BROWN
Orlando Bravo Center
for Economic Research

Mediated (Anti)Persuasive Communication*

Bravo Working Paper # 2023-001

Zeky Murra-Anton† Roberto Serrano‡

First version: January 2023; This Version: July 2023

Abstract

Can private information or mediation change a sender's behavior and improve the receiver's expected utility in persuasive communication games? In a mediated Bayesian persuasion model, private information cannot improve the receiver's expected utility when the sender communicates it. When the intermediary communicates the private information, the receiver's expected utility improves only with a positive autarky value of the intermediary's private information (AVIPI), a novel information accuracy measure we propose. Finally, the sender's strategic behavior is generally affected by the intermediary's presence as he tries to persuade the intermediary to, in turn, persuade the receiver.

JEL Classifications: D82, C83, C72

Keywords: communication, intermediation, private information, mitigation, fake news

*We thank Antonio Cabrales, Vincent Crawford, Emir Kamenica, Sam Kapon, Teddy Mekonnen, and Jesse Shapiro for their helpful comments. Serrano thanks Universidad Carlos III in Madrid for its warm hospitality and Fundacion ONCE for research support.

†Market Development, ISO New England. Contact: zmurraanton@iso-ne.com. This paper has not been funded or endorsed by ISO New England. All the views herein expressed are the authors' only.

‡Economics Department, Brown University. Contact: roberto_serrano@brown.edu

Mediated (Anti)Persuasive Communication*

Zeky Murra-Anton† Roberto Serrano‡

First version: January 2023; This Version: July 2023

Abstract

Can private information or mediation change a sender's behavior and improve the receiver's expected utility in persuasive communication games? In a mediated Bayesian persuasion model, private information cannot improve the receiver's expected utility when the sender communicates it. When the intermediary communicates the private information, the receiver's expected utility improves only with a positive *autarky value of the intermediary's private information* (AVIPI), a novel information accuracy measure we propose. Finally, the sender's strategic behavior is generally affected by the intermediary's presence as he tries to persuade the intermediary to, in turn, persuade the receiver.

JEL Classifications: D82, D83, C72

Keywords: communication, intermediation, private information, mitigation, fake news

*We thank Antonio Cabrales, Vincent Crawford, Emir Kamenica, Sam Kapon, Teddy Mekonnen, and Jesse Shapiro for their helpful comments. Serrano thanks Universidad Carlos III in Madrid for its warm hospitality and Fundacion ONCE for research support.

†Market Development, ISO New England. Contact: zmurraanton@iso-ne.com. This paper has not been funded or endorsed by ISO New England. All the views herein expressed are the authors' only.

‡Economics Department, Brown University. Contact: roberto_serrano@brown.edu

1. Introduction

In 2005, an influential anti-vaccine opinion leader published the article "Deadly Immunity," claiming that vaccines cause autism.[1] To persuade care-givers not to vaccinate their children, the author claimed to build a compelling argument, but in fact what he did was to exaggerate spurious medical evidence while suppressing other evidence against his point.[2] It was not until 2011 that the publisher retracted "Deadly Immunity" after journalist Seth Mnookin's published the evidence-based book *The Panic Virus* (Mnookin, 2011) pointing at "Deadly Immunity"'s numerous flaws.

"Deadly Immunity" is an example of a Bayesian Persuasion problem: an agent, the sender (he), strategically provides information to another agent, the receiver (she), to persuade her to take an action that may not be in her best interest. However, like many other applications, including political lobbying, clinical research design, and recommendation-based hiring, "Deadly Immunity" possesses two distinctive features that escape a growing body of research in information design. First, a communication intermediary representing the receiver's interests, à la Seth Mnookin. Second, a source of (possibly imperfect) objective information, like the results of clinical trials, that the receiver learns through other agents.

This paper studies a mediated persuasive communication model with private information unavailable to the receiver. Without an intermediary or the use of private information, the sender can persuade the receiver by selectively obfuscating information, an outcome that typically extracts all the surplus from the receiver (Kamenica and Gentzkow (2011), Bergemann and Morris (2019)). That is, the receiver does not benefit from communication as she is typically left indifferent to just relying on her prior. Can newly available private information or the presence of an intermediary improve the receiver's expected utility? Does the private information's effectiveness depend on who communicates it to the receiver? Can private information or an intermediary change the sender's communication strategy? Our analysis provides an answer to all these questions.

