Persuasion via Weak Institutions*

Elliot Lipnowski†
Columbia University

Doron Ravid
University of Chicago

Denis Shishkin
Princeton University

December 5, 2019

Abstract

A sender commissions a study to persuade a receiver, but influences the report with some state-dependent 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 show the sender especially prefers her report to be immune to influence in bad states. 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, Jon Eguia, Emir Kamenica, Navin Kartik, Stephen Morris, Pietro Ortoleva, Wolfgang Pesendorfer, Marzena Rostek, Zichang Wang, Richard van Weelden, and Leeat Yariv for useful suggestions. We also thank audience members at Princeton, ESEM 2017, CETC 2018, METC 2018, NASMES 2018, and Stony Brook 2019. We would also like to thank Chris Baker and Takuma Habu for excellent research assistance.

†E-mail: `e.lipnowski@columbia.edu`, `dravid@uchicago.edu`, `shishkin@princeton.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, capable of delivering both good and bad news. Delivering bad news might be especially difficult, requiring the institution to withstand pressure exerted by 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.

We study a persuasion game between a receiver (R, he) and a sender (S, she) who cares only about R's chosen action. The game begins with S publicly announcing an official reporting protocol, which is a Blackwell experiment about the state. After the announcement, S privately learns the state and whether her reporting protocol is credible. If credible, R observes a message drawn from the announced reporting protocol. Otherwise, S can freely choose the message that R sees. R then takes an action, not knowing the message's origin. Given state θ, reporting is credible with probability $\chi(\theta)$, a probability that we interpret as the strength of S's institution in said state.

As in the recent Bayesian persuasion literature (e.g., Kamenica and Gentzkow, 2011; Alonso and Câmara, 2016; Ely, 2017), we view S as a principal, capable of steering R toward her preferred equilibrium. Our main result (Theorem 1) characterizes S's highest equilibrium payoff . The characterization is geometric and is based on S's *value function*, which specifies the highest value S can obtain from R responding optimally to a given posterior belief. Under full credibility ($\chi(\theta) = 1$ for all θ), our model is equivalent to the one studied by Kamenica and Gentzkow (2011). As such, in this case, S's highest equilibrium value is given by the concave envelope of S's value function. The value function's quasiconcave envelope gives S's highest value under cheap talk (see Lipnowski and Ravid (2019)), and therefore S's highest equilibrium value under no credibility ($\chi(\theta) = 0$ for all θ). For intermediate credibility values, Theorem 1's characterization combines the quasiconcave envelope of S's value function and the concave envelope of S's *capped* value function, which captures S's incentive constraints.

Using our characterization, we analyze how S's and R's values change with $\chi(\cdot)$. To illustrate, consider a multinational firm (R) that can make a large investment ($a = 1$), a small investment ($a = \frac{1}{2}$), or no investment ($a = 0$) in a small open economy. Profits from each investment level depend on the state of the economy, θ, which can be good ($\theta = 1$), or bad ($\theta = 0$) with equal probability. In particular,[1]

$$u_R(a, \theta) = a\theta - \tfrac{1}{2}a^2.$$

Because the state of the world is binary, the firm's beliefs can be identified with the probability that the economy is good, μ. Given the above preferences, no investment is optimal when $\mu \leq \frac{1}{4}$; a large investment is optimal when $\mu \geq \frac{3}{4}$; and a small investment is optimal when $\mu \in \left[\frac{1}{4}, \frac{3}{4}\right]$. A local policymaker (S) wants to maximize the firm's investment, and receives a payoff of 0, 1, and 2 from no, small, and large investments, respectively. To persuade the firm, the policymaker publicly commissions a report by the central bank. Formally, a report is a Blackwell experiment producing a stochastic investment recommendation conditional on the economy's state.[2] The reliability of this recommendation is questionable, as it is produced by the announced experiment only with probability x, independent of the state. With probability $1 - x$, the bank succumbs to the policymaker's pressure, producing the policymaker's recommendation of choice.

Proposition 1 shows R is often better off with a less credible S. The proposition applies to the above example. To see this, suppose first that the bank's report is fully credible, that is $x = 1$. In this case, the optimal report recommends either a large or a small investment with equal ex-ante probability in a way that makes the firm just willing to accept each recommendation. In other words, the firm's posterior belief that the state is good is uniformly distributed on $\{\frac{1}{4}, \frac{3}{4}\}$, with the firm making a large investment when its belief is $\frac{3}{4}$, and a small investment otherwise. In this case, the firm's expected utility is $\frac{1}{8}$. Consider now a weaker central bank, capable of resisting the policymaker's pressure with a lower probability of $x = \frac{2}{3}$. Take any report that leads to an incentive-compatible large investment recommendation with positive probability. Because the policymaker gets to secretly influence the report with probability $1 - x = \frac{1}{3}$, the report produces a large investment recommendation with a probability of at least $\frac{1}{3}$, *regardless of the state*. By Bayes' rule, conditional on such a recommendation, the firm's posterior belief that the state is good is no greater than $\frac{3}{4}$. Note this upper bound can be achieved only if the bank's official report fully reveals the

[1] An alternative, behaviorally equivalent specification has $u_R(a, \theta) = -(a - \theta)^2$.

[2] Restricting attention to such experiments in this example turns out to be without loss.

state. Hence, the report must generate a "no investment" recommendation whenever the economy is bad and reporting is uninfluenced (which happens with probability $\frac{1}{3}$), and a "large investment" recommendation otherwise. This policy is strictly better for the policymaker than conveying no information (which yields a small investment with certainty), and so is the policymaker's unique preferred equilibrium. Thus, when $x = \frac{2}{3}$, the firm's expected utility is $\frac{1}{6}$. In particular, the firm strictly benefits from a weaker central bank; that is, productive mistrust occurs.

Our next result, Proposition 2, shows that small decreases in credibility lead to large drops in the sender's value for all interesting 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 clearly present in our example: Given the preceding analysis, $\frac{2}{3}$ is the lowest credibility level that allows the bank to credibly recommend a large investment. For any $x < \frac{2}{3}$, the policymaker can do no better than have the bank provide no information to the firm, giving the policymaker a payoff of $\frac{1}{2}$. Because $\frac{2}{3}$ is the policymaker's payoff when $x = \frac{2}{3}$, even an infinitesimal decrease in credibility results in a discrete drop in the policymaker's value.

One can construct examples in which S's value collapses at full credibility. For example, suppose the firm can make a very large investment, which yields a payoff of 10 to the policymaker, and is optimal if and only if the firm is *certain* the economy is good. Under full credibility, the policymaker can obtain a payoff of 5 by revealing the state and having the central bank recommend no investment when the economy is bad and a very large investment when the economy is good. A very large investment recommendation, however, is never credible for any $x < 1$. If it were, the policymaker would always send it when influencing the bank's report, regardless of the economy's state, and so the firm could never be completely certain that the economy's state is good. As such, the policymaker's optimal equilibrium policy for any $x \in [\frac{3}{4}, 1)$ remains as it was in the unmodified example, giving her a payoff of $\frac{3}{4}$. Thus, even a tiny imperfection in the central bank's credibility causes the policymaker's payoff to drop from 5 to $\frac{3}{4}$.

One may suspect the non-robustness of the full-credibility solution in the above modified example is rather special. Proposition 3 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 violated in the above modified example: The firm is willing to make a very large investment only if it assigns a zero probability to the economy's state being bad. 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, which enables us to assess the relative value of credibility in different states. 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 vying for the support of her populace using state-owned media. 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 is always sent when S influences the report. We also show that concentrating the credibility of S's institution in low states uniformly increases S's payoffs across all type distributions (Claim 2). Hence, S especially prefers her institution to be resistant to pressure in bad states.

To conclude our analysis, we 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 in trust are a byproduct of rigid institutional structures. Finally, we show that in the public-persuasion setting of Section 5, S always chooses an institution that is immune to influence in low states, and perfectly amenable otherwise.

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),[3] which studies sender-receiver games in which a sender commits to an information-transmission strategy. By contrast, our no-credibility case reduces to

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

cheap talk (Crawford and Sobel, 1982; Green and Stokey, 2007). In particular, we build on Lipnowski and Ravid (2019), who use the belief-based approach to study cheap talk under state-independent sender preferences.

Two recent papers (Min, 2018; Fréchette, Lizzeri, and Perego, 2019) study closely related models. Fréchette, Lizzeri, and Perego (2019) test experimentally the connection between the informativeness of the sender's communication and her credibility in the binary-state, binary-action, independent-credibility version of our model. Min (2018) looks at a generalization of the independent-credibility version 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) insights, Min (2018) also shows allowing the sender to commit with positive probability strictly helps both players in Crawford and Sobel's (1982) uniform-quadratic example.

Our paper is related to the literature on cheap talk with lying costs. In Kartik (2009), each message includes a reported state, and the cost of a message is measured via the distance between the reported and true states; as the cost increases, the sender's strategy becomes (in some sense) more truthful. In Guo and Shmaya (2019a), each communicated message is a distribution of states, and the sender faces a miscalibration cost that increases in the distance between the message and its induced equilibrium posterior belief. They obtain a surprising result: When costs are sufficiently large, the sender attains her full-commitment payoff under any extensive-form rationalizable play. Therefore, like our work, Guo and Shmaya's (2019a) model bridges the cheap talk and the Bayesian persuasion models.

Another related paper is Nguyen and Tan (2019). In Nguyen and Tan (2019), a sender has the opportunity to privately change the publicly observed outcome of a previously announced experiment. Such a change comes at a cost that may depend on the outcome. They find conditions under which the sender does not alter the experiment's outcome in the sender-optimal equilibrium, and identify examples under which the sender obtains her commitment payoff.

We also speak to the literature that studies Bayesian persuasion under additional sender incentive constraints. In Salamanca (2019), a sender can use a mediator to design a communication protocol, but cannot commit to her own reporting strategy, and therefore must satisfy truth-telling constraints. Best and Quigley (2017) and Mathevet, Pearce, and Stacchetti (2019) both study a long-lived sender who interacts with a sequence of short-lived receivers via cheap talk. Each shows how enriching the environment can restore the sender's commitment value: in Best and Quigley (2017), by coarsening receivers' information via a review aggregator, and in Mathevet, Pearce, and Stacchetti (2019), via a reputational concern for the sender. A number of papers (Perez-Richet, 2014; Hedlund, 2017; Alonso and Câmara, 2018) study persuasion by a privately informed sender who might face exogenous constraints in her choice of signal. Perez-Richet (2014) studies the information-design analogue of an informed-principal (Myerson, 1983) problem. In Alonso and Câmara (2018) and Hedlund (2017), the sender is imperfectly informed. The former compares the value of expertise with the uninformed case and shows that private information cannot be beneficial if the sender's private information is (sequentially) redundant relative to the set of available signals. The latter shows that in a two-state model with state-independent preferences, the sender's behavior in any D1 equilibrium reveals either the sender's private information or the state. Perez-Richet and Skreta (2018) introduce the possibility of falsification in the context of test design, where a sender can make each state produce the conditional signal distribution associated with the other. Thus, their sender can manipulate a Blackwell experiment's input, whereas our sender manipulates the experiment's output.

Our productive-mistrust result relates to Ichihashi (2019), who analyzes the effect of bounding the informativeness of the sender's experiment in the binary-action specialization of Kamenica and Gentzkow (2011). Ichihashi's (2019) main result characterizes the equilibrium outcome set as a function of said upper bound. He also shows that, whereas such a bound often helps the receiver, the receiver is always harmed from such a bound when the state is binary. By contrast, productive mistrust can occur with any number of states.

The model we analyze in Section 5 concerns persuasion of a population, and so relates thematically to the literature on persuasion with multiple receivers (e.g., Alonso and Câmara, 2016; Bardhi and Guo, 2018; Chan et al., 2019). Because our sender's motive is separable across audience members, the model in that section can be reinterpreted as communication to a single receiver who holds private information. Consequently, it relates to work by Kolotilin (2018), Guo and Shmaya (2019b), and Kolotilin et al. (2017), all of whom study information design under full commitment. We contribute to this literature by studying the effects of limited credibility.

Whereas our sender derives credibility through an institution, credibility can also arise via hard evidence. The effect of evidence on communication has been the subject of many studies (Glazer and Rubinstein, 2006; Sher, 2011; Hart, Kremer, and Perry, 2017; Ben-Porath, Dekel, and Lipman, 2019; Rappoport, 2017). Many such studies share our assumption of sender state-independent preferences but focus on receiver-

(rather than sender-) optimal equilibria. The equivalence between such equilibria and the receiver's commitment outcome is a common point of inquiry.

Weak institutions often serve as a justification for examining mechanism design under limited commitment (Bester and Strausz, 2001; Skreta, 2006; Deb and Said, 2015; Liu et al., 2019). We complement this literature by relaxing a principal's commitment power in the control of information rather than of mechanisms.

2 A Weak Institution

There are 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 $u_S : A \to \mathbb{R}$ and $u_R : A \times \Theta \to \mathbb{R}$, respectively, and each aims to maximize expected payoffs.

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

We now introduce some notation, which we use throughout. For a compact metrizable space, Y, we denote by ΔY the set of all Borel probability measures over Y, endowed with the weak* topology. If $f : Y \to \mathbb{R}$ is bounded and measurable and $\zeta \in \Delta Y$, define the measure $f\zeta$ on Y via $f\zeta(\hat{Y}) := \int_{\hat{Y}} f \, \mathrm{d}\zeta$ for each Borel $\hat{Y} \subseteq Y$. When the domain is not ambiguous, we use $\mathbf{1}$ and $\mathbf{0}$ to denote constant functions taking value 1 and 0, respectively.

We impose some technical restrictions on our model. Both A and Θ are compact metrizable spaces with at least two elements, the objectives u_R and u_S are continuous, and $\chi : \Theta \to [0, 1]$ is measurable. We say the model is **finite** when referring to the special case in which both A and Θ are finite. The state, θ, is assumed to follow some full-support prior distribution $\mu_0 \in \Delta\Theta$, which is known to both players. Finally, we assume the message space M is an uncountable compact metrizable space.[4]

[4] This richness condition enables our complete characterization of equilibrium outcomes (Lemma 1). If Θ is finite, our characterization of sender-optimal equilibrium values (Theorem 1) and our applied propositions hold without change for all M such that $|M| \geq 2|\Theta|$.

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

1. $\pi : M \to \Delta\Theta$ is derived from μ_0 via Bayes' rule, given message policy
$$\chi\xi + (1 - \chi)\,\sigma : \Theta \to \Delta M,$$
whenever possible;
2. $\alpha(m)$ is supported on $\operatorname{argmax}_{a\in A} \int_\Theta u_R\,(a, \cdot)\ \mathrm{d}\pi(\cdot|m)$ for all $m \in M$;
3. $\sigma(\theta)$ is supported on $\operatorname{argmax}_{m\in M} \int_A u_S\ \mathrm{d}\alpha(\cdot|m)$ for all $\theta \in \Theta$.

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

3 Persuasion with Partial Credibility

This section presents Theorem 1, which geometrically characterizes S's optimal χ-equilibrium value. To prove the theorem, we adopt a belief-based approach by using R's ex-ante belief distribution, $p \in \Delta\Delta\Theta$, to summarize equilibrium communication. When communication is sufficiently flexible, the sole restriction imposed on an induced belief distribution is Bayes plausibility: R's average posterior belief equals his prior belief; that is, $\int_{\Delta\Theta} \mu\ \mathrm{d}p(\mu) = \mu_0$. We refer to any such p as an **information policy** and denote the set of all information policies by $\mathcal{R}(\mu_0)$.

We represent each of S's messages with the posterior belief it induces in equilibrium and use S's **value correspondence**,

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

to account for R's incentive constraints. In words, $V(\mu)$ is the set of payoffs that S can attain when R behaves optimally given posterior belief μ. Note that (appealing to Berge's theorem) V is a Kakutani correspondence, that is, a nonempty-compact-convex-valued, upper hemicontinuous correspondence. As such, S's **value function**, $v(\mu) := \max V(\mu)$, which identifies S's highest continuation payoff from inducing posterior μ, is a well-defined, upper semicontinuous function.

When S is fully credible ($\chi(\cdot) = \mathbf{1}$), only S's official reporting protocol matters. Because S publicly commits to this rule at the beginning of the game, Bayes plausibility is the only constraint imposed on equilibrium communication. Hence, R may as well break ties in S's favor, reducing the maximization of S's equilibrium value to the maximization of v's expected value across all information policies. Aumann and Maschler (1995) and Kamenica and Gentzkow (2011) show the highest such value is given by the pointwise lowest concave upper semicontinuous function that majorizes v.[5] This function, which we denote by $\hat{v}$, is known as v's **concave envelope**.

