Persuasion via Weak Institutions

Elliot Lipnowski

Columbia University

Doron Ravid

University of Chicago

Denis Shishkin

University of California San Diego

A sender commissions a study to persuade a receiver but influences the report with some probability. We show that increasing this probability can benefit the receiver and can lead to a discontinuous drop in the sender's payoffs. To derive our results, we geometrically characterize the sender's highest equilibrium payoff, which is based on the concavification of a capped value function.

I. Introduction

Many institutions routinely collect and disseminate information. Although the collected information is instrumental to its consumers, often the main goal of dissemination is to persuade. Persuading one's audience, however,

Lipnowski and Ravid acknowledge support from the National Science Foundation (grant SES-1730168). We thank Roland Bénabou, Ben Brooks, Joyee Deb, Eddie Dekel, Wouter Dessein, Jon Eguia, Emir Kamenica, Navin Kartik, Stephen Morris, Pietro Ortoleva, Wolfgang Pesendorfer, Carlo Prato, Marzena Rostek, Evan Sadler, Zichang Wang, Richard Van Weelden, Leeat Yariv, and various audiences for useful suggestions. We also thank

Electronically published August 18, 2022

Journal of Political Economy, volume 130, number 10, October 2022.

https://doi.org/10.1086/720462

requires the audience to believe what one says. In other words, the institution must be credible, meaning it must be capable of delivering both good and bad news. Yet if the institution is not independent from its superiors, delivering unfavorable news might be especially difficult. This paper studies how an institution's credibility influences its persuasiveness and the quality of information it provides.

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

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

When the central bank is fully credible, $\chi = 1$, it is committed to the official report. As such, the leader can communicate any information she chooses, and so this example falls within the framework of Kamenica and Gentzkow (2011). Using their results, one can deduce that the policy maker optimally chooses a symmetric binary signal,

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

Chris Baker, Sulagna Dasgupta, Takuma Habu, and Elena Istomina for excellent research assistance. This paper was edited by Emir Kamenica.

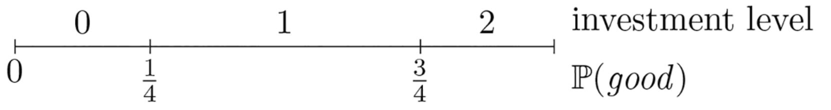

FIG. 1.—Firm's best response in central bank example.

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

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

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

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

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

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

The corner cases of our model lend themselves to geometric analysis. Let the sender's *value function* be the highest value the sender can obtain from the receiver responding optimally at a given posterior belief. Kamenica and Gentzkow (2011) show that concavifying this function gives the sender's largest equilibrium payoff in the Bayesian persuasion model. More recently, Lipnowski and Ravid (2020) observe that as long as the sender cares about only the receiver's actions, quasiconcavifying the sender's value function delivers her highest equilibrium payoff under cheap talk.

Our theorem 1 uses the aforementioned geometric approach to characterize the sender's maximal equilibrium value in the intermediate credibility case, $\chi \in (0, 1)$. To do so, the theorem partitions the sender's equilibrium messages into two sets: messages the sender willingly sends when influencing the report (e.g., g in the above example) and messages communicated only by the official report. One might guess that concavification and quasiconcavification characterize the sender's payoffs from official and influenced reporting, respectively. However, we show that whereas quasiconcavification characterizes the sender's payoffs from influenced reporting, one cannot find the sender's utility from official reporting using concavification alone. The reason is that the sender's payoff from a message cannot surpass the utility she obtains under compromised reporting: if it did, the sender would have a profitable deviation. To account for this incentive constraint, one must cap the sender's value function at her utility from influenced reporting before concavifying it.

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

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

One may wonder if such collapses may occur at full credibility. Our proposition 3 shows that such a discontinuity can occur but only in knife-edge cases. Thus, although the sender's value often drops at some prior and some χ because of small decreases in credibility, it rarely does so at $\chi = 1$.

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), which studies sender-receiver games in which a sender commits to an information transmission strategy. The no-credibility specialization of our model reduces to cheap talk (Crawford and Sobel 1982; Green and Stokey 2007). In particular, we build on Lipnowski and Ravid (2020), who use the belief-based approach to study cheap talk under state-independent sender preferences.

Two recent papers (Fréchette, Lizzeri, and Perego 2022; Min 2021) study closely related models. Fréchette, Lizzeri, and Perego (2022) test experimentally the connection between the informativeness of the sender's communication and her credibility in the binary state, binary action version of our model. Min (2021) looks at a generalization of our model in which the sender's preferences can be state dependent. He shows that the sender weakly benefits from a higher commitment probability. Applying Blume, Board, and Kawamura's (2007) results on noisy communication, Min (2021) also shows that allowing the sender to commit with positive rather than zero probability strictly helps both players in Crawford and Sobel's (1982) uniform quadratic example.

Other thematically related work studies games of information transmission while varying the (exogenous or endogenous) limits to communication. Some such work focuses on games of direct communication, showing how some manner of commitment power can be sustained (for either a sender or a receiver) via lying costs (e.g., Kartik 2009; Guo and Shmaya 2021; Nguyen and Tan 2021), repeated interactions (e.g., Mathevet, Pearce, and Stacchetti 2022; Best and Quigley 2022), verifiable information (e.g., Glazer and Rubinstein 2006; Sher 2011; Hart, Kremer, and Perry 2017; Ben-Porath, Dekel, and Lipman 2019), informational control (e.g., Ivanov 2010; Luo and Rozenas 2018), or mediation (e.g., Goltsman et al. 2009; Salamanca 2021). Other work considers models in which a sender chooses an experiment ex ante, asking how persuasion can be shaped by exogenous experiment constraints (e.g., Ichihashi 2019; Perez-Richet and Skreta 2022) or by signaling motives (e.g., Perez-Richet 2014; Hedlund 2017; Alonso and Câmara 2018).

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

II. A Weak Institution

We analyze a game with two players: a sender (she) and a receiver (he). Whereas both players' payoffs depend on the receiver's action, $a \in A$, the receiver's payoff also depends on an unknown state, $\theta \in \Theta$. Thus, the sender and the receiver 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 the sender commissioning a report, $\xi : \Theta \to \Delta M$, to be delivered by a research institution. The state then realizes, and the receiver sees a message $m \in M$ (without observing θ). Given any θ, the sender is credible with probability χ, meaning m is drawn according to the official reporting protocol, $\xi(\cdot|\theta)$. With probability $1 - \chi$, the sender is not credible, in which case the sender decides which message to send after privately observing θ. Only the sender learns her credibility type, and she learns it only after announcing the official reporting protocol.[1]

[1] In the appendix, we show that our payoff results are unchanged if the sender learns her credibility type before choosing the official report.

We impose some technical restrictions on our model.[2] Both A and Θ are finite spaces with at least two elements. The state, θ, follows some prior distribution $\mu_0 \in \Delta\Theta$, which is known to both players. Finally, we assume that M is rich enough to ensure that the sender faces no exogenous constraints on communication.[3]

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

1. *Bayesian updating*: the belief map $\pi : M \to \Delta\Theta$ satisfies Bayes's rule given prior μ_0 and the message policy
$$\chi\xi + (1 - \chi)\sigma : \Theta \to \Delta M.$$
2. *Receiver optimality*: every $m \in M$ has $\alpha(m)$ supported on
$$\operatorname*{argmax}_{a \in A} \sum_{\theta \in \Theta} u_R(a, \theta)\pi(\theta | m).$$
3. *Sender optimality*: every $\theta \in \Theta$ has $\sigma(\theta)$ supported on
$$\operatorname*{argmax}_{m \in M} \sum_{a \in A} u_S(a)\alpha(a | m).$$

We view the sender as a principal capable of steering the receiver toward her favorite χ-equilibria. In Lipnowski, Ravid, and Shishkin (2022), we define the notion of perfect Bayesian χ-equilibrium in which we explicitly model the sender's incentives at the experiment choice stage. By appropriately completing off-path play, that paper shows that the sender's highest χ-equilibrium payoff coincides with her highest perfect Bayesian χ-equilibrium payoff.

[2] We view every topological space as a measurable space with its Borel field. For any measurable space Y, we denote by ΔY the set of all probability measures over Y. For any measurable spaces X, Y, a map $X \to Y$ is a measurable function $X \to Y$.

[3] For example, we could take $M = [0, 1]$ (see appendix). Moreover, corollary 1 in the appendix implies that the sender's optimal equilibrium payoff would remain unchanged if M were instead finite with $|M| \geq \min\{|A|, 2|\Theta| - 1\}$.

III. Persuasion with Partial Credibility

In this section, we characterize the sender's maximal χ-equilibrium payoff. Our analysis applies the belief-based approach (Kamenica 2019; Forges 2020). Within an equilibrium, each message m that the sender communicates to the receiver induces a posterior belief $\mu = \pi(m) \in \Delta\Theta$ and an expected sender utility from the receiver's (potentially mixed) action $s = \Sigma_{a\in A} u_S(a)\alpha(a|m) \in \mathbb{R}$. By replacing each message with its associated μ and s, one can transform the equilibrium distribution of messages into its induced joint distribution $\mathbf{P}$ of the receiver's beliefs and the sender's continuation payoffs. We refer to $(\mu, s) \in \Delta\Theta \times \mathbb{R}$ as an *outcome*, and to a distribution $\mathbf{P} \in \Delta(\Delta\Theta \times \mathbb{R})$ as an *outcome distribution*, and we define a *χ-equilibrium outcome distribution* to be an outcome distribution induced by a χ-equilibrium.

A. *The Extreme Cases*

We now review existing results that cover the extreme cases of our model. These cases serve as building blocks for proving our main theorem, which covers the case in which χ is intermediate.

1. Full Credibility

When $\chi = 1$, the sender's official announcement is binding, and so our model reduces to the Bayesian persuasion model of Kamenica and Gentzkow (2011). We now review some of their results. With full credibility, the sender is hampered by only two constraints. The first constraint is that the receiver updates his beliefs using Bayes's rule, which is equivalent to the receiver's posterior belief averaging to his prior. That is, $\mathbf{P}$ must satisfy

$$\int \mu \, \mathrm{d}\mathbf{P}(\mu, s) = \mu_0. \qquad \text{(Bayes)}$$

The second constraint is that the receiver must be best responding: for any belief the receiver holds, he must take only actions he finds optimal. To formalize this requirement, define the sender's *value correspondence* to be the correspondence mapping each posterior belief to the set of payoffs the sender can attain from the receiver-optimal behavior,[4]

$$\begin{aligned} V : \Delta\Theta &\rightrightarrows \mathbb{R} \\ \mu &\mapsto \operatorname{co} u_S\left(\operatorname*{argmax}_{a\in A} \sum_{\theta\in\Theta} u_R(a,\theta)\mu(\theta)\right). \end{aligned}$$

[4] The reason for the convex hull in V's definition is that the receiver may choose to mix in the event that he has multiple best responses to a given belief.

Then, **P** is compatible with the receiver's incentive constraint if and only if **P** is supported on the graph of V; that is, a message can induce an outcome (μ, s) only if $s \in V(\mu)$. Letting $\text{gr } V := \{(\mu, s) : s \in V(\mu)\}$ denote the graph of V, we can state this constraint formally as

$$\mathbf{P}(\text{gr } V) = 1. \tag{R-IC}$$

As noted by Kamenica and Gentzkow (2011), the conditions (R-IC) and (Bayes) are together necessary and sufficient for an outcome distribution **P** to arise from some 1-equilibrium. Denote the subset of $\Delta(\Delta\Theta \times \mathbb{R})$ that satisfy these conditions for a prior μ_0 and value correspondence V by

$$\text{BP}(\mu_0, V) = \{\mathbf{P} \in \Delta(\Delta\Theta \times \mathbb{R}) : \mathbf{P} \text{ satisfies (Bayes) and (R-IC)}\}.$$

One can characterize the sender's highest 1-equilibrium payoff using her *value function*,

$$v : \Delta\Theta \to \mathbb{R}$$
$$\mu \mapsto \max V(\mu),$$

which maps every belief to the utility the sender obtains if the receiver chooses optimally and breaks ties in the sender's favor given multiple best responses. Specifically, one can show that the sender's utility in her favorite 1-equilibrium equals $\hat{v}(\mu_0)$, where

$$\hat{v} := \text{cav}(v)$$

is the lowest concave function that is everywhere above v (e.g., Aumann and Maschler 1995; Kamenica and Gentzkow 2011). The function $\hat{v}$ is known as v's concavification.

