Persuasion via Weak Institutions

Elliot Lipnowski*†
University of Chicago

Doron Ravid
University of Chicago

Denis Shishkin
Princeton University

February 15, 2018

Preliminary and incomplete. Please do not circulate.

Abstract

A sender (S) publicly commissions a study by an institution to persuade a receiver (R). A study consists of a research plan and an official reporting rule. S privately learns the research's outcome, and also whether she can influence the report. Under influenced reporting, S can privately change the report to a message of her choice. Otherwise, the official reporting rule applies. We geometrically characterize S's highest equilibrium value, and examine how optimal persuasion varies with the probability that reporting is uninfluenced – S's "credibility". We identify two phenomena: (1) R can strictly benefit from a reduction in S's credibility; (2) small decreases in credibility often lead to large payoff losses for S, but this typically will not happen when S is almost fully credible.

*E-mail: lipnowski@uchicago.edu, dravid@uchicago.edu, shishkin@princeton.edu.
†To be written.

1 Introduction

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

To fix ideas, consider a central bank that releases reports about the state of the domestic economy. These reports are read by foreign investors, who decide how much to invest there. Higher investment levels are preferred both by the central bank and by the incumbent politician. Not caring about the bank's credibility, the politician pressures the central bank to release more favorable reports. The weaker the central bank is, the more likely is the politician to succeed in influencing the report. Knowing that the report may be influenced, foreign investors take its contents with a grain of salt. Our model allows us to analyze the central banks' gains from releasing its reports, and how these gains vary with the strength of the central bank.

More generally, we study a persuasion game between a receiver (R, he) and a sender (S, she) who cares only about R's actions. The game begins with S publicly announcing a research plan and an official reporting rule. Formally, these are a pair of Blackwell experiments: the research plan is an experiment about the state, and the reporting rule is an experiment about the research's outcome. After the announcement, S privately learns the research's outcome as well as whether her reporting rule is credible. If credible, R observes a message drawn from the announced reporting rule. Otherwise, S can freely choose which message R will receive. R then takes an action, not knowing the message's origin. We assume that reporting is credible with fixed and known probability χ, independent of any other aspect of the game. We interpret the average credibility, χ, as the strength of S's institutions.

As in the recent Bayesian persuasion literature (e.g., Kamenica and Gentzkow (2011), Alonso and Câmara (2016) and Ely (2017)), we view S as a principal, capable of coordinating R towards her preferred equilibrium. Our main result (Theorem 1) characterizes S's favorite equilibrium payoff . The characterization is geometric, and is based on S's *value function*, which specifies the highest value S can obtain from R best responding given his posterior belief. Under full credibility ($\chi = 1$), our model is equivalent to that studied by Kamenica and Gentzkow (2011). As such, S's highest equilibrium value in this case is given by the concave envelope of S's value function. The value function's quasiconcave envelope, which gives S's highest value under cheap talk (see Lipnowski and Ravid (2017)), also delivers S's highest equilibrium value under no credibility – i.e., when $\chi = 0$. For intermediate credibility values, the theorem's characterization combines the quasiconcave envelope of S's value function, and the concave envelope of S's value function *with a cap*, which captures S's incentive constraints.

Using our characterization, we analyze how S's and R's values change with χ. To illustrate, consider the following instance of the central bank example. An incumbent politician wants to use the central bank's reports to lure a a multinational firm to enter its market. The firm can make a large investment ($a = 2$), small investment ($a = 1$), or no investment ($a = 0$). Profits from each investment level depend on state of the economy, θ, which can be good ($\theta = 1$), or bad ($\theta = 0$) with equal probability. To make its decision, the firm uses the reports provided by the central bank.

Suppose that the firm's profit from investment level a in state θ is given by $a\theta - \frac{1}{4}a^2$. As the state of the world is binary, the firm's beliefs can be identified with the probability that the economy is good. Given the above preferences: no investment is optimal if the firm's belief is lower than $\frac{1}{4}$, a large investment is optimal if the firm's belief is at least $\frac{3}{4}$, and a small investment is optimal for intermediate beliefs.

The politician always prefers larger investments. Suppose that her payoffs from no, a small, and a large investment are 0, 1, and 2, respectively. To persuade the firm, she commissions a research report to be released by the central bank. In this example, it is without loss to assume that the politician publicly announces a Blackwell experiment which produces a stochastic investment recommendation conditional on the economy's state. The reliability of this recommendation is questionable, as it is only produced by the announced experiment with probability χ. With probability $1 - \chi$, the bank succumbs to the politician's pressure, producing the politician's preferred recommendation.

Proposition 1 shows that R is often better off with a less credible sender. Specifically, the proposition provides sufficient conditions for there to be a full-support prior and two credibility levels, $\chi < \chi'$, such that R is strictly better under χ than under χ'. Our proposition applies to the above example. To see that, suppose first that the bank's report is fully credible – i.e., $\chi = 1$. In this case, the optimal report recommends either a large or a small investment with equal ex-ante probability in a way that makes the firm just willing to accept said recommendation. In other words, the firm's posterior belief that the state is good is uniformly distributed on $\{\frac{1}{4}, \frac{3}{4}\}$, with the firm making a large investment when her belief is $\frac{3}{4}$, and a small investment otherwise. In this case, the firm's expected utility is $\frac{1}{4}$.

Suppose now that the central bank is not as strong. Specifically, suppose that $\chi = \frac{2}{3}$, and consider a report that leads to an incentive compatible large investment recommendation with positive probability. Since the bank is weaker, the politician gets to secretly influence the report with probability $1 - \chi = \frac{1}{3}$. Therefore, the report will produce a large investment recommendation with probability of at least $\frac{1}{3}$ regardless of the state. By Bayes' rule, conditional on such a recommendation, the firm's posterior belief that the state is good is at most $\frac{3}{4}$. Notice that this upper bound can *only* be achieved if the bank's official report fully reveals the state. As such, the firm must observe a "no investment" recommendation whenever the economy is bad and reporting is uninfluenced (which happens with probability $\frac{1}{3}$), and a "large investment" recommendation otherwise. This policy is strictly better for the politician than conveying no information (which yields small investment with certainty), and so is her unique preferred equilibrium. Thus, when $\chi = \frac{2}{3}$, the firm's expected utility is $\frac{1}{3}$. In other words, the firm strictly benefits from a weaker central bank – i.e., there is productive mistrust.

Our next result, Proposition 2, shows that small decreases in credibility can lead to large drops in the sender's value. More precisely, such a collapse occurs at some full-support prior and some credibility level if and only if S can benefit from persuasion. Such a collapse is clearly present in our example: given the preceding analysis, $\frac{2}{3}$ is the lowest credibility level that allows the bank to credibly recommend a large investment. For any $\chi < \frac{2}{3}$ the politician can do no better than have the bank provide no information to the firm, giving the politician a payoff of 1. Since $\frac{4}{3}$ is the politician's payoff when $\chi = \frac{2}{3}$, even an infinitesimal decrease in credibility results in a discrete drop in the value.

It is also possible to construct examples in which S's value collapses at full credibility. To see that, suppose that the firm can alternatively decide to make a very large investment, which is optimal only if the firm is *certain* that the economy is good. The government's payoff in this case is 10. Under full credibility, the politician can obtain a payoff of 5 by revealing the state, having the central bank recommend no investment when the economy is bad, and a very large investment when the economy is good. A very large investment recommendation, however, is never credible for any $\chi < 1$. If it were, the politician would always send it when influencing the bank's report, meaning that the firm can never be completely certain that the economy's state is good. As such, the politician's optimal equilibrium policy for any $\chi \in [\frac{3}{4}, 1)$ remains as it was in the unmodified example. Thus, reducing the central bank's credibility from 1 to any $\chi < 1$ results in the politician's payoff dropping from 5 to $\frac{3}{2}$.

One may suspect that the non-robustness of the full credibility solution in the above modified example is rather special. Proposition 3 confirms this suspicion. In particular, it shows that S's value collapses at full credibility if and only if R does not give S the benefit of the doubt – i.e., to obtain her best feasible payoff, S must persuade R that some state is impossible. This property is clearly violated in the above modified example: the firm is only willing to make a very large investment if she assigns a zero probability to the economy's state being bad. Thus, while S's value often collapses due to small decreases in credibility, such collapses rarely occur at full credibility.

Related Literature. To be written.

2 Credible Persuasion

In this section, we present our model and our main characterization result.