Under no credibility ($\chi(\cdot) = \mathbf{0}$), the official reporting protocol plays no role, because S always influences the report. Therefore, S's messages must satisfy her incentive constraints, which take a very simple form due to S's state-independent payoffs: All on-path messages must give S the same continuation payoff. Lipnowski and Ravid (2019) show the maximal value that S can attain subject to this constraint is given by v's **quasiconcave envelope**, which is the lowest quasiconcave upper semicontinuous function that majorizes v. We denote this function by $\bar{v}$.

Theorem 1 shows that for intermediate $\chi(\cdot)$, S's highest χ-equilibrium value is characterized by an object that combines the concave and quasiconcave envelopes. For $\gamma \in \Delta\Theta$, define

$$\begin{aligned} v_{\wedge\gamma} : \Delta\Theta &\to \mathbb{R} \\ \mu &\mapsto \min\{\bar{v}(\gamma), v(\mu)\}. \end{aligned}$$

Theorem 1's characterization is based on the concave envelope of $v_{\wedge\gamma}$, which we denote by $\hat{v}_{\wedge\gamma}$. Figure 1 below visualizes the construction of $\hat{v}_{\wedge\gamma}$ in the binary-state case.

With the relevant building blocks in hand, we now state our main result.

[5]In the case in which Θ is finite, the qualifier "upper semicontinuous" may be ommited in the definition of the (quasi)concave envelope. For instance, see Lipnowski and Ravid (2019), Corollary 4.

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

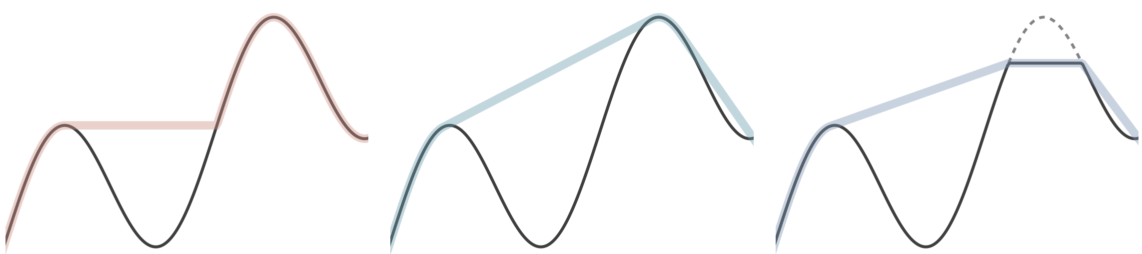

Figure 1: Quasiconcave envelope, concave envelope, and concave envelope with a cap.

Theorem 1. *A sender-optimal χ-equilibrium exists and yields ex-ante sender payoff*

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

To understand Theorem 1, note that every χ-equilibrium partitions the messages R sees into two sets: the messages that are sometimes sent under influenced reporting, M_γ (messages that are "good" for S), and the messages that are not, M_β (those that are "bad" for S). Official reporting can send messages from either set. The theorem follows from maximizing S's expected payoffs from M_γ and M_β, holding R's expected posterior conditional on M_γ and M_β fixed at γ and β, respectively. As we explain below, this maximization yields a value of $k\hat{v}_{\wedge\gamma}(\beta) + (1-k)\bar{v}(\gamma)$, where k is the probability that the realized message is in M_β. All that remains is to maximize this value over the set of feasible triplets, (β, γ, k), which are constrained by Bayes plausibility in two ways, corresponding to (R-BP) and (χ-BP), respectively. First, the average posteriors must be equal to the prior, yielding (R-BP). Second, the ex-ante probability that R sees a message from M_γ and an event $\hat{\Theta}$ occurs is at least the ex-ante probability that $\hat{\Theta}$ occurs and reporting is influenced.

We now explain the characterization of S's optimal values from M_γ and M_β, which is based on the no-credibility and full-credibility cases, respectively. Because all messages in M_γ are sent under influenced reporting, they must satisfy the same constraints as in the no-credibility case. By Lipnowski and Ravid's (2019) arguments, $\bar{v}(\gamma)$ is the highest payoff that S can obtain from sending a message under these constraints. For

S to send such messages, though, S's payoff from M_γ must be above her continuation payoff from any message in M_β. This requirement restricts M_β in two ways: (1) It caps S's continuation payoff from any feasible posterior, and (2) it restricts the set of feasible posteriors in M_β, precluding posteriors from which S must obtain too high a continuation payoff. In the proof, we argue the second constraint is automatically satisfied at the optimum. As such, one can apply the same arguments as in the full-credibility case, but with v replaced by $v_{\wedge\gamma}$. That S's highest payoff from M_β is given by $\hat{v}_{\wedge\gamma}(\beta)$ follows.

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 often lead to a large drop in S's payoffs. Finally, we note that these drops rarely occur at full credibility. In other words, the full credibility value is robust to small decreases in S's commitment power.

Productive Mistrust We now study how a decrease in S's credibility impacts 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.[6] 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 credibility can result in a more informative equilibrium (à la Blackwell, 1953). Moreover, 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 μ, an information policy $p \in \mathcal{R}(\mu)$ is a **show-or-best** (SOB) policy if it is supported on $\{\delta_\theta\}_{\theta\in\Theta} \cup \text{argmax}_{\mu'\in\Delta[\text{supp}(\mu)]} v(\mu')$. 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, $\int_{\Delta\Theta} v \, \mathrm{d}p \leq \int_{\Delta\Theta} v \, \mathrm{d}p'$.

[6]See Lemma 1 in the appendix.

Figure 2 depicts an example in which S is a two-faced SOB. Note that productive mistrust cannot occur in this example. Indeed, one can show that, if S's favorite equilibrium policy changes as credibility declines, it must switch to no information. As such, R prefers a more 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.[7]

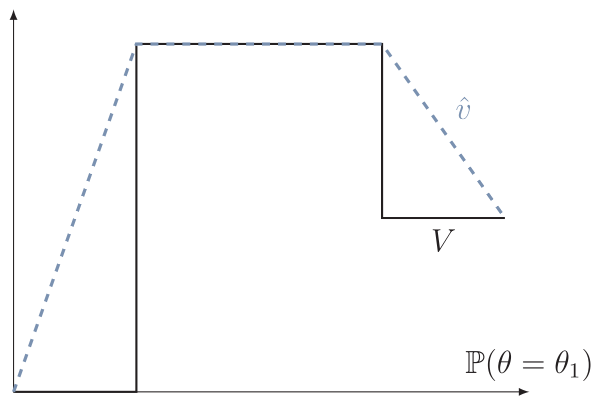

Figure 2: Sender is a two-faced SOB

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 prior. Intuitively, S being an SOB means that 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 that S's bad messages optimally hide some instrumental information. By reducing S's credibility just enough to make the full-credibility solution infeasible, one can push S 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.

[7]Given a fixed finite A and Θ, 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 $\theta_1, \theta_2 \in \Theta$.

Proposition 1. *Consider a finite and generic model in which S is not a two-faced SOB. Then, a full-support prior and credibility functions $\chi'(\cdot) < \chi(\cdot)$ exist such that every sender-optimal χ'-equilibrium is strictly better for R than every sender-optimal χ-equilibrium.*[8]

We should emphasize that Proposition 1's conditions are not necessary. We provide a necessary and sufficient condition for productive mistrust to occur at a given prior for the binary-state, finite-action case in the appendix. In particular, we weaken the SOB condition by requiring only that S wants to withhold information at the lowest credibility level at which she can beat her no-credibility payoff. We refer the reader to Lemma 2 in the appendix for precise details. We do not know an analogous tight characterization of when productive mistrust occurs in the many-state model.

Collapse of Trust Theorem 1 immediately implies lowering S's credibility can only decrease her value.[9] 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:[10]

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

for every $\chi(\cdot) \in [0,1]^\Theta$ and every full-support prior μ_0.

(ii) Commitment is of no value: $v^*_{\mathbf{1}} = v^*_{\mathbf{0}}$.

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

Proposition 2 establishes that, in most finite examples, S's value collapses discontinuously when credibility decreases. In particular, such collapses are absent for all priors if and only if S wants to tell R all that she knows, or if, equivalently, commitment is immaterial to S.

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

[9]It also implies value *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.

[10]Convergence of $\chi'(\cdot) \to \chi(\cdot)$ is in the Euclidean topology on $\mathbb{R}^\Theta$.

Robustness of the Commitment Case Given the large and growing literature on optimal persuasion with commitment, wondering whether the commitment solution is robust to small decreases in S's credibility is natural. The answer turns out to be almost never. Thus, although small decreases in credibility often lead to a collapse in S's value, these collapses rarely occur at $\chi(\cdot) = \mathbf{1}$.

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

(i) The full commitment value is robust: $\lim_{\chi(\cdot)\nearrow\mathbf{1}} v^*_\chi(\mu_0) = v^*_{\mathbf{1}}(\mu_0)$ *for every full-support* μ_0.

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

Proposition 3 shows that the full-credibility value is robust if and only if S can persuade R to take her favorite action without ruling out any states. In other words, robustness of the commitment solution is equivalent to S getting the benefit of the doubt.

5 Persuading the Public

This section considers a single sender 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, and identify the states at which credibility is most valuable for S.

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 μ_0, and $\omega_i \in \mathbb{R}$ is receiver i's type. The mass of receivers whose type is below ω 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 $\xi : \Theta \to \Delta M$, $\sigma : \Theta \to \Delta M$, and $\pi : M \to \Delta\Theta$ 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 $\alpha : M \to [0,1]$ 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 θ, and E maps beliefs to their associated expectations.[11] As such, given a posterior μ, the proportion of receivers taking action 1 is given by $H(E\mu)$. Thus, a **χ-equilibrium** is a tuple $(\xi, \sigma, \alpha, \pi)$ where π is derived from μ_0 via Bayes' rule, $\alpha(\cdot) = H(E\pi(\cdot))$, and $\sigma(\theta)$ is supported on $\arg\max_{m \in M} \alpha(m)$ for all θ.

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 μ 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 χ-equilibrium. We begin with the extreme credibility levels. Suppose first S has no credibility; that is, $\chi = \mathbf{0}$. 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, $\chi = \mathbf{1}$. In this case, S's maximal χ-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 expectations 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.[12] Formally, we say $\mu \in \Delta\Theta$ is a **mean-preserving**

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

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

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 is well known,[13] μ_0 being a mean-preserving spread of μ is both necessary and sufficient for μ to arise as the posterior mean distribution of some information policy. Thus, S's 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, an $\omega^* \in [0,1]$ 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 θ^* upper censorship, which we now formally define. Under full credibility, θ^* upper censorship arises whenever S's official report reveals (pools) all states below (above) θ^*. Given such an official reporting protocol, it is optimal for S to say the state is above θ^* whenever she influences the report. Thus, we say (ξ,σ) is a θ^***-upper-censorship pair** if every $\theta \in \Theta$ has $\sigma(\cdot|\theta) = \delta_1$ and

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

Given a θ^*-upper-censorship pair, we refer to the resulting posterior mean distribution,[14]

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

as a θ^* **upper censorship** of μ_0. 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.

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

[13] See Gentzkow and Kamenica (2016) and references therein.

[14] Recall our notational convention: For bounded measurable $f : \Theta \to \mathbb{R}_+$ and $\mu \in \Delta\Theta$, we let $f\mu$ represent the measure defined via $f\mu(\hat{\Theta}) = \int_{\hat{\Theta}} f\ \mathrm{d}\mu$.

pair induces a χ-equilibrium. The reason is that under partial credibility, the posterior 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, θ^*, which is equivalent to[15]

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

Observe that with intermediate credibility,[16] the left-hand side of (θ^*-IC) is continuous and strictly decreasing in θ^*, strictly positive for $\theta^* = 0$, and strictly negative for $\theta^* = 1$.[17] As such, (θ^*-IC) holds whenever θ^* 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 θ^*-upper-censorship pair does not typically yield an upper censorship of μ_0 as its posterior mean distribution. Instead, every θ^*-upper-censorship pair with $\theta^* \leq \bar{\theta}_\chi$ turns out to yield a θ^* 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.

Claim 1 below shows that upper censorship always yields an S optimal χ-equilibrium. Moreover, to find the optimal censorship cutoff, one can solve the full-credibility problem with the modified prior $\bar{\mu}_\chi$.

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

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

Moreover, the corresponding θ^*-upper-censorship pair is an S-optimal χ-equilibrium that induces μ_{χ,θ^*} as its posterior mean distribution.

Using Claim 1, we can compare the value of credibility in different states. Indeed, the claim makes it obvious that, regardless of the population's type distribution, S prefers the credibility distribution χ over $\tilde{\chi}$ whenever $\bar{\mu}_\chi$ is a mean-preserving spread of $\bar{\mu}_{\tilde{\chi}}$. One can then show by construction that the converse is also true; that is, S

---

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

[16] That is, if $\mu_0\{\chi = 0\}, \mu_0\{\chi = 1\} < 1$.

[17] Recall μ_0 is assumed to be an atomless, full-support distribution over $[0, 1]$.

prefers χ to $\tilde{\chi}$ for all population type distributions *only if* $\bar{\mu}_{\tilde{\chi}}$ is a mean-preserving spread of $\bar{\mu}_{\chi}$. We present this result in Claim 2 below.

Claim 2. *$v^*_{\chi}(\mu_0) \geq v^*_{\tilde{\chi}}(\mu_0)$ for all type distributions*[18] *if and only if $\bar{\mu}_{\chi} \succeq \bar{\mu}_{\tilde{\chi}}$.*

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. Indeed, it is useful to notice that $\bar{\mu}_{\chi} \succeq \bar{\mu}_{\tilde{\chi}}$ holds if and only if[19]

$$\int_0^{\hat{\theta}} \int_0^{\theta} (\chi - \tilde{\chi}) \, \mathrm{d}\mu_0 \, \mathrm{d}\theta \geq 0 \text{ for all } \theta \in [0, \bar{\theta}_{\tilde{\chi}}].$$

Thus, the claim shows a sense in which S prefers to have more credibility in low states. Intuitively, low states are those which 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 is S's value.

6 Investing in Credibility

In this section, we extend our model to endogenize S's credibility χ. Specifically, suppose S can choose any measurable $\chi : \Theta \to [0, 1]$ at a cost of $c\left(\int \chi \, \mathrm{d}\mu_0\right)$ prior to the persuasion game, where $c : [0, 1] \to \mathbb{R}_+$ is continuous and strictly increasing. Then, S chooses χ to solve

$$v^{**}_c(\mu_0) = \max_{\chi} \left[v^*_{\chi}(\mu_0) - c\left(\int \chi \, \mathrm{d}\mu_0 \right) \right].$$

Clearly, S never invests in greater credibility than is necessary to induce her equilibrium information. As such, S always chooses (χ, k, β, γ) so that (χ-BP) holds with equality. Combining this observation with (R-BP) yields

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

[18] That is, for all H admitting a continuous, quasiconcave density.

[19] To see the equivalence, one can verify that $\hat{\theta} \in [0, \bar{\theta}_{\chi}]$ has $\int_0^{\hat{\theta}} \bar{\mu}_{\chi}[0, \theta] \, \mathrm{d}\theta = \int_0^{\hat{\theta}} \int_0^{\theta} \chi \, \mathrm{d}\mu_0 \geq \theta - E\mu_0$, and each $\hat{\theta} \in [\bar{\theta}_{\chi}, 1]$ has $\int_0^{\hat{\theta}} \bar{\mu}_{\chi}[0, \theta] \, \mathrm{d}\theta = \theta - E\mu_0$ — and similarly for $\tilde{\chi}$. Therefore, the ranking (MPS) holds vacuously above $\bar{\theta}_{\tilde{\chi}}$ and reduces to the given equation below $\bar{\theta}_{\tilde{\chi}}$.

S's problem therefore reduces to

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

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 has a unique optimal investment choice, leading to equilibrium distribution of posteriors

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

It is straightforward that this equilibrium information structure is Blackwell-monotone in λ — higher λ leads to a mean-preserving spread in posterior beliefs. Consequently, R's equilibrium payoff $(\frac{1}{4} - \frac{1}{4\lambda})$ is increasing in λ.

Whereas reducing χ 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 (β, γ, 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.

In our public-persuasion application (Section 5), we saw that optimal communication takes an upper-censorship form and S especially benefits from credibility in low states. These observations, together with the observation that S never invests in extraneous credibility, lead us to simple institutions when credibility is endogenous. In particular, S's optimal institution is fully immune to influence below a cutoff state, fully susceptible above, and fully informative in its official report. See Appendix B.6 for the formal result.