Figure 2 illustrates the above in the context of the central bank example from the introduction. Because the state is binary, we identify the receiver's posterior belief μ with the probability it assigns to the economy being good. The left panel in figure 2 plots the sender's value correspondence, taking μ

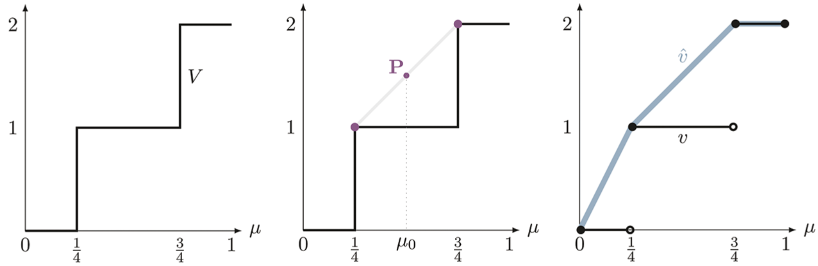

FIG. 2.—Value correspondence V, sender's best 1-equilibrium outcome **P**, and value function v with its concavification $\hat{v}$ in central bank example.

as an input. For $\mu < 1/4$, the sender can only get a payoff of 0, whereas when $\mu \in (1/4, 3/4)$, she can only get 1, and when $\mu > 3/4$, she can only get 2. The sender can attain any payoff between 0 and 1 when $\mu = 1/4$ and any payoff between 1 and 2 when $\mu = 3/4$. The middle panel depicts the sender's best 1-equilibrium outcome distribution **P**, which assigns equal weight to the points $(\mu, s) = (1/4, 1)$ and $(\mu, s) = (3/4, 2)$. As can be seen, both points lie on the graph of V, meaning that this distribution satisfies (R-IC). This distribution also satisfies (Bayes) because the average probability assigned to $\theta = good$ equals $1/2$, which is the probability assigned to that state by the prior. One can visually verify that this distribution is indeed sender optimal by examining the right panel, which shows the sender's value function along with its concave envelope,

$$v(\mu) = \begin{cases} 0 & \text{if } \mu \leq 1/4, \\ 1 & \text{if } \mu \in [1/4, 3/4], \\ 2 & \text{if } \mu \geq 3/4, \end{cases} \quad \hat{v}(\mu) = \begin{cases} 4\mu & \text{if } \mu \leq 1/4, \\ 1 + 2(\mu - 1/4) & \text{if } \mu \in [1/4, 3/4], \\ 2 & \text{if } \mu \geq 3/4. \end{cases} \quad (1)$$

As seen in the figure, the outcome distribution **P** gives the sender an expected payoff of $3/2$, which is also the value of $\hat{v}(\mu_0)$, thereby confirming that **P** is indeed sender optimal.

2. No Credibility

We now turn to the $\chi = 0$ case, in which the receiver knows the sender is choosing m after observing the state. Being freely chosen, the sender's communication is cheap talk (Crawford and Sobel 1982; Green and Stokey 2007) and thus needs to satisfy the sender's incentive constraints. Our assumption that the sender's preferences are state independent simplifies these constraints considerably: the sender must be indifferent between all on-path messages. The reason is that if the sender's payoffs across two distinct messages differ, the sender will never (in any state) want to send the lower-payoff message. As such, the sender's payoff from all outcomes in the support of a 0-equilibrium outcome distribution must be the same. In other words, every 0-equilibrium outcome distribution **P** must satisfy

$$\mathbf{P}\{\Delta\Theta \times \{s_i\}\} = 1 \text{ for some } s_i \in \mathbb{R}. \quad \text{(CP)}$$

Combining (CP) with the restrictions imposed by Bayesian updating (Bayes) and the receiver incentives (R-IC), one obtains a full characterization of the attainable outcome distributions under no credibility (see Aumann and Hart 2003; Lipnowski and Ravid 2020). It follows that the sender's highest 0-equilibrium payoff is given by

$$\max_{\mathbf{P}\in \mathrm{BP}(\mu_0, V)} \int s\, \mathrm{d}\mathbf{P}(\mu, s) \text{ subject to (CP).} \tag{CT}$$

Lipnowski and Ravid (2020) show that this maximal payoff is equal to $\bar{v}(\mu_0)$, where

$$\bar{v} = \mathrm{qcav}(v)$$

is v's quasiconcavification, that is, the lowest quasiconcave function that is everywhere above v.

Figure 3 depicts v's quasiconcavification and concavification, respectively, for some function v. These functions describe the sender's ability to benefit from communication by connecting points on the graph of the sender's value correspondence. With full credibility, the sender can connect such points using any affine segment. When $\chi = 0$, the sender's incentive constraints dictate that her payoff coordinate must remain constant; that is, the sender can use only flat segments.

Let us revisit the example from the introduction when $\chi = 0$. Observe that the optimal 1-equilibrium outcome distribution in this example does not satisfy (CP), because it generates two outcomes with different sender payoffs and so cannot be induced by a 0-equilibrium (see fig. 2, middle panel). We now argue that the sender cannot attain any value above 1 in any 0-equilibrium. One way of seeing this fact is to observe that the sender's value function in this example is quasiconcave and is therefore equal to its quasiconcavification. Alternatively, observe that (Bayes) requires every 0-equilibrium outcome distribution $\mathbf{P}$ to induce at least one outcome with $\mu \leq 1/2$, whereas (R-IC) requires the sender's payoff from all such beliefs to be below 1. Because the sender's payoff must be constant over $\mathbf{P}$'s support by (CP), it follows that $\mathbf{P}$ cannot induce a sender payoff strictly larger than 1.

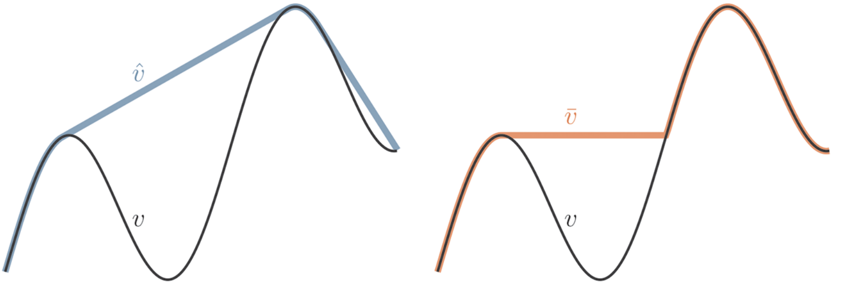

FIG. 3.—Value function v and its quasiconcavification $\bar{v}$ and concavification $\hat{v}$.

B. The Intermediate Credibility Case

This section presents theorem 1, which geometrically characterizes the sender's optimal χ-equilibrium value for our general model.

Suppose that credibility is not extreme ($0 < \chi < 1$) so that both the official reporting protocol and the sender's influencing strategy are relevant, and let $\mathbf{P}$ be a χ-equilibrium outcome distribution. Notice that the receiver optimality and the Bayesian-updating conditions are as in the full- and no-credibility cases, and so $\mathbf{P}$ must satisfy (Bayes) and (R-IC); that is, $\mathbf{P} \in \mathrm{BP}(\mu_0, V)$. We now use these conditions to derive an upper bound on the sender's value from $\mathbf{P}$.

We begin by decomposing $\mathbf{P}$ into two distributions. To do so, let

$$s_{\max} := \max\{s : (\mu, s) \in \mathrm{supp}(\mathbf{P})\}$$

be the highest payoff in the support of $\mathbf{P}$, and let $k \in [0, 1]$ denote the $\mathbf{P}$-probability of sender payoffs strictly below $s_{\max}$. In what follows, we focus on the case in which $0 < k < 1$.[5] Let $\mathbf{G}$ be the distribution over outcomes induced by $\mathbf{P}$ conditional on $s = s_{\max}$, and let $\mathbf{B}$ be the outcome distribution conditional on $s < s_{\max}$. By construction,

$$\mathbf{P} = (1 - k)\mathbf{G} + k\mathbf{B}.$$

For an example, consider the optimal 1-equilibrium outcome distribution $\mathbf{P}$ from the central bank example, which generates the outcomes $(\mu, s) = (1/4, 1)$ and $(\mu, s) = (3/4, 2)$ with equal probability. In this case, $s_{\max} = 2$ and $k = 1/2$, whereas $\mathbf{G}$ and $\mathbf{B}$ are degenerate on $(3/4, 2)$ and $(1/4, 1)$, respectively.

We now bound the sender's payoff from $\mathbf{P}$ from above by applying the results of the extreme cases of our model to the above decomposition. We begin by bounding the value the sender obtains from $\mathbf{G}$. To do so, note that because $\mathbf{P}$ satisfies (R-IC), $\mathbf{G}$ is supported on the graph of V. It follows that $\mathbf{G} \in \mathrm{BP}(\gamma, V)$, where $\gamma = \int \mu \, \mathrm{d}\mathbf{G}(\mu, s)$ is the receiver's expected posterior under $\mathbf{G}$. Moreover, observe that $\mathbf{G}$ satisfies the constant sender payoff condition (CP): by construction, $\mathbf{G}$ only induces outcomes that give the sender a payoff of $s_{\max}$. Hence, given the above characterization of feasible distributions for the no-credibility case, $\mathbf{G}$ is compatible with a 0-equilibrium for the game with modified prior γ. Therefore, we can bound the sender's expected payoff from $\mathbf{G}$ using the quasiconcavification of the sender's value function:

$$s_{\max} = \int s \, \mathrm{d}\mathbf{G}(\mu, s) \le \bar{v}(\gamma).$$

[5] It will be apparent that in the cases of $k = 0$ and $k = 1$, the payoff upper bound we derive will remain an upper bound.

Next, we use concavification to bound from above the sender's expected payoff from **B**. Toward this goal, for every payoff $\bar{s}$, define the correspondence $V_{\wedge\bar{s}} : \Delta\Theta \rightrightarrows \mathbb{R}$ that censors $V(\mu)$ from above by $\bar{s}$:

$$V_{\wedge\bar{s}}(\mu) = \{\min\{s, \bar{s}\} : s \in V(\mu)\}.$$

Figure 4 illustrates $V_{\wedge\bar{s}}$. The graph of this correspondence is constructed by reducing to $\bar{s}$ the payoff coordinate of every outcome (μ, s) in V's graph whose s is above $\bar{s}$. Other outcomes in V's graph are kept unchanged.

To understand why $V_{\wedge\bar{s}}$ is a useful correspondence, observe that **B** is supported on the graph of V and that, by definition, **B** never yields a sender payoff above $s_{\max}$. In other words, for any $\bar{s}$ larger than $s_{\max}$, **B** only generates outcomes from the graph of V that are also in the graph of $V_{\wedge\bar{s}}$. Hence, whenever $\bar{s} \geq s_{\max}$, the outcome distribution **B** is in the set $\mathrm{BP}(\beta, V_{\wedge\bar{s}})$, where $\beta = \int \mu \, \mathrm{d}\mathbf{B}(\mu, s)$ is the receiver's average posterior under **B**. Therefore, **B** must give the sender a utility below the maximal payoff that the sender can get from some distribution in this set. As we explained in section III.A, one can find this maximal payoff using concavification. Specifically, let

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

be the function that assigns every belief μ with the highest sender utility in $V_{\wedge\bar{s}}(\mu)$, and let $\hat{v}_{\wedge\bar{s}}$ be the concavification of $v_{\wedge\bar{s}}$. Then, $\hat{v}_{\wedge\bar{s}}(\beta)$ is the highest payoff the sender can obtain from any distribution in $\mathrm{BP}(\beta, V_{\wedge\bar{s}})$. Because $\bar{v}(\gamma) \geq s_{\max}$, setting $\bar{s} = \bar{v}(\gamma)$ delivers that **B** gives the sender an expected payoff below $\hat{v}_{\wedge\bar{v}(\gamma)}$. To ease notational burden, we use

$$\hat{v}_{\wedge\gamma} := \mathrm{cav}(v_{\wedge\bar{v}(\gamma)})$$

as shorthand for $\hat{v}_{\wedge\bar{v}(\gamma)}$.

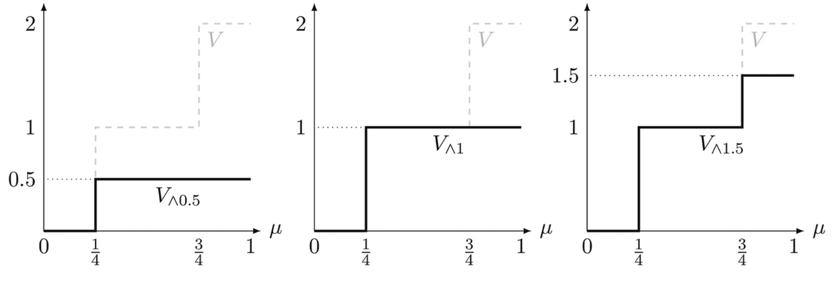

FIG. 4.—Construction of $V_{\wedge\bar{s}}$ for $\bar{s} = 0.5, 1, 1.5$ in central bank example.

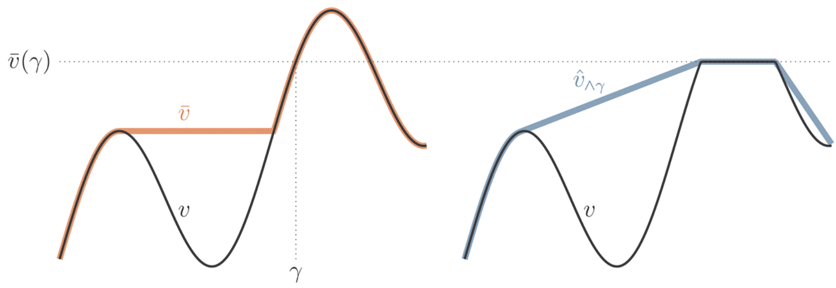

FIG. 5.—Construction of concavification of value function capped at some γ.