There are two players: a sender (S, she) and a receiver (R, he). The game begins with S commissioning study to be conducted by a research institution. Formally, she announces a primary research protocol (ρ) and an official reporting protocol (ξ). An unknown state ($\theta \in \Theta$) then realizes, from which a research outcome ($\omega \in \Omega$) results. R does not observe ω, but S does. S then sends R a report, $m \in M$. With probability $\chi \in [0, 1]$ (the "credibility level"), the report is drawn according to ξ – i.e., S is committed. With probability $1 - \chi$, S is not committed and can decide which report to send after observing the research outcome. R then observes m (but not θ or ω) and decides which action, $a \in A$, to take. While R's payoffs from a depend on the state, S's payoffs do not. We refer to whether S is committed or not as S's credibility type, and assume that it is independent of θ. Only S learns her credibility type, and this happens after announcing the primary research protocol and official reporting protocol.

We impose a few technical restrictions on our model. Both Θ and A are finite sets with at least two elements. The state, θ, is assumed to follow some full-support prior distribution $\mu_0 \in \Delta\Theta$, which is known to both players.[1] S and R have objectives $u_S : A \to \mathbb{R}$ and $u_R : A \times \Theta \to \mathbb{R}$, respectively. Finally, we assume that the research outcome space and message space (Ω and M, respectively) are compact metric spaces which contain Θ and $\{0, 1\} \times \Delta\Theta$, respectively.[2]

[1] For a Polish space Y we denote by ΔY the set of all Borel probability measures over Y, endowed with the weak* topology. If $\gamma \in \Delta Y$, we let $\text{supp}(\gamma)$ denote the support of γ. If $f : Y \to \mathbb{R}$ is bounded and measurable, let $f(\gamma) := \int_Y f \, d\gamma$.

[2] These richness conditions enable our complete characterization of equilibrium outcomes (Lemma 1). In our finite setting, however, our characterization of sender-optimal equilibrium values (Theorem 1) and our applied propositions will hold without change if $|\Omega| \geq |\Theta|$ and $|M| \geq 2|\Theta|$.

The behavior of S is captured by three objects. The first is an announcement of primary research, $\rho : \Theta \to \Delta\Omega$, to which S is committed. The second is an official reporting protocol, $\xi : \Omega \to \Delta M$, which is executed as is whenever S cannot influence reporting. The third object is the strategy S will employ when she is not committed, $\sigma : \Omega \to \Delta M$. The probability that S is committed is χ. R's behavior is captured only by her strategy, $\alpha : M \to \Delta A$. We define a χ-equilibrium as a research and official reporting policy (ρ, ξ) together with a perfect Bayes equilibrium of the above game with S being committed to ξ with probability χ. Thus, a **χ-equilibrium** is a tuple $(\rho, \xi, \sigma, \alpha, \pi)$ of measurable maps such that:

1. $\pi : M \to \Delta\Theta$ is derived from μ_0 via Bayes' rule, given message policy

$$\theta \mapsto \int_\Omega [\chi\xi + (1-\chi)\,\sigma]\,\mathrm{d}\rho(\cdot|\theta) \in \Delta M,$$

   whenever possible.

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

3. $\sigma(\omega)$ is supported on $\mathrm{argmax}_{m\in M} u_S\,(\alpha(m))$ for all $\omega \in \Omega$.

Notice that we did not impose any incentive constraints on S's research and official report, (ρ, ξ). This is because we view S as a principal capable of coordinating R towards her favorite χ-equilibrium. For such equilibria, S's incentive constraints with respect to (ρ, ξ) are automatically satisfied. We therefore omit said constraints for the sake of brevity.

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

Thus, we identify each of S's messages with the posterior belief it induces in equilibrium, and use S's value correspondence,

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

to summarize the effect of R's incentive constraints on S. More precisely, $V(\mu)$ is the set of continuation payoffs that S could attain when R best responds to a report giving him the posterior μ. Notice that (appealing to Berge's theorem) V is a Kakutani correspondence, i.e. a nonempty-, compact-, and convex-valued upper hemicontinuous correspondence. As such, S's **value function**, $v(\mu) := \max V(\mu)$, which identifies S's highest continuation payoff from inducing posterior μ, is a well-defined, upper semicontinuous function.

Suppose first that S is fully credible ($\chi = 1$). In this case, only S's official reporting rule matters. Since S publicly commits to this rule at the beginning of the game, and since S can always choose to observe the state, Bayes plausibility is the only constraint imposed on equilibrium communication. As such, R may as well break ties in S's favor, reducing the maximization of S's equilibrium value to the maximization of v's expected value across all information policies. Kamenica and Gentzkow (2011) show that the highest such value is given by the pointwise lowest concave function that majorizes v.[3] This function, which we denote by $\hat{v}$, is known as v's **concave envelope**.

Under no credibility ($\chi = 0$), the official reporting protocol plays no role, as S can influence the report. As such, S's messages must satisfy her incentive constraints, which take a very simple form due to S's state-independent payoffs: All messages must give S the same continuation payoff. Lipnowski and Ravid (2017) show that the maximal value S can attain subject to this constraint is given by v's **quasiconcave envelope** – i.e., the lowest quasiconcave function that majorizes v. We denote this function by $\bar{v}$.

Theorem 1 shows that, for intermediate χ, S's highest χ-equilibrium value is characterized by an object which combines the concave and quasiconcave envelopes. For $\gamma \in \Delta\Theta$, define $v_{\wedge\gamma} := \bar{v}(\gamma) \wedge v = \min\{\bar{v}(\gamma), v\} : \Delta\Theta \to \mathbb{R}$. Theorem 1's characterization is based on the concave envelope of $v_{\wedge\gamma}$, which we denote by $\hat{v}_{\wedge\gamma}$. Figure 1 below visualizes the construction of $\hat{v}_{\wedge\gamma}$. With the relevant building blocks in hand, we may now state the theorem.

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

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

*We say that $(k, \beta, \gamma) \in [0,1] \times (\Delta\Theta)^2$ is **χ-feasible** if it satisfies R-BP and χ-BP above.*

[3] While we assume in the paper that Θ and A are finite, we prove both Theorem 1 and Lemma 1 for the case in which both Θ and A are compact metrizable, and u_R, u_S are continuous. In those cases, the (quasi)concave envelope is defined as the pointwise lowest (quasi)concave and *upper semicontinuous* function that majorizes v; one can show these definitions are equivalent in the finite case, because v is itself upper semicontinuous. For instance, see (Lipnowski and Ravid, 2017, Fact 1).

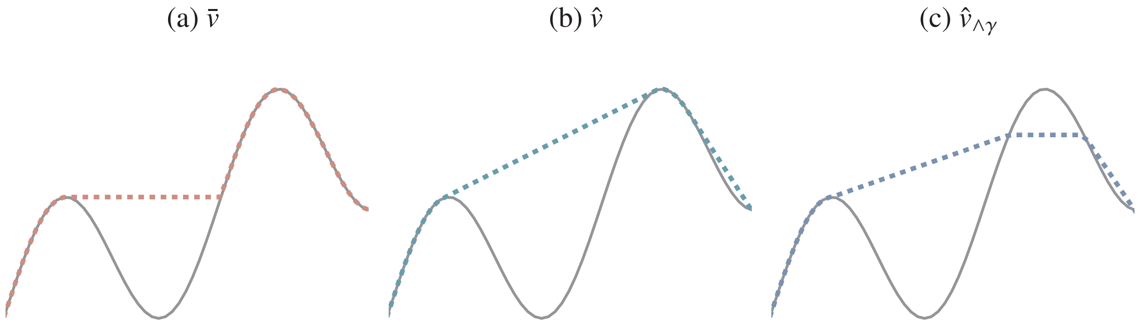

Figure 1: Quasiconcave envelope, concave envelope, and concave envelope with a cap

To understand Theorem 1, notice that every χ-equilibrium partitions the messages R sees into two sets, M_β and M_γ, where messages in M_γ are those that are sometimes sent by influenced reporting; official reporting can send messages either in this set or the complementary set M_β. The theorem follows from maximizing S's expected payoffs from M_γ and M_β, holding R's expected posterior conditional on M_γ and M_β fixed at γ and β, respectively. Letting k be the probability that the realized message is in M_β, these posteriors must respect two instances of Bayes plausibility. First, the two posterior beliefs γ, β must average back to the prior. Second, the joint probability S sending a message in M_γ and the state being θ is at least the ex-ante probability of state θ prevailing and reporting being influenced – i.e., $(1-k)\gamma(\theta) \geq (1-\chi)\mu_0(\theta)$ for every $\theta \in \Theta$.