7 Conclusion

This paper studies a sender who uses a weak institution to disseminate information with the aim of persuading a receiver. An institution is weaker if it succumbs to external pressures with higher probability. Specifically, the weaker the institution is, the higher is the probability that its report reflects the sender's agenda rather than the truth. We analyze the value that the sender derives from communication through such an institution, as well as the information that it provides to the receiver.

Our analysis shows an institution's weakness reduces the sender's value through two channels: Restricting the kind of information the institution can disseminate, and reducing the value that the sender can extract from said information. Together, these channels lead to collapses of trust, whereby a slight decrease in an institution's strength yields a large drop in the sender's value. Moreover, these channels often result in productive mistrust, whereby the receiver benefits from the sender employing a weaker institution. Intuitively, to credibly communicate the information the sender wishes to convey, a weaker institution must reveal information the sender would otherwise hide. Through these effects, our model highlights the role that weak institutions play in persuasion.

Our model also allows us to analyze the value of an institution's strength in different states. As a demonstration, we study a public-persuasion setting where a single sender attempts to persuade a population of receivers to take a favorable action. In this setting, the sender commissions her institution to reveal bad states, but hides those states when influencing the report. Accordingly, the sender prefers institutions that are immune to pressure in bad states, where the conflict between her ex-post and ex-ante incentives is largest.

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

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

Bardhi, Arjada and Yingni Guo. 2018. "Modes of persuasion toward unanimous consent." *Theoretical Economics* 13 (3):1111–1149.

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

Best, James and Daniel Quigley. 2017. "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.

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

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

Deb, Rahul and Maher Said. 2015. "Dynamic screening with limited commitment." *Journal of Economic Theory* 159:891 – 928. Symposium Issue on Dynamic Contracts and Mechanism Design.

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

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

Fréchette, Guillaume, Alessandro Lizzeri, and Jacopo Perego. 2019. "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.

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. 2019a. "Costly Miscalibration." Working paper.

———. 2019b. "The interval structure of optimal disclosure." *Econometrica* 87 (2):653–675.

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.

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.

Kolotilin, Anton, Tymofiy Mylovanov, Andriy Zapechelnyuk, and Ming Li. 2017. "Persuasion of a privately informed receiver." *Econometrica* 85 (6):1949–1964.

Lipnowski, Elliot and Doron Ravid. 2019. "Cheap Talk with Transparent Motives." Working Paper, Economics Department, University of Chicago.

Liu, Qingmin, Konrad Mierendorff, Xianwen Shi, and Weijie Zhong. 2019. “Auctions with limited commitment.” *American Economic Review* 109 (3):876–910.

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

Min, Daehong. 2018. “Bayesian Persuasion under Partial Commitment.” Working paper.

Myerson, Roger B. 1983. “Mechanism design by an informed principal.” *Econometrica: Journal of the Econometric Society* :1767–1797.

Nguyen, Anh and Yong Teck Tan. 2019. “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. 2018. “Test Design under Falsification.” Working paper.

Rappoport, Daniel. 2017. “Evidence and Skepticism in Verifiable Disclosure Games.” Working paper.

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

Salamanca, Andrés. 2019. “The Value of Mediated Communication.” Working paper.

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.

A Online Appendix: Proof Exposition

This appendix provides exposition for the paper's proofs. The exposition is not formally necessary, and so a reader interested solely in our formal arguments may proceed directly to Appendix B.

We begin by explaining how to visualize Theorem 1's program. Using this visualization, we provide intuitions for Proposition 1, Proposition 2 and Proposition 3. Finally, we elaborate on the main text's exposition for Claim 1.

A.1 Visualizing Theorem 1

We now explain how to use Theorem 1 to graphically solve for S's optimal equilibrium value when Θ is binary, $\Theta = \{\theta_1, \theta_2\}$. Consider Figure 3, which visualizes constraints (R-BP) and (χ-BP) for the binary-state case. In this figure, the horizontal axis is the mass on θ_1, and the vertical axis is the mass on θ_2. Because μ_0, β, and γ assign a total probability of 1 to both states, each of them can be represented as a point on the line connecting the two atomistic beliefs δ_{θ_1} and δ_{θ_2}. Every point underneath this line represents the product $(1-k)\gamma$ for some k and γ. The drawn box represents the constraints in Theorem 1's program. By (χ-BP), $(1-k)\gamma$ must be pointwise larger than $[\mathbf{1} - \chi(\cdot)]\mu_0$, which is the box's bottom-left corner. The box's top-right corner, which corresponds to the prior, must be pointwise larger than $(1-k)\gamma$ by (R-BP).[20] In other words, $(1-k)\gamma$ must lie within the drawn box. Once $(1-k)\gamma$ is chosen, one can recover γ and β by finding the unique points on the line $[\delta_{\theta_1}, \delta_{\theta_2}]$ that lie in the same direction as $(1-k)\gamma$ and $\mu_0 - (1-k)\gamma$, respectively.

Figure 4 shows how to simultaneously visualize the constraint illustrated in Figure 3 and S's value for the introduction's example, where χ is a constant $x\mathbf{1}$. Such a visualization enables us to solve for S's optimal equilibrium value. To do so, we start by drawing $\bar{v}$, the quasiconcave envelope of S's value function. For each feasible candidate $(1-k)\gamma$, we find the corresponding β, as in Figure 3. To calculate S's value from the resulting (β, γ, k), we simply find the value above μ_0 of the line connecting the points $(\beta, \hat{v}_{\wedge\gamma}(\beta))$ and $(\gamma, \bar{v}(\gamma))$.

[20] To see this requirement, rearrange (R-BP) to obtain that $\mu_0 - (1-k)\gamma = k\beta \geq 0$.

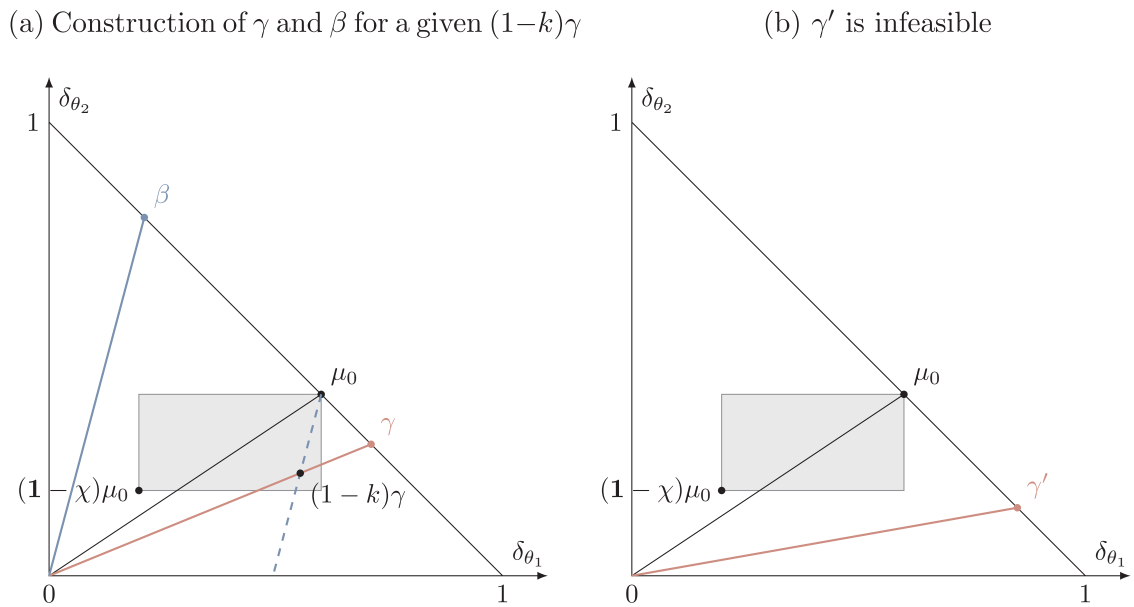

Figure 3: Constraints (R-BP) and (χ-BP) and construction of γ and β for a given $(1-k)\gamma$.

A.2 Exposition for Proposition 1

This section sketches the argument behind Proposition 1. The proposition builds on the binary-state case. In this case, genericity implies $\bar{v}$ has a non-degenerate interval of maximizers, and S not being an SOB implies $\hat{v}$ has a kink somewhere outside of this interval. Fixing a prior near this interval, but toward the nearest kink, we then find the lowest constant $x \in [0,1]$ such that S still obtains her full credibility value at $\chi(\cdot) = x\mathbf{1}$. At this $\chi(\cdot)$, S's favorite equilibrium information policy is unique and is supported on the beliefs (γ, β) that solve Theorem 1's program. These beliefs are interior, and $\hat{v}$ has a kink at β. Although γ remains optimal in Theorem 1's program for any additional small reduction in credibility, (χ-BP) forces the optimal β to move away from the prior. Relying on the set of beliefs being one-dimensional, we show the only incentive-compatible way of attaining S's new optimal value is to spread the original β between γ and a further posterior that gives S an even lower continuation value than under β. Hence, S provides R with more information. The reduction in S's value indicates a change in R's optimal behavior. In other words, the additional information is instrumental, strictly increasing R's utility. Figure 5 illustrates the argument using our introductory example.

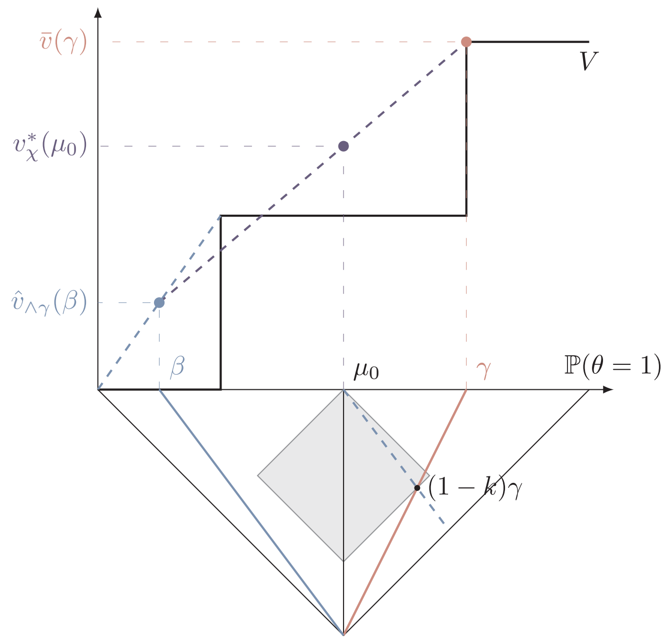

Figure 4: An illustration of the solution to Theorem 1's program for the example from the introduction, with a constant credibility level between $\frac{2}{3}$ and $\frac{3}{4}$.

A.3 Exposition for Proposition 2

This section describes the proof of Proposition 2. Notice that 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 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 that (i) implies (iii), we show that any failure of (iii) implies the failure of (i). To do so, we fix a full-support prior μ_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 some feasible triplet (β, γ, 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 a constant credibility $\chi(\cdot) = x\mathbf{1}$ if and only if x is weakly greater than some strictly positive x^*. We thus have that for all $x < x^*$,

$$v^*_{x^*\mathbf{1}}(\mu_0) \geq k\bar{v}(\mu_0) + (1-k)\bar{v}(\gamma) > \bar{v}(\mu_0) = v^*_{x\mathbf{1}}(\mu_0),$$

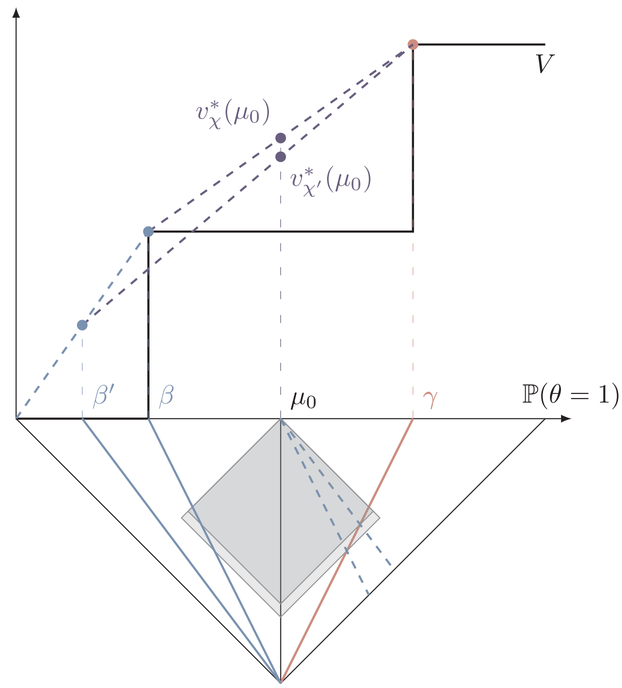

Figure 5: An illustration of Proposition 1's proof for two states. The argument begins by identifying a μ_0 as above. Given μ_0, we find two constant $\chi(\cdot) > \chi'(\cdot)$ as above, yielding the constraints depicted by the light and dark boxes, respectively. Whereas γ is optimal under both credibility levels, β is optimal under χ, whereas β' is optimal under χ'. One can then deduce R is strictly better off under χ' than under χ.

where the first inequality follows from μ_0 minimizing $\bar{v}$; that is, a collapse of trust occurs. Figure 6 below illustrates the argument in the context of our leading example. The figure depicts a prior that minimizes S's payoff under no credibility. The depicted constraint set is drawn for $\chi^* = x^*\mathbf{1}$, the lowest constant credibility for which a (β, γ, k) satisfying both $k < 1$ and $\bar{v}(\gamma) > \bar{v}(\mu_0)$ is feasible. In other words, $\chi^*(\cdot)$ is the lowest constant credibility at which S's value is strictly above $\bar{v}(\mu_0)$. Therefore, $v^*_{x^*\mathbf{1}}(\mu_0) > v^*_{x\mathbf{1}}(\mu_0)$ for any x strictly below x^*.

A.4 Exposition for Proposition 3

This section discusses the proof of Proposition 3 that is based on establishing a four-way equivalence between (a) S getting the benefit of the doubt, (b) $\bar{v}$ being maximized by a full-support prior γ, (c) a full-support γ existing such that $\hat{v}_{\wedge\gamma}$ and $\hat{v}$ agree over all full-support prior(s), (d) robustness to limited credibility. That (a) is equivalent to (b) follows from the arguments of Lipnowski and Ravid (2019). For the equivalence of

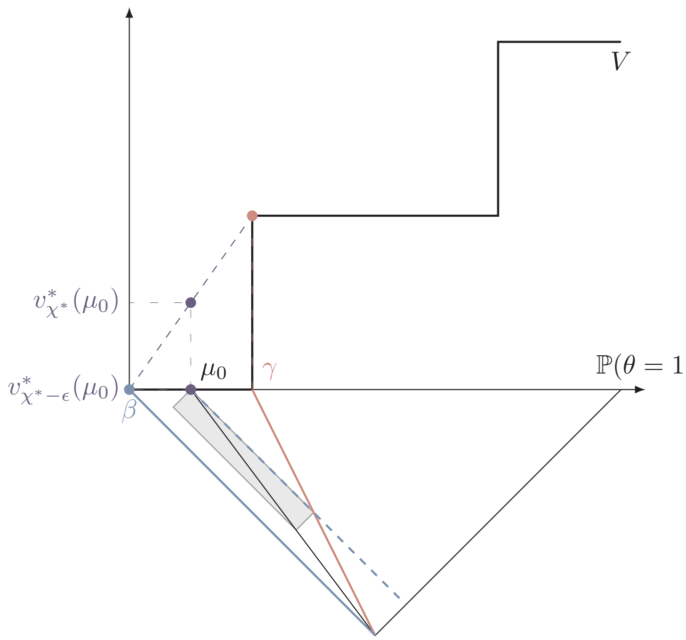

Figure 6: An illustration of the Proposition 2's proof in the context of the introduction's example. The proof starts with choosing a prior minimizing the payoff S obtains under no credibility. We then identify x^*, the lowest credibility level for which a (β, γ, k) attaining a value strictly above $\bar{v}(\mu_0)$ is feasible at $\chi^* = x^*\mathbf{1}$. By choice of x^*, $v^*_{x^*\mathbf{1}}(\mu_0) > v^*_{x\mathbf{1}}(\mu_0)$ must hold for any $x = x^* - \epsilon < x^*$; that is, S's value collapses.