Figure 5 illustrates the construction of $\hat{v}_{\wedge\gamma}$. The first step in the construction is to find $\bar{v}(\gamma)$, the value of the quasiconcavification of v at an arbitrary γ. Using this value, one then caps the sender's value function so that no belief results in a payoff higher than $\bar{v}(\gamma)$. The result is the function $v_{\wedge\gamma}(\cdot) = \min\{v(\cdot), \bar{v}(\gamma)\}$, which is the same function one obtains by mapping every belief μ to the maximal value in $V_{\wedge\bar{v}(\gamma)}$. Concavifying this function delivers $\hat{v}_{\wedge\gamma}$.

Collecting the above observations allows us to bound the sender's payoff from a fixed χ-equilibrium outcome distribution $\mathbf{P}$,

$$\int s\,\mathrm{d}\mathbf{P}(\mu, s) = k\int s\,\mathrm{d}\mathbf{B}(\mu, s) + (1-k)\int s\,\mathrm{d}\mathbf{G}(\mu, s) \\ \leq k\hat{v}_{\wedge\gamma}(\beta) + (1-k)\bar{v}(\gamma).$$

Of course, the above bound holds only for $\mathbf{P}$, the χ-equilibrium outcome distribution we started from. To attain an upper bound across all χ-equilibria, we maximize the right-hand side of the above equation over all (β, γ, k) satisfying two restrictions necessary for a χ-equilibrium outcome distribution. For the first restriction, recall that $\mathbf{P}$ must satisfy the Bayesian updating constraint (Bayes), and so

$$\mu_0 = \int \mu\,\mathrm{d}\mathbf{P}(\mu, s) = k\int \mu\,\mathrm{d}\mathbf{B}(\mu, s) + (1-k)\int \mu\,\mathrm{d}\mathbf{G}(\mu, s).$$

Because $\int \mu\,\mathrm{d}\mathbf{B}(\mu, s) = \beta$ and $\int \mu\,\mathrm{d}\mathbf{G}(\mu, s) = \gamma$, it follows that (β, γ, k) must satisfy the Bayesian splitting constraint

$$k\beta + (1-k)\gamma = \mu_0. \tag{BS}$$

For the second restriction, observe that an influencing sender only sends messages whose induced outcome results in a sender payoff of $s_{\max}$. Indeed, she never attains a higher payoff, since no on-path message leads to a payoff above $s_{\max}$, and she cannot find sending a message yielding a lower payoff optimal, because then she would prefer to deviate to a message generating a payoff of $s_{\max}$. Hence, for each state θ, the probability the state is θ and the sender obtains a payoff of $s_{\max}$ is at least the probability the state is θ and reporting is influenced—that is, $(1-\chi)\mu_0(\theta)$. Expressing this inequality directly in terms of $\mathbf{P}$ and using the definitions of k and $\mathbf{G}$ gives

$$(1-\chi)\mu_0(\theta) \le \int_{\{(\mu,s):s=s_{\max}\}} \mu(\theta)\,\mathrm{d}\mathbf{P}(\mu,s) = (1-k)\int \mu(\theta)\,\mathrm{d}\mathbf{G}(\mu,s).$$

Recalling that $\int \mu\,\mathrm{d}\mathbf{G}(\mu,s) = \gamma$ delivers that (β,γ,k) must satisfy the credibility constraint

$$(1-k)\gamma(\theta) \ge (1-\chi)\mu_0(\theta)\ \forall\theta\in\Theta. \tag{χC}$$

Thus, we have obtained the following upper bound on the sender's maximal χ-equilibrium value:

$$v_\chi^*(\mu_0) := \max_{\beta,\gamma\in\Delta\Theta,\ k\in[0,1]} \{k\hat{v}_{\wedge\gamma}(\beta) + (1-k)\bar{v}(\gamma)\} \quad \text{subject to (BS) and } (\chi\text{C}). \tag{$*$}$$

Our main theorem shows that this bound is also tight when χ is intermediate.

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

Our proof uses a (β,γ,k) that solves the program $(*)$ to construct a χ-equilibrium yielding the sender a value of $v_\chi^*(\mu_0)$. Intuitively, one pastes together a sender-optimal equilibrium of a cheap talk game with prior γ and a Bayesian persuasion solution with prior β. We give an informal description of this construction in appendix A and a formal proof in appendix B.

We now apply the theorem to the introduction's central bank example. To solve the program for $v_\chi^*(\mu_0)$, first note that setting $(\beta,\gamma,k) = (\mu_0,\mu_0,0)$ is always feasible, and hence $v_\chi^*(\mu_0) \ge \bar{v}(\mu_0) = 1$. But what form must a solution (β,γ,k) take if $v_\chi^*(\mu_0) > 1$? First, because the objective is bounded above by $\bar{v}(\gamma)$, it must be that $\bar{v}(\gamma) > 1$. Equivalently, $\gamma \ge 3/4$. Constraint (BS) then requires $\beta \le 1/2$ and further gives us an exact formula for k in terms of (β,γ):

$$k = k_{\beta,\gamma} := \frac{\gamma-\mu_0}{\gamma-\beta}.$$

In what follows, we treat the program as an optimization over (β,γ), taking for granted that k will be set to $k_{\beta,\gamma}$.

Observe that we can (still under the hypothesis that $v_\chi^*(\mu_0) > \bar{v}(\mu_0)$) take $\gamma = 3/4$. Indeed, moving $\gamma\in[3/4,1]$ closer to the prior—hence, lowering k to preserve (BS)—always preserves (χC).[6] Meanwhile, because $\hat{v}_{\wedge\gamma}(\beta) \le \bar{v}(\gamma)$ by definition, such a modification raises the program's objective if the modification does not alter the value of $\bar{v}(\gamma)$. Therefore, because $\bar{v}$ is constant on [3/4, 1], any solution (β, γ, k) such that $\gamma \ge 3/4$ can be replaced with one that has $\gamma = 3/4$.

Thus, we have argued that the program (*) always admits a solution of the form $(\beta, 3/4, k_{\beta,3/4})$ for $\beta \in [0, 1/2]$. Restricted to solutions of this form, the program (*) reduces to a univariate constrained maximization program, which can be solved in three exhaustive cases. If $\chi \ge 3/4$, the triplet (1/4, 3/4, 1/2) is a feasible (β, γ, k) that delivers the sender her full commitment value of $v_\chi^*(\mu_0) = 3/2$, meaning that said triplet is optimal. If $2/3 \le \chi < 3/4$, it is optimal to set β equal to

$$\beta_\chi^* := \frac{3\chi - 2}{4\chi - 2},$$

which is the highest β for which $k_{\beta,3/4}$ and $\gamma = 3/4$ satisfy the constraint (χC). The sender's utility in this case is $v_\chi^*(\mu_0) = 2\chi$. Finally, if $\chi < 2/3$, no $\beta \in [0, 1/2)$ can satisfy the constraints required to support $\gamma = 3/4$, and so we cannot improve upon feasible solution $(\beta, \gamma, k) = (1/2, 1/2, 0)$, which yields value $v_\chi^*(\mu_0) = 1$; that is, the sender can do no better than a babbling equilibrium. To summarize, the sender's maximal equilibrium payoff is given by

$$v_\chi^*(\mu_0) = \begin{cases} 1 & \text{if } \chi < 2/3, \\ 2\chi & \text{if } \chi \in [2/3, 3/4], \\ 3/2 & \text{if } \chi \ge 3/4. \end{cases}$$

Figure 6 illustrates the calculation of this value for some $\chi \in (2/3, 3/4)$.

The way the sender obtains the above value—following the construction described after the proof of theorem 1—depends on χ. When $\chi < 2/3$, it is best for the sender to leave the receiver uninformed. When $\chi = 1$, the sender is best commissioning the report described in the introduction, ξ_1^*. To obtain her full-credibility payoff when $\chi \in [3/4, 1)$, the sender commissions a report that induces the same information about θ in equilibrium, but the official report is itself more informative than ξ_1 to compensate for the fact that an influencing sender always sends the high message. When $\chi \in (2/3, 3/4)$, the sender commissions a report that sends three different messages. The low and medium messages, which induce posterior beliefs 0 and 1/4, respectively, are only ever sent under

[6] In the presence of (BS), the constraint (χC) is equivalent to requiring $k\beta(\theta) \le \chi\mu_0(\theta)$ for every state θ, a constraint that relaxes as k decreases.

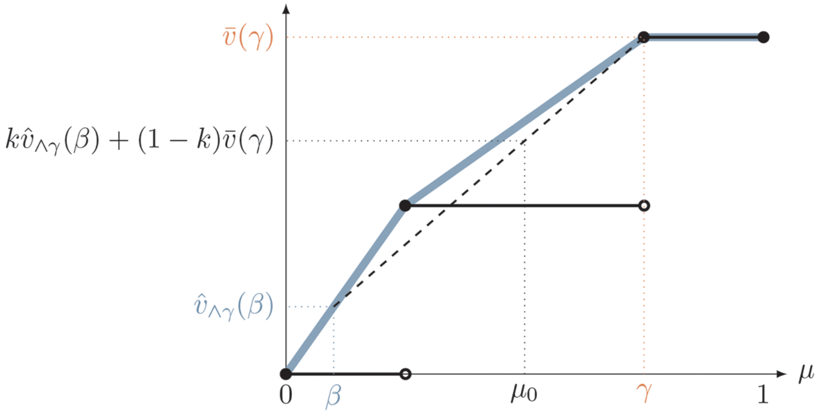

FIG. 6.—Calculating sender value for feasible β and γ in central bank example.

official reporting. The high message would induce a belief strictly higher than 3/4 if it were known to come from official reporting, but when taking into account that influenced reporting sends this message in either state, its induced receiver belief is exactly 3/4. Finally, the case of $\chi = 3/4$ is a limiting version of the latter case in which the medium message is never sent; in this case, the official report is fully informative.

IV. Varying Credibility

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

A. *Productive Mistrust*

We now study how a decrease in the sender's credibility affects the receiver's value and the informativeness of the sender's equilibrium communication. In general, the less credible the sender, the smaller the set of equilibrium outcome distributions.[7] However, that the set of outcome distributions shrinks does not mean that less information is transmitted in the sender's preferred equilibrium. Our introductory example is a

[7] Given credibility levels $\chi' < \chi$ and a χ'-equilibrium $(\xi, \sigma, \alpha, \pi)$, one can construct a χ-equilibrium that generates the same outcome distribution, e.g., $((\chi'/\chi)\xi + [1 - (\chi'/\chi)]\sigma, \sigma, \alpha, \pi)$.

case in point, showing that lowering the sender's credibility can result in a more informative equilibrium (à la Blackwell 1953). Moreover, in that example, the receiver uses this additional information, obtaining a strictly higher value when the sender'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 the sender's optimal outcome distribution under full credibility. For a state θ, let $\delta_\theta \in \Delta\Theta$ be the degenerate belief that generates θ with probability 1. Given prior μ, an outcome distribution $\mathbf{P} \in \mathrm{BP}(\mu, V)$ is a *show-or-best* (SOB) outcome distribution if every supported receiver belief lies in

$$\{\delta_\theta\}_{\theta\in\Theta} \cup \operatorname*{argmax}_{\mu'\in\Delta[\mathrm{supp}(\mu)]} v(\mu').$$

In words, $\mathbf{P}$ is an SOB distribution if it either reveals the state to the receiver or brings the receiver to a posterior belief that attains the sender's best feasible value. Say the sender is a *two-faced SOB* if for every binary support prior $\mu \in \Delta\Theta$, every $\mathbf{P} \in \mathrm{BP}(\mu, V)$ is outperformed by an SOB distribution $\mathbf{P}' \in \mathrm{BP}(\mu, V)$; that is, $\int s\,\mathrm{d}\mathbf{P}(\mu', s) \leq \int s\,\mathrm{d}\mathbf{P}'(\mu', s)$. Figure 7 depicts an example in which the sender is a two-faced SOB. Note that productive mistrust cannot occur in this example: one can show that if the sender's favorite equilibrium outcome distribution changes as credibility

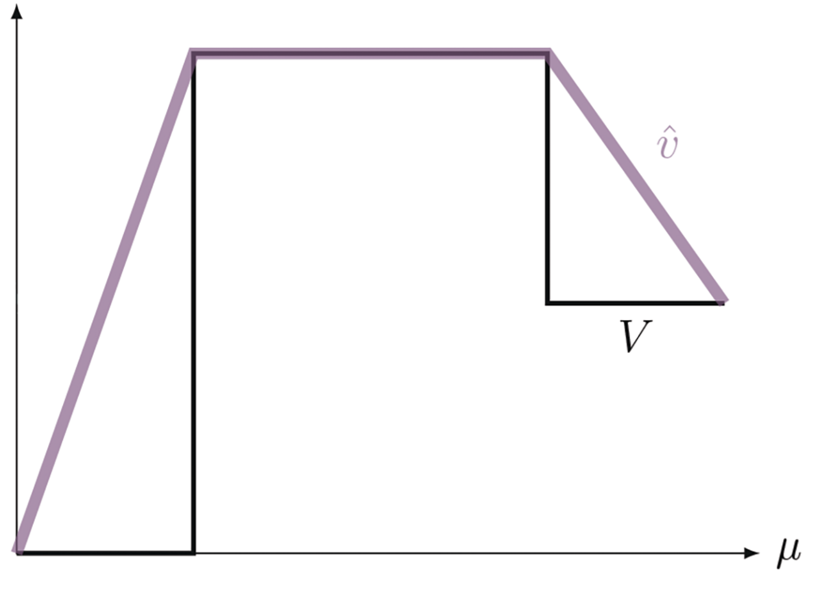

FIG. 7.—Sender is a two-faced SOB.

declines, no information must become sender optimal.[8] As such, the receiver need not benefit from a less credible sender.

