Proposition 1 from Kamenica and Gentzkow (2011) delivers some $b \in \mathcal{R}(\beta)$ such that $|\text{supp}(b)| \leq |A \setminus A_+|$.[50] Hence, some S-optimal $\boldsymbol{\chi}$-equilibrium $(\xi, \sigma, \alpha, \pi)$ exists in which some measurable $M^* \subseteq M$ with $|M^*| \leq |A|$ has $\xi(M^*|\cdot) = \sigma(M^*|\cdot) = \mathbf{1}$.

Now, supposing $n \coloneqq |\Theta| < \infty$, we prove the bound based on the number of states. Lemma 1 of Lipnowski and Ravid (2020) implies $\gamma$ is in the convex hull of the compact set $\{V \ni s_i\}$, and then Caratheodory's theorem says $\gamma$ is in the convex hull of some affinely independent subset $D \subseteq \{V \ni s_i\}$. Clearly, $|D| \leq n$, so nothing remains to be shown in case 1; let us now focus on case 2.

As $|D| < \infty$, we can without loss remove elements from $D$ to ensure $\gamma$ is a proper convex combination of all elements of $D$. By Choquet's theorem, $\tilde{b}$ is the barycenter of extreme points of $\mathcal{R}(b)$, which themselves must then be solutions to $\max_{b \in \mathcal{R}(\beta)} \int v_{\wedge\gamma} \, \mathrm{d}b$. Taking one such extreme point yields $b \in \text{ext}\mathcal{R}(\beta)$ such that $b\{v < s_i\} = 1$ and $\int v_{\wedge\gamma} \, \mathrm{d}b = \hat{v}_{\wedge\gamma}(\beta)$. Because extreme points of $\mathcal{R}(\beta)$ have affinely independent support, it follows that $|\text{supp}(b)| \leq n$. Hence, some S-optimal $\boldsymbol{\chi}$-equilibrium $(\xi, \sigma, \alpha, \pi)$ exists in which some $M^* \subseteq M$ with $|M^*| \leq n + |D|$ has $\xi(M^*|\cdot) = \sigma(M^*|\cdot) = \mathbf{1}$. The corollary then follows if we can establish (in case 2) that $|D| < n$.

Assume for a contradiction that $|D| = n$. Then the set of proper convex combinations of all elements of $|D|$ is an open subset of $\Delta\Theta$ that contains $\gamma$. In particular, some proper convex combination $\gamma'$ of $\gamma$ and $\mu_0$ lies in the convex hull of $|D|$. Observe three properties of $\gamma'$. First, by construction, some $k' \in (0, k)$ exists such that $k'\beta + (1 - k')\gamma' = \mu_0$. Second, quasiconcavity of $\bar{v}$ implies $\bar{v}(\gamma') \geq \min \bar{v}(D) \geq s_i$. Third,

$$(1 - k')\gamma' = \mu_0 - k'\beta \geq \mu_0 - k\beta = (1 - k)\gamma,$$

so that $(\beta, \gamma', k')$ is feasible in program (12). Hence,

$$k'\hat{v}_{\wedge\gamma'}(\beta) + (1 - k')\bar{v}(\gamma') \geq k'\hat{v}_{\wedge\gamma}(\beta) + (1 - k')s_i > k\hat{v}_{\wedge\gamma}(\beta) + (1 - k)s_i,$$

contradicting the optimality of $(\beta, \gamma, k)$. □

[50]In both of the cited propositions, the result we use is proven in the cited paper, but not written in the proposition's statement. The proof of Proposition 2 from Lipnowski and Ravid (2020) shows that any attainable equilibrium S payoff of the cheap talk game is a attainable in an equilibrium in which every on-path message is a pure-action recommendation, and the recommended action is S's preferred action in the support of R's (possibly mixed-action) response to that recommendation. The proof of Proposition 1 from Kamenica and Gentzkow (2011) shows, given a communication protocol with R best responding to Bayesian beliefs, that communication can be garbled to an incentive compatible direct recommendation producing the same joint distribution of states and actions.