The characterization of S's optimal values from M_γ and M_β is based on the no credibility and full credibility cases, respectively. Since messages in M_γ are sent under influenced reporting, they must satisfy the same constraints as in the no credibility case: S must be indifferent across all such messages. As such, one can apply Lipnowski and Ravid's (2017) argument to get that $\bar{v}(\gamma)$ is the highest payoff S can obtain from sending a message in M_γ. For S to send such messages, though, it must be that S's payoff from M_γ is above S's continuation payoff from any message in M_β. This results in two restrictions on M_β: (1) It caps S's continuation payoff from any feasible posterior; and (2) It restricts the set of feasible posteriors in M_β, precluding posteriors from which S must obtain too high of a continuation payoff. In the proof, we argue (with reasoning similar in spirit to the proof of Lipnowski and Ravid's (2017) securability theorem) that the second constraint is automatically satisfied at the optimum. As such, one can apply the same arguments as in the full credibility case, but with v replaced by $v_{\wedge\gamma}$. That S's highest payoff from M_β is given by $\hat{v}_{\wedge\gamma}(\beta)$ follows.

The program in Theorem 1 reduces the task of finding S's favorite χ-equilibrium value to finding the two posteriors, γ and β, and a probability, k. Alternatively, one chooses γ

and k, with (R-BP) pinning down the implied posterior $\beta = \frac{1}{k}[\mu_0 - (1-k)\gamma]$. One can then reduce the problem even further. In particular, fixing γ, one can always choose k to be as large as possible, making the constraint (χ-BP) bind.[4] To see why, suppose that both (k,β) and (k',β') are feasible for some fixed γ, and that $k' > k$. Constraint (R-BP) implies that $\beta' = \frac{k}{k'}\beta + \left(1 - \frac{k}{k'}\right)\gamma$. Since $\hat{v}_{\wedge\gamma}$ is concave and is equal to $\bar{v}$ at γ, we have

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

Thus, using (k',β') rather than (k,β) weakly increases the program's value. Intuitively, (χ-BP) being slack means that one can increase k, the probability S sends a messages from M_β, without changing the informative content of messages in M_γ. Such a substitution can only increase S's value since messages in M_β are less constrained: They must respect only an upper bound on S's continuation value, while messages in M_γ must also respect a lower bound.

With only two states, $\Theta = \{\theta_1, \theta_2\}$, one can use Theorem 1 to graphically solve for S's optimal equilibrium value. Consider Figure 2 below, which visualizes constraints (R-BP) and (χ-BP) for the binary-state case. In this figure, the horizontal axis is the probability of θ_1, while the vertical axis is the probability of θ_2. Since μ_0, β and γ assign a total probability of 1 to both states, each one of them can be represented as a point on the line connecting the two atomistic beliefs δ_{θ_1} and δ_{θ_2}. Every point underneath this line represents the product $(1-k)\gamma$ for some unique $k > 0$ and γ. Fixing a prior and a credibility level, χ, the constraints (R-BP) and (χ-BP) require $(1-k)\gamma$ to fall within the drawn box. The constraint (χ-BP) requires $(1-k)\gamma$ to be pointwise above $(1-\chi)\mu_0$, which is the box's bottom left corner. In contrast, constraint (R-BP) implies that the prior, which represents the box's top right corner, has to be pointwise larger than $(1-k)\gamma$.[5] Once $(1-k)\gamma$ is chosen, one can recover γ and β by finding the unique points on the line $[\delta_{\theta_1}, \delta_{\theta_2}]$ that lie in the same direction as $(1-k)\gamma$ and $\mu_0 - (1-k)\gamma$, respectively.

Figure 3 below shows how to simultaneously visualize the constraint from Figure 2 and S's value for the introduction's example. Such a visualization enables us to solve for S's

[4] It is worth noting that constraint (χ-BP) binding does not imply $(1-k)\gamma = (1-\chi)\mu_0$, since the inequality is an inequality of measures. Rather, in our finite state setting, a binding constraint means that $(1-k)\gamma(\theta) = (1-\chi)\mu_0(\theta)$ for some state $\theta \in \Theta$.

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

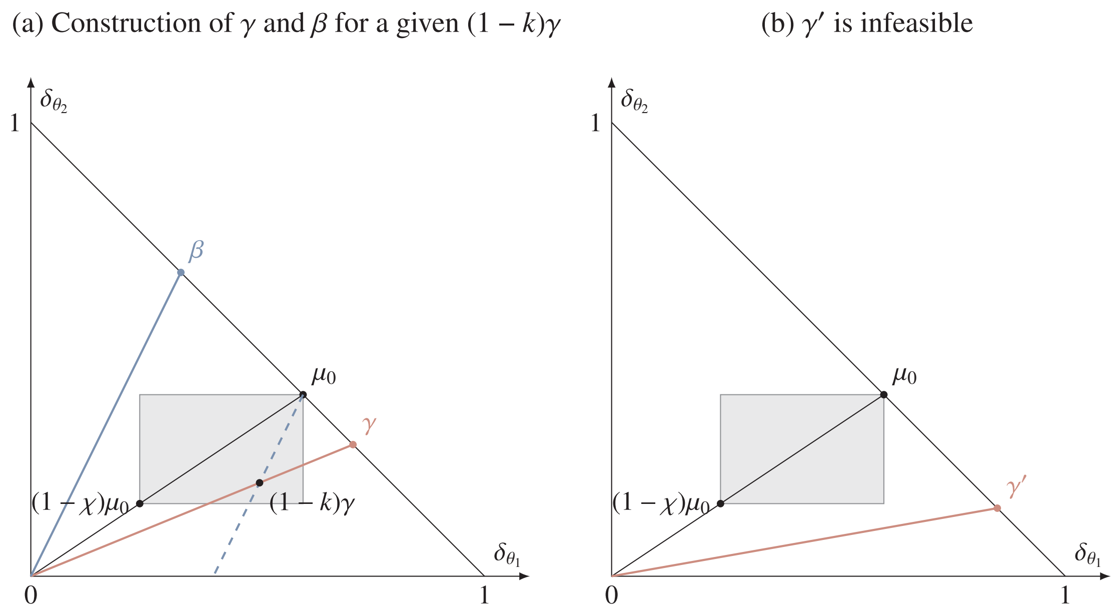

Figure 2: Constraints (R-BP) and (χ-BP) and construction of γ and β for a given $(1-k)\gamma$

optimal equilibrium value. To do so, one starts by drawing $\bar{v}$, the quasiconcave envelope of S's value function. For each candidate γ, one can find the unique k for which constraint χ-BP binds by finding the intersection of the box's lowest edges (i.e. those closest to the origin) with the line connecting γ to the origin. Once $(1-k)\gamma$ is identified, one finds the corresponding β as in Figure 2. To calculate S's value from the resulting (k, β, γ), one simply finds the value above μ_0 of the line connecting the points $(\beta, \hat{v}_{\wedge\gamma}(\beta))$ and $(\gamma, \bar{v}(\gamma))$.

3 Productive Mistrust

This section studies how a decrease in S's credibility impacts R's value and the informativeness of S's equilibrium communication. In general, the less credible is the sender, the smaller is the set of equilibrium information policies (see Lemma 1). However, that the set of equilibrium policies shrinks does not mean that less information is transmitted in S's preferred equilibrium. Our introductory example is a case in point, showing that lowering S's credibility can actually result in more information (in Blackwell's (1953) sense) being transmitted in equilibrium. Moreover, this information is useful for R: His value is strictly higher when S's credibility is lower. In other words, there is productive mistrust. The current section provides sufficient conditions for such productive mistrust to occur.