Finally, say a model is *generic* if the receiver is (1) not indifferent between any two actions at any degenerate belief and (2) not indifferent between any three actions at any binary support belief.[9]

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

PROPOSITION 1. Consider a generic model in which the sender is not a two-faced SOB. Then, a full-support prior and credibility levels $\chi' < \chi$ exist such that every sender-optimal χ'-equilibrium is strictly better for the receiver than every sender-optimal χ-equilibrium.[10]

The proposition builds on the binary state case, extending to the general case via a continuity argument. We now sketch the binary state argument. To follow the argument, consulting figure 8, which depicts the relevant objects for the central bank example, is useful. Because the model is generic, $\bar{v}$ has a nondegenerate interval of maximizers (which correspond to beliefs in [3/4, 1] in fig. 8). Fixing a prior near this interval but toward the nearest kink, we then find the lowest $\chi \in [0, 1]$ at which the sender still obtains her full-credibility value. In the central bank example, one can use any prior in (1/4, 3/4). If we choose $\mu_0 = 1/2$, we

[8] For an explanation, observe that the claim is obvious for priors that allow the sender to attain her first-best under no information. For other priors, a feasible (β, γ, k) exists that improves on the sender's no-information payoff if and only if a feasible (β, γ, k) exists that gives the sender her full-credibility payoff.

[9] 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 $(u_R(a_1, \theta_1) - u_R(a_2, \theta_1))/(u_R(a_1, \theta_2) - u_R(a_2, \theta_2)) \neq (u_R(a_2, \theta_1) - u_R(a_3, \theta_1))/(u_R(a_2, \theta_2) - u_R(a_3, \theta_2))$ for all distinct $a_1, a_2, a_3 \in A$ and all distinct $\theta_1, \theta_2 \in \Theta$.

[10] Two additional remarks are in order. First, when $|\Theta| = 2$, every sender-optimal χ'-equilibrium is more Blackwell informative than every sender-optimal χ-equilibrium.

Second, with more than two states, one can also find payoff environments in which every sender-optimal 0-equilibrium is strictly better for the receiver than every sender-optimal 1-equilibrium.

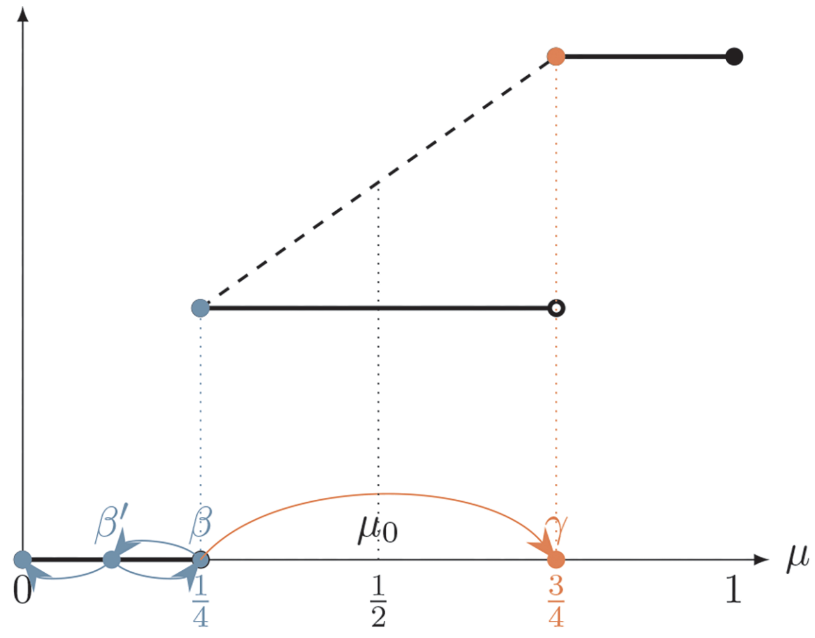

FIG. 8.—Productive mistrust in central bank example.

take χ to be 3/4, which is the lowest credibility level that delivers the sender's full-commitment payoff. At this χ, the sender's favorite equilibrium outcome distribution **P** is unique, generating the outcome $(\gamma, \bar{v}(\gamma))$ with probability $(1 - k)$ and the outcome $(\beta, \bar{v}_{\wedge\gamma}(\beta))$ with probability k, where (β, γ, k) is a solution to theorem 1's program (see $\gamma = 3/4$ and $\beta = 1/4$ in fig. 8). The beliefs γ and β 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, (χC) means that one must replace β with a new belief β' (β^*_χ in the central bank example) that is further from the prior. Relying on the set of beliefs being one-dimensional, we show that this new solution results in an outcome distribution $\mathbf{P}'$ whose marginal distribution p' over the receiver's posterior belief (so $p' \in \Delta\Delta\Theta$) is strictly more informative than the corresponding marginal p for **P**. Intuitively, one can attain p' from p using two consecutive splittings, each of which involves an increase in informativeness: First, β is split between γ and β', and then β' is split between β and another posterior (0 in fig. 8). This posterior lies even further from the prior than β' does and gives the sender a strictly lower continuation value than β. Hence, the additional information p' provides to the receiver over p is instrumental, strictly increasing the receiver's utility.

B. Collapse of Trust

Theorem 1 immediately implies that lowering the sender's credibility can only decrease her value.[11] Below, we show that this decrease is discontinuous for many payoff specifications of our model. In other words, small decreases in the sender's credibility can result in a large drop in the sender's benefits from communication.

PROPOSITION 2. The following are equivalent:

i. A collapse of trust never occurs:

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

for every $\chi \in [0, 1]$ and every full-support prior μ_0.

ii. Commitment is of no value: $v^*_1 = v^*_0$.

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

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

To show that point i implies point iii, we show that any failure of point iii implies the failure of point 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 a 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 that such a triplet is feasible for credibility χ if and only if χ is weakly greater than some strictly positive χ^*. We thus have

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

where the first inequality follows from μ_0 minimizing $\bar{v}$; that is, a collapse of trust occurs.

[11] In app. sec. B.1.4, we show that credibility increases have a continuous payoff effect: a sufficiently small increase in the sender's credibility never results in a large gain in the sender's benefits from communication. Thus, the sender's value is an upper-semicontinuous function of χ. Proposition 2 implies that lower semicontinuity is frequently violated.

C. *Robustness of the Commitment Case*

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

Proposition 3. The following are equivalent:

i. The full-commitment value is robust: $\lim_{\chi \nearrow 1} v_\chi^*(\mu_0) = v_1^*(\mu_0)$ for every full-support μ_0.
ii. The sender receives the benefit of the doubt: every $\theta \in \Theta$ is in the support of some member of $\operatorname{argmax}_{\mu \in \Delta\Theta} v(\mu)$.

Thus, the proposition shows that the sender's full-credibility value is robust if and only if the sender can persuade the receiver to take her favorite action without ruling out any states. A sufficient condition for the latter is that the receiver is willing to take the sender's preferred undominated action at some full-support belief, a property that holds generically.[12] Hence, although small decreases in credibility often lead to a collapse in the sender's value, these collapses rarely occur at $\chi = 1$.

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

a. the sender 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 priors, and
d. robustness to limited credibility.

To see that point a implies point b, notice that whenever the sender receives the benefit of the doubt, one can find a full-support prior in the convex hull of the beliefs in which the receiver is willing to give the sender her first-best action. Splitting this prior across those beliefs gives an outcome distribution in $\mathrm{BP}(\mu_0, V)$ that delivers the sender her highest feasible payoff for every supported outcome, meaning the sender can attain this payoff using cheap talk. For the converse direction, one can use the fact that $\max \bar{v}(\Delta\Theta) = \max v(\Delta\Theta)$. Specifically, this fact implies $\bar{v}$ is maximized at a full-support prior γ if and only if one can split γ in a way

[12] More precisely, proposition 3 implies that the sender's full-credibility value is robust whenever a sender-best action among those not strictly dominated for the receiver is a best reply for some full-support belief. It follows from lemma 1 in Lipnowski, Ravid, and Shishkin (2022) that this property holds for Lebesgue-almost every preference specification.

that attains v's maximal value at all posteriors, because $\bar{v}$ gives the sender's highest cheap-talk payoff for every prior. The sender receiving the benefit of the doubt then follows from γ having full support.

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

To see why point c is equivalent to point d, fix some full-support μ_0 and consider two questions about theorem 1's program. First, which beliefs can serve as γ for $\chi < 1$ large enough? Second, how do the optimal (β, k) for a given γ change as χ goes to 1? For the first question, the answer is that γ is feasible for some $\chi < 1$ if and only if γ has full support.[13] For the second question, one can show that it is always optimal to choose (β, k) so as to make (χC) bind while still satisfying (BS).[14] Direct computation reveals that as χ goes to 1, every such (β, k) must converge to $(\mu_0, 1)$. Combined, one obtains that as χ increases, the sender's optimal value converges to $\max_{\gamma\in\text{int}(\Delta\Theta)} \hat{v}_{\wedge\gamma}(\mu_0)$. Thus, the sender'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, point c is equivalent to point d. The proposition follows.

V. Conclusion

This paper studies a model of persuasion through a weak institution whose messages are compromised. Our model has certain features that are worth further discussion.

Throughout the paper, we assumed that the sender's credibility is independent of the state of the world. However, in many scenarios, it is natural for the sender's credibility to be correlated with the state. For example, an autocrat may be more likely to influence the media in a rich economy with abundant resources than in a country where resources are scarce (e.g., Egorov, Guriev, and Sonin 2009). One can capture such correlation by supposing that when the state is θ, the message is drawn

[13] It is easy to see that every full-support γ admits some β and $k < 1$ that make (BS) hold. Moreover, (χC) is also satisfied at (β, γ, k) for all sufficiently high χ, because (χC)'s right-hand side converges to zero as $\chi \to 1$. Conversely, observe that if $\gamma(\theta) = 0$, (χC) is violated at θ for all $\chi < 1$, because μ_0 has full support.

[14] To see why, for any feasible (β, γ, k), a (β', k') exists such that (β', γ, k') is feasible, (χC) binds, and $k' \geq k$. By (BS), $\beta' = (k/k')\beta + (1 - k/k')\gamma$. Because $\hat{v}_{\wedge\gamma}$ is concave and $\hat{v}_{\wedge\gamma}(\gamma) = \bar{v}(\gamma)$,

$$k'\hat{v}_{\wedge\gamma}(\beta') + (1-k')\bar{v}(\gamma) = k'\hat{v}_{\wedge\gamma}\left(\frac{k}{k'}\beta + \left(1 - \frac{k}{k'}\right)\gamma\right) + (1-k')\bar{v}(\gamma)$$

$$\geq k\hat{v}_{\wedge\gamma}(\beta) + (k'-k)\hat{v}_{\wedge\gamma}(\gamma) + (1-k')\hat{v}(\gamma) = k\hat{v}_{\wedge\gamma}(\beta) + (1-k)\hat{v}(\gamma).$$

from the sender's official report with probability $\chi(\theta)$. Theorem 1 generalizes to this case with a minor modification. For a bounded and measurable $f:\Theta\to\mathbb{R}$ and $\mu\in\Delta\Theta$, let $f\mu$ denote the measure on Θ given by $f\mu(\hat{\Theta}):=\int_{\hat{\Theta}} f\,\mathrm{d}\mu$. Then, appendix B shows that some sender-favorite equilibrium exists, and the sender's value in this equilibrium is given by

$$v_\chi^*(\mu_0) = \max_{\beta,\gamma\in\Delta\Theta,\ k\in[0,1]} k\hat{v}_{\wedge\gamma}(\beta) + (1-k)\bar{v}(\gamma) \tag{2}$$

$$\text{subject to } k\beta + (1-k)\gamma = \mu_0,$$

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

With the above characterization in hand, the propositions of section IV extend to the state-dependent credibility model in a straightforward manner; see the appendix for precise statements.

We also assumed that the sender announces her official report before knowing whether the announcement is credible. In practice, the sender may be privy to institutional features that affect her chances of influencing the report before she commissions it. To understand such situations, appendix C considers a modified model in which the sender learns her credibility type before announcing the official reporting protocol. We show that this modification has no impact on the sender's equilibrium payoffs, and so the sender's maximal equilibrium value remains unchanged.

Finally, we formulated our model as having a finite number of actions and states. However, many applications admit infinite states, infinite actions, or both (e.g., Gentzkow and Kamenica 2016; Kolotilin et al. 2017; Dworczak and Martini 2019). To accommodate such applications, the appendix considers a more general model in which both the action and the state space are compact metrizable. As we show there, our characterization of sender-optimal equilibrium payoffs generalizes to this case in a straightforward manner.