(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, γ maximizes $\bar{v}$. To see why (c) is equivalent to (d), fix some full-support μ_0, and consider two questions about Theorem 1's program. First, which beliefs can serve as γ for $\chi(\cdot) \ll \mathbf{1}$ large enough?[21] Second, how do the optimal (k, β) for a given γ change as $\chi(\cdot)$ goes to $\mathbf{1}$? Figure 7 illustrates the answer to both questions for the two-state case. For the first question, the answer is that γ is feasible for some $\chi(\cdot) \ll \mathbf{1}$ if and only if γ has full support. For the second question, one can show it is always optimal to choose (k, β) so as to make (χ-BP) bind while still satisfying (R-BP).[22] Direct computation reveals that, as $\chi(\cdot)$ goes to $\mathbf{1}$, every such (k, β) must converge to $(1, \mu_0)$. Combined, one obtains that, as

[21] By $\chi(\cdot) \ll \mathbf{1}$, we mean $\chi(\theta) < 1$ for all $\theta \in \Theta$.

[22] To see why, for any feasible (k, β, γ), a (k', β') exists such that (k', β', γ) is feasible, (χ-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 $\hat{v}_{\wedge\gamma}(\gamma) = \bar{v}(\gamma)$,

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

$\chi(\cdot)$ 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 γ exists for which $\hat{v}_{\wedge\gamma} = \hat{v}$ for all full-support priors; that is, (c) is equivalent to (d). The proposition follows.

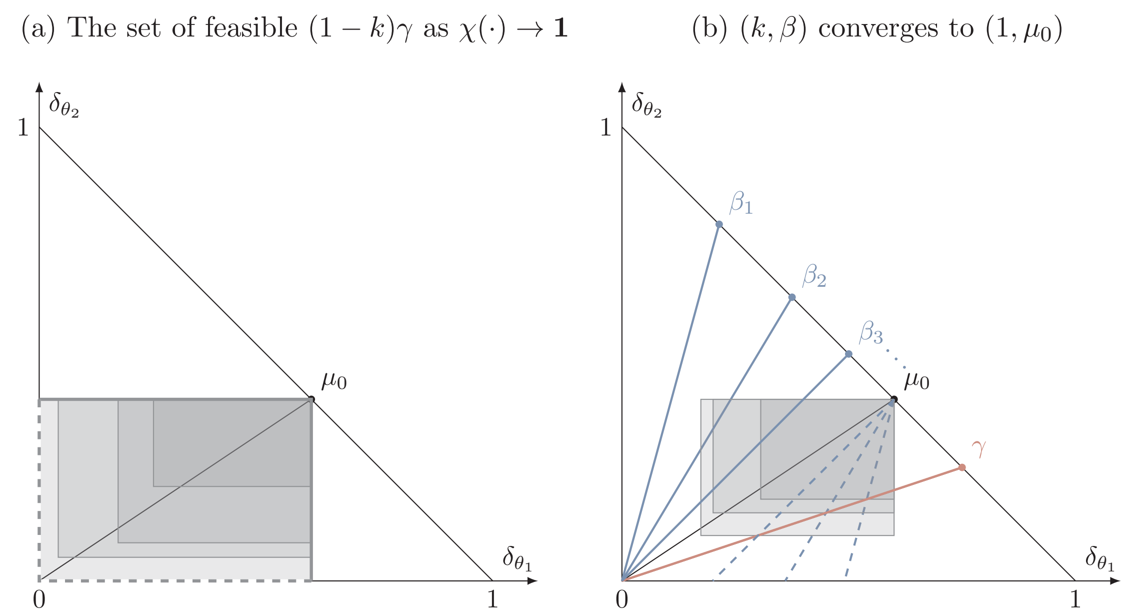

Figure 7: Robustness to limited credibility

A.5 Exposition for Claim 1

This section provides some intuition for Claim 1. Let us first explain why $v^*_\chi(\mu_0) \geq \hat{v}(\bar{\mu}_\chi)$. As explained in the main text, $\hat{v}(\bar{\mu}_\chi) = \int H \, \mathrm{d}\mu_{\chi,\theta^*}$, where μ_{χ,θ^*} is a θ^* 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 θ upper censorship of $\bar{\mu}_\chi$ for a θ above $\bar{\theta}_\chi$ is just $\bar{\mu}_\chi$ itself. Thus, assuming θ^* is in $[0,\bar{\theta}_\chi]$ is without loss. Given such a θ^*, one can induce the posterior mean distribution μ_{χ,θ^*} in a χ-equilibrium (with the original prior μ_0) using a θ^*-upper-censorship pair. As such, S's maximal χ-equilibrium value is at least as high as the value generated by μ_{χ,θ^*}; that is, $v^*_\chi(\mu_0) \geq \int H \, \mathrm{d}\mu_{\chi,\theta^*} = \hat{v}(\bar{\mu}_\chi)$.

We now sketch the reasoning behind $v^*_\chi(\mu_0) \leq \hat{v}(\bar{\mu}_\chi)$. Suppose (β,γ,k) solves Theorem 1's program. Because cheap talk is equivalent to no information (as explained earlier in this section), one can attain $\bar{v}(\gamma)$ with a single message that induces a posterior mean of $E\gamma$. Therefore, $v_{\wedge\gamma}(\mu) = H(E\gamma) \wedge H(E\mu)$, meaning $\hat{v}_{\wedge\gamma}(\beta)$ is given by

$$\hat{v}_{\wedge\gamma}(\beta) = \max_{\tilde{\beta} \preceq \beta} \int H(E\gamma) \wedge H(\cdot) \, \mathrm{d}\tilde{\beta}.$$

Using optimality of (β, γ, k), one can show the above program is solved by a $\tilde{\beta}$ whose support lies in $[0, E\gamma]$. As such, H's expected value according to $\tilde{\mu} := k\tilde{\beta} + (1-k)\delta_{E\gamma}$ equals S's maximal χ-equilibrium value; that is, $v^*_\chi(\mu_0) = \int H \, \mathrm{d}\tilde{\mu}$. Hence, a sufficient condition for $v^*_\chi(\mu_0) \leq \hat{v}(\bar{\mu}_\chi)$ is that $\tilde{\mu} \preceq \bar{\mu}_\chi$. In other words, it suffices to establish that (MPS) holds for $\bar{\mu}_\chi$ and $\tilde{\mu}$ for all $\hat{\theta}$. To establish (MPS) for $\hat{\theta} \geq E\gamma$, we use two facts. First, $\tilde{\mu}[0,\theta] = 1 \geq \bar{\mu}_\chi[0,\theta]$ holds for all $\theta \geq E\gamma$. And, second, both $\bar{\mu}_\chi$ and $\tilde{\mu}$ admit μ_0 as a mean-preserving spread. As such, $\int_0^{\tilde{\theta}} (\tilde{\mu}[0,\theta] - \bar{\mu}_\chi[0,\theta]) \, \mathrm{d}\theta$ decreases in $\hat{\theta}$ over $[E\gamma, 1]$ and reaches a value of zero at $\tilde{\theta} = 1$. It follows that (MPS) holds for $\bar{\mu}_\chi$ and $\tilde{\mu}$ for all $\tilde{\theta} \geq E\gamma$. To establish (MPS) for $\hat{\theta} < E\gamma$, notice that $\tilde{\mu}[0,\theta] = k\tilde{\beta}[0,\theta]$ whenever $\theta < E\gamma$. Therefore, if $\hat{\theta} < E\gamma$,

$$\begin{aligned}
\int_0^{\hat{\theta}} \tilde{\mu}[0,\theta] \, \mathrm{d}\theta = k \int_0^{\hat{\theta}} \tilde{\beta}[0,\theta] \, \mathrm{d}\theta &\leq k \int_0^{\hat{\theta}} \beta[0,\theta] \, \mathrm{d}\theta \\
&= \int_0^{\hat{\theta}} (\mu_0 - (1-k)\gamma)[0,\theta] \, \mathrm{d}\theta \leq \int_0^{\hat{\theta}} \chi\mu_0[0,\theta] \, \mathrm{d}\theta \leq \int_0^{\hat{\theta}} \bar{\mu}_\chi[0,\theta] \, \mathrm{d}\theta,
\end{aligned}$$

where the first inequality follows from $\beta \succeq \tilde{\beta}$, the second equality from (R-BP), and the second inequality from (χ-BP).

B Online Appendix: Proofs

We first introduce some convenient notation that we will use below. For a compact metrizable space, Y, and $f : Y \to \mathbb{R}$ bounded and measurable, let $f(\gamma) := \int_Y f \, \mathrm{d}\gamma$.

B.1 Toward the Proof of the Main Theorem

To present unified proofs, we adopt the notational convention that $\frac{0}{0} = 1$ wherever it appears.

B.1.1 Characterization of All Equilibrium Outcomes

En route to our characterization of the sender-preferred equilibrium outcomes, we characterize the full range of equilibrium outcomes.

Definition 1. $(p, s_o, s_i) \in \Delta\Delta\Theta \times \mathbb{R} \times \mathbb{R}$ *is a* χ-***equilibrium outcome*** *if there exists a* χ*-equilibrium* $(\xi, \sigma, \alpha, \pi)$ *such that, letting* $P_o := \frac{1}{\chi(\mu_0)} \int_\Theta \chi\xi \, \mathrm{d}\mu_0$ *and* $P_i := \frac{1}{1-\chi(\mu_0)} \int_\Theta (\mathbf{1} - \chi)\sigma \, \mathrm{d}\mu_0$ *be the equilibrium distributions over* M *conditional on official and influenced reporting, respectively, we have:* $p = [\chi(\mu_0)P_o + [1 - \chi(\mu_0)]P_i] \circ \pi^{-1}$, $s_o = u_S\left(\int_M \alpha \, \mathrm{d}P_o\right)$, *and* $s_i = u_S\left(\int_M \alpha \, \mathrm{d}P_i\right)$.

The following lemma adopts a belief-based approach, directly characterizing χ-equilibrium outcomes of our game.

Lemma 1. *Fix* $(p, s_o, s_i) \in \Delta\Delta\Theta \times \mathbb{R} \times \mathbb{R}$. *Then* (p, s_o, s_i) *is a* χ*-equilibrium outcome if and only if there exists* $k \in [0, 1]$, b, $g \in \Delta\Delta\Theta$ *such that*

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

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

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

(iv) $[1 - \chi(\mu_0)] \, s_i + \chi(\mu_0)s_o \in (1 - k) \, s_i + k \int_{\mathrm{supp}(b)} s_i \wedge V \, \mathrm{d}b$.[23]

[23] 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 $\mathrm{supp}(b)$. The integral is the (Aumann) integral of a correspondence:

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

Proof. As M is an uncountable Polish space, Kuratowski's theorem says M is isomorphic (as a measurable space) to $\{0,1\} \times \Delta\Theta$. We can therefore assume without loss that $M = \{0,1\} \times \Delta\Theta$.

First, suppose $k \in [0,1]$, $g, b \in \Delta\Delta\Theta$ satisfy the four listed conditions. Let ϕ be a measurable selector of $s_i \wedge V|_{\mathrm{supp}(b)}$ with $s_o = \left(1 - \frac{k}{\chi(\mu_0)}\right) s_i + \frac{k}{\chi(\mu_0)} \int_{\mathrm{supp}(b)} \phi \ \mathrm{d}b$.

Define $D := \mathrm{supp}(p)$, $\beta := \int_{\Delta\Theta} \mu \ \mathrm{d}b(\mu)$, and $\gamma := \int_{\Delta\Theta} \mu \ \mathrm{d}g(\mu)$. Let measurable $\eta_g, \eta_b : \Theta \to \Delta\Delta\Theta$ be signals that induce belief distribution g for prior γ and belief distribution b for prior β, respectively.[24] 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 μ_0-null set, we may assume that $\mathbf{0} \leq \frac{k}{\chi} \frac{\mathrm{d}\beta}{\mathrm{d}\mu_0} \leq \mathbf{1}$ since $(1-k)\gamma \geq (\mathbf{1} - \chi)\mu_0$.

Next, define the sender's influenced strategy and reporting protocol $\sigma, \xi : \Theta \to \Delta M$ by letting, for every Borel $\hat{M} \subseteq M$,

$$\begin{aligned}
\sigma(\hat{M}|\cdot) &:= \eta_g\left(\left\{\mu \in D : \ (0,\mu) \in \hat{M}\right\} \ \middle| \ \cdot\right), \\
\xi(\hat{M}|\cdot) &:= \left[\mathbf{1} - \tfrac{k}{\chi}\tfrac{\mathrm{d}\beta}{\mathrm{d}\mu_0}\right] \eta_g\left(\left\{\mu \in D : \ (0,\mu) \in \hat{M}\right\} \ \middle| \ \cdot\right) \\
&\qquad + \tfrac{k}{\chi}\tfrac{\mathrm{d}\beta}{\mathrm{d}\mu_0} \eta_b\left(\left\{\mu \in D : \ (1,\mu) \in \hat{M}\right\} \ \middle| \ \cdot\right).
\end{aligned}$$

Now, fix some $\hat{\mu} \in D$ and $\hat{a} \in \mathrm{argmax}_{a \in A} u_R(a, \hat{\mu})$ with $u_S(\hat{a}) \leq s_i$; we can then define a receiver belief map as

$$\begin{aligned}
\pi : M &\to \Delta\Theta \\
m &\mapsto \begin{cases} \mu & : \ m \in \{0,1\} \times \{\mu\} \text{ for } \mu \in D \\ \hat{\mu} & : \ m \notin \{0,1\} \times D. \end{cases}
\end{aligned}$$

Finally, by Lipnowski and Ravid (2019, Lemma 2), there are some measurable $\alpha_b, \alpha_g : \mathrm{supp}(p) \to \Delta A$ such that:[25]

[24] These are the partially informative signals about $\theta \in \Theta$ such that it is Bayes-consistent for the listener's posterior belief to equal the message.

[25] The cited lemma will exactly deliver $\alpha_b|_{\mathrm{supp}(b)}, \alpha_g|_{\mathrm{supp}(g)}$. 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 $\mathrm{supp}(p) \backslash [\mathrm{supp}(b) \cap \mathrm{supp}(g)]$.

- $\alpha_b(\mu), \alpha_g(\mu) \in \text{argmax}_{\tilde{\alpha}\in\Delta A} u_R(\tilde{\alpha}, \mu)\ \forall \mu \in \text{supp}(p)$;
- $u_S(\alpha_b(\mu)) = \phi(\mu)\ \forall \mu \in \text{supp}(b)$, and $u_S(\alpha_g(\mu)) = s_i\ \forall \mu \in \text{supp}(g)$.

From these, we can define a receiver strategy as

$$\begin{aligned} \alpha : M &\to \Delta A \\ m &\mapsto \begin{cases} \alpha_b(\mu) & : \ m = (1, \mu) \text{ for } \mu \in D \\ \alpha_g(\mu) & : \ m = (0, \mu) \text{ for } \mu \in D \\ \delta_{\hat{a}} & : \ m \notin \{0, 1\} \times D. \end{cases} \end{aligned}$$

We want to show that the tuple $(\xi, \sigma, \alpha, \pi)$ is a χ-equilibrium resulting in outcome (p, s_o, s_i). It is immediate from the construction of (σ, α, π) that sender incentive compatibility and receiver incentive compatibility hold, and that the expected sender payoff is s_i given influenced reporting.