Our sufficient condition is about S's optimal information policy under full credibility. We

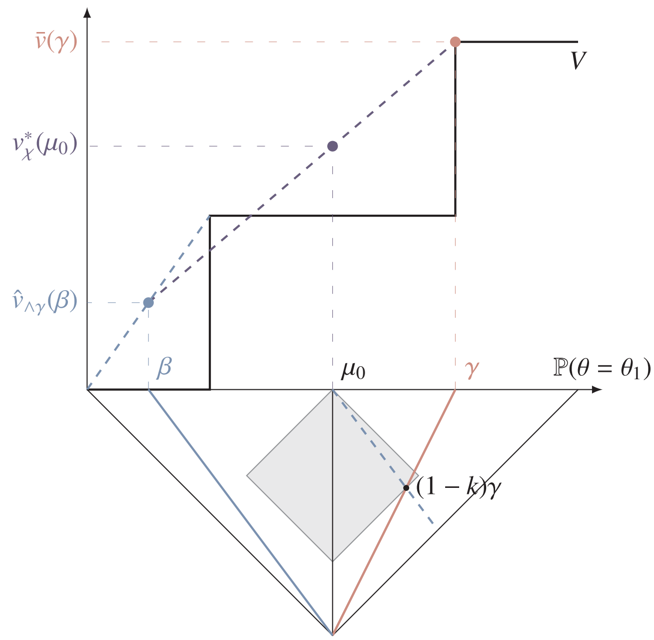

Figure 3: Sender's value

say that an information policy $p \in \mathcal{I}(\mu)$ is a **show or best** (SOB) policy if it is supported on $\{\delta_\theta\}_{\theta\in\Delta\Theta} \cup \text{argmax}_{\mu'\in\Delta[\text{supp}(\mu)]} v(\mu')$. So p is an SOB policy if it either shows the state to R, or brings R to a posterior that attains S's best feasible value. S is **an SOB** at some prior $\mu \in \Delta\Theta$ if every $p \in \mathcal{I}(\mu)$ is outperformed by an SOB policy $p' \in \mathcal{I}(\mu)$—i.e. $\int_{\Delta\Theta} v \, \mathrm{d}p \leq \int_{\Delta\Theta} v \, \mathrm{d}p'$. Figure 4 depicts an example in which S is an SOB for all priors. This example is identical to our introductory example, but with S preferring action 1 to action 2. Notice that productive mistrust cannot happen in this example. This is because, as credibility declines, S's favorite equilibrium policy switches from a binary message policy in which the bad message shows the state to no information. Given this, R prefers a more credible S.

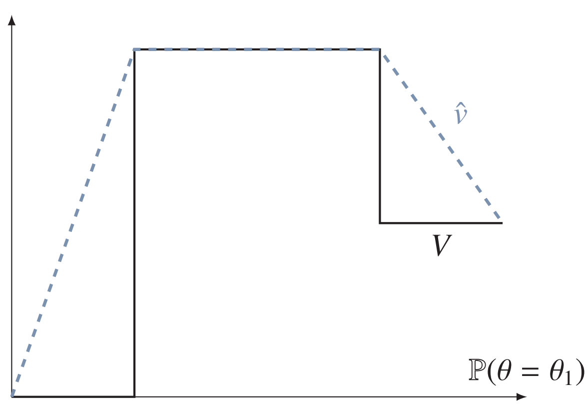

Figure 4: Sender is an SOB for all priors

Proposition 1 below shows that S not being an SOB for some binary-support belief, combined with a technical condition, is sufficient for productive mistrust to occur for some full-support prior.[6] Intuitively, S being an SOB means that a highly credible S has no information to hide: under full credibility, S's bad messages are maximally informative subject to keeping R's posterior following S's good messages fixed. S not being an SOB at some prior means that S's bad messages optimally hide some instrumental information. By reducing S's credibility just enough to make the full credibility solution infeasible, one can force S to reveal some of that information to R. In other words, S commits to potentially revealing more extreme bad information in order to preserve the perceived credibility of her good messages. Proposition 1 below formalizes this intuition.

Proposition 1. *Suppose there are two beliefs with identical binary support, μ, $\mu' \in \Delta\{\theta_1, \theta_2\}$, such that S is not an SOB at μ, and $V(\mu') = \{\max v\,(\Delta\{\theta_1, \theta_2\})\}$. Then there exists a full-support prior μ_0 and credibility levels $\chi' < \chi$ such that every sender-optimal χ'-equilibrium is both strictly better for R—and, when $|\Theta| = 2$, more Blackwell-informative—than every sender-optimal χ-equilibrium.*

To understand the proof, consider first the binary-state case. In this case, our sufficient conditions imply that $\bar{v}$'s maxmizers include some interior beliefs, but do not include all beliefs at which $\hat{v}$ has a kink. Fixing a prior near $\bar{v}$'s maximizers but toward the nearest kink, we then find the lowest χ under which S still obtains her full credibility value. At this χ, S's favorite equilibrium information policy is unique and is supported on the unique β and γ that solve Theorem 1's problem. While γ remains optimal in Theorem 1's problem for any additional small reduction in credibility, the optimal β moves away from the prior—which it must do in order to preserve constraint (χ-BP) given γ. Relying on the set of beliefs being one-dimensional, we show that the only incentive compatible way of attaining S's new optimal value is to spread the original β between γ and an even further posterior that gives S a lower continuation value than under β; this is an increase in informativeness. Finally, to see the payoff ranking, S's continuation value being strictly lower means that R's optimal behavior is different. That is, the additional information is instrumental, strictly increasing R's utility. Figure 5 illustrates the argument using our introductory example.

The proposition's statement for the many-state model builds on the result for only two states. In particular, using the binary-state logic, one can always obtain a binary-support prior

[6] The technical condition holds if $u_R(a,\theta) \neq u_R(a',\theta)$ for all distinct $a, a' \in A$ and all $\theta \in \Theta$, and $\frac{u_R(a_1,\theta_1)-u_R(a_2,\theta_1)}{u_R(a_1,\theta_2)-u_R(a_2,\theta_2)} \neq \frac{u_R(a_2,\theta_1)-u_R(a_3,\theta_1)}{u_R(a_2,\theta_2)-u_R(a_3,\theta_2)}$ for every distinct $a_1, a_2, a_3 \in A$ and distinct $\theta_1, \theta_2 \in \Theta$. In particular, this holds for any $u_S \in \mathbb{R}^A$ and generic $u_R \in \mathbb{R}^{A\times\Theta}$.

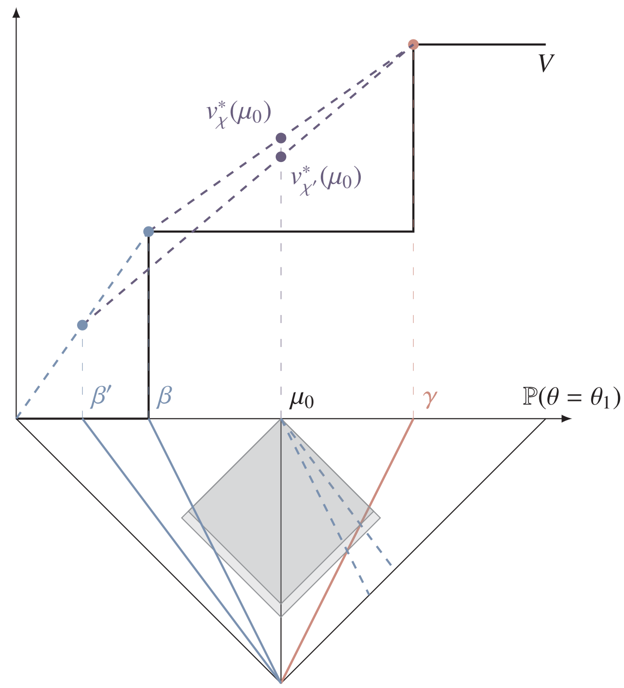

Figure 5: Productive mistrust

μ_0^∞ and credibility levels $\chi' < \chi$ such that R strictly prefers every S-optimal χ'-equilibrium to every S-optimal χ-equilibrium. Using our technical condition, we find an interior direction through which to approach μ_0^∞ while keeping S's optimal equilibrium value under both credibility levels continuous; some care is required here because $\bar{v}$ is discontinuous. The continuity in S's value from this direction then ensures upper hemicontinuity of S's optimal equilibrium policy set—i.e., the limit of every sequence of S-optimal equilibrium policies from said direction must be optimal under μ_0^∞ as well. Now, if the proposition were false, one could construct a convergent sequence of S-optimal equilibrium policies from said direction for each credibility level, $\{p_n^\chi, p_n^{\chi'}\}_n$, such that R weakly prefers p_n^χ to $p_n^{\chi'}$. As R's payoffs are continuous, R being better off under χ than under χ' along the sequences implies the same at the sequences' limits. Notice, though, these limits are S-optimal equilibrium policies for the prior μ_0^∞ by the choice of direction. But this would contradict the productive mistrust result at prior μ_0^∞.

We should emphasize that both of the proposition's conditions are far from necessary. We provide necessary and sufficient condition for productive mistrust to occur at a given prior for the binary-state case in the appendix. In particular, we weaken the SOB condition by only requiring S to want to withhold information at the lowest possible credibility level that allows S to beat her no credibility payoff. We refer the reader to Lemma 2 in the appendix for precise details.