References

Alonso, Ricardo, and Odilon Câmara. 2018. "On the Value of Persuasion by Experts." *J. Econ. Theory* 174:103–23.

Aumann, Robert J., and Sergiu Hart. 2003. "Long Cheap Talk." *Econometrica* 71 (6): 1619–60.

Aumann, Robert J., and Michael Maschler. 1995. *Repeated Games with Incomplete Information.* Cambridge, MA: MIT Press.

Ben-Porath, Elchanan, Eddie Dekel, and Barton L. Lipman. 2019. "Mechanisms with Evidence: Commitment and Robustness." *Econometrica* 87 (2): 529–66.

Best, James, and Daniel Quigley. 2022. "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): 1077–98.

Blackwell, David. 1953. “Equivalent Comparisons of Experiments.” *Ann. Math. Statis.* 24 (2): 265–72.

Blume, Andreas, Oliver J. Board, and Kohei Kawamura. 2007. “Noisy Talk.” *Theoretical Econ.* 2 (4): 395–440.

Chakraborty, Archishman, and Rick Harbaugh. 2010. “Persuasion by Cheap Talk.” *A.E.R.* 100 (5): 2361–82.

Crawford, Vincent P., and Joel Sobel. 1982. “Strategic Information Transmission.” *Econometrica* 50 (6): 1431–51.

Dworczak, Piotr, and Giorgio Martini. 2019. “The Simple Economics of Optimal Persuasion.” *J.P.E.* 127 (5): 1993–2048.

Egorov, Georgy, Sergei Guriev, and Konstantin Sonin. 2009. “Why Resource-Poor Dictators Allow Freer Media: A Theory and Evidence from Panel Data.” *American Polit. Sci. Rev.* 103 (4): 645–68.

Forges, Françoise. 2020. “Games with Incomplete Information: From Repetition to Cheap Talk and Persuasion.” *Ann. Econ. and Statis.* (137): 3–30.

Fréchette, Guillaume, Alessandro Lizzeri, and Jacopo Perego. 2022. “Rules and Commitment in Communication: An Experimental Analysis.” *Econometrica*, forthcoming.

Gentzkow, Matthew, and Emir Kamenica. 2016. “A Rothschild-Stiglitz Approach to Bayesian Persuasion.” *A.E.R.* 106 (5): 597–601.

Glazer, Jacob, and Ariel Rubinstein. 2006. “A Study in the Pragmatics of Persuasion: A Game Theoretical Approach.” *Theoretical Econ.* 4 (1): 395–410.

Goltsman, Maria, Johannes Hörner, Gregory Pavlov, and Francesco Squintani. 2009. “Mediation, Arbitration and Negotiation.” *J. Econ. Theory* 144 (4): 1397–420.

Green, Jerry R., and Nancy L. Stokey. 2007. “A Two-Person Game of Information Transmission.” *J. Econ. Theory* 135 (1): 90–104.

Guo, Yingni, and Eran Shmaya. 2021. “Costly Miscalibration.” *Theoretical Econ.* 16 (2): 477–506.

Hart, Sergiu, Ilan Kremer, and Motty Perry. 2017. “Evidence Games: Truth and Commitment.” *A.E.R.* 107 (3): 690–713.

Hedlund, Jonas. 2017. “Bayesian Persuasion by a Privately Informed Sender.” *J. Econ. Theory* 167:229–68.

Ichihashi, Shota. 2019. “Limiting Sender’s Information in Bayesian Persuasion.” *Games and Econ. Behavior* 117:276–88.

Ivanov, Maxim. 2010. “Informational Control and Organizational Design.” *J. Econ. Theory* 145 (2): 721–51.

Kamenica, Emir. 2019. “Bayesian Persuasion and Information Design.” *Annual Rev. Econ.* 11 (1): 249–72.

Kamenica, Emir, and Matthew Gentzkow. 2011. “Bayesian Persuasion.” *A.E.R.* 101 (6): 2590–615.

Kartik, Navin. 2009. “Strategic Communication with Lying Costs.” *Rev. Econ. Studies* 76 (4): 1359–95.

Kolotilin, Anton, Tymofiy Mylovanov, Andriy Zapechelnyuk, and Ming Li. 2017. “Persuasion of a Privately Informed Receiver.” *Econometrica* 85 (6): 1949–64.

Lipnowski, Elliot, and Doron Ravid. 2020. “Cheap Talk with Transparent Motives.” *Econometrica* 88 (4): 1631–60.

Lipnowski, Elliot, Doron Ravid, and Denis Shishkin. 2022. “Perfect Bayesian Persuasion.” Working paper.

Luo, Zhaotian, and Arturas Rozenas. 2018. "Strategies of Election Rigging: Trade-Offs, Determinants, and Consequences." *Q. J. Polit. Sci.* 13 (1): 1–28.

Mathevet, Laurent, David Pearce, and Ennio Stacchetti. 2022. "Reputation for a Degree of Honesty." Working paper.

Min, Daehong. 2021. "Bayesian Persuasion under Partial Commitment." *Econ. Theory* 72:743–64.

Nguyen, Anh, and Teck Yong Tan. 2021. "Bayesian Persuasion with Costly Messages." *J. Econ. Theory* 193:105212.

Perez-Richet, Eduardo. 2014. "Interim Bayesian Persuasion: First Steps." *A.E.R.* 104 (5): 469–74.

Perez-Richet, Eduardo, and Vasiliki Skreta. 2022. "Test Design under Falsification." *Econometrica* 90 (3): 1109–42.

Salamanca, Andrés. 2021. "The Value of Mediated Communication." *J. Econ. Theory* 192:105191.

Sher, Itai. 2011. "Credibility and Determinism in a Game of Persuasion." *Games and Econ. Behavior* 71 (2): 409–19.

Skreta, Vasiliki. 2006. "Sequentially Optimal Mechanisms." *Rev. Econ. Studies* 73 (4): 1085–111.

Online Appendix

A Constructing an S-optimal Equilibrium

In this appendix, we informally explain how to use a (β, γ, k) that solves the program $(*)$ to construct a χ-equilibrium yielding S a value of $v^*_\chi(\mu_0)$. As a first step, let $(\sigma_\gamma, \alpha_\gamma, \pi_\gamma)$ denote an equilibrium of the cheap-talk game with modified prior γ that generates S payoff $\bar{v}(\gamma)$; some such equilibrium exists as we outlined in discussing the no-credibility case. If $k = 0$ (implying $\gamma = \mu_0$ by (χC)), then $(\xi, \sigma, \alpha, \pi) = (\sigma_\gamma, \sigma_\gamma, \alpha_\gamma, \pi_\gamma)$ is a χ-equilibrium delivering the desired S payoff.

Given the above observation, we can focus on the case in which every solution (β, γ, k) to the program has $k > 0$—or, equivalently, that $v^*_\chi(\mu_0) > \bar{v}(\mu_0)$. Let $\mathbf{B} \in \text{BP}(\beta, V_{\wedge \bar{v}(\gamma)})$ be such that $\int (\mu, s) \, \mathrm{d}\mathbf{B}(\mu, s) = (\beta, \hat{v}_{\wedge\gamma}(\beta))$. Lemma 3 in Appendix B.1.2 uses the geometry of concavification and quasiconcavification to prove $\mathbf{B}$ is supported only on outcomes in V that are left untouched by moving from V to $V_{\wedge \bar{v}(\gamma)}$. It follows $\mathbf{B}$ is in $\text{BP}(\beta, V)$, and so one can use the results from the full-credibility case to obtain some triple $(\xi_\beta, \alpha_\beta, \pi_\beta)$ that—when the prior is β and credibility level is $\chi = 1$—induces the outcome distribution $\mathbf{B}$, is consistent with Bayesian updating, and satisfies R's incentive constraints. Moreover, because the message space is rich, we may assume without loss that the messages M_β used by ξ_β have no overlap with the messages M_γ used by σ_γ.

Now, let us describe how the the above objects can be "pasted" together to deliver a χ-equilibrium with the relevant S payoff. Because constraint (BS) is satisfied, a binary signal can be used to "split" the prior into beliefs γ and β: concretely, some $\lambda : \Theta \to [0, 1]$ exists such that if message "high" and "low" are respectively sent with probability $1 - \lambda(\theta)$ and $\lambda(\theta)$ in state θ, the would-be posterior distribution from hearing message "high" is γ and from "low" is β. Further, constraint (χC) implies $\lambda(\theta) \leq \chi$ for every state θ. We can therefore construct a χ-equilibrium as follows. The influencing S strategy σ is σ_γ; the official reporting protocol is given by

$$\xi(\theta) \coloneqq \xi^*(\theta)\xi_\beta(\theta) + [1 - \xi^*(\theta)]\sigma_\gamma(\theta)$$

, where

$$\xi^*(\theta) \coloneqq 1 - \lambda(\theta)/\chi \in [0, 1];$$

and the R strategy α and belief map π agree with $(\alpha_\beta, \pi_\beta)$ for messages in M_β and $(\alpha_\gamma, \pi_\gamma)$ for messages in M_γ. In the appendix, we show $(\xi, \sigma, \alpha, \pi)$ inherits the Bayesian and R incentive properties from its constituent pieces and generates an S payoff of $v^*_\chi(\mu_0)$. Moreover, because S is indifferent between all messages in M_γ and receives payoffs from $V_{\wedge \bar{v}(\gamma)}$ (hence, below $\bar{v}(\gamma)$) from messages in M_β, S's incentive constraints are also satisfied. Hence, we have found a χ-equilibrium generating S payoff $v^*_\chi(\mu_0)$, delivering the theorem.

A byproduct of the theorem's construction is the following result, which bounds the number of on-path messages required for an S-optimal equilibrium.

Corollary 1. *Some S-optimal χ-equilibrium exists with no more than* $\min\{|A|, 2|\Theta|-1\}$ *distinct messages sent on path.*

Existing literature has already established the above bounds hold when credibility is extreme. Specifically, Kamenica and Gentzkow (2011) and Lipnowski and Ravid (2020) note that when $\chi \in \{0, 1\}$, an S-optimal χ-equilibrium exists that uses only $\min\{|A|, |\Theta|\}$ messages. Applying these bounds separately to **G** and **B** delivers that no χ-equilibrium S-value requires more than twice as many messages, that is, $\min\{2|A|, 2|\Theta|\}$. The corollary shows one can tighten these bounds by utilizing Theorem 1's construction. See Appendix B.1.3 for more details.

B Main Results

B.1 Toward the Proof of Theorem 1

Throughout this subsection, we work with a more general setting of the model in which both Θ and A are compact metrizable spaces with at least two elements, and the objectives u_R and u_S are continuous.[14] Finally, we assume M is an uncountable compact metrizable space.[15] To generalize the definition of a χ-equilibrium and the value correspondence V, the sums are replaced with the corresponding integrals with respect to measures $\pi(m)$, $\alpha(m)$, and μ over Θ, A, and Θ, respectively. Further, throughout the

[14]We view any compact metrizable space Y as a measurable space with its Borel field; let ΔY denote the set of all probability measures on Y; and endow ΔY with its weak* topology, so that ΔY is itself a compact metrizable space.

[15]In the special case in which A and $|\Theta|$ are finite, our characterization of sender-optimal equilibrium values (Theorem 1) and the propositions of section 4 hold if $|M| \geq \min\{|A|, 2|\Theta|-1\}$; see Corollary 1.

appendix, we modify the definitions of the value function's concavification $\hat{v}$ (resp. quasiconcavification $\bar{v}$), letting it be the lowest (quasi)concave *and upper semicontinuous* function that dominates v.[16]

In addition, we allow for the possibility that credibility is state dependent, given by some measurable function $\boldsymbol{\chi} : \Theta \to [0, 1]$. Throughout this appendix, we adopt the following notational convention. For a compact metrizable space Y, a probability measure $\mu \in \Delta Y$, and a function $f\colon Y \to \mathbb{R}$ that is bounded and measurable, let $f(\mu) \coloneqq \int_Y f \,\mathrm{d}\mu \in \mathbb{R}$ denote the average value of f. In particular, for any credibility function $\boldsymbol{\chi}$, the scalar $\boldsymbol{\chi}(\mu_0)$ is simply the total probability that the report is not subject to influence.

Although accommodating this more general model entails some notational cost, all conceptual content of the proof is identical in the special case of constant credibility, and so the generalization requires no additional arguments. We therefore encourage the reader to read the entire proof while keeping in mind with the special case in which the function $\boldsymbol{\chi}$ is a constant χ.

We now provide a brief overview of the proof. Formalizing a form of *equilibrium summary* that is sufficient to calculate players' payoffs, the proof begins by showing an equivalence between the set of $\boldsymbol{\chi}$-equilibrium summaries, the set of $\boldsymbol{\chi}$-nonical equilibrium summaries, and the existence of a particular decomposition of the equilibrium distribution of R beliefs. This decomposition makes it easy to see program (2) is a relaxation of the program that maximizes S's value across all $\boldsymbol{\chi}$-equilibrium summaries. In particular, program (2) enables S to induce posteriors that would generate too high a continuation payoff for S. The proof's next part establishes this constraint is nonbinding at the optimum. We then conclude by explicitly writing the program that finds S's favorite equilibrium summary and showing its value is identical to that of (2).