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

$$\begin{aligned} [(\mathbf{1} - \chi)\sigma + \chi\xi](\{1\} \times \hat{D}|\cdot) &= k\tfrac{\mathrm{d}\beta}{\mathrm{d}\mu_0}\eta_b(\hat{D}|\cdot) \\ [(\mathbf{1} - \chi)\sigma + \chi\xi](\{0\} \times \hat{D}|\cdot) &= \left[(\mathbf{1} - \chi) + \chi\left(\mathbf{1} - \tfrac{k}{\chi}\tfrac{\mathrm{d}\beta}{\mathrm{d}\mu_0}\right)\right]\eta_g(\hat{D}|\cdot) \\ &= \left(\mathbf{1} - k\tfrac{\mathrm{d}\beta}{\mathrm{d}\mu_0}\right)\eta_g(\hat{D}|\cdot). \end{aligned}$$

Now, take any Borel $\hat{M} \subseteq M$ and $\hat{\Theta} \subseteq \Theta$, and let $D_z := \left\{\mu \in D :\ (z, \mu) \in \hat{M}\right\}$ for $z \in \{0, 1\}$. Observe that

$$\begin{aligned} &\int_\Theta \int_{\hat{M}} \pi(\hat{\Theta}|\cdot)\ \mathrm{d}[(1 - \chi(\theta))\sigma + \chi(\theta)\xi](\cdot|\theta)\ \mathrm{d}\mu_0(\theta) \\ = &\int_\Theta \int_{\hat{M}\cap[\{0,1\}\times D]} \pi(\hat{\Theta}|\cdot)\ \mathrm{d}[(1 - \chi(\theta))\sigma + \chi(\theta)\xi](\cdot|\theta)\ \mathrm{d}\mu_0(\theta) \\ = &\int_\Theta \left(\int_{\{1\}\times D_1} + \int_{\{0\}\times D_0}\right) \pi(\hat{\Theta}|\cdot)\ \mathrm{d}[(1 - \chi(\theta))\sigma + \chi(\theta)\xi](\cdot|\theta)\ \mathrm{d}\mu_0(\theta) \\ = &\int_\Theta \left[k\tfrac{\mathrm{d}\beta}{\mathrm{d}\mu_0}(\theta)\int_{D_1} \mu(\hat{\Theta})\ \mathrm{d}\eta_b(\mu) + \left(1 - k\tfrac{\mathrm{d}\beta}{\mathrm{d}\mu_0}(\theta)\right)\int_{D_0} \mu(\hat{\Theta})\ \mathrm{d}\eta_g(\mu)\right]\ \mathrm{d}\mu_0(\theta) \\ = &\ k\int_\Theta \int_{D_1} \mu(\hat{\Theta})\ \mathrm{d}\eta_b(\mu)\ \mathrm{d}\beta + \int_\Theta \int_{D_0} \mu(\hat{\Theta})\ \mathrm{d}\eta_g(\mu)\ \mathrm{d}[\mu_0 - k\beta] \end{aligned}$$

$$
\begin{aligned}
&= k\int_\Theta\int_{D_1}\mu(\hat\Theta)\ \mathrm{d}\eta_b(\mu)\ \mathrm{d}\beta(\theta)+(1-k)\int_\Theta\int_{D_0}\mu(\hat\Theta)\ \mathrm{d}\eta_g(\mu)\ \mathrm{d}\gamma(\theta)\\
&= k\int_{D_1}\int_\Theta\mu(\hat\Theta)\ \mathrm{d}\mu(\theta)\ \mathrm{d}b+(1-k)\int_{D_0}\int_\Theta\mu(\hat\Theta)\ \mathrm{d}\mu(\theta)\ \mathrm{d}b(\mu)\\
&= k\int_{D_1}\mu(\hat\Theta)\ \mathrm{d}b(\mu)+(1-k)\int_{D_0}\mu(\hat\Theta)\ \mathrm{d}g(\mu)\\
&= k\int_{\hat\Theta}\eta_b(D_1|\cdot)\ \mathrm{d}\beta+(1-k)\int_{\hat\Theta}\eta_g(D_0|\cdot)\ \mathrm{d}\gamma\\
&= \int_{\hat\Theta}\eta_b(D_1|\cdot)\ \mathrm{d}[k\beta]+\int_{\hat\Theta}\eta_g(D_0|\cdot)\ \mathrm{d}[\mu_0-k\beta]\\
&= \int_{\hat\Theta}k\tfrac{\mathrm{d}\beta}{\mathrm{d}\mu_0}\eta_b(D_1|\cdot)\ \mathrm{d}\mu_0+\int_{\hat\Theta}\left(\mathbf{1}-k\tfrac{\mathrm{d}\beta}{\mathrm{d}\mu_0}\right)\eta_g(D_0|\cdot)\ \mathrm{d}\mu_0\\
&= \int_{\hat\Theta}[(\mathbf{1}-\chi)\sigma+\chi\xi]\left(\hat M\cap[\{0,1\}\times D]\ \middle|\ \cdot\right)\ \mathrm{d}\mu_0\\
&= \int_{\hat\Theta}[(\mathbf{1}-\chi)\sigma+\chi\xi](\hat M|\cdot)\ \mathrm{d}\mu_0,
\end{aligned}
$$

verifying the Bayesian property. So (ξ,σ,α,π) is a χ-equilibrium. Moreover, for any Borel $\hat D\subseteq\Delta\Theta$, the equilibrium probability of the receiver posterior belief belonging to $\hat D$ is exactly (specializing the above algebra to $D_1=D_0=\hat D$ and $\hat\Theta=\Theta$)

$$\int_\Theta[(\mathbf{1}-\chi)\sigma+\chi\xi](\{0,1\}\times\hat D|\cdot)\ \mathrm{d}\mu_0=k\int_{\hat D}\mathbf{1}\ \mathrm{d}b+(1-k)\int_{\hat D}\mathbf{1}\ \mathrm{d}g=p(\hat D).$$

Finally, the expected sender payoff conditional on reporting not being influenced—note the conditional distribution $\frac{\chi}{\chi(\mu_0)}\mu_0\in\Delta\Theta$—is given by:

$$
\begin{aligned}
&\int_\Theta\int_M u_S\left(\alpha(m)\right)\ \mathrm{d}\xi(m|\cdot)\ \mathrm{d}\left[\tfrac{\chi}{\chi(\mu_0)}\mu_0\right]\\
&= \int_\Theta\left[\left(1-\tfrac{k}{\chi}\tfrac{\mathrm{d}\beta}{\mathrm{d}\mu_0}\right)\int_{\Delta\Theta}u_S\left(\alpha(0,\mu)\right)\ \mathrm{d}\eta_g(\mu|\cdot)+\tfrac{k}{\chi}\tfrac{\mathrm{d}\beta}{\mathrm{d}\mu_0}\int_{\Delta\Theta}u_S\left(\alpha(1,\mu)\right)\ \mathrm{d}\eta_b(\mu|\cdot)\right]\ \mathrm{d}\left[\tfrac{\chi}{\chi(\mu_0)}\mu_0\right]\\
&= \int_\Theta\left[\left(\mathbf{1}-\tfrac{k}{\chi}\tfrac{\mathrm{d}\beta}{\mathrm{d}\mu_0}\right)\int_{\Delta\Theta}s_i\ \mathrm{d}\eta_g(\mu|\cdot)+\tfrac{k}{\chi}\tfrac{\mathrm{d}\beta}{\mathrm{d}\mu_0}\int_{\mathrm{supp}(b)}\phi(\mu)\ \mathrm{d}\eta_b(\mu|\cdot)\right]\ \mathrm{d}\left[\tfrac{\chi}{\chi(\mu_0)}\mu_0\right]\\
&= s_i+\tfrac{k}{\chi(\mu_0)}\int_\Theta\left[-s_i+\int_{\mathrm{supp}(b)}\phi(\mu)\ \mathrm{d}\eta_b(\mu|\theta)\right]\ \mathrm{d}\beta(\theta)\\
&= \left[1-\tfrac{k}{\chi(\mu_0)}\right]s_i+\tfrac{k}{\chi(\mu_0)}\int_{\Delta\Theta}\int_\Theta\phi(\mu)\ \mathrm{d}\mu(\theta)\ \mathrm{d}b(\mu)\\
&= \tfrac{(1-k)-[1-\chi(\mu_0)]}{\chi(\mu_0)}s_i+\tfrac{k}{\chi(\mu_0)}\int_{\mathrm{supp}(b)}\phi\ \mathrm{d}b\\
&= s_o,
\end{aligned}
$$

as required.

Conversely, suppose $(\xi, \sigma, \alpha, \pi)$ is a χ-equilibrium resulting in outcome (p, s_o, s_i). Let

$$\tilde{G} := \int_\Theta \sigma \, \mathrm{d}\mu_0 \text{ and } P := \int_\Theta [\chi\xi + (\mathbf{1} - \chi)\sigma] \, \mathrm{d}\mu_0 \in \Delta M$$

denote the probability measures over messages induced by non-committed behavior and by average sender behavior, respectively.

Let $M^* := \{m \in M : u_S(\alpha(m)) = s_i\}$ and $k := 1 - P(M^*)$. Sender incentive compatibility (which implies that $\sigma(M^*|\cdot) = \mathbf{1}$) tells us that $k \in [0, \chi(\mu_0)]$. Let $G := \frac{1}{1-k} P(\cdot \cap M^*)$ if $k < 1$; and let $G := \tilde{G}$ otherwise. Let $B := \frac{1}{k}[P - (1-k)G]$ if $k > 0$; and let $B := \int_\Theta \xi \, \mathrm{d}\mu_0$ otherwise. Both G and B are in ΔM because $(1-k)G \leq P$. Let $g := G \circ \pi^{-1}$ and $b := B \circ \pi^{-1}$, both in $\Delta\Delta\Theta$. By construction, $kb + (1-k)g = P \circ \pi^{-1} = p \in \mathcal{R}(\mu_0)$. Moreover,

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

where the last inequality follows from the Bayesian property of π, together with the fact that σ almost surely sends a message from M^* on the path of play.

Next, for any $m \in M$ sender incentive compatibility tells us that $u_S(\alpha(m)) \leq s_i$, and receiver incentive compatibility tells us that $\alpha(m) \in V(\pi(m))$. If follows directly that $g\{V \ni s_i\} = b\{\min V \leq s_i\} = 1$.

Now viewing π, α as random variables on the probability space $\langle M, P \rangle$, define the conditional expectation $\phi_0 := \mathbb{E}_B[u_S(\alpha)|\pi] : M \to \mathbb{R}$. By Doob-Dynkin, there is 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 ϕ on a b-null set, we may assume without loss that ϕ is a measurable selector of $s_i \wedge V$.

Observe now that $\tilde{G}(M^*) = G(M^*) = 1$ and

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

Therefore,

$$s_o = \int_M u_S \circ \pi \ \mathrm{d}\frac{P-[1-\chi(\mu_0)]\tilde{G}}{\chi} = \int_M u_S \circ \pi \ \mathrm{d}\frac{P-[1-\chi(\mu_0)]G}{\chi(\mu_0)} = \int_M u_S \circ \pi \ \mathrm{d}\frac{kB+(1-k)G-[1-\chi(\mu_0)]G}{\chi(\mu_0)}$$
$$= \int_M u_S \circ \pi \ \mathrm{d}\left[\left(1-\frac{k}{\chi(\mu_0)}\right)G + \frac{k}{\chi(\mu_0)}B\right] = \left(1-\frac{k}{\chi(\mu_0)}\right)s_i + \frac{k}{\chi(\mu_0)}\int_{\mathrm{supp}(b)} \phi \ \mathrm{d}b,$$

as required. ∎

B.1.2 Proof of Theorem 1

Proof. By Lemma 1, the supremum sender value over all χ-equilibrium outcomes is

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

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

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

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 (2019, Lemma 1 and Theorem 2) tell us that any $\gamma \in \Delta\Theta$ has

$$\max_{g \in \mathcal{R}(\gamma), s_i \in \mathbb{R}:\ g\{V \ni s_i\}=1} s_i = \bar{v}(\gamma),$$

where $\bar{v}$ is the quasiconcave envelope of v.[26] Therefore,

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

<u>Claim:</u> If $b \in \Delta\Delta\Theta$, $\gamma \in \Delta\Theta$, and $k \in [0,1]$ satisfy $k \int_{\Delta\Theta} \mu \ \mathrm{d}b(\mu) + (1-k)\gamma = \mu_0$ and $(1-k)\gamma \geq (\mathbf{1}-\chi)\mu_0$, then there exists (b^*, γ^*, k^*) feasible in the above program[27] such that $(1-k^*)\bar{v}(\gamma^*) + k^* \int_{\Delta\Theta} \bar{v}(\gamma^*) \wedge v \ \mathrm{d}b^* \geq (1-k)\bar{v}(\gamma) + k \int_{\Delta\Theta} \bar{v}(\gamma) \wedge v \ \mathrm{d}b$.

To prove the claim, let $\beta := \int_{\Delta\Theta} \mu \ \mathrm{d}b(\mu)$, and consider three exhaustive cases.

<u>Case 1:</u> $\bar{v}(\gamma) \leq v(\mu_0)$.

In this case, $(b^*, \gamma^*, k^*) := (\delta_{\mu_0}, \mu_0, 0)$ will work.

<u>Case 2:</u> $v(\mu_0) < \bar{v}(\gamma) \leq v(\beta)$.

In this case, Lipnowski and Ravid (2019, Lemma 3) delivers some $\beta^* \in \mathrm{co}\{\beta, \mu_0\}$ such that $V(\beta^*) \ni \bar{v}(\gamma)$. But then $\mu_0 \in \mathrm{co}\{\beta^*, \gamma\}$. As $\bar{v}$ is quasiconcave, $\bar{v}(\mu_0) \geq \min\{\bar{v}(\beta^*), \bar{v}(\gamma)\} \geq \min\{v(\beta^*), \bar{v}(\gamma)\} = \bar{v}(\gamma)$.

Therefore, $(b^*, \gamma^*, k^*) := (\delta_{\mu_0}, \mu_0, 0)$ will again work.

<u>Case 3:</u> $v(\beta) < \bar{v}(\gamma)$.

In this case, our aim is to show that there exists a $b^* \in \Delta\Delta\Theta$ such that:

- $b^* \in \mathcal{R}(\beta)$ and $b\{\min V \leq \bar{v}(\gamma)\} = 1$;
- $\int_{\Delta\Theta} \bar{v}(\gamma) \wedge v \ \mathrm{d}b^* \geq \int_{\Delta\Theta} \bar{v}(\gamma) \wedge v \ \mathrm{d}b$.

Given such a measure, (b^*, γ, k) will be as required. We explicitly construct such a b^*.

[26] 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.

[27] That is, (b^*, γ^*, k^*) satisfy the same constraints, and further have $b^*\{\min V \leq \bar{v}(\gamma)\} = 1$.

Let $D := \operatorname{supp}(b)$, and define the measurable function,

$$\begin{aligned} \lambda : D &\to [0,1] \\ \mu &\mapsto \begin{cases} 1 & : \ v(\mu) \le \bar{v}(\gamma) \\ \inf\left\{ \hat{\lambda} \in [0,1] : \ v\left((1-\hat{\lambda})\gamma + \hat{\lambda}\mu \right) \ge \bar{v}(\gamma) \right\} & : \ \text{otherwise.} \end{cases} \end{aligned}$$

Lipnowski and Ravid (2019, Lemma 3) tells us that $\bar{v}(\gamma) \in V([1-\lambda(\mu)]\gamma + \lambda(\mu)\mu)$ for every $\mu \in D$ for which $v(\mu) > \bar{v}(\gamma)$. This implies that $\min V([1-\lambda(\mu)]\gamma+\lambda(\mu)\mu) \le \bar{v}(\gamma)$ for *every* $\mu \in D$.

There must some number $\epsilon > 0$ such that $\lambda \ge \epsilon$ uniformly, because v is upper semicontinuous and $\bar{v}(\gamma) > v(\beta)$; and so $\frac{1}{\lambda} : D \to [1, \infty)$ is bounded. Moreover, by construction, $\lambda(\mu) < 1$ only for $\mu \in D$ with $v(\mu) > v(\gamma)$.

Now, define $b^* \in \Delta\Delta\Theta$ via

$$b^*(\hat{D}) := \left(\int_{\Delta\Theta} \tfrac{1}{\lambda} \ \mathrm{d}b \right)^{-1} \cdot \int_{\Delta\Theta} \frac{1}{\lambda(\mu)} \mathbf{1}_{[1-\lambda(\mu)]\mu_0 + \lambda(\mu)\mu \in \hat{D}} \ \mathrm{d}b(\mu), \ \forall \text{ Borel } \hat{D} \subseteq \Delta\Theta.$$

Direct computation shows that $\int_{\Delta\Theta} \mu \ \mathrm{d}b^*(\mu) = \int_{\Delta\Theta} \mu \ \mathrm{d}b(\mu)$, i.e. $b^* \in \mathcal{R}(\beta)$. Moreover, by construction, $\min V([1-\lambda(\mu)]\gamma + \lambda(\mu)\mu) \le \bar{v}(\gamma) \ \forall \mu \in D$. All that remains, then, is the value comparison.

$$\begin{aligned}
&\left(\int_{\Delta\Theta} \tfrac{1}{\lambda} \ \mathrm{d}b \right) \int_{\Delta\Theta} \bar{v}(\gamma) \wedge v \ \mathrm{d}[b^* - b] \\
&= \int_{\Delta\Theta} \left[\tfrac{1}{\lambda(\mu)} \bar{v}(\gamma) \wedge v\left([1-\lambda(\mu)]\mu_0 + \lambda(\mu)\mu \right) - \left(\int_{\Delta\Theta} \tfrac{1}{\lambda} \ \mathrm{d}b \right) \bar{v}(\gamma) \wedge v(\mu) \right] \ \mathrm{d}b(\mu) \\
&= \int_{\Delta\Theta} \left(\tfrac{1}{\lambda(\mu)} - \int_{\Delta\Theta} \tfrac{1}{\lambda} \ \mathrm{d}b \right) \left[v(\mu)\mathbf{1}_{v(\mu) \le \bar{v}(\gamma)} + \bar{v}(\gamma)\mathbf{1}_{v(\mu) > \bar{v}(\gamma)} \right] \ \mathrm{d}b(\mu) \\
&= \int_{\Delta\Theta} \left(\tfrac{1}{\lambda(\mu)} - \int_{\Delta\Theta} \tfrac{1}{\lambda} \ \mathrm{d}b \right) \left\{ \bar{v}(\gamma) - [\bar{v}(\gamma) - v(\mu)]\mathbf{1}_{v(\mu) \le \bar{v}(\gamma)} \right\} \ \mathrm{d}b(\mu) \\
&= 0 + \int_{\Delta\Theta} \left(\int_{\Delta\Theta} \tfrac{1}{\lambda} \ \mathrm{d}b - \tfrac{1}{\lambda(\mu)} \right) [\bar{v}(\gamma) - v(\mu)]\mathbf{1}_{v(\mu) \le \bar{v}(\gamma)} \ \mathrm{d}b(\mu) \\
&= \int_{\{\mu \in \Delta\Theta : \ v(\mu) \le \bar{v}(\gamma)\}} \left(\int_{\Delta\Theta} \tfrac{1}{\lambda} \ \mathrm{d}b - 1 \right) [\bar{v}(\gamma) - v(\mu)] \ \mathrm{d}b(\mu) \\
&= \left(\int_{\Delta\Theta} \tfrac{1-\lambda}{\lambda} \ \mathrm{d}b \right) \int_{\{\mu \in \Delta\Theta : \ v(\mu) \le \bar{v}(\gamma)\}} [\bar{v}(\gamma) - v] \ \mathrm{d}b \\
&\ge 0,
\end{aligned}$$

proving the claim.