To highlight the main points of our analysis, consider the following stylized version of our model. A sponsor (the sender) wants the public (the receiver) to approve an anti-vaccine bill in a referendum. The public only wishes to approve the bill when it is

[1] Numerous clinical trials have failed to find a causal relationship between vaccines and autism (FDA, 2018).

[2] "Deadly Immunity" was fact-checked by the publisher at least five times.

good, namely, if vaccines cause autism, as approving it otherwise would be harmful. There is a source of objective information, like clinical trials, that noisily signals the true quality of the bill. However, there are information frictions, and the public cannot directly verify the results of those trials, so it must learn about it from other agents, making it private information.[3] To persuade the public to approve the bill, the sponsor writes an article that may contain part of the evidence. Finally, the media (intermediary), representing the public interest, interprets the sponsor's article and recommends to the public how to vote after independently researching the evidence, hence realizing the private information.

To begin with, Theorem 1 shows that, independently of the evidence, without the media, the sponsor persuades the public to approve the bill and leaves it indifferent between following a recommendation and the default action. In other words, private information is not beneficial to the public when directly communicated by the sponsor. The reason is that the sponsor can always find a way to *bend the facts* by suitably picking a communication strategy that obfuscates the private information. As a result, the sponsor simply mimics the case where the private information does not exist: with probability one, the sponsor reveals that the bill is good, when it is actually good for the public, whereas with positive probability he engages in *fake news* (suggesting that the bill is good when it is actually bad).

Our main result, Theorem 2 characterizes the public's equilibrium expected utility as a function of the accuracy of the private information the media privately researches. To that end, we construct a novel measure of private information, the *Autarky Value of the Intermediary's Private Information* (AVIPI). In a nutshell, the AVIPI measures the maximum equilibrium expected utility gain the media can attain for the public over allowing the sender to communicate the private information by making a recommendation based only on the media's private information.

In particular, Theorem 2 shows that the public's expected utility is only as good as the value of the intermediary's private information. A strictly positive AVIPI endows the media with two powerful tools to improve the public's expected utility over direct communication. First, the media can conceive of a mechanism for identifying cases where complementing its private information with the sponsor's recommendation is strictly beneficial to the public, compared to direct communication. Second, the media

[3]For instance, imagine that the typical citizen does not know how to find or interpret scientific evidence.

can credibly consider an experiment that is beneficial for the public consisting of completely shutting off the sponsor's persuasive efforts, namely, ignoring him. As a result, the sponsor understands that part of his recommendation will only reach the public—the media will not shut it down—only if it offers value to the public. An important remark is that the media does not necessarily shut down the sponsor's recommendation in equilibrium, even though the public's expected utility is only equal to the AVIPI. Indeed, we show that when the public's expected utility is strictly greater than the AVIPI, the sponsor can always find an experiment that extracts the surplus over the AVIPI.

Our final complementary results detail a variety of equilibrium behaviors. In Proposition 3, we establish that there is always an equilibrium, independently of the accuracy of the private information, in which the media ignores the sponsor. However, under general conditions, there are no equilibria where the media always recommends accepting the bill, as it creates channels for the sponsor to persuade the public. For similar reasons, there are no equilibria where the media passes the sponsor's recommendation intact to the public, as Proposition 4 establishes. Finally, Proposition 5 finds that, except for the extreme cases in which the private information is uninformative or when the media can perfectly identify a good bill, there are no equilibria in which the sponsor's behavior is unaffected by the media.[4]

Our results have implications for the design of policy to combat misinformation. An important application is social media. For instance, Twitter deals with misleading communication by labeling content, warning users about its flaws, and *pre-bunking*.[5] First, we highlight the relevance of independent communication of private information while holding the receiver's interests. Without it, agents engaging in persuasive communication can find ways to bend the facts and harm the receiver. Second, we highlight the importance of accurate private information. Even independent anti-persuasive efforts fail at protecting the receiver from persuasive communication if the private information is not informative enough in the AVIPI sense. Without precise private information, the sender knows that the intermediary has limited power in identifying persuasive communication and can find a way to extract the receiver's whole surplus over the AVIPI through fake news.