B.1.1 Characterization of All Equilibrium Summaries

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

[16] When Θ is finite, it follows from Carathéodory's theorem that the lowest (quasi)concave majorant of v is upper semicontinuous because v is. Hence, the present definition generalizes the one in the main text.

S receives conditional on having the opportunity to influence.

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

We now define a convenient class of equilibria.

Definition 1. *A* ***$\boldsymbol{\chi}$-nonical equilibrium*** *is a $\boldsymbol{\chi}$-equilibrium $(\xi, \sigma, \alpha, \pi)$ such that every Borel $\hat{M} \subseteq M^*_\alpha$ has $\xi(\hat{M}|\cdot) = \xi(M^*_\alpha|\cdot)\ \sigma(\hat{M}|\cdot)$, where $M^*_\alpha := \operatorname{argmax}_{m\in M} u_S(\alpha(m))$.*

The above definition imposes further structure on a $\boldsymbol{\chi}$-equilibrium. The requirement pertains to the set M^*_α of the highest-payoff messages for S, which are necessarily the only messages an influencing S chooses. The condition says the conditional distribution of messages in M^*_α is identical for the official experiment and for an influencing sender's choices, in any state for which the official report sometimes sends messages in M^*_α. Informally, the condition says all differences in how the official and influenced report communicate are through whether they send a message in M^*_α in a given state.

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

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

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

Observe that knowing a $\boldsymbol{\chi}$-equilibrium's summary is sufficient for recovering each player's expected payoff: given a summary (p, s_o, s_i), S earns a payoff of $\boldsymbol{\chi}(\mu_0)s_o + [1 - \boldsymbol{\chi}(\mu_0)]\, s_i$, whereas R's expected utility is $\int_{\Delta\Theta} \max_{a\in A} \int_\Theta u_R(a, \cdot)\, \mathrm{d}\mu\, \mathrm{d}p(\mu)$.

The following lemma adopts a belief-based approach, directly characterizing the range of $\boldsymbol{\chi}$-equilibrium summaries in our game. To state the characterization, let $\mathcal{P}(\mu) \coloneqq \{p \in \Delta\Delta\Theta\colon\ \int \tilde{\mu}\, \mathrm{d}p(\tilde{\mu}) = \mu\}$ denote the set of **information policies** corresponding to prior $\mu \in \Delta\Theta$.

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

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

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

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

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

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

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

(iv) $s_i - s_o \in \frac{k}{\boldsymbol{\chi}(\mu_0)}\left[s_i - \int_{\mathrm{supp}(b)} s_i \wedge V \, \mathrm{d}b\right]$.[17]

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

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

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

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

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

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

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

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

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

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

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

where the inequality follows from the Bayesian property of π, together with the fact that σ almost surely sends a message from M^*_α on the path of play. Next, observe that for any $m \in M$, sender incentive compatibility tells us $u_S(\alpha(m)) \leq s_i$, and receiver incentive compatibility implies $\alpha(m) \in V(\pi(m))$. It follows directly that $g\{V \ni s_i\} = b\{\min V \leq s_i\} = 1$; that is, the third condition holds. Toward the fourth and final condition, let us view π, α as random variables on the probability space $\langle M, P\rangle$. Defining the conditional expectation $\phi_0 \coloneqq \mathbb{E}_B[u_S(\alpha)|\pi] : M \to \mathbb{R}$, the Doob-Dynkin lemma delivers a measurable function $\phi : \Delta\Theta \to \mathbb{R}$ such that $\phi \circ \pi =_{B-\text{a.e.}} \phi_0$. Because $u_S(\alpha(m)) \in s_i \wedge V(m)$ 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

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

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

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

as required.

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

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

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

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

Henceforth, we focus on the remaining case in which $s_o < s_i$. Without loss of generality, we may further assume $b\{\phi < s_i\} = 1$.[18] Define $\beta := \int_{\Delta\Theta} \mu \, \mathrm{d}b(\mu)$ and $\gamma := \int_{\Delta\Theta} \mu \, \mathrm{d}g(\mu)$. Let measurable $\eta_g : \Theta \to \Delta[\text{supp}(g)] \subseteq \Delta\Delta\Theta$ and $\eta_b : \Theta \to \Delta[\text{supp}(b)] \subseteq \Delta\Delta\Theta$ be signals that induce belief distribution g for prior γ and belief distribution b for prior β, respectively, such that for each such signal the induced

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

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

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

Take some Radon-Nikodym derivative $\frac{\mathrm{d}\beta}{\mathrm{d}\mu_0} : \Theta \to \mathbb{R}_+$; changing it on a μ_0-null set, we may assume $\mathbf{0} \leq \frac{k}{\boldsymbol{\chi}} \frac{\mathrm{d}\beta}{\mathrm{d}\mu_0} \leq \mathbf{1}$ because $(1-k)\gamma \geq (\mathbf{1} - \boldsymbol{\chi})\mu_0$. With the above ingredients in hand, we can define the sender's influenced strategy and reporting protocol

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

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

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

Finally, by Lipnowski and Ravid (2020, Lemma 2), some measurable $\alpha_b, \alpha_g : \Delta\Theta \to \Delta A$ exist such that[19]

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

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

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

From these selectors, we can define a receiver strategy as

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

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

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

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

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

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

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

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

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

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

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

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

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

as required.

□

B.1.2 Proof of Theorem 1

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

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

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

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

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

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

Lemma 3. *If* (β, γ, k) *solve program* (2) *and have* $\hat{v}_{\wedge\gamma}(\beta) < \bar{v}(\gamma)$, *and* $b \in \mathcal{P}(\beta)$ *has* $\int v_{\wedge\gamma}\,\mathrm{d}b = \hat{v}_{\wedge\gamma}(\beta)$, *then* $v_{\wedge\gamma}(\mu) \in V(\mu)$ *for every* $\mu \in \operatorname{supp}(b)$. *In particular,* $b\{\min V \leq \bar{v}(\gamma)\} = 1$.

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

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

We now prove our main theorem: an S-optimal $\boldsymbol{\chi}$-equilibrium exists, giving S payoff $v^*_{\boldsymbol{\chi}}(\mu_0)$.

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

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

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

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

Given any feasible (b, g, k, s_i) in the latter program, replacing (g, s_i) with any (g^*, s_i^*) such that $\int_{\Delta\Theta}\mu \,\mathrm{d}g^*(\mu) = \int_{\Delta\Theta}\mu \,\mathrm{d}g(\mu)$, $g^*\{V \ni s_i^*\} = 1$, and $s_i^* \geq s_i$ will preserve all constraints and weakly raise the objective. Moreover, Lipnowski and Ravid (2020, Lemma 1 and Theorem 2) tell us that any $\gamma \in \Delta\Theta$ has $\max_{g\in\mathcal{P}(\gamma),s_i\in\mathbb{R}:\ g\{V\ni s_i\}=1} s_i = \bar{v}(\gamma)$.[20] Therefore,

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

Trivially, the program (2) that defines $v^*_{\boldsymbol{\chi}}(\mu_0)$ is a relaxation of the above program; that is, for every feasible (β, γ, k, b) for the above program, (β, γ, k) is feasible in (2) and generates a weakly higher objective there; that is, $\tilde{v}^*_{\boldsymbol{\chi}}(\mu_0) \leq v^*_{\boldsymbol{\chi}}(\mu_0)$. We now prove the opposite inequality also holds, thereby completing the theorem's proof. Notice the program (2) has an upper-semicontinuous objective and compact constraint set, and so admits some solution (β, γ, k). We now argue some $(\tilde{\beta}, \tilde{\gamma}, \tilde{k}, b)$ exists that is feasible for the above program and such that

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

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

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

B.1.3 Simple Communication: Proof of Corollary 1

We begin with a lemma showing program (2) always admits a solution with additional structure. In particular, whenever S-optimal $\boldsymbol{\chi}$-equilibrium requires the official reporting protocol to differ from an influencing S's behavior, we can assume without loss that every message sent by official reporting is *strictly* suboptimal for an influencing S.

Lemma 4. *One of the following holds:*

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

Proof. As observed in (the SDC generalization of) Theorem 1, program (2) admits some solution (β, γ, k). Further, some $b \in \mathcal{P}(\beta)$ exists with $\int v_{\wedge\gamma}\,\mathrm{d}b = \hat{v}_{\wedge\gamma}(\beta)$ because $\mathcal{P}(\beta)$ is compact and $b \mapsto \int v_{\wedge\gamma}\,\mathrm{d}b$ is upper semicontinuous. Letting $D := \{v \geq \bar{v}(\gamma)\} \subseteq \Delta\Theta$, we have nothing to show if $b(D) = 0$, so suppose $b(D) > 0$.

Now, let $k' := k[1 - b(D)] \in [0, 1)$; let $\gamma' := \frac{1}{1-k'}\left[(1-k)\gamma + k\int_D \mu\,\mathrm{d}b(\mu)\right] \in \Delta\Theta$; and let $\beta' := \frac{1}{1-b(D)}\int_{(\Delta\Theta)\setminus D} \mu\,\mathrm{d}b(\mu)$ if $b(D) < 1$, and $\beta' := \mu_0$ if $b(D) = 1$. Because $k'\beta' + (1-k')\gamma' = k\beta + (1-k)\gamma$ and $(1-k')\gamma' \geq (1-k)\gamma$ by construction, (β', γ', k') is feasible in (2). In what follows, we show (β', γ', k') is an optimal solution to (2) with the desired features.

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

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

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

Optimality of (β, γ, k) in (2) then implies (β', γ', k') is optimal too. Therefore, the inequalities in the above chain must hold with equality, from which the first line of the above chain yields $\hat{v}_{\wedge\gamma'}(\beta') = \int v_{\wedge\gamma'} \, \mathrm{d}b'$. Thus, (β', γ', k') and b' are as required. □

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

We now proceed to prove the corollary. Our proof applies to the general model (not assuming A and Θ are finite, and not assuming $\boldsymbol{\chi}$ is state independent). Specifically, we show two things. First, some S-optimal $\boldsymbol{\chi}$-equilibrium entails no more than $|A|$ on-path messages. Second, if Θ is finite, some S-optimal $\boldsymbol{\chi}$-equilibrium entails no more than $2|\Theta| - 1$ on-path messages. The central-bank example, for which every S-optimal equilibrium requires at least three on-path messages when $2/3 < \chi < 3/4$, demonstrates both bounds are tight.

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

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

First, we prove the bound based on the number of actions. Letting $A_+ := \{a \in A : u_S(a) \geq s_i\}$, (the proof of) Proposition 2 from Lipnowski and Ravid (2020) delivers some $g \in \mathcal{P}(\gamma)$ such that $g\{V \ni s_i\} = 1$ and $|\mathrm{supp}(g)| \leq |A_+|$. In case 1, nothing remains to be shown, so we now focus on case 2. Because $b \in \mathcal{P}(\beta)$ is such that $\mathrm{argmax}_{a \in A} \int u_R(a, \cdot) \, \mathrm{d}\mu \subseteq A \backslash A_+$ a.s.-$b(\mu)$, (the proof of) Proposition 1 from Kamenica and Gentzkow (2011) delivers some $b \in \mathcal{P}(\beta)$ such that $|\mathrm{supp}(b)| \leq |A \setminus A_+|$.[21] Hence,

[21] In both of the cited propositions, the result we use is proven in the cited paper, but not written in the proposition's statement. The proof of Proposition 2 from Lipnowski and Ravid (2020) shows any attainable equilibrium S payoff of the cheap-talk game is attainable in an equilibrium in which every on-path message is a pure-action recommendation, and the recommended action is S's preferred action in the support of R's (possibly mixed-action) response to that recommendation. The proof of

some S-optimal $\boldsymbol{\chi}$-equilibrium $(\xi, \sigma, \alpha, \pi)$ exists in which some measurable $M^* \subseteq M$ with $|M^*| \leq |A|$ has $\xi(M^*|\cdot) = \sigma(M^*|\cdot) = \mathbf{1}$.

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

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

Assume for a contradiction that $|D| = n$. Then, the set of proper convex combinations of all elements of $|D|$ is an open subset of $\Delta\Theta$ that contains γ. In particular, some proper convex combination γ' of γ and μ_0 lies in the convex hull of $|D|$. Observe three properties of γ'. First, by construction, some $k' \in (0, k)$ exists such that $k'\beta + (1 - k')\gamma' = \mu_0$. Second, quasiconcavity of $\bar{v}$ implies $\bar{v}(\gamma') \geq \min \bar{v}(D) \geq s_i$. Third,

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

so that (β, γ', k') is feasible in program (2). Hence,

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

contradicting the optimality of (β, γ, k). □

B.1.4 Further Consequences of Lemma 1 and Theorem 1

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

Proposition 1 from Kamenica and Gentzkow (2011) shows, given a communication protocol with R best responding to Bayesian beliefs, that communication can be garbled to an incentive-compatible direct recommendation producing the same joint distribution of states and actions.

Corollary 2. *The set of χ-equilibrium summaries (p, s_o, s_i) at prior μ_0 is a compact-valued, upper-hemicontinuous correspondence of (μ_0, χ) on $\Delta\Theta \times [0,1]$.*

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

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

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

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

$\square$

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

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

is weakly increasing.