4 Collapse of Trust

The current section examines what happens to S's value as her credibility decreases. As one might suspect, Theorem 1 immediately implies that, the less credible S is, the lower her value.[7] We show that this decrease is often discontinuous. In other words, small decreases in S's credibility often result in a large drop in S's benefits from communication (Proposition 2). Given the prevalence of such discontinuities, it is natural to wonder when such discontinuities happen at full credibility, i.e., at $\chi = 1$. The answer turns out to be almost never (Proposition 3). Thus, while small decreases in credibility often lead to a collapse in S's value, these collapses rarely occur at full credibility.

The example in the introduction demonstrates that S's value can collapse discontinuously when credibility decreases. Formally, such collapses correspond to $\lim_{\epsilon \searrow 0} v^*_{\chi-\epsilon}(\mu_0) < v^*_{\chi}(\mu_0)$. Proposition 2 below establishes that such collapses occur in most examples. In particular, such collapses are absent if and only if S wants to tell R all that she knows, or if, equivalently, commitment is immaterial to S.

Proposition 2. *The following are equivalent:*

(i) A collapse of trust never occurs: $\lim_{\epsilon \searrow 0} v^*_{\chi-\epsilon}(\mu_0) = v^*_{\chi}(\mu_0)$ *for every* $\chi \in (0, 1]$ *and every full-support prior* μ_0.

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

(iii) There is no conflict: $v(\delta_\theta) = \max v(\Delta\Theta)$ *for every* $\theta \in \Theta$.

Two of the proposition's three implications are immediate. First, whenever there is no conflict, S can reveal the state while obtaining her first best payoff (given R's incentives) in an incentive compatible way, meaning that commitment is of no value: (iii) implies (i). Second, since S's highest equilibrium value increases with her credibility, commitment being useless means that S's best equilibrium value is constant (and, a fortiori, continuous) in the credibility level: (ii) implies (i).

To show that (i) implies (iii), we show that any failure of (iii) implies the failure of (i). Consider some full-support prior μ_0 at which $\bar{v}$ is minimized. Notice that the existence of

[7] It also implies that value *increases* have a continuous payoff effect: a small increase in S's credibility never results in a large gain in S's benefits from communication.

conflict implies $\bar{v}$ is nonconstant, and so takes values strictly greater than $\bar{v}(\mu_0)$. Appealing to Theorem 1, one has that $v^*_\chi(\mu_0) > \bar{v}(\mu_0)$ if and only if there is some feasible (k, β, γ) with $k < 1$ such that $\bar{v}(\gamma) > \bar{v}(\mu_0)$. Using upper semicontinuity of $\bar{v}$, we show that such a (k, β, γ) is feasible if and only if S's credibility level is weakly greater than some strictly positive χ^*.[8] We therefore obtain $v^*_{\chi^*}(\mu_0) \geq k\bar{v}(\mu_0) + (1-k)\bar{v}(\gamma) > \bar{v}(\mu_0) = v^*_{\chi^*-\epsilon}$ for all $\epsilon > 0$—i.e., a collapse of trust occurs.

Figure 6 below illustrates the argument in the context of our leading example. The figure depicts a prior as described in the above proof, i.e., a prior which minimizes the payoff S would obtain under no credibility. The depicted constraint set is drawn for χ^*, the lowest credibility level for which there is a feasible (k, β, γ) satisfying both $k < 1$ and $\bar{v}(\gamma) > \bar{v}(\mu_0)$. At χ^*, S's value is strictly above her cheap talk payoff. At any credibility level lower than χ^*, however, S's payoff is equal to the no credibility payoff, since such credibility levels imply that any feasible γ minimizes $\bar{v}$. Thus, at μ_0, S's best equilibrium value collapses whenever her credibility decreases below χ^*.

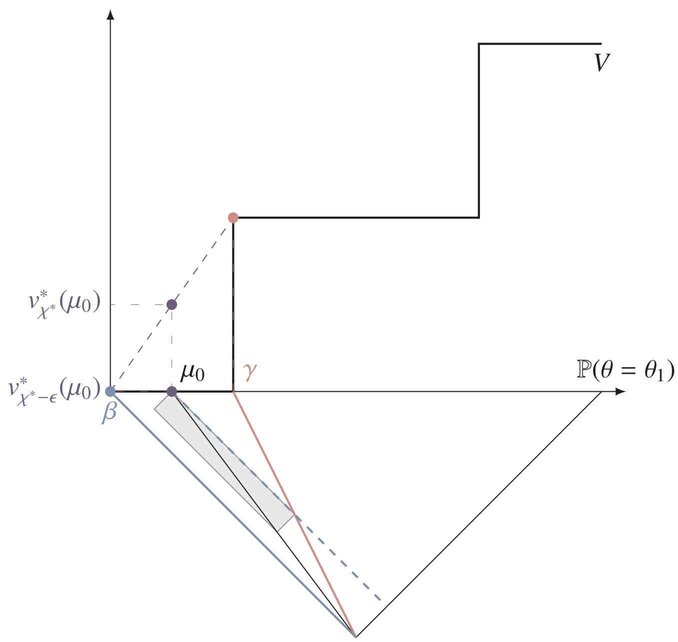

Figure 6: Collapse of trust

Given the large and growing literature on optimal persuasion with full commitment, one may wonder to what extent S's value relies on full credibility. Said differently, when is the full commitment solution robust to a small decrease in S's credibility? Proposition 3 below

[8] If χ^* were not strictly positive we would have $\bar{v}(\mu_0) < v^*_{\chi^*}(\mu_0) = v^*_0(\mu_0) = \bar{v}(\mu_0)$, a contradiction.

shows that the full commitment value is robust (that is, a collapse of trust never occurs at full credibility) if and only if S can persuade R to take her favorite action without ruling out any states – i.e., S gets the benefit of the doubt.

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) S gets the benefit of the doubt: Every $\theta \in \Theta$ *admits some* $\mu \in \operatorname{argmax}_{\mu \in \Delta\Theta} v(\mu)$ *which has* θ *in its support.*

For intuition, we first note that Lipnowski and Ravid's (2017) Theorem 2 allows us to obtain an equivalence between S getting the benefit of the doubt and $\bar{v}$ being maximized at some full-support γ. Now fix some full-support μ_0, and consider two questions about Theorem 1's program. First, which beliefs can serve as γ for all $\chi < 1$ large enough? Second, how do the optimal (k, β) for a given γ change as χ goes to 1? Figure 7 illustrates the answer to both questions for the two state case.

(a) The set of feasible $(1-k)\gamma$ as $\chi \to 1$ (b) (k, β) converges to $(1, \mu_0)$