[4]By "unaffected," we mean that the sponsor sticks to Kamenica and Gentzkow's (2011) canonical Bayesian persuasion solution in the absence of a mediator and private information.

[5]Pre-bunking consists of timely or proactively releasing informative messages to counter misleading narratives. It can be done by fact-checking, logic-checking, or source-checking, to name a few.

We organize the rest of the paper as follows. Section 2 presents our model. In Section 3 we present our benchmark results: when the sender communicates the private information to the receiver directly and when an intermediary intervenes in the sender's recommendation by surprise. Section 4 is the core of our paper; in it, we present our equilibrium model, construct the AVIPI, and state our main results. Related literature and our concluding remarks are contained in Section 5. Proofs are relegated to an appendix.

2. Model

Actions and Payoffs

We consider Bergemann and Morris's (2019) two-state, binary-action communication model. The state of the world can either be good or bad, denoted by $t \in T \equiv \{t_g, t_b\}$. A sender (he), a receiver (she), and an intermediary (she) who are uninformed about the true state of the world share a common uniform prior. The receiver must choose an action $a \in A \equiv \{a_1, a_2\}$; we sometimes refer to these two actions as the *nondefault action* and the *default action*, respectively. In Section 5, we discuss how our model can be used in various practical problems and how our main results extend to a model with an arbitrary number of states and actions.

We assume that the intermediary and the receiver share the same preferences. Table 1 summarizes the receiver's (first entry) and sender's (second entry) payoffs from each state-action combination:

Table 1: Receiver's and Sender's Ex-post Payoffs.

	t_g	t_b
a_1	$x \in (0,1), 1$	$-1, 1$
a_2	$0, 0$	$0, 0$

Our framework is standard in information design:[6] the sender has state-independent preferences and under the prior, the receiver's default action a_2 is the sender's least

[6]Our model belongs to the class studied by Gentzkow and Kamenica (2016b) if we represent state t_g by $\omega = 1$, state t_b by $\omega = 0$, and write the receiver's preferences as $v(\omega) = \omega x + (1-\omega)(-1)$.

preferred one. We use the following running example throughout the paper:

Running Example, Part 1 (Fake News) *A specific bill or piece of policy is proposed by a sponsor (sender) to the public (receiver) in a referendum. The quality of the bill can be good (t_g) or bad (t_b).*[7] *When the public approves the bill in the referendum (a_1), it gets a reward of x when it is of good quality (t_g) or a penalty of -1 when it is of bad quality (t_b). When the public rejects the bill (a_2), it gets a reward of zero. Finally, the sponsor is focused on his own agenda, so her objective is to maximize the probability that the bill is approved, regardless of its quality. This opens the door to the possibility of "fake news" as a vehicle to implement that agenda.*

Private Information

There is a source of *private information* concerning the state of the world interpreted as a noisy signal about the true state, denoted by s. Formally, $s : T \to \triangle(T)$, with $s(t' \mid t)$ representing the probability that the realized signal is t' when the true state is t. We denote by $\hat{s}$ a realization of signal s, and sometimes refer to it as the outcome of *researching the evidence*.[8]

The *accuracy* of the private information is the probability that the signal accurately shows the true state of the world, namely $\epsilon_t \equiv s(\hat{s} = t \mid t)$. The private information is a *perfectly accurate signal* when $\epsilon_g = \epsilon_b = 1$, denoted by $\epsilon^A = (1, 1)$. Likewise, it is a *perfectly inaccurate signal* when $\epsilon_g = \epsilon_b = 0$, denoted by $\epsilon^I = (0, 0)$. Useful in our results is the *combined informativeness of the intermediary's signal* (i.e., its departure from a completely uninformative signal), defined as

$$I(\epsilon_g, \epsilon_b) = |\epsilon_g + \epsilon_b - 1|. \tag{1}$$

[7] In this context, "good quality" can mean that the bill is aligned with the social incentives (perhaps with the wishes of the majority of the electorate), whereas "bad quality" can mean that the proposer has private interests, misaligned with the social majority.