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

Corollary 4. *For any $\mu_0 \in \Delta\Theta$, the map*

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

is weakly increasing and right-continuous.

Proof. That it is weakly increasing is a specialization of Corollary 3. That it is upper semicontinuous (and so, since nondecreasing, it is right-continuous) follows directly from Corollary 2. □

Corollary 5. *For any $\chi \in [0,1]$, the map $v^*_\chi : \Delta\Theta \to \mathbb{R}$ is upper semicontinuous.*

Proof. This result is immediate from Corollary 2. □

B.2 Varying Credibility: Proofs for Section 4

In this section, we provide proofs for the results reported in section 4. We note these results are stated for the version of the model developed in the main text (with finite action space, finite state space, and state-independent credibility). In contrast to the proof of Theorem 1, finiteness plays a nontrivial role in the proofs of these propositions. As our proofs make clear, the same results would hold with state-dependent credibility.

B.2.1 Productive Mistrust: Proof of Proposition 1

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

Corollary 6. *Consider a finite and generic model in which S is not a two-faced SOB. Then, a full-support prior and state-dependent credibility levels $\boldsymbol{\chi}' < \boldsymbol{\chi}$ exist such that every S-optimal $\boldsymbol{\chi}'$ equilibrium is strictly better for R than every S-optimal $\boldsymbol{\chi}$-equilibrium.*

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

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

To this end, to introducing a more detailed language for our key SOB condition is useful. Given a prior $\mu \in \Delta\Theta$, say S is **an SOB at** μ if every $p \in \mathcal{P}(\mu)$ is outperformed by an SOB policy $p' \in \mathcal{P}(\mu)$, that is, has $\int v \, \mathrm{d}p' \geq \int v \, \mathrm{d}p$.

Lemma 5. *Suppose $|\Theta| = 2$, the model is finite and generic, and a full-support belief $\mu \in \Delta\Theta$ exists such that the sender is not an SOB at μ. Then, a full-support prior μ_0 and credibility levels $\chi' < \chi$ exist such that every S-optimal χ'-equilibrium is both strictly better for R and more Blackwell-informative than every S-optimal χ-equilibrium.*

Moreover, some full-support belief μ_+ exists such that any solution (β, γ, k) to the program in Theorem 1 at prior μ_0 and credibility level in $\{\chi, \chi'\}$ has $\gamma = \mu_+$.

Proof. First, note 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 upper semicontinuous and piecewise constant, which implies its concavification v_1^* is piecewise affine. That is, some $n \in \mathbb{N}$ and $\{\mu^i\}_{i=0}^n$ exist such that $0 = \mu^0 \leq \cdots \leq \mu^n = 1$ and $v_1^*|_{[\mu^{i-1},\mu^i]}$ is affine for every $i \in \{1, \ldots, n\}$. Taking n to be minimal, we can assume $\mu^0 < \cdots < \mu^n$ and the slope of $v_1^*|_{[\mu^{i-1},\mu^i]}$ is strictly decreasing in i. Therefore, some $i_0, i_1 \in \{0, \ldots, n\}$ exist such that $i_1 \in \{i_0, i_0 + 1\}$ and $\operatorname{argmax}_{\tilde\mu\in[0,1]} v_1^*(\tilde\mu) = [\mu^{i_0}, \mu^{i_1}]$. That the sender is not an SOB at μ implies $i_0 > 1$ or $i_1 < n - 1$. Without loss of generality, say $i_0 > 1$. Now let $\mu_- := \mu^{i_0-1}$ and $\mu_+ := \mu^{i_0}$.

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

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

where the last inequality follows from $\bar v|_{[0,\mu_+)} \leq v_1^*|_{[0,\mu_+)} < v_1^*(\mu_+)$. Thus, the desired properties are satisfied by any $\mu_0 \in (\max\{\mu_-, \mu_+ - \epsilon, \mu_+ - \tilde\epsilon\}, \mu_+)$. Let μ_0 be one such belief.

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

Let $\chi \in [0, 1]$ be the smallest credibility level such that $v_\chi^*(\mu_0) = v_1^*(\mu_0)$, which exists by Corollary 4. That $v_0^*(\mu_0) < v_1^*(\mu_0)$ implies $\chi > 0$. Notice μ_+ has full support, because $0 \leq \mu_- < \mu_+ \leq \mu' < 1$. It follows that $\chi < 1$. Consider now the following claim.

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

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

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

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

We now prove the claim.

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

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

Because

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

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

Next, note $\bar{v}$—as a weakly quasiconcave function that 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 $\gamma' = \mu_+$. A direct implication is that

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

Let us now see why we cannot have $\beta' \in (\mu_-, \mu_0)$. Because $\hat{v}_{\wedge\mu_+}$ is affine on $[\mu_+, \mu_-]$, replacing such (k', β') with (k, μ_-) that satisfies $k\mu_- + (1-k)\mu_+ = \mu_0$ necessarily has

$(1-k)(\mu_+, 1-\mu_+) \gg (1-\chi')(\mu_0, 1-\mu_0)$. This would contradict minimality of χ. Therefore, $\beta' \le \mu_-$.

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

With the claim in hand, we can now prove the lemma. The claim implies that, for credibility level χ, any solution (β^*, γ^*, k^*) of the program (3) is such that $\gamma^* = \mu_+$, $k^* = \frac{\mu_+ - \mu_0}{\mu_+ - \beta^*}$, and β^* solves

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

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

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

by concavity of $\hat{v}_{\wedge\mu_+}$. Moreover, the inequality is strict for $\beta < \mu_-$ but holds with equality for $\beta = \mu_-$, because $\hat{v}_{\wedge\mu_+}$ is affine on $[\mu_-, \mu_+]$ and on no larger interval. Hence, the unique solution to (3) is (μ_-, μ_+, k^*), where $k^*\mu_- + (1-k^*)\mu_+ = \mu_0$. Moreover, the minimality property defining χ implies $(1-k^*)(1-\mu_+) = (1-\chi)(1-\mu_0)$.

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

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

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

Let p' be any S-optimal χ'-equilibrium information policy. Appealing to the claim, some $b' \in \mathcal{P}(\beta') \cap \Delta[0, \mu_-]$ exists such that $p' \in \text{co}\{b', \delta_{\mu_+}\}$. Therefore, p' is weakly more

Blackwell-informative than p^*. Finally, because $(1-k^*)(1-\mu_+) = (1-\chi)(1-\mu_0)$ and $\chi' < \chi$, feasibility of p' tells us $p' \neq p^*$. Therefore (the Blackwell order being antisymmetric), p' is strictly more informative than p^*.

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

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

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

Productive Mistrust with Many States: Proof of Proposition 1 Given Lemma 5, 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 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 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 $p_n^{\chi'}$. Because 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, such limits must be S-optimal for the prior μ_0^∞ by the choice of direction, meaning productive mistrust fails at μ_0^∞; that is, we have a contradiction. Below, we proceed with the formal proof.

Proof. Suppose some prior with binary support $\Theta_2 = \{\theta_1, \theta_2\}$ exists at which S is not an SOB. Let $\bar{s} := \max v(\Delta\Theta_2)$, and define the R value function $v_R\colon \Delta\Delta\Theta \to \mathbb{R}$ via $v_R(p) := \int_{\Delta\Theta} \max_{a\in A} u_R(a,\mu)\, \mathrm{d}p(\mu)$. Lemma 5 delivers 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.
<u>Claim:</u> Some sequence $\{\mu_0^n\}$ of full-support priors exists that converges to μ_0^∞ with

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

Before proving the claim, let us argue 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 $\Delta\Delta\Theta \times [\text{co } u_S(A)]^2$) to some $\Psi' = (p', s'_i, s'_o)$ and $\Psi'' = (p'', s''_i, s''_o)$, respectively. By Corollary 2, 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 $v_R(p') \geq v_R(p'')$. 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, some $n \in \mathbb{N}$ exists such that the full-support prior μ_0^n is as required for the proposition.

So all that remains is to prove the claim, which we do by constructing the desired sequence.

First, Lemma 5 delivers some $\gamma^\infty \in \Delta\Theta$ with support Θ_2 such that $\bar{v}(\gamma^\infty) = \bar{s}$ and, for $\chi \in \{\chi', \chi''\}$, any solution (β, γ, k) to the program in Theorem 1 at prior μ_0^∞ and credibility level χ has $\gamma = \gamma^\infty$.

Let us now show a closed convex set $D \subseteq \Delta\Theta$ exists that contains γ^∞, has a nonempty interior, and satisfies $\bar{v}|_D = \bar{s}$. Notice, first, that the genericity assumption delivers μ' with support Θ_2 such that $V(\mu') = \{\bar{s}\}$. Then, for any $n \in \mathbb{N}$, let $B_n \subseteq \Delta\Theta$ be the closed ball (say, with respect to the Euclidean metric) of radius $\frac{1}{n}$ around μ', and let $D_n := \text{co}\,[\{\gamma^\infty\} \cup B_n]$. Because $v|_{\Delta\Theta_2} \leq \bar{s}$ and $\bar{v} = \max_{p\in\mathcal{P}(\cdot)} \inf v(\text{supp}(p))$ (see Lipnowski and Ravid 2020, Theorem 2), it follows $\bar{v}|_{\Delta\Theta_2} \leq \bar{s}$ as well. Because V is upper hemicontinuous, the hypothesis on μ' ensures $\bar{v}|_{B_n} \geq v|_{B_n} = \bar{s}$ for sufficiently large $n \in \mathbb{N}$; quasiconcavity then tells us $\bar{v}|_{D_n} \geq \bar{s}$. Assume now, for a contradiction, that every $n \in \mathbb{N}$ has $\bar{v}|_{D_n} \not\leq \bar{s}$. That is, each $n \in \mathbb{N}$ admits some $\lambda_n \in [0,1]$ and $\mu'_n \in B_n$ such that $\bar{v}\left((1-\lambda_n)\gamma^\infty + \lambda_n \mu'_n\right) > \bar{s}$. In this case, each $n \in \mathbb{N}$ has $\bar{v}\left((1-\lambda_n)\gamma^\infty + \lambda_n \mu'_n\right) \geq \hat{s} := \min\left[\bar{v}(\Delta\Theta) \cap (\bar{s}, \infty)\right]$ (observe $\hat{s}$ is well defined because $|\bar{v}(\Delta\Theta)| < \infty$ due to the model being finite). Dropping to a subsequence, we get a strictly increasing sequence $(n_\ell)_{\ell=1}^\infty$ of natural numbers such that (because $[0,1]$ is compact) $\lambda_{n_\ell} \xrightarrow{\ell\to\infty} \lambda \in [0,1]$

and $\bar{v}\left((1-\lambda_{n_\ell})\gamma^\infty + \lambda_{n_\ell}\mu'_{n_\ell}\right) \geq \hat{s}$ for every $\ell \in \mathbb{N}$. Because $\bar{v}$ is upper semicontinuous, the sequence of inequalities would imply $\bar{v}\left((1-\lambda)\gamma^\infty + \lambda\mu'\right) \geq \hat{s} > \bar{s}$, contradicting the definition of $\bar{s}$ and μ'. Therefore, some $D \in \{D_{n_\ell}\}_{\ell=1}^\infty$ is as desired.

In what follows, let $\gamma_1 \in D$ be some interior element with full support. Then, for each $n \in \mathbb{N}$, define $\mu_0^n := \frac{n-1}{n}\mu_0^\infty + \frac{1}{n}\gamma_1$. We show 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,

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

and

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

Therefore, (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 hence the proposition. □

B.2.2 Collapse of Trust: Proof of Proposition 2

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

$(i)'$ Every $\boldsymbol{\chi} \in [0,1]^\Theta$ and full-support prior μ_0 have $\lim_{\boldsymbol{\chi}' \nearrow \boldsymbol{\chi}} v^*_{\boldsymbol{\chi}'}(\mu_0) = v^*_{\boldsymbol{\chi}}(\mu_0)$, where convergence of $\boldsymbol{\chi}'(\cdot) \to \boldsymbol{\chi}(\cdot)$ is in the Euclidean topology on $\mathbb{R}^\Theta$.

Three of four implications are easy given Corollary 3. First, $(i)'$ trivially implies (i). Second ((iii) implies (ii)), in the absence of conflict, Lemma 1 from Lipnowski and Ravid (2020) tells us a 0-equilibrium exists with full information that generates sender value $\max v(\Delta\Theta) \geq v^*_1$; in particular, $v^*_0 = v^*_1$. Third ((ii) implies $(i)'$), if $v^*_0 = v^*_1$, Corollary 3 implies $v^*_{\boldsymbol{\chi}}$ is constant in $\boldsymbol{\chi}$, ruling out a collapse of trust (even under state-dependent credibility). Below, we show that any conflict implies a collapse of trust; that is, a failure of (iii) implies a failure of (i).