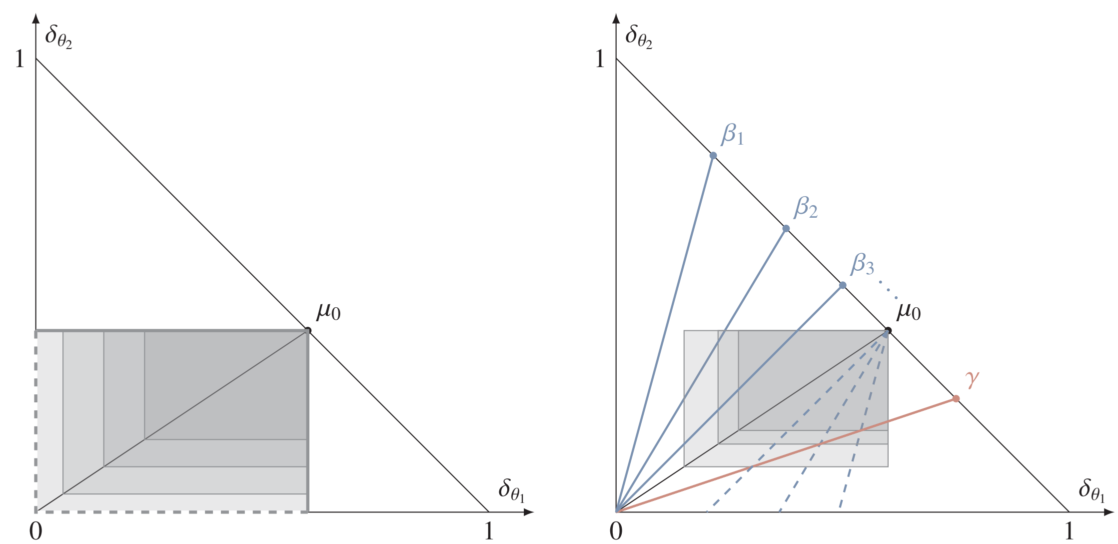

Figure 7: Robustness to limited credibility

For the first question, the answer is that γ is feasible for some $\chi < 1$ if and only if γ has full support. For the second question, recall that it is always optimal to choose (k, β) so as to make (χ-BP) bind while still satisfying (R-BP). Direct computation reveals that (k, β) must converge to $(1, \mu_0)$. Combined, one obtains that, as χ goes to 1, S's optimal value converges to $\max_{\gamma\in \text{int}(\Delta\Theta)} \hat{v}_{\wedge\gamma}(\mu_0)$. Thus, S's value is robust to limited credibility if and only if there is some full-support γ for which $\hat{v}_{\wedge\gamma} = \hat{v}$ for all full-support priors. Clearly, $\hat{v}_{\wedge\gamma}$ and $\hat{v}$ agree whenever γ maximizes $\bar{v}$. Moreover, for any γ that does not maximize $\bar{v}$, one can approach a global maximizer of $\bar{v}$ from $\Delta\Theta$'s interior to find a full-support prior over which $\hat{v}_{\wedge\gamma}$ and $\hat{v}$ disagree. This is illustrated in Figure 7a. To conclude, there is a four-way equivalence between: (1) robustness to limited credibility; (2) $\hat{v}_{\wedge\gamma}$ and $\hat{v}$ agreeing for all full-support priors; (3) $\bar{v}$ being maximized by a full-support γ; and (4) S getting the benefit of the doubt. The proposition follows.

5 Other Equilibrium Values

Section to be written.

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

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

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

(ii) $(1-k) \int_{\Delta\Theta} \mu \, dg(\mu) \geq (1-\chi)\mu_0$;

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

(iv) $(1-\chi)\, s_i + \chi s_o \in (1-k)\, s_i + k \int_{supp(b)} s_i \wedge V \, db$.[9]

6 Final Remarks

Section to be written.

[9]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 supp(b). The integral is the (Aumann) integral of a correspondence:

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

7 Appendix: Proofs

7.1 Proofs from Sections 5 and 2: Equilibrium Values

The results of Sections 5 and 2 hold in a generalization of our model. For this subsection, we do not assume A and Θ to be finite; we instead assume that they are compact metrizable and the objectives u_S and u_R are continuous.[10]

To present unified proofs, we adopt the notational convention that $\frac{k}{\chi} = 1$ if $k = \chi = 0$.

7.1.1 Proof of Lemma 1

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

Define $D := \text{supp}(p)$, $\beta := \int_{\Delta\Theta} \mu \, \mathrm{d}b(\mu)$, and $\gamma := \int_{\Delta\Theta} \mu \, \mathrm{d}g(\mu)$; and let primary research be given by $\rho : \Theta \to \Delta\Omega$ with $\rho(\theta|\theta) = 1 \; \forall \theta \in \Theta$. Let measurable $\eta_g, \eta_b : \Theta \to \Delta\Delta\Theta$ be signals that induce belief distribution g for prior γ and belief distribution b for prior β, respectively.[11] That is, for every Borel $\hat{\Theta} \subseteq \Theta$ and $\hat{D} \subseteq \Delta\Theta$,

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

Take some Radon-Nikodym derivative $\frac{\mathrm{d}\beta}{\mathrm{d}\mu_0} : \Theta \to \mathbb{R}_+$; changing it on a μ_0-null set, we may assume that $0 \leq \frac{k}{\chi} \frac{\mathrm{d}\beta}{\mathrm{d}\mu_0} \leq 1$. Then, fixing $\hat{\theta} \in \Theta$, extend every map $f \in \left\{\eta_g, \; \eta_b, \; \frac{\mathrm{d}\beta}{\mathrm{d}\mu_0}\right\}$ to a map $\Omega \to \Delta\Delta\Theta$ by letting $f(\omega) = f(\hat{\theta})$ for every $\omega \in \Omega \setminus \Theta$.

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

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

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

[10] Of course, continuity would hold vacuously with A and Θ being (discrete) finite spaces.

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

receiver belief map as

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

Finally, by Lipnowski and Ravid (2017, Lemma 2), there are some measurable $\alpha_b, \alpha_g : \text{supp}(p) \to \Delta A$ such that:[12]

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

From these, we can define a receiver strategy as

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

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

To see that the Bayesian property holds, observe that every Borel $\hat{D} \subseteq D$ satisfies

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

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

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

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

verifying the Bayesian property. So $(\rho, \xi, \sigma, \alpha, \pi)$ is a χ-equilibrium. Moreover, for any Borel $\hat{D} \subseteq \Delta\Theta$, the equilibrium probability of the receiver posterior belief belonging to $\hat{D}$ is exactly (specializing the above algebra)

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

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

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

as required.

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

Let $\mathbb{O} := \int_\Theta \rho\, \mathrm{d}\mu_0 \in \Delta\Omega$ be the ex-ante distribution over research outcomes, and

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

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

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

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

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

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

Now viewing π, α as random variables on the probability space $\langle M, P \rangle$, define the conditional expectation $\phi_0 := \mathbb{E}_B[u_S(\alpha)|\pi] : M \to \mathbb{R}$. By Doob-Dynkin, there is a measurable function $\phi : \Delta\Theta \to \mathbb{R}$ such that $\phi \circ \pi =_{B\text{–a.e.}} \phi_0$. As $u_S(\alpha(m)) \in s_i \wedge V(m)$ for every $m \in M$, and the correspondence $s_i \wedge V$ is compact- and convex-valued, it must be that $\phi_0 \in_{B\text{–a.e.}} s_i \wedge V(\pi)$. Therefore, $\phi \in_{b\text{–a.e.}} s_i \wedge V$. Modifying ϕ on a b-null set, we may assume without loss that ϕ is a measurable selector of $s_i \wedge V$.

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

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

Therefore, if $\chi > 0$,

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

If $\chi = 0$, then

$$s_o = \int_M u_S \circ \pi \, \mathrm{d}B = \int_{\mathrm{supp}(b)} \phi \, \mathrm{d}b = \left(1-\tfrac{k}{\chi}\right)s_i + \tfrac{k}{\chi}\int_{\mathrm{supp}(b)} \phi \, \mathrm{d}b,$$

as required. □

7.1.2 Proof of Theorem 1

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

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

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

Therefore,

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

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 (2017, Lemma 1 and Theorem 2) tell us that any $\gamma \in \Delta\Theta$ has

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

where $\bar{v}$ is the quasiconcave envelope of v.[13] Therefore,

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

<u>Claim:</u> If $b \in \Delta\Delta\Theta$, $\gamma \in \Delta\Theta$, and $k \in [0, 1]$ satisfy $k\int_{\Delta\Theta} \mu \, \mathrm{d}b(\mu) + (1-k)\gamma = \mu_0$ and $(1-k)\gamma \geq (1-\chi)\mu_0$, then there exists (b^*, γ^*, k^*) feasible in the above program[14] such that $(1-k^*)\bar{v}(\gamma^*) + k^*\int_{\Delta\Theta} \bar{v}(\gamma^*) \wedge v \, \mathrm{d}b^* \geq (1-k)\bar{v}(\gamma) + k\int_{\Delta\Theta} \bar{v}(\gamma) \wedge v \, \mathrm{d}b$.

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

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

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

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

In this case, Lipnowski and Ravid (2017, Lemma 3) delivers some $\beta^* \in \text{co}\{\beta, \mu_0\}$ such

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

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

that $V(\beta^*) \ni \bar{v}(\gamma)$. But then $\mu_0 \in \text{co}\{\beta^*, \gamma\}$. As $\bar{v}$ is quasiconcave, $\bar{v}(\mu_0) \geq \min\{\bar{v}(\beta^*), \bar{v}(\gamma)\} \geq \min\{v(\beta^*), \bar{v}(\gamma)\} = \bar{v}(\gamma)$.

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

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

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

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

Given such a measure, (b^*, γ, k) will be as required. We now construct such a b^* by modifying the proof of Lipnowski and Ravid (2017, Theorem 1).

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

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

Lipnowski and Ravid (2017, Lemma 3) tells us that $\bar{v}(\gamma) \in V([1-\lambda(\mu)]\gamma + \lambda(\mu)\mu)$ for every $\mu \in D$ for which $v(\mu) > \bar{v}(\gamma)$. This implies that $\min V([1-\lambda(\mu)]\gamma + \lambda(\mu)\mu) \leq \bar{v}(\gamma)$ for *every* $\mu \in D$.

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

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

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

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

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

proving the claim.