[8] Our use of the word "evidence" in this paper is different from the traditional use in the mechanism design literature (Ben-Porath et al., 2021; Perez-Richet and Skreta, 2022). We use the term to refer to scientific evidence in the context of our example, whereas the mechanism design use pertains to verifiable information in the context of games of imperfect information.

Recommendation Experiments

Without loss of generality, we study a game where the sender and intermediary choose recommendation experiments (Perez-Richet and Skreta, 2022).[9] As usual in Bayesian persuasion games, the sender and intermediary choose and commit to recommendation experiments before uncertainty is realized. In our case, this is before agents' private information and the state of the world realize. The experiments have the following characteristics:

- The sender chooses an experiment $\sigma : T \times T \to \triangle(A)$, where $\sigma(\hat{\sigma} \mid \hat{s}, t)$ is the probability of recommending $\hat{\sigma}$ when the sender's realization of the private information is $\hat{s}$ and the true state of the world is t.
- The intermediary chooses an experiment $\mu : T \times A \to \triangle(A)$, where $\mu(\hat{\mu} \mid \hat{s}, \hat{\sigma}')$ is the probability of recommending $\hat{\mu}$ when the intermediary's realization of the private information is $\hat{s}$ and the sender's recommendation is $\hat{\sigma}'$.

We introduce the timing of the interaction as needed. Moreover, we make two assumptions throughout:

Assumption 1 (Information Assumption) *The receiver knows the distribution of the private information but does not observe its realizations.*

Assumption 2 (Independence Assumption) *Conditional on the state of the world, the private information realizations and the sender's recommendation are pairwise independent.*

Assumption 2 amounts to saying two things, conditional on the true state: (i) that the intermediary's private information is not swayed by the sender's recommendation, and (ii) that differences in the sender's message and the intermediary's private information are idiosyncratic.

In the sequel, when we condition an experiment's distribution only on the state of the world, we mean "the total probability that the experiment makes a recommendation, conditional on the state of the world". Specifically, under Assumptions 1 and 2,

[9] A recommendation experiment is a communication experiment with the action space as a message space.

the probabilities that the sender and the intermediary recommend $\hat{\sigma}$ and $\hat{\mu}$ in state t are

$$\sigma(\hat{\sigma} \mid t) \equiv \mathbb{P}(\hat{\sigma} \mid t) = \sum_{\hat{s}} \sigma(\hat{\sigma} \mid \hat{s}, t) s(\hat{s} \mid t), \tag{2}$$

and

$$\mu(\hat{\mu} \mid t) = \sum_{\hat{s},\hat{\sigma}} \mu(\hat{\mu} \mid \hat{s}, \hat{\sigma}) \sigma(\hat{\sigma} \mid t) s(\hat{s} \mid t). \tag{3}$$

Expected Payoffs Under Obedient Recommendation Experiments

Consider an arbitrary recommendation experiment $e \in \{\sigma, \mu\}$. We denote by e_t the probability of experiment e recommending a_1 when the state of the world is t.[10] We say that an experiment is *obedient* when it is optimal for the receiver to follow a recommendation. The receiver follows a recommendation a_1 and a_2 only when, respectively,[11]

$$e_g x - e_b \geq 0. \tag{OC}$$

$$(1 - e_g)x - (1 - e_b) \leq 0. \tag{4}$$

As $x < 1$, equation 4 is redundant to OC, which we subsequently call *the obedience constraint*. When OC is satisfied, the probability that the receiver chooses a_1 after a recommendation is the probability that the experiment recommends a_1. Thus, the sender's expected utility is the total probability of recommending a_1:

$$U_e = \frac{1}{2}(e_g + e_b). \tag{5}$$

Meanwhile, the receiver's expected utility is

$$V_e = \frac{1}{2}(x e_g - e_b). \tag{6}$$

Condition OC is equivalent to $V_e \geq 0$, which we use when suitable. In other words, the receiver obeys experiment e only when the expected utility of doing it is greater than the utility from the default action.

The sender's and receiver's expected utilities show a partial incentive misalign-

[10] For instance, σ_t is the probability of σ recommending a_1 in state t.