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

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

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

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

The only remaining case is the one in which $\chi = 1$. In this case, some $\epsilon \in (0,1)$ and $\mu \in \Delta\Theta \setminus D$ exist such that $\epsilon\mu \leq \mu_0$. Then,

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

So, again, a collapse of trust occurs at credibility level χ. □

B.2.3 Robustness: Proof of Proposition 3

Before proving the proposition, let us briefly observe that the proposition as stated is equivalent to the analogous statement for state-dependent credibility. Indeed, given Corollary 3, any prior μ_0 and state-dependent credibility $\boldsymbol{\chi}$ has $v_\chi^*(\mu_0) \leq v_{\boldsymbol{\chi}}^*(\mu_0) \leq v_1^*(\mu_0)$ for $\chi = \min_{\theta\in\Theta} \boldsymbol{\chi}(\theta) \in [0,1]$. It follows immediately that $\lim_{\chi\nearrow 1} v_\chi^*(\mu_0) = v_1^*(\mu_0)$ if and only if $\lim_{\boldsymbol{\chi}\nearrow\mathbf{1}} v_{\boldsymbol{\chi}}^*(\mu_0) = v_{\mathbf{1}}^*(\mu_0)$, where convergence of $\boldsymbol{\chi} \to \mathbf{1}$ is in the Euclidean topology on $\mathbb{R}^\Theta$. That is, the stronger property of robustness of the commitment value to small *state-dependent* departures from perfect credibility is equivalent to that stated in the proposition.

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

Proof. By Lipnowski and Ravid (2020, Lemma 1 and Theorem 2), S receives the benefit of the doubt (i.e., every $\theta \in \Theta$ is in the support of some member of $\text{argmax}_{\mu\in\Delta\Theta} v(\mu)$) if and only if some full-support $\gamma \in \Delta\Theta$ exists 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 $\hat{v}_{\wedge\gamma} = \hat{v} = v_1^*$. Let $r_0 := \min_{\theta\in\Theta} \frac{\mu_0\{\theta\}}{\gamma\{\theta\}} \in (0,\infty)$

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

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

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

Conversely, suppose S does not receive the benefit of the doubt (which of course implies v is nonconstant). Taking an affine transformation of u_S, we may assume without loss that $\max v(\Delta\Theta) = 1$ and (because $v(\Delta\Theta) \subseteq u_S(A)$ is finite) $\max[\bar{v}(\Delta\Theta) \setminus \{1\}] = 0$. Fix any full-support prior μ_0 and consider any credibility level $\chi \in [0, 1)$. For any $\beta, \gamma \in \Delta\Theta,\ k \in [0, 1]$ with $k\beta + (1-k)\gamma = \mu_0$ and $(1-k)\gamma \geq (1-\chi)\mu_0$, that S does not get the benefit of the doubt implies (see Lipnowski and Ravid, 2020, Theorem 1) that $\bar{v}(\gamma) \leq 0$, and therefore that $k\hat{v}_{\wedge\gamma}(\beta) + (1-k)v(\gamma) \leq 0$. Theorem 1 then implies $v^*_\chi(\mu_0) \leq 0$.

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

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

which is strictly positive for sufficiently small ϵ. Persuasion is therefore not robust to limited commitment at prior μ_ϵ. □

C Extension on Signaling Credibility

In this section, we consider the modified version of our model in which S learns her credibility type before announcing the official reporting protocol. By letting S commission a different official report based on her credibility, the modified model allows S to signal whether she can influence the report's message. We show such signaling has no impact on S's attainable payoffs. More precisely, every interim S-payoff profile (i.e., every pair specifying S's payoffs conditional on each credibility type) is attainable in a pooling equilibrium in which both credibility types choose the same official experiment. It follows that pooling equilibria are without loss as far as S payoffs are concerned. We also show an S-payoff profile is attainable in a pooling equilibrium if and only if it is attainable in a χ-equilibrium. Our definition will make the fact that every pooling-equilibrium payoff profile is attainable in a χ-equilibrium immediate: a pooling equilibrium of the modified game requires the same conditions as a χ-equilibrium, except S must also be willing to announce the equilibrium experiment conditional on her credibility type. For the converse direction, we show every χ-equilibrium can be implemented as a pooling equilibrium of the signaling game by appropriately constructing R's behavior off path. Thus, we show a three-way equivalence between S's payoffs in all equilibria of the signaling game, all pooling equilibria of the signaling game, and χ-equilibria of the original game. It follows that informing S of her ability to influence the report before its announcement has no impact on S's achievable payoffs.

C.1 On S's Equilibrium Payoff Sets

We begin by providing results on the space of S payoffs that will be of use in the extension that follows and may be of independent use. We return to the general specification of our model in which the state and action spaces may be finite or infinite, and the credibility level may or may not depend on the payoff state.

First, we characterize the set of payoffs attainable in a $\boldsymbol{\chi}$-equilibrium by an influencing S, in particular showing this payoff set is an interval. Then, we show the set of ex-ante S payoffs attainable in a $\boldsymbol{\chi}$-equilibrium is an interval as well.

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

Fact 1. *The range of a Kakutani correspondence from a nonempty, compact, convex space to $\mathbb{R}$ is a nonempty compact interval.*

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

Next, we establish convexity and compactness of the sets of S's possible $\boldsymbol{\chi}$-equilibrium ex-ante payoffs and payoffs from influencing. To do so, we now provide a characterization of the set

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

Lemma 6. *Let $s_i \in \mathbb{R}$. Then $s_i \in S_i^{\boldsymbol{\chi}}$ if and only if some $k \in [0,1], \gamma, \beta \in \Delta\Theta$ exist such that*

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

(ii) $(1-k)\gamma \geqslant (\mathbf{1} - \boldsymbol{\chi})\mu_0$,

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

Moreover, the set $S_i^{\boldsymbol{\chi}}$ is a nonempty compact interval.

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

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

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

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

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

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

Thus, we obtain the desired characterization.

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

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

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

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

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

We now proceed to show $S_i^{\boldsymbol{\chi}}$ is a nonempty compact interval. First, observe that κ is upper semicontinuous and quasiconcave—because both $\bar{v}$ and $-\underline{w}$ are, and therefore so are $\tilde{v}$ and $-\tilde{w}$. Hence, the set $\kappa^{-1}([0,\infty)) = \mathcal{G}^*$ is compact and convex, and it is also nonempty because it contains μ_0. Second, note τ is a Kakutani correspondence because it is compact-convex-valued by definition, nonempty-valued by the definition of $\mathcal{G}^*$, and upper hemicontinuous by upper (resp. lower) semicontinuity of $\tilde{v}$ ($\tilde{w}$). Hence, the result follows from Fact 1. □

Building on the previous two lemmas, the following result shows the set of ex-ante $\boldsymbol{\chi}$-equilibrium payoffs for S is convex.

Lemma 7. *The set* $\{\boldsymbol{\chi} s_o + (\mathbf{1} - \boldsymbol{\chi}) s_i \colon (p, s_o, s_i)$ *is a* $\boldsymbol{\chi}$*-equilibrium summary*$\}$ *of ex-ante* $\boldsymbol{\chi}$*-equilibrium payoffs is a nonempty compact interval.*

Proof. Define the correspondence

$$\begin{aligned} \varsigma \colon & S_i^{\boldsymbol{\chi}} \rightrightarrows \mathbb{R} \\ & s_i \mapsto \left\{ \boldsymbol{\chi}(\mu_0) s_o + [1 - \boldsymbol{\chi}(\mu_0)]\, s_i \colon \ (p, s_o, s_i) \text{ is a } \chi\text{-equilibrium summary} \right\}. \end{aligned}$$

We show ς is a Kakutani correspondence, which will give the desired result in light of Fact 1 and Lemma 6.

First, ς is nonempty-valued by the definition of $S_i^{\boldsymbol{\chi}}$. Second, the graph of ς is compact as a continuous image of the compact space X defined in the proof of Corollary 2. Therefore, ς is compact-valued and upper hemicontinuous.

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

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

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

C.2 Signaling Credibility

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

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

We focus on perfect Bayesian equilibria in which R's off-path beliefs satisfy a standard *"no signaling what you don't know"* restriction. To formalize the relevant solution concept, let Ξ denote the set of all official reporting protocols, that is, measurable maps $\xi\colon \Theta \to \Delta M$; endow Ξ with some measurable structure such that singletons are measurable. Then, let $(\xi_o, \xi_i) \in \Xi^T$ denote S's signaling strategy;[22] let the measurable maps $\boldsymbol{\sigma}\colon \Theta \times T \times \Xi \to \Delta M$, $\boldsymbol{\alpha}\colon M \times \Xi \to \Delta A$, and $\boldsymbol{\pi}\colon M \times \Xi \to \Delta\Theta$ denote S's influencing strategy, R's strategy, and R's belief map, respectively, that take into account the announced reporting protocol $\xi \in \Xi$; and let $\tilde{\boldsymbol{\chi}} : \Theta \times \Xi \to [0, 1]$ denote R's measurable belief mapping from an announced official reporting protocol to S's posterior credibility. Then, a **χ signaling PBE (χ-SPBE)** is a tuple $(\xi_o, \xi_i, \boldsymbol{\sigma}, \boldsymbol{\alpha}, \tilde{\boldsymbol{\chi}}, \boldsymbol{\pi})$ such that (letting $\boldsymbol{\sigma}_\xi \coloneqq \boldsymbol{\sigma}(\cdot, \xi)$ and similarly for $\boldsymbol{\alpha}$, $\tilde{\boldsymbol{\chi}}$, and $\boldsymbol{\pi}$):

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

2. $(\xi, \boldsymbol{\sigma}_\xi, \boldsymbol{\alpha}_\xi, \boldsymbol{\pi}_\xi)$ is a $\tilde{\boldsymbol{\chi}}_\xi$-equilibrium (for prior μ_0) for each $\xi \in \Xi$.

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

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

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

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

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

[22] To simplify notation, here we focus on pure signaling strategies. Analogous results holds for mixed signaling strategies.

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

Proposition 4. *Fixing $(s_o, s_i) \in \mathbb{R}^2$, the following are equivalent:*

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

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

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

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

Now, let us show (c) implies (b). To do so, consider some χ-equilibrium $(\xi, \sigma, \alpha, \pi)$ generating summary (p, s_o, s_i). Observe that for each $\xi' \in \Xi \setminus \{\xi\}$, some uncountable Borel $M_{\xi'} \subset M$ exists such that $\int_\Theta \xi'(M_{\xi'}|\cdot)\,\mathrm{d}\mu_0 = 0$.[23] It then follows readily from Theorem 2 of Lipnowski and Ravid (2020) that some 0-equilibrium $(\xi', \sigma_{\xi'}, \alpha_{\xi'}, \pi_{\xi'})$ exists giving S payoff $\underline{w}(\mu_0)$ with messages restricted to $M_{\xi'}$, that is, with $\boldsymbol{\sigma}_{\xi'}(M_{\xi'}|\cdot) = \mathbf{1}$. We now proceed to construct a pooling $\boldsymbol{\chi}$-SPBE. Define an influencing sender strategy $\boldsymbol{\sigma}$ and credibility belief function $\tilde{\boldsymbol{\chi}}$ by letting, for each $\xi' \in \Xi$,

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

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

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

By construction, $(\xi, \xi, \boldsymbol{\sigma}, \boldsymbol{\alpha}, \tilde{\boldsymbol{\chi}}, \boldsymbol{\pi})$ satisfies conditions 1 and 2 of the definition of $\boldsymbol{\chi}$-SPBE. Moreover, observe that, by Lemma 6, some $\gamma, \beta \in \Delta\Theta$ exist such that $s_i \geq$

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

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

$\max\{\underline{w}(\beta), \underline{w}(\gamma)\}$ and $\mu_0 \in \text{co}\{\gamma, \beta\}$. Hence, $s_i \geq \underline{w}(\mu_0)$ because $\underline{w}$ is quasiconvex. Therefore, condition 3 of the definition of a $\boldsymbol{\chi}$-SPBE is satisfied because $\boldsymbol{s_i}(\xi) = s_i \geqslant \underline{w}(\mu_0) = \boldsymbol{s_i}(\xi')$ and $\boldsymbol{s_o}(\xi) = s_o \geqslant \min_{\Delta\Theta} w = \boldsymbol{s_o}(\xi')$ for all $\xi' \in \Xi \setminus \{\xi\}$. Therefore, $(\xi, \xi, \boldsymbol{\sigma}, \boldsymbol{\alpha}, \tilde{\boldsymbol{\chi}}, \boldsymbol{\pi})$ is a pooling $\boldsymbol{\chi}$-SPBE with S's payoff vector (s_o, s_i) as desired.

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

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

But then consider

$$
\begin{aligned}
k &:= \boldsymbol{\chi}(\mu_0) k_{\xi_o} + [1 - \boldsymbol{\chi}(\mu_0)]\, k_{\xi_i} \in [0,1), \\
b &:= \tfrac{\boldsymbol{\chi}(\mu_0) k_{\xi_o}}{k} b_{\xi_o} + \left[1 - \tfrac{\boldsymbol{\chi}(\mu_0) k_{\xi_o}}{k}\right] b_{\xi_i} \in \Delta\Delta\Theta, \\
g &:= \left(1 - \tfrac{[1-\boldsymbol{\chi}(\mu_0)](1-k_{\xi_i})}{1-k}\right) g_{\xi_o} + \tfrac{[1-\boldsymbol{\chi}(\mu_0)](1-k_{\xi_i})}{1-k} g_{\xi_i} \in \Delta\Delta\Theta.
\end{aligned}
$$

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