In light of the claim, the optimal value is

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

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

7.1.3 Consequences of Lemma 1 and Theorem 1

Corollary 1. *The set of χ-equilibrium outcomes (p, s_o, s_i) at prior μ_0 is an upper hemicontinuous correspondence of (μ_0, χ).*

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

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

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

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

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

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

is weakly increasing and right-continuous.

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

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

Proof. This is immediate from Corollary 1. □

7.2 Proofs from Section 3: Productive Mistrust

Toward verifying our sufficient conditions for productive mistrust to occur, we study in some depth the possibility of productive mistrust in the binary-state world.

7.2.1 Productive Mistrust with Binary States

Given binary states and a full-support prior μ_0, we know that the quasiconcave envelope function $\bar{v} : \Delta\Theta \to \mathbb{R}$ is upper semicontinuous, weakly quasiconcave, and piecewise constant. Therefore, if $\mu_0 \notin \operatorname{argmax}_{\mu\in\Delta\Theta}\bar{v}(\mu)$, there is then a unique $\mu_+ = \mu_+(\mu_0)$ closest to μ_0 with the property that $\bar{v}(\mu_+) > \bar{v}(\mu_0)$, and a unique $\theta = \theta(\mu_0) \in \Theta$ with $\mu_0 \in \text{co}\{\mu_+(\mu_0), \delta_\theta\}$. In this case, for the rest of the subsection, we identify $\Delta\Theta \cong [0, 1]$ by identifying $\nu \in \Delta\Theta$ with $1 - \nu(\theta(\mu_0))$.[15]

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

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

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

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

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

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

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

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

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

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

For case (b), we need only consider the case of $\chi = 1$. In case (b), that $\bar{v}$ is weakly quasiconcave implies it is monotonic. So $\mu_+ = 1$, and $\bar{v} : [0, 1] \to \mathbb{R}$ is nondecreasing with $\bar{v}|_{[\mu_0,1)} = \bar{v}(\mu_0) < \bar{v}(1)$. As v_1^* is the concave envelope of $\bar{v}$, it must be that the support of any S-optimal 1-equilibrium information policy is contained in $[0,\ \min\{\mu \in [0, 1] :\ v(\mu) = v(\mu_0))\}] \cup \{1\}$, so that (1) fails as well.

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

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

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

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

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

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

Conversely, suppose (2) holds.

The function $v : [0,1] \to \mathbb{R}$ is upper semicontinuous and piecewise constant, which implies that its concave envelope v^*_1 is piecewise affine. We may then define

$$\mu^*_- := \min\{\mu \in [0, \mu_0] : \ v^*_1 \text{ is affine over } [\mu, \mu_0]\}.$$

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

There are thus beliefs $\mu_-, \mu_+ \in [0,1]$ such that: $0 < \mu_- < \mu_0 < \mu_+$; v^*_1 is affine on $[\mu_-, \mu_+]$ and on no larger interval; and v^*_1 is strictly increasing on $[0, \mu_+]$. It follows that $\hat{v}_{\wedge\mu_+} = v^*_1$ on $[0, \mu_+]$. By definition of $\mu_+ = \mu_+(\mu_0)$, we know that $\bar{v}$ is constant on $[\mu_0, \mu_+)$. That is, (appealing to Lipnowski and Ravid (2017, Theorem 2)) v^*_0 is constant on $[\mu_0, \mu_+)$. Then, since v^*_1 strictly decreases there, it must be that $v^*_1 > v^*_0$ on (μ_0, μ_+).

Let $\chi \in [0,1]$ be the smallest credibility level such that $v^*_\chi(\mu_0) = v^*_1(\mu_0)$, which exists by Corollary 2. That $v^*_0(\mu_0) < v^*_1(\mu_0)$ implies $\chi > 0$. That μ_+ has full support, which follows from (2), implies that $\chi < 1$.[16]

Consider now the following claim.

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

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

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

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

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

We now prove the claim.

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

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

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

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

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

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

We now prove the second bullet. First, every $\mu < \mu_+$ satisfies $v(\mu) \leq v_1^*(\mu) < v_1^*(\mu_+) = v(\mu_+)$. This implies that δ_{μ_+} is the unique $\ell \in \mathcal{I}(\mu_+)$ with $\inf v(\operatorname{supp}\ell) \geq v(\mu_+)$. Therefore, $\ell' = \delta_{\mu_+}$.

Second, the measure $h' \in \mathcal{I}(\beta')$ can be expressed as $h' = (1-\gamma)h_L + \gamma h_R$ for $h_L \in \Delta[0, \mu_-]$, $h_R \in \Delta(\mu_-, 1]$, and $\gamma \in [0, 1)$. Notice that $(\mu_-, v(\mu_-))$ is an extreme point of the subgraph of

v_1^*, and therefore an extreme point of the subgraph of $\hat{v}_{\wedge\mu_+}$. Taking the unique $\hat{\gamma} \in [0, \gamma]$ such that $\hat{h} := (1-\hat{\gamma})h_L + \hat{\gamma}\delta_{\mu_-} \in \mathcal{I}(\beta')$, it follows that $\int_{[0,1]} \hat{v}_{\wedge\mu_+} \, \mathrm{d}\hat{h} \geq \int_{[0,1]} \hat{v}_{\wedge\mu_+} \, \mathrm{d}h'$, strictly so if $\hat{\gamma} < \gamma$. But $\hat{\gamma} < \gamma$ necessarily if $\gamma > 0$, since $\int_{[0,1]} \mu \, \mathrm{d}h_R(\mu) > \mu_-$. Optimality of h' then implies that $\gamma = 0$, i.e. $h'[0, \mu_-] = 1$. This completes the proof of the claim.

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

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

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

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

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

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

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

Having shown that (2) implies (1), all that remains is to show that the receiver's optimal payoff is strictly higher given p' than given p^*. To that end, fix sender-preferred receiver best responses a_- and a_+ to μ_- and μ_+, respectively. As the receiver's optimal value given p^* is attainable using only actions $\{a_-, a_+\}$, and the same value is feasible given only information p' and using only actions $\{a_-, a_+\}$, it suffices to show that there beliefs in the support of p' to which neither of $\{a_-, a_+\}$ is a receiver best response. But, at every $\mu \in [0, \mu_-)$ satisfies

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

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

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

Lemma 3. *Suppose $|\Theta| = 2$, and there are two full-support beliefs $\mu, \mu' \in \Delta\Theta$, such that the sender is not an SOB at μ, and $V(\mu') = \{\max v(\Delta\Theta)\}$. Then there exists a full-support prior μ_0 and credibility levels $\chi' < \chi$ such that every S-optimal χ'-equilibrium is both strictly better for R and more Blackwell-informative than every S-optimal χ-equilibrium.*

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

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

7.2.2 Productive Mistrust with Many States: Proof of Proposition 1

Given Lemma 3, we need only prove the proposition for the case of $|\Theta| > 2$, which we do below.

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

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

Consider the following claim.

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

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

Before proving the claim, let us argue that it implies the proposition. Given the claim, assume for contradiction that: for every $n \in \mathbb{N}$, prior μ_0^n admits some S-optimal χ'-equilibrium and χ''-equilibrium, $\Psi'_n = (p'_n, s'_{in}, s'_{on})$ and $\Psi''_n = (p''_n, s''_{i\,n}, s''_{o\,n})$, respectively, such that $v_R(p'_n) \ge$

$v_R(p''_n)$. Dropping to a subsequence if necessary, we may assume by compactness that $(\Psi'_n)_n$ and $(\Psi''_n)_n$ converge (in $\Delta\Delta\Theta \times \mathbb{R} \times \mathbb{R}$) to some $\Psi' = (p', s'_i, s'_o)$ and $\Psi'' = (p'', s''_i, s''_o)$ respectively. By Corollary 1, for every credibility level χ, the set of χ-equilibria is an upper hemicontinuous correspondence of the prior. Therefore, Ψ' and Ψ'' are χ'- and χ''-equilibria, respectively, at prior μ_0^∞. Continuity of v_R (by Berge's theorem) then implies that $v_R(p') \geq v_R(p'')$. Finally, by the claim, it must be that Ψ' and Ψ'' are S-optimal χ'- and χ''-equilibria, respectively, contradicting the definition of μ_0^∞. Therefore, there is some $n \in \mathbb{N}$ for which the full-support prior μ_0^n is as required for the proposition.

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

First, the proof of Lemma 3 delivers some $\gamma^\infty \in \Delta\Theta$ such that $\bar{v}(\gamma^\infty) = u$ and, for both $\chi \in \{\chi', \chi''\}$, some $(\beta, \gamma, k) \in \Delta\Theta \times \{\gamma^\infty\} \times [0,1]$ solves the program in Theorem 1 at prior μ_0^∞.