In light of the claim, the optimal value is

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

Finally, observe that the supremum is in fact a maximum because the constraint set is a compact subset of $(\Delta\Theta)^2 \times [0,1]$ and the objective upper semicontinuous. □

B.1.3 Consequences of Lemma 1 and Theorem 1

Corollary 1. *As x ranges over $[0,1]$, the set of $x\mathbf{1}$-equilibrium outcomes (p, s_o, s_i) at prior μ_0 is a compact-valued, upper hemicontinuous correspondence of (μ_0, x).*

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 $(\mu_0, p, g, b, x, k, s_o, s_i) \in (\Delta\Theta)\times(\Delta\Delta\Theta)^3\times[0,1]^2\times[\text{co } u_S(A)]^2$ such that:

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

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*

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

is weakly increasing and right-continuous.

Proof. That it is weakly increasing is immediate from Theorem 1, given that increasing credibility expands the constraint set. That it is upper semicontinuous (and so, since nondecreasing, it is right-continuous) follows directly from Corollary 1. $\square$

Corollary 3. *For any $x \in [0,1]$, the map $v^*_{x\mathbf{1}} : \Delta\Theta \to \mathbb{R}$ is upper semicontinuous.*

Proof. This is immediate from Corollary 1. $\square$

B.2 Productive Mistrust: Proofs

Toward verifying our sufficient conditions for productive mistrust to occur, we first study in some depth the possibility of productive mistrust in the binary-state world. We then leverage that analysis to study the same in many-state environments.

To this end, it 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** μ if every $p \in \mathcal{R}(\mu)$ is outperformed by an SOB policy $p' \in \mathcal{R}(\mu)$.

B.2.1 Productive Mistrust with Binary States

Given binary states, finitely many actions, and a full-support prior μ_0, we know that the quasiconcave envelope function $\bar{v} : \Delta\Theta \to \mathbb{R}$ is upper semicontinuous, weakly quasiconcave, and piecewise constant. Therefore, if $\mu_0 \notin \text{argmax}_{\mu\in\Delta\Theta}\bar{v}(\mu)$, there is then a unique $\mu_+ = \mu_+(\mu_0)$ closest to μ_0 with the property that $\bar{v}(\mu_+) > \bar{v}(\mu_0)$, and a unique $\theta = \theta(\mu_0) \in \Theta$ with $\mu_0 \in \text{co}\{\mu_+(\mu_0), \delta_\theta\}$. In this case, for the rest of the subsection, we identify $\Delta\Theta \approxeq [0,1]$ by identifying $\nu \in \Delta\Theta$ with $1 - \nu(\theta(\mu_0))$.[28]

Lemma 2. *Given finite A, binary Θ, and a full-support prior $\mu_0 \in \Delta\Theta$, the following are equivalent:*

[28] So, under this normalization, $0 = \theta < \mu_0 < \mu_+$.

1. *There exist credibility levels $\chi' < \chi$ such that, for every S-optimal χ-equilibrium outcome (p, s) and S-optimal χ'-equilibrium outcome (p', s'), the policy p' is strictly more Blackwell-informative than p.*

2. $\mu_0 \notin \text{argmax}_{\mu \in \Delta\Theta:\ \mu \textit{ full-support}} \bar{v}(\mu)$, *and there exists* $\mu_- \in [0, \mu_0]$ *such that* $v(\mu_-) > v(0) + \frac{\mu_-}{\mu_+}\left[v(\mu_+) - v(0)\right]$.

Moreover, in this case, every S-optimal χ'-equilibrium outcome gives the receiver a strictly higher payoff than any S-optimal χ-equilibrium.

Proof. First, suppose (2) fails. There are three ways it could fail:

(a) With $\mu_0 \in \text{argmax}_{\mu \in \Delta\Theta} \bar{v}(\mu)$;

(b) With $\mu_0 \in \text{argmax}_{\mu \in \Delta\Theta:\ \mu \text{ full-support}} \bar{v}(\mu) \setminus \text{argmax}_{\mu \in \Delta\Theta} \bar{v}(\mu)$;

(c) With $\mu_0 \notin \text{argmax}_{\mu \in \Delta\Theta:\ \mu \text{ full-support}} \bar{v}(\mu)$;

In case (a) or (b), pick some S-optimal 0-equilibrium information policy p_0. For any $\hat{x} \in [0, 1)$, we know $(p_0, \bar{v}(\mu_0))$ is a S-optimal 0-equilibrium outcome; and in case (a) it is also a S-optimal $\mathbf{1}$-equilibrium outcome.

For case (a), there is nothing left to show.

For case (b), we need only consider the case of $\chi = \mathbf{1}$. In case (b), that $\bar{v}$ is weakly quasiconcave implies it is monotonic. So $\mu_+ = 1$, and $\bar{v} : [0, 1] \to \mathbb{R}$ is non-decreasing with $\bar{v}|_{[\mu_0, 1)} = \bar{v}(\mu_0) < \bar{v}(1)$. As $v^*_{\mathbf{1}}$ is the concave envelope of $\bar{v}$, it must be that the support of any S-optimal $\mathbf{1}$-equilibrium information policy is contained in $[0,\ \min\{\mu \in [0, 1] :\ v(\mu) = v(\mu_0))\}] \cup \{1\}$, so that (1) fails as well.

In case (c), failure of (2) tells us $v(\mu) \le v(0) + \frac{\mu}{\mu_+}\left[v(\mu_+) - v(0)\right],\ \forall \mu \in [0, \mu_0]$. As $\bar{v}|_{[0, \mu_+)} \le \bar{v}(\mu_0)$, it follows that

$$
\begin{aligned}
v^*_{\hat{x}\mathbf{1}}(\mu_0) = & \max_{\beta, \gamma, k \in [0,1]} \left\{ k \hat{v}_{\wedge\gamma}(\beta) + (1-k)\bar{v}(\gamma) \right\} \\
& \text{s.t.} \qquad k\beta + (1-k)\gamma = \mu_0,\ (1-k)(\gamma, 1-\gamma) \ge (1-\hat{x})(\mu_0, 1-\mu_0) \\
= & \max_{\gamma \in [\mu_0, 1], k \in [0,1]} \left\{ k v(0) + (1-k)\bar{v}(\gamma) \right\} \\
& \text{s.t.} \qquad k0 + (1-k)\gamma = \mu_0,\ (1-k)(\gamma, 1-\gamma) \ge (1-\hat{x})(\mu_0, 1-\mu_0) \\
= & \max_{\gamma \in [\mu_0, 1]} \left\{ \left(1 - \frac{\mu_0}{\gamma}\right) v(0) + \frac{\mu_0}{\gamma}\bar{v}(\gamma) \right\} \\
& \text{s.t.} \qquad \frac{\mu_0}{\gamma}(1-\gamma) \ge (1-\hat{x})(1-\mu_0).
\end{aligned}
$$

In particular, defining $\gamma(\hat{x})$ to be the largest argmax in the above optimization problem, it follows that

$$p_{\hat{x}} = \left(1 - \tfrac{\mu_0}{\gamma(\hat{x})}\right)\delta_0 + \tfrac{\mu_0}{\gamma}\delta_{\gamma(\hat{x})}$$

is a S-optimal $\hat{x}\mathbf{1}$-equilibrium information policy for any $\hat{x} \in [0,1]$, so that (1) does not hold.

Conversely, suppose (2) holds.

The function $v : [0,1] \to \mathbb{R}$ is upper semicontinuous and piecewise constant, which implies that its concave envelope $v^*_{\mathbf{1}}$ is piecewise affine. We may then define

$$\mu^*_- := \min\{\mu \in [0,\mu_0] : \; v^*_{\mathbf{1}} \text{ is affine over } [\mu,\mu_0]\}.$$

That (2) holds tells us that $\mu^*_- \in (0,\mu_0)$. It is then without loss to take $\mu_- = \mu^*_-$.

There are thus beliefs $\mu_-, \mu_+ \in [0,1]$ such that: $0 < \mu_- < \mu_0 < \mu_+$; $v^*_{\mathbf{1}}$ is affine on $[\mu_-,\mu_+]$ and on no larger interval; and $v^*_{\mathbf{1}}$ is strictly increasing on $[0,\mu_+]$. It follows that $\hat{v}_{\wedge\mu_+} = v^*_{\mathbf{1}}$ on $[0,\mu_+]$. By definition of $\mu_+ = \mu_+(\mu_0)$, we know that $\bar{v}$ is constant on $[\mu_0,\mu_+)$. That is, (appealing to Lipnowski and Ravid (2019, Theorem 2)) $v^*_{\mathbf{0}}$ is constant on $[\mu_0,\mu_+)$. Then, since $v^*_{\mathbf{1}}$ strictly decreases there, it must be that $v^*_{\mathbf{1}} > v^*_{\mathbf{0}}$ on (μ_0,μ_+).

Let $x \in [0,1]$ be the smallest credibility level such that $v^*_{x\mathbf{1}}(\mu_0) = v^*_{\mathbf{1}}(\mu_0)$, which exists by Corollary 2. That $v^*_{\mathbf{0}}(\mu_0) < v^*_{\mathbf{1}}(\mu_0)$ implies $\chi > 0$. That μ_+ has full support, which follows from (2), implies that $x < 1$.[29]

Consider now the following claim.

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

$$\begin{aligned}(\beta',\gamma',k') \in \operatorname{argmax}_{(\beta,\gamma,k)\in[0,1]^3} & \left\{k\hat{v}_{\wedge\gamma}(\beta) + (1-k)\bar{v}(\gamma)\right\} \\ \textit{s.t.} \qquad & k\beta + (1-k)\gamma = \mu_0, \; (1-k)(\gamma,1-\gamma) \geq (1-x')(\mu_0,1-\mu_0),\end{aligned}$$

for a value strictly higher than $\bar{v}(\mu_0)$. Then:

- *$\gamma' = \mu_+$ and $\beta' \leq \mu_-$.*

- *If $h' \in \mathcal{R}(\beta')$ and $\ell' \in \mathcal{R}(\gamma')$ are such that $p' = k'h' + (1-k')\ell'$ is the information*

[29] In particular, this follows from the hypothesis that there exists some *full-support* belief at which $\bar{v}$ takes a strictly higher value than $v(\mu_0)$. This implies $x < 1$ by the same argument employed to prove Proposition 3.

policy of a S-optimal $x'\mathbf{1}$-equilibrium, then $h'[0,\mu_-] = \ell'\{\mu_+\} = 1$.

We now prove the claim.

If $\gamma' > \mu_+$, then let $k'' \in (0, k')$ be the unique solution to $k''\beta' + (1-k'')\mu_+ = \mu_0$. As $(1-k'')(\mu_+, 1-\mu_+) \geq (1-x')(\mu_0, 1-\mu_0)$ and

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

the feasible solution (β', μ_+, k'') would strictly outperform (β', γ', k'). So optimality implies $\gamma' \leq \mu_+$.

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 (β', γ', k') strictly outperforms $\bar{v}(\mu_0)$, it follows that $\gamma' = \mu_+$. One direct implication is that

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

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', β') with (k, μ_-) which satisfies $k\mu_- + (1-k)\mu_+ = \mu_0$ necessarily has $(1-k)(\mu_+, 1-\mu_+) \gg (1-x')(\mu_0, 1-\mu_0)$. This would contradict minimality of x. Therefore, $\beta' \leq \mu_-$.

We now prove the second bullet. First, every $\mu < \mu_+$ satisfies $v(\mu) \leq v^*_{\mathbf{1}}(\mu) < v^*_{\mathbf{1}}(\mu_+) = v(\mu_+)$. This implies that δ_{μ_+} is the unique $\ell \in \mathcal{R}(\mu_+)$ with $\inf v(\operatorname{supp}\ell) \geq v(\mu_+)$. Therefore, $\ell' = \delta_{\mu_+}$.

Second, the measure $h' \in \mathcal{R}(\beta')$ can be expressed as $h' = (1-\gamma)h_L + \gamma h_R$ for $h_L \in \Delta[0, \mu_-]$, $h_R \in \Delta(\mu_-, 1]$, and $\gamma \in [0, 1)$. Notice that $(\mu_-, v(\mu_-))$ is an extreme point of the subgraph of $v^*_{\mathbf{1}}$, and therefore an extreme point of the subgraph of $\hat{v}_{\wedge\mu_+}$. Taking the unique $\hat{\gamma} \in [0, \gamma]$ such that $\hat{h} := (1-\hat{\gamma})h_L + \hat{\gamma}\delta_{\mu_-} \in \mathcal{R}(\beta')$, it follows that $\int_{[0,1]} \hat{v}_{\wedge\mu_+}\, \mathrm{d}\hat{h} \geq \int_{[0,1]} \hat{v}_{\wedge\mu_+}\, \mathrm{d}h'$, strictly so if $\hat{\gamma} < \gamma$. But $\hat{\gamma} < \gamma$ necessarily if $\gamma > 0$, since $\int_{[0,1]} \mu\, \mathrm{d}h_R(\mu) > \mu_-$. Optimality of h' then implies that $\gamma = 0$, i.e. $h'[0, \mu_-] = 1$. This completes the proof of the claim.

With the claim in hand, we can now prove the proposition. Letting $k^* \in (0, 1)$ be the solution to $k^*\mu_- + (1-k^*)\mu_+ = \mu_0$, the claim implies that (μ_-, μ_+, k^*) is the unique solution to

$$\max_{(\beta,\gamma,k)\in[0,1]^3} \left\{ k\hat{v}_{\wedge\gamma}(\beta) + (1-k)\bar{v}(\gamma) \right\}$$
$$\text{s.t.} \quad k\beta + (1-k)\gamma = \mu_0, \ (1-k)(\gamma, 1-\gamma) \geq (1-x)(\mu_0, 1-\mu_0),$$

and that $p^* = k^*\delta_{\mu_-} + (1-k^*)\delta_{\mu_+}$ is the uniquely S-optimal $x\mathbf{1}$-equilibrium information policy. Moreover, the minimality property defining x implies that $(1-k^*)(1-\mu_+) = (1-x)(1-\mu_0)$.

Given $x' < x$ sufficiently close to x, one can verify directly that (β', μ_+, k') is feasible, where

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

As $\hat{v}_{\wedge\mu_+}$ is a continuous function, it follows that $v^*_{x'\mathbf{1}}(\mu_0) \nearrow v^*_{x\mathbf{1}}(\mu_0)$ as $x' \nearrow x$. In particular, $v^*_{x'\mathbf{1}}(\mu_0) > v^*_0(\mu_0)$ for $x' < x$ sufficiently close to x. Fix such a x'.

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

Having shown that (2) implies (1), 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 μ_- and μ_+, 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, at 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. $\square$

The following Lemma is the specialization of Proposition 1 to the binary-state world. In addition to being a special case of the proposition, it will also be an important lemma for proving the more general result.

Lemma 3. *Suppose $|\Theta| = 2$, the model is finite and generic, a full-support belief $\mu \in \Delta\Theta$ exists such that the sender is not an SOB at μ. Then there exists a full-support prior μ_0 and credibility levels $\chi' < \chi$ such that every S-optimal χ'-equilibrium is both strictly better for R and more Blackwell-informative than every S-optimal χ-equilibrium.*

Proof. First, notice that the genericity assumption delivers full-support μ', 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 piecewise constant, which implies that its concave envelope $v^*_{\mathbf{1}}$ is piecewise affine. That is, there exist $n \in \mathbb{N}$ and $\{\mu^i\}_{i=0}^n$ such that $0 = \mu^0 \leq \cdots \leq \mu^n = 1$ and $v^*_{\mathbf{1}}|_{[\mu^{i-1},\mu^i]}$ 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 $v^*_{\mathbf{1}}|_{[\mu^{i-1},\mu^i]}$ is strictly decreasing in i. Therefore, there exist $i_0, i_1 \in \{0, \ldots, n\}$ such that $i_1 \in \{i_0, i_0 + 1\}$ and $\text{argmax}_{\mu\in[0,1]} v(\mu) = [\mu^{i_0}, \mu^{i_1}]$. That the sender is not an SOB at μ implies that $i_0 > 1$ or $i_1 < n - 1$. Without loss of generality, say $i_0 > 1$. Now let $\mu_- := \mu^{i_0-1}$ and $\mu_+ := \mu^{i_0}$.

Finally, that $V(\mu') = \{\max v(\Delta\Theta)\}$, and V is (by Berge's theorem) upper hemicontinuous implies $\text{argmax}_{\mu\in\Delta\Theta:\ \mu \text{ full-support}} \bar{v}(\mu) = \text{argmax}_{\mu\in\Delta\Theta} \bar{v}(\mu)$. Therefore, considering any prior of the form $\mu_0 = \mu_+ - \epsilon$ for sufficiently small $\epsilon > 0$, Lemma 2 applies. □

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

Given Lemma 3, 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 μ_0^∞ and constant credibility levels $\chi' < \chi$ such that R strictly prefers every S-optimal χ'-equilibrium to every S-optimal χ-equilibrium. We then find an interior direction through which to approach μ_0^∞, 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 μ_0^∞. 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^χ to $p_n^{\chi'}$. As

R's payoffs are continuous, R being weakly better off under χ than under χ' along the sequences would imply the same at the sequences' limits. Notice, though, that such limits must be S-optimal for the prior μ_0^∞ by the choice of direction, meaning that productive mistrust fails at μ_0^∞; that is, we have a contradiction. Below, we proceed with the formal proof.

Proof. Let $\Theta_2 := \{\theta_1, \theta_2\}$ and $u := \max v\,(\Delta\Theta_2)$, and define the receiver value function $v_R : \Delta\Delta\Theta \to \mathbb{R}$ via $v_R(p) := \int_{\Delta\Theta} \max_{a\in A} u_R(a,\mu)\, \mathrm{d}p(\mu)$.

Appealing to Lemma 3, there is some $\mu_0^\infty \in \Delta\Theta$ with support Θ_2 and credibility levels $\chi'' < \chi'$ such that every S-optimal χ''-equilibrium is strictly better for R than every S-optimal χ'-equilibrium.

Consider the following claim.

Claim: There exists a sequence $\{\mu_0^n\}$ of full-support priors converging to μ_0^∞ such that

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

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 μ_0^n admits some S-optimal χ'-equilibrium and χ''-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 $\Delta\Delta\Theta \times \mathbb{R} \times \mathbb{R}$) to some $\Psi' = (p', s'_i, s'_o)$ and $\Psi'' = (p'', s''_i, s''_o)$ respectively. By Corollary 1, for every credibility level χ, the set of χ-equilibria is an upper hemicontinuous correspondence of the prior. Therefore, Ψ' and Ψ'' are χ'- and χ''-equilibria, respectively, at prior μ_0^∞. 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 Ψ' and Ψ'' are S-optimal χ'- and χ''-equilibria, respectively, contradicting the definition of μ_0^∞. Therefore, there is some $n \in \mathbb{N}$ for which the full-support prior μ_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, the proof of Lemma 3 delivers some $\gamma^\infty \in \Delta\Theta$ such that $\bar{v}(\gamma^\infty) = u$ and, for both $\chi \in \{\chi', \chi''\}$, some $(\beta, \gamma, k) \in \Delta\Theta \times \{\gamma^\infty\} \times [0,1]$ solves the program in Theorem 1 at prior μ_0^∞.

Let us now show that there exists a closed convex set $D \subseteq \Delta\Theta$ which contains γ^∞, has nonempty interior, and satisfies $\bar{v}|_D = u$. Indeed, 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 μ', and let $D_n := \text{co}\,[\{\gamma^\infty\} \cup B_n]$. As $v|_{\Delta\Theta_2} \leq u$ and constant functions are quasiconcave, Lipnowski and Ravid (2019, Theorem 2) tells us $\bar{v}|_{\Delta\Theta_2} \leq u$ as well. As V is upper hemicontinuous, the hypothesis on μ' ensures that $\bar{v}|_{B_n} \geq v|_{B_n} = u$ for sufficiently large $n \in \mathbb{N}$; quasiconcavity then tells us $\bar{v}|_{D_n} \geq u$. Assume now, for a contradiction, that every $n \in \mathbb{N}$ has $\bar{v}|_{D_n} \not\leq u$. That is, there is some $\lambda_n \in [0,1]$ and $\mu'_n \in B_n$ such that $\bar{v}\left((1-\lambda_n)\mu + \lambda_n \mu'_n\right) > u$. Dropping to a subsequence, we get a strictly increasing sequence $(n_\ell)_\ell^\infty$ of natural numbers such that (since $[0,1]$ is compact and $\bar{v}(\Delta\Theta)$ is finite) $\lambda_{n_\ell} \xrightarrow{\ell\to\infty} \lambda \in [0,1]$ and $\bar{v}\left((1-\lambda_{n_\ell})\mu + \lambda_{n_k}\mu'_{n_k}\right) = \hat{u}$ for some number $\hat{u} \in (u, \infty)$ and every $\ell \in \mathbb{N}$. As $\bar{v}$ is upper semicontinuous, this would imply that $\bar{v}\left((1-\lambda)\mu + \lambda\mu'\right) \geq \hat{u} > u$, contradicting the definition of u. 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.

Now, for each $n \in \mathbb{N}$, define $\mu_0^n := \frac{n-1}{n}\mu_0^\infty + \frac{1}{n}\gamma_1$. We will show that the sequence $(\mu_0^n)_{n=1}^\infty$—a sequence of full-support priors converging to μ_0^∞—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 μ_0^∞. 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,

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

Therefore, (β, γ_n, k_n) is χ-feasible at prior μ_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. □

B.3 Collapse of Trust: Proof of Proposition 2

Proof. Two of three implications are easy given Corollary 2. First, if there is no conflict, then Lipnowski and Ravid (2019, Lemma 1) tells us that there is a 0-equilibrium with full information that generates sender value $\max v(\Delta\Theta) \geq v_{\mathbf{1}}^*$; in particular, $v_{\mathbf{0}}^* = v_{\mathbf{1}}^*$. Second, if $v_{\mathbf{0}}^* = v_{\mathbf{1}}^*$, then v_χ^* is constant in χ, ruling out a collapse of trust. Below we show that any conflict whatsoever implies a collapse of trust.

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

$$\Gamma(\hat{x}) := \left\{(\beta, \gamma, k) \in \Delta\Theta \times (\Delta\Theta \setminus D) \times [0, 1] : \; k\beta + (1-k)\gamma = \mu_0, \; (1-k)\gamma \geq (1-\hat{x})\mu_0\right\},$$

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

Given any scalar $x' \in [0, x)$, it must be that $K(x') = \emptyset$. That is, if $\beta, \gamma \in \Delta\Theta$ and $k \in [0, 1]$ with $k\beta + (1-k)\gamma = \mu_0$ and $(1-k)\gamma \geq (1-\hat{x})\mu_0$, then $\gamma \in D$. By Theorem 1, then, $v_{x'\mathbf{1}}^*(\mu_0) = v(\mu_0) = 0$.

There is, however, some $k \in K(x)$. By Theorem 1 and the definition of Γ, there is therefore a $x\mathbf{1}$-equilibrium generating ex-ante sender payoff of at least $k\cdot 0 + (1-k)\cdot 1 = (1-k) \geq (1-x)$. If $x < 1$, a collapse of trust occurs at credibility level x.

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

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

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

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

B.4 Robustness: Proof of Proposition 3

Proof. By Lipnowski and Ravid (2019, 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 μ_0, suppose $\gamma \in \Delta\Theta$ is full-support with $\bar{v}(\gamma) = \max v(\Delta\Theta)$. It follows immediately that $\hat{v}_{\wedge\gamma} = \hat{v} = v^*_{\mathbf{1}}$.

Let $r_0 := \min_{\theta\in\Theta} \frac{\mu_0\{\theta\}}{\gamma\{\theta\}} \in (0,\infty)$ and $r_1 := \max_{\theta\in\Theta} \frac{\mu_0\{\theta\}}{\gamma\{\theta\}} \in [r_0,\infty)$. Then Theorem 1 tells us that, for $\chi \in \left[\frac{r_1 - r_0}{r_1}, 1\right)^\Theta$, letting $\underline{x} := \min_{\theta\in\Theta} \chi(\theta) \in \left[\frac{r_1-r_0}{r_1}, 1\right)$:

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

But notice that $v^*_{\mathbf{1}}$, being a concave function on a finite-dimensional space, is continuous on the interior of its domain. Therefore, $v^*_{\mathbf{1}}\left(\frac{\mu_0 - (1-\underline{x})r_1\gamma}{1-(1-\underline{x})r_1}\right) \to v^*_{\mathbf{1}}(\mu_0)$ as $\chi \to \mathbf{1}$,[30] implying $\liminf_{\chi\nearrow 1} v^*_\chi(\mu_0) \geq v^*_{\mathbf{1}}(\mu_0)$. Finally, monotonicity of $\chi \mapsto v^*_\chi(\mu_0)$ implies

[30] Note that Θ is finite, so that $\chi(\cdot) \to \mathbf{1}$ is equivalent to $\underline{x} \to 1$.

$v^*_\chi(\mu_0) \to v^*_{\mathbf{1}}(\mu_0)$ as $\chi \to \mathbf{1}$. 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) $\max[\bar{v}(\Delta\Theta) \setminus \{1\}] = 0$.

Consider any full-support prior μ_0. We will now prove a slightly stronger robustness result, that $v^*_\chi(\mu_0) \not\to v^*_{\mathbf{1}}(\mu_0)$ as $\chi \to \mathbf{1}$ *even if* we restrict attention to imperfect credibility which is independent of the state. To that end, take any constant $\chi \in [0,1)$. For any $\beta, \gamma \in \Delta\Theta,\ k \in [0,1]$ with $k\beta + (1-k)\gamma = \mu_0$ and $(1-k)\gamma \geq (\mathbf{1}-\chi)\mu_0$, that S does not get the benefit of the doubt implies (say by Lipnowski and Ravid (2019, 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 $\mu_\epsilon := (1-\epsilon)\gamma + \epsilon\mu_1$ has full support and satisfies

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

For sufficiently small ϵ, then, $v^*_{\mathbf{1}}(\mu_\epsilon) > 0$. Persuasion is therefore not robust to limited commitment at prior μ_ϵ. □

B.5 Persuading the Public: Proofs from Section 5

B.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 $\theta_0 := E\mu_0$ be the prior mean; let

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

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

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

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

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

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 $1 - I_\eta(1)$.*

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

B.5.2 Characterizing S-optimal equilibrium

Lemma 4. *Suppose $\bar{I} \in \mathcal{I}$, $I \in \mathcal{I}(\bar{I})$, and $\omega \in [0,1]$. Then there exist $\theta^* \in [0,\omega]$, $\theta^{**} \in [\omega, 1]$ and $I^* \in \mathcal{I}(\bar{I})$ such that:*

- *$I^* = \bar{I}$ on $[0,\theta^*]$, I is affine on $[\theta^*, \theta^{**}]$, and ${I^*}'(\theta) = 1$ on $[\theta^{**}, 1]$;*

- *$I^* - I$ is nonnegative on $[0,\omega]$ and nonpositive on $[\omega, 1]$.*

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 ω: this point will be our θ^*. The same line intersects the tangent line $\theta \mapsto \theta - \theta_0$ to the right of ω: this will be our θ^{**}. Finally, we construct I^* to coincide with upper bound function $\bar{I}$ to the left of θ^*, the θ^* tangent line on $[\theta^*, \theta^{**}]$, and the 1 tangent line $\theta \mapsto \theta - \theta_0$ to the right of θ^{**}.

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 Λ is closed because $\bar{I}$ is continuous, and it contains $I'(\omega)$ because I is convex and below $\bar{I}$. So let $\lambda := \min \Lambda$.

Let us now show that there is some $\theta^* \in [0,\omega]$ such that $I(\omega) - \lambda(\omega - \theta^*) = \bar{I}(\theta^*)$. First, if $\lambda = 0$, then $0 \leq I(\omega) \leq \bar{I}(0) = 0$; and so $\theta^* = 0$ is as desired. Focus now on the case that $\lambda > 0$. The compact subset $\{\bar{I}(\theta) - [I(\omega) - \lambda(\omega - \theta)] :\ 0 \leq \theta \leq \omega\}$ of $\mathbb{R}_+$

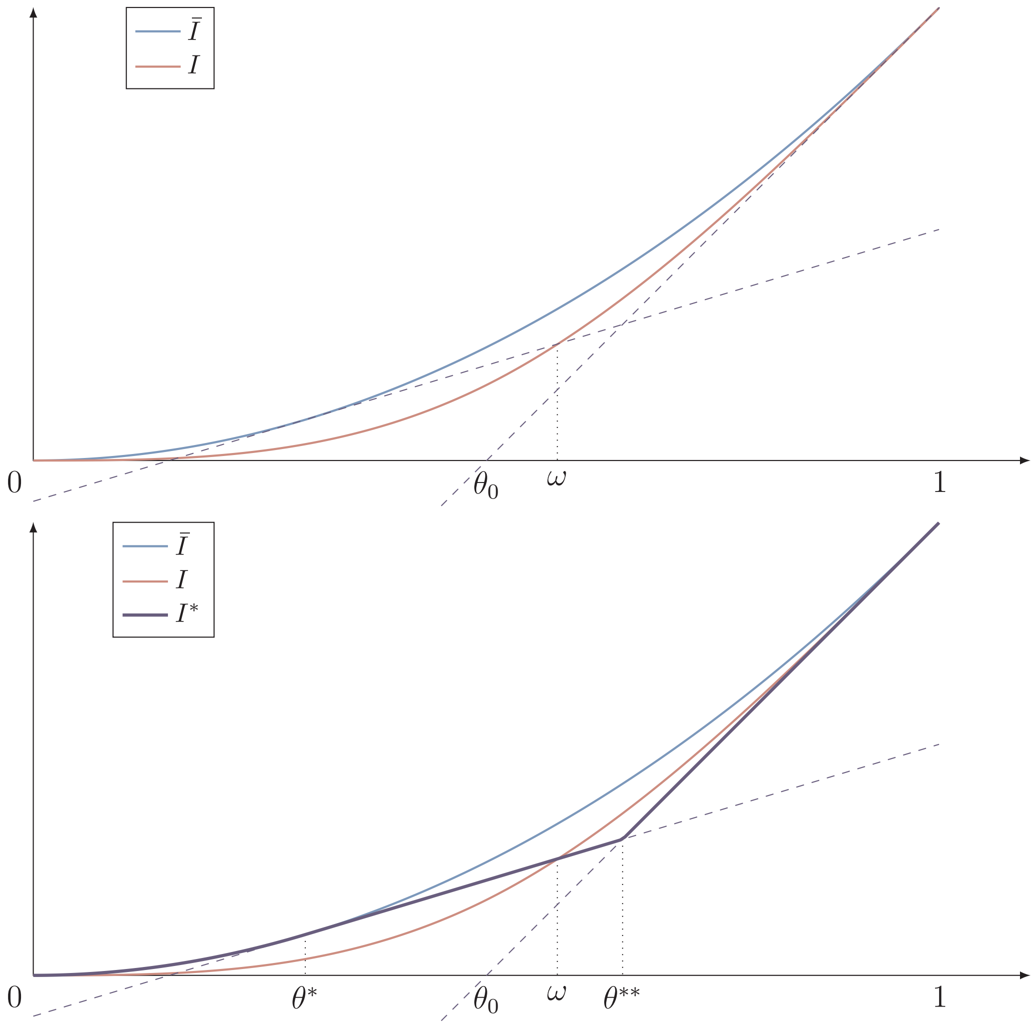

Figure 8: Construction of $(\theta^*, \theta^{**}, I^*)$ in Lemma 4

attains a minimum, which we wish to show is zero. If the minimum were $\epsilon > 0$, then $\max\{\lambda - \epsilon, 0\} \in \Lambda$ too, a contradiction to $\lambda = \min \Lambda$. So 0 is in the set as desired.