Let us now show that there exists a closed convex set $D \subseteq \Delta\Theta$ which contains γ^∞, has nonempty interior, and satisfies $\bar{v}|_D = u$. Indeed, for any $n \in \mathbb{N}$, let $B_n \subseteq \Delta\Theta$ be the closed ball (say with respect to the Euclidean metric) of radius $\frac{1}{n}$ around μ', and let $D_n := \text{co}\,[\{\gamma^\infty\} \cup B_n]$. As $v|_{\Delta\Theta_2} \leq u$ and constant functions are quasiconcave, Lipnowski and Ravid (2017, Theorem 2) tells us $\bar{v}|_{\Delta\Theta_2} \leq u$ as well. As V is upper hemicontinuous, the hypothesis on μ' ensures that $\bar{v}|_{B_n} \geq v|_{B_n} = u$ for sufficiently large $n \in \mathbb{N}$; quasiconcavity then tells us $\bar{v}|_{D_n} \geq u$. Assume now, for a contradiction, that every $n \in \mathbb{N}$ has $\bar{v}|_{D_n} \not\leq u$. That is, there is some $\lambda_n \in [0,1]$ and $\mu'_n \in B_n$ such that $\bar{v}\left((1-\lambda_n)\mu + \lambda_n \mu'_n\right) > u$. Dropping to a subsequence, we get a strictly increasing sequence $(n_\ell)_\ell^\infty$ of natural numbers such that (since $[0,1]$ is compact and $\bar{v}(\Delta\Theta)$ is finite) $\lambda_{n_\ell} \xrightarrow{\ell\to\infty} \lambda \in [0,1]$ and $\bar{v}\left((1-\lambda_{n_\ell})\mu + \lambda_{n_k}\mu'_{n_k}\right) = \hat{u}$ for some number $\hat{u} \in (u, \infty)$ and every $\ell \in \mathbb{N}$. As $\bar{v}$ is upper semicontinuous, this would imply that $\bar{v}\left((1-\lambda)\mu + \lambda\mu'\right) \geq \hat{u} > u$, contradicting the definition of u. Therefore, some $D \in \{D_{n_\ell}\}_{\ell=1}^\infty$ is as desired. In what follows, let $\gamma_1 \in D$ be some interior element with full support.

Now, for each $n \in \mathbb{N}$, define $\mu_0^n := \frac{n-1}{n}\mu_0^\infty + \frac{1}{n}\gamma_1$. We will show that the sequence $(\mu_0^n)_{n=1}^\infty$—a sequence of full-support priors converging to μ_0^∞—is as desired. To that end, fix $\chi \in \{\chi', \chi''\}$ and some $(\beta, k) \in \Delta\Theta \times [0,1]$ such that $(\beta, \gamma^\infty, k)$ solves the program in Theorem 1 at prior μ_0^∞. Then, for any $n \in \mathbb{N}$, let:

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

Given these definitions,

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

Therefore, (β,γ_n,k_n) is χ-feasible at prior μ_0^n. As a result,

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

This proves the claim, and so too the proposition. □

7.3 Proofs from Section 4: Collapse of Trust

7.3.1 Proof of Proposition 2

Proof. Two of three implications are easy given Corollary 2. First, if there is no conflict, then Lipnowski and Ravid (2017, Lemma 1) tells us that there is a 0-equilibrium with full information that generates sender value $\max v(\Delta\Theta) \geq v_1^*$; in particular, $v_0^* = v_1^*$. Second, if $v_0^* = v_1^*$, then v_χ^* is constant in χ, ruling out a collapse of trust. Below we show that any conflict whatsoever implies a collapse of trust.

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

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

and $K(\hat{\chi})$ be its projection onto its last coordinate. As the correspondence Γ is upper hemicontinuous and decreasing (with respect to set containment), K inherits the same properties.

Next, notice that $K(1) \ni 1$ (as v is nonconstant by hypothesis, so that $\Delta\Theta \neq D$) and $K(0) = \emptyset$ (as $\mu_0 \in D$). Therefore, $\chi := \min\{\hat{\chi} \in [0,1] : K(\hat{\chi}) \neq \emptyset\}$ exists and belongs to $(0,1]$.

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

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

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

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

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

7.3.2 Proof of Proposition 3

Proof. By Lipnowski and Ravid (2017, Lemma 1 and Theorem 2), S gets the benefit of the doubt (i.e. every $\theta \in \Theta$ is in the support of some member of $\text{argmax}_{\mu \in \Delta\Theta} v(\mu)$) if and only if there is some full-support $\gamma \in \Delta\Theta$ such that $\bar{v}(\gamma) = \max v(\Delta\Theta)$.

First, given a full-support prior μ_0, suppose $\gamma \in \Delta\Theta$ is full-support with $\bar{v}(\gamma) = \max v(\Delta\Theta)$. It follows immediately that $\hat{v}_{\wedge\gamma} = \hat{v} = v^*_1$.

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

$$\begin{aligned}
v_\chi^*(\mu_0) &\geq \sup_{\beta \in \Delta\Theta,\ k \in [0,1]} \left\{ k v_1^*(\beta) + (1-k) v(\gamma) \right\} \\
&\qquad \text{s.t.} \quad k\beta + (1-k)\gamma = \mu_0,\ (1-k)\gamma \geq (1-\chi)\mu_0 \\
&= \sup_{k \in [0,1]} \left\{ k v_1^* \left(\tfrac{\mu_0 - (1-k)\gamma}{k} \right) + (1-k) v(\gamma) \right\} \\
&\qquad \text{s.t.} \ (1-\chi)\mu_0 \leq (1-k)\gamma \leq \mu_0 \\
&\geq \sup_{k \in [0,1]} \left\{ k v_1^* \left(\tfrac{\mu_0 - (1-k)\gamma}{k} \right) + (1-k) v(\gamma) \right\} \\
&\qquad \text{s.t.} \ (1-\chi) r_1 \leq (1-k) \leq r_0 \\
&\geq \sup_{k \in [0,1]} \left\{ k v_1^* \left(\tfrac{\mu_0 - (1-k)\gamma}{k} \right) + (1-k) v(\gamma) \right\} \\
&\qquad \text{s.t.} \ (1-\chi) r_1 = (1-k) \\
&= \left[1 - (1-\chi) r_1\right] v_1^* \left(\tfrac{\mu_0 - (1-\chi) r_1 \gamma}{1 - (1-\chi) r_1} \right) + (1-\chi) r_1 v(\gamma).
\end{aligned}$$

But notice that v_1^*, being a concave function on a finite-dimensional space, is continuous on the interior of its domain. It follows that $v_\chi^*(\mu_0) \to v_1^*(\mu_0)$ as $\chi \to 1$. That is, persuasion is robust to limited commitment.

Conversely, suppose that S does not get the benefit of the doubt. Taking an affine transformation of u_S, we may assume without loss that $\max v(\Delta\Theta) = 1$ and (since $v(\Delta\Theta) \subseteq u_S(A)$ is finite) $\max[\bar{v}(\Delta\Theta) \setminus \{1\}] = 0$.

Consider any full-support prior μ_0 and any $\chi \in [0,1)$. For any $\beta, \gamma \in \Delta\Theta,\ k \in [0,1]$ with $k\beta + (1-k)\gamma = \mu_0$ and $(1-k)\gamma \geq (1-\chi)\mu_0$, that S does not get the benefit of the doubt implies (say by Lipnowski and Ravid (2017, Theorem 1)) that $\bar{v}(\gamma) \leq 0$, and therefore that $k\hat{v}_{\wedge\gamma}(\beta) + (1-k)v(\gamma) \leq 0$. Theorem 1 then implies that $v_\chi^*(\mu_0) \leq 0$.

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

$$v_1^*(\mu_\epsilon) \geq (1-\epsilon)v(\gamma) + \epsilon v(\mu_1) \geq (1-\epsilon) + \epsilon \cdot \min v(\Delta\Theta).$$

For sufficiently small ϵ, then, $v_1^*(\mu_\epsilon) > 0$. Persuasion is therefore not robust to limited commitment at prior μ_ϵ. □

References

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

Blackwell, David. 1953. "Equivalent comparisons of experiments." *The annals of mathematical statistics* 24 (2):265–272.

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

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

Lipnowski, Elliot and Doron Ravid. 2017. "Cheap Talk with Fixed Agenda." Working Paper, Economics Department, University of Chicago.