Construct now the function

$$I^* : \mathbb{R}_+ \to \mathbb{R}_+$$

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

The definition of θ^* ensures $I^*(\theta^*)$ is well-defined. 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 $\max\{I(\omega) + \lambda(1-\omega) \leq I(1) - I'(1)(1-1)$. So there is some $\theta^{**} \in [\omega, 1]$ such that $I^*(\theta)$ is equal to $I(\omega) + \lambda(\theta - \omega)$ for $\theta \in [\omega, \theta^{**}]$ and equal to $I(1) - I'(1)(1-\theta)$ for $\theta \in [\theta^{**}, \infty)$. This verifies the first bullet.

It remains to verify that $I^* - I$ is nonpositive on $[\omega, 1]$ and nonnegative on $[0, \omega]$, and that $I^* \in \mathcal{I}(\bar{I})$.

To see that $I^* - I$ is nonpositive above ω, consider any $\theta \in [\omega, 1]$ and use convexity of I. Specifically, first observe that $I(\theta) \geq I(1) - I'(1)(1-\theta) = \bar{I}(1) - (1-\theta)\bar{I}'(1)$. Next, that $\lambda \leq I'(\omega)$ implies $I(\theta) \geq I(\omega) + \lambda(\theta - \omega)$. So $I(\theta) \geq I^*(\theta)$. Moreover, $I^{*'} = 1 = I'$ on $[1, \infty)$, so the ranking holds everywhere above ω.

It is immediate that $I^* - I$ is nonnegative on $[0, \theta^*]$, so we turn to showing it is nonnegative on $(\theta^*, \omega]$ too; focus on the nontrivial case with $\theta^* < \omega$. That $I^* \leq \bar{I}$ on $(\theta^*, \omega]$ by definition of λ implies $\lambda = \bar{I}'(\theta^*)$. Assume then, for a contradiction, that some $\theta \in (\theta^*, \omega]$ has $I(\theta) > I^*(\theta)$. Then

$$\tfrac{I(\theta) - I(\theta^*)}{\theta - \theta^*} > \tfrac{I^*(\theta) - \bar{I}(\theta^*)}{\theta - \theta^*} = \lambda.$$

But then, I being convex, $I(\omega) > I(\theta) + \lambda(\omega - \theta) > I^*(\theta) + \lambda(\omega - \theta) = I(\omega)$, a contradiction. Thus $I^* - I$ is nonnegative on $[0, \theta^*]$ as desired.

All that remains is to show that $I^* \in \mathcal{I}(\bar{I})$. Letting $\underline{I} : \mathbb{R}_+ \to \mathbb{R}_+$ be given by $\underline{I}(\theta) := \max\{\bar{I}(1) - \bar{I}'(1)(1-\theta),\ 0\}$, 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 λ. 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 θ^*. But λ is a subdifferential of $\bar{I} \geq I^*$ at θ^*, and the two functions coincide at θ^*. It is therefore a subdifferential for I^* at the same, as required. □

Lemma 5. *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*

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

Proof.

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

□

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 θ^* upper censorship of μ is

$$\mathbf{1}_{[0,\theta^*)}\mu + \mu[\theta^*, 1]\delta_{\frac{1}{\mu[\theta^*,1]}\int_{[\theta^*,1]}\theta\ \mathrm{d}\mu(\theta)} \in \Delta\Theta$$

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

Lemma 6. *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 θ^* upper censorship of $\bar{\mu}$ for some $\theta^* \in [0, \omega]$. Moreover, this θ^* upper censorship puts probability* 1 *on* $[0, \theta^*] \cup [\omega, 1]$.

Proof. Let μ be a solution to the given program. Taking $\bar{I} := I_{\bar{\mu}}$ and $I := I_\mu$, note that the conditions of Lemma 4 are satisfied. Let $I^* \in \mathcal{I}$, $\theta^* \in [0, \omega]$, and $\theta^{**} \in [\omega, 1]$ be as delivered by Lemma 4 and $\mu^* \in \Delta\Theta$ be such that $I^* = I_{\mu^*}$. Then, by Lemma 5 (letting $\tilde{h} := \tilde{H}'$),

$$\begin{aligned}
\int_0^1 \tilde{H}\ \mathrm{d}\mu^* - \int_0^1 \tilde{H}\ \mathrm{d}\mu &= \tilde{H}(0)(I^* - I)'(0) + \int_0^1 \tilde{H}\ \mathrm{d}(I^* - I)' \\
&= \int_0^\omega (I^* - I)\ \mathrm{d}\tilde{h} + \int_\omega^1 (I - I^*)\ \mathrm{d}(-\tilde{h}).
\end{aligned}$$

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 μ then tells us that μ^* is optimal (and equal to μ).

By construction, $\mu^*[0,\theta]=\bar{\mu}[0,\theta]$ for every $\theta\in[0,\theta^*)$, and (since, by hypothesis, $\bar{\mu}\{\theta^*\}=0$ if $\bar{\mu}(\theta^*,1]>0$) we have $|[\theta^*,1]\cap\mathrm{supp}(\mu^*)|=1$. But these properties—which will clearly also be satisfied by a θ^* upper censorship of $\bar{\mu}$—characterize a unique distribution of any given mean. Therefore, μ^* is a θ^* upper censorship of $\bar{\mu}$.

Finally, the "moreover" point follows from $\theta^{**}\geq\omega$, as guaranteed by Lemma 4. $\square$

Lemma 7. *There is a unique $\bar{\theta}_\chi\in[0,1]$ such that*[31] $\int_0^{\bar{\theta}} \chi\mu_0[0,\theta]\ \mathrm{d}\theta$ *is*

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

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

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

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

with the last inequality being strict if $\chi\mu_0\neq\mu_0$. The result then follows from the intermediate value theorem. $\square$

In what follows, recall the mean distribution $\bar{\mu}_\chi$ as defined in Section 5.

Lemma 8. *For any $\theta\in[0,1]$, we have*

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

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

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

Proof. That $I_{\bar{\mu}_\chi}$ coincides with $I_{\chi\mu_0}$ on $[0, \bar{\theta}_\chi]$ and has derivative 1 on $(\bar{\theta}_\chi, 1]$ follows directly from the definition of $\bar{\mu}_\chi$. Noting that $I_{\chi\mu_0}(\bar{\theta}_\chi) = \bar{\theta}_\chi - \theta_0$ by Lemma 7, it follows that $I_{\bar{\mu}_\chi}(\theta) = \theta - \theta_0$ for $\theta \in [\bar{\theta}_\chi, 1]$.

Next, recall that $I_{\chi\mu_0}(\theta) - (\theta - \theta_0)$ is nonnegative for $\theta \in [0, \bar{\theta}_\chi]$ and nonpositive for $\theta \in [\bar{\theta}_\chi, 1]$ by Lemma 7. Consequently, $I_{\bar{\mu}_\chi}(\theta) = \max\{I_{\chi\mu_0}(\theta),\ \theta - \theta_0\}$ for every $\theta \in [0, 1]$.

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

We now prove Claim 1.

Proof. First, we show that $\hat{v}(\bar{\mu}_\chi) = \max_{\theta^* \in [0, \bar{\theta}_\chi]} \int H \, \mathrm{d}\mu_{\chi,\theta^*}$, and that the maximum on the RHS is attained. By Lemma 6, there is some $\theta^* \in [0, 1]$ such that $\hat{v}(\bar{\mu}_\chi) = \int H \, \mathrm{d}\mu_{\chi,\theta^*}$. As $\bar{\mu}_\chi[0, \bar{\theta}_\chi] = 1$, we have $\mu_{\chi,\theta} = \mu_{\chi,\bar{\theta}_\chi}$ for every $\theta \in [\bar{\theta}_\chi, 1]$; so we may without loss take $\theta^* \leq \bar{\theta}_\chi$. Furthermore, since

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

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

Next, given $\theta^* \in [0, \bar{\theta}_\chi]$, we exhibit an equilibrium in which S communicates via a θ^*-upper-censorship pair, and observe that this induces S value $\int H \, \mathrm{d}\mu_{\chi,\theta^*}$—in particular showing $\int H \, \mathrm{d}\mu_{\chi,\theta^*} \leq v^*_\chi(\mu_0)$. 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 $\gamma := \frac{[1 - \chi \mathbf{1}_{[0,\theta^*)}]\mu_0}{1 - \chi\mu_0[0,\theta^*)}$ (with $\gamma := \delta_1$ if $\chi\mu_0[0, \theta^*) = 1$). Then let R behavior be given by $\alpha := H \circ E \circ \pi$. The Bayesian property is now straightforward, and the R incentive condition holds by construction. To verify that this is a χ-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 show that $E\gamma \geq \theta^*$. This holds vacuously if $\gamma = \delta_1$, so focus on the alternative case in which $\bar{\mu}_\chi[\theta^*, \bar{\theta}_\chi] > 0$. In this case,

$$\begin{aligned}
\bar{\mu}_\chi[\theta^*, \bar{\theta}_\chi]\ \ (E\gamma - \theta^*) &= \int_{[\theta^*, \bar{\theta}_\chi]} (\theta - \theta^*)\ \mathrm{d}\bar{\mu}_\chi(\theta) \\
&= -\int_{[\theta^*, 1]} (\theta^* - \theta)\ \mathrm{d}\bar{\mu}_\chi(\theta) \\
&= \int_{[0, \theta^*)} (\theta^* - \theta)\ \mathrm{d}\bar{\mu}_\chi(\theta) - (\theta^* - \theta_0) \text{ (by Lemma 8)} \\
&= [(\theta^* - \theta)\bar{\mu}_\chi[0, \theta]]_0^{(\theta^*)^-} - \int_{[0, \theta^*)} (-1)\bar{\mu}_\chi[0, \theta]\ \mathrm{d}\theta - (\theta^* - \theta_0) \\
&= [0 - 0] + I_{\chi\mu_0}(\theta^*) - (\theta^* - \theta_0) \\
&\geq 0 \text{ by Lemma 8.}
\end{aligned}$$

S incentive-compatibility follows. To show this equilibrium generates the required payoff, it suffices to show that the induced distribution μ of posterior means is equal to μ_{χ,θ^*}. For any $\theta \in [0, \theta^*)$, notice that

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

Moreover, $|[\theta^*, 1] \cap \mathrm{supp}(\mu)| = 1 = |[\theta^*, 1] \cap \mathrm{supp}(\mu_{\chi,\theta^*})|$. Equality then follows from equality of their means (Lemma 8).

Finally, we show that $v^*_\chi(\mu_0) \leq \hat{v}(\bar{\mu}_\chi)$. To that end, let (β, γ, k) solve the program of Theorem 1 – and, without loss, say $\beta = \mu_0$ if $k = 0$. Let $\omega := \omega^* \wedge E\gamma$, and see that $H(E\gamma) \wedge H$ is continuous, convex on $[0, \omega]$, and concave on $[\omega, 1]$. Therefore, by Lemma 6, there is some $\theta^* \in [0, \omega]$ such that the θ^* upper censorship of β belongs to $\mathrm{argmax}_{\hat{\beta} \preceq \beta} \int H(E\gamma) \wedge H\ \mathrm{d}\hat{\beta}$. Let $\lambda := \beta[0, \theta^*) \in [0, 1]$, $\eta := \frac{\mathbf{1}_{[\theta^*,1]}\beta}{1-\lambda} \in \Delta\Theta$, $\hat{\gamma} := \frac{(1-k)\gamma + (1-\lambda)k\eta}{1-\lambda k} \in \Delta\Theta$, and $\hat{\beta} := \frac{\mathbf{1}_{[0,\theta^*)}\beta}{\lambda} \in \Delta\Theta$.[32] 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 6 tells us

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

$E\eta \geq \omega$, so that $H(E\gamma) \wedge H$ is concave on $\text{co}\{E\gamma, E\eta\}$. Now, observe that

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

Letting $\hat{\mu} := k\lambda\hat{\beta} + (1-\lambda k)\delta_{E\hat{\gamma}}$, the payoff ranking (and so too the claim) will follow if we show that $\hat{\mu} \preceq \bar{\mu}_\chi$. As (appealing to Lemma 8) $E\bar{\mu}_\chi = \theta_0 = E\hat{\mu}$, it suffices to show that $I_{\hat{\mu}} \leq I_{\bar{\mu}_\chi}$.

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

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

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 $I_{\hat{\mu}}(\theta) \leq \max\{I_{\chi\mu_0}(\theta), \theta - \theta_0\}$ for $\theta \in [0, E\hat{\gamma}] \cup [E\hat{\gamma}, 1]$. Lemma 8 then tells us that $I_{\hat{\mu}} \leq I_{\bar{\mu}_\chi}$. □

B.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 $\langle\mu_0, \chi\rangle$ and $\langle\tilde{\mu}_0, \tilde{\chi}\rangle$ such that $E\mu_0 = E\tilde{\mu}_0 = \theta_0$, we show that $v_\chi^*(\mu_0) \geq v_{\tilde{\chi}}^*(\tilde{\mu}_0)$ if and only if $\bar{\mu}_\chi \succeq \bar{\tilde{\mu}}_{\tilde{\chi}}$.

Proof. Appealing to Claim 1 and Lemma 5,

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

Let $I_* := I_{\bar\mu_\chi}$ and $\tilde I_* := I_{\bar{\tilde\mu}_{\tilde\chi}}$. 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 (θ_1, θ_2). If h is increasing, then

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

As $(I_* - \tilde I_*)$ is strictly positive over (θ_1, θ_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 $v^*_\chi(\mu_0) \geq v^*_{\tilde\chi}(\tilde\mu_0)$.

□

B.6 Proofs from Section 6: Investing in Credibility

In this section, we prove the following formal claim concerning the public persuasion application with costly endogenous credibility.

Claim 3. *There exists an optimal credibility choice. Moreover, any optimal choice (along with S-optimal equilibrium) is a cutoff credibility choice, and entails full revelation by the official reporting protocol.*

Toward the proof, we first establish the following lemma.

Lemma 9. *For any non-cutoff credibility choice (i.e. any χ such that there is no $\theta^* \in [0,1]$ with $\chi = \mathbf{1}_{[0,\theta^*)}$ μ_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 χ not of the desired form. In particular, this implies that χ is not μ_0-a.s. equal to 1, so that $\chi\mu_0(\Theta) < 1$.

As μ_0 is atomless, there is some $\theta^* \in [0,1)$ such that $\mu_0[0,\theta^*) = \chi\mu_0(\Theta)$. That $\mathbf{1}_{[0,\theta^*)}\mu_0 \neq \chi\mu_0$ but the two have the same total measure implies that $\text{supp}[(\mathbf{1}-\chi)\mu_0]$ intersects $[0,\theta^*)$. For each $\theta_* \in [0,\theta^*]$, define the function $\eta_{\theta_*} := I_{\mathbf{1}_{[0,\theta_*)}\mu_0} - I_{\chi\mu_0} : \mathbb{R}_+ \to \mathbb{R}$. By construction, its right-hand-side derivative at any θ is given by $\eta'_{\theta_*}(\theta) = \int_0^\theta (\mathbf{1}_{[0,\theta_*)} - \chi)\, \mathrm{d}\mu_0$. In particular, this implies (since $\chi\mu_0$ strictly first-order stochastically dominates $\mathbf{1}_{[0,\theta^*)}$) that η'_{θ^*} is globally nonnegative, weakly quasiconcave with peak at θ^*, 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 η_{θ^*} continuous, there is some $\theta_* \in [0,\theta^*)$ close enough to θ^* to ensure that $\eta_{\theta_*}(\theta_*) \geq \epsilon$ and $\mu_0(\theta_*,\theta^*] \leq \epsilon$. Let $\eta := \eta_{\theta_*}$.

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

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

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

Let us now observe that η 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 $I_{\mathbf{1}_{[0,\theta_*)}\mu_0} \geq I_{\chi\mu_0}$ globally. Lemma 8 then implies that $\bar{\mu}_{\mathbf{1}_{[0,\theta_*)}} \succeq \bar{\mu}_\chi$. Finally, Claim 2 tells us that $v^*_{\mathbf{1}_{[0,\theta_*)}}(\mu_0) \geq v^*_\chi(\mu_0)$. Meanwhile, the cost of credibility $\mathbf{1}_{[0,\theta_*)}$, is strictly below that of credibility χ. $\square$

Now, we prove Claim 3

Proof. Consider any credibility choice χ and accompanying χ-equilibrium. Lemma 9 shows that χ 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 χ-equilibrium entails influenced θ^* 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 θ_* can be lowered to θ^* for a strict gain to the objective. But then, since $\chi|_{[\theta_*,1]} = 0$, 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 θ^*, 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. $\square$