[11] The receiver's posterior belief of the state t following a recommendation a_1 is $\frac{e_t}{e_t + e_{-t}}$.

ment. Both agents benefit from a higher probability of recommending a_1 when the state is t_g, whereas the sender benefits and the receiver suffers from a higher probability of recommending a_1 when the state is t_b. Finally, the maximum possible sender's and receiver's expected utility under condition OC are:

$$U^{\max} = \frac{1+x}{2}, \tag{7}$$

and

$$V^{\max} = \frac{x}{2}. \tag{8}$$

Bayesian Persuasion Case

To add perspective to our results, it is helpful to outline the Kamenica and Gentzkow's (2011) canonical Bayesian persuasion (BP) problem that arises when there is neither an intermediary nor private information, and the sender can directly communicate with the receiver. In this case, the sender's problem is

$$\max_{\sigma} \frac{1}{2}(\sigma_g + \sigma_b) \text{ s.t. } OC \tag{BP}$$

The unique solution to the BP problem is recommending a_1 with probability one when the state is t_g, and with probability x when the state is t_b (Bergemann and Morris, 2019). By recommending a_1 in state t_g, the sender maximizes the probability that the receiver is willing to tolerate, while still satisfying OC, for recommending a_1 in state t_b. As a result, the sender leaves the receiver indifferent between following a recommendation and choosing the default action: the sender's and receiver's expected utilities are $U^{BP} = U^{max}$ and $V^{BP} = 0$. In the sequel, we call this solution the *Bayesian Persuasion Communication Policy (BPCP)*, denoted by σ^{BP}, with $\sigma_g^{BP} = 1$ and $\sigma_b^{BP} = x$.

Running Example, Part 2 *We return to the running example. To support the bill, the sponsor prepares a report containing details about the proposal, which signals its quality. The sponsor is strategic in how his report is designed to maximize the probability that the bill is approved. Specifically, when the bill is good, the sponsor always reveals it to the public with certainty ($\sigma_g^{BP} = 1$). If the bill is bad, however, the sponsor engages in fake news: with positive probability ($\sigma_b^{BP} = x$), he suggests that the bill is good when it is, in fact, bad for the public. By following this strategy, the*

sponsor convinces the public to give him "the benefit of the doubt" and approve the bill regardless. In the process, the sponsor maximizes the probability of acceptance at the expense of the public's expected utility.

3. Two Preliminary Benchmarks

Direct Communication

The first preliminary benchmark for our mediated communication model is the case where the intermediary does not exist, and the sender directly communicates the private information to the receiver. We consider the following timing of events:

1. The sender chooses an experiment $\sigma : T \times T \to \triangle(A)$ and commits to it.
2. The sender's private information and the state of the world are privately realized, and the sender makes a recommendation.
3. The receiver observes the sender's recommendation and updates her beliefs to make a decision.

The following theorem summarizes the sender's optimal communication strategy:

Theorem 1 *When the sender directly communicates to the receiver, any experiment* σ^{DC} *such that* $\sum_{\hat{s}} \sigma^{DC}(a_1 \mid \hat{s}, t)s(\hat{s} \mid t) = \sigma_t^{BP}$ *for all* t *is sender-optimal. Moreover, under any sender-optimal experiment,* $U^{DC} = U^{max}$ *and* $V^{DC} = 0$.

Theorem 1 establishes that under direct communication, the sender's optimal experiment mimics the BPCP. As a result, the sender extracts the receiver's whole expected utility, leaving her indifferent between following a recommendation and choosing the default action. Strikingly, the sender does so independently of the accuracy of the private information.

The result is a consequence of two facts. First, the sender's expected utility and the obedience constraint depend only on the total probability of recommending a_1 in state t. Second, Lemma 1, stated below and key to the message of the theorem, establishes that the sender can *bend the facts*: he can manipulate his communication of the private information by suitably choosing $\sigma(a_1 \mid \hat{s}, t)$ to achieve any desired total probability of recommending a_1 in state t. As a result, the sender can target a probability $\sigma(a_1 \mid t)$ to solve the BP problem.

Lemma 1 (private information Suppression) *For any state t and any $y \in [0,1]$, there are probabilities $\sigma(a_1 \mid \hat{s}, t)$ such that*

$$y = \sum_{\hat{s}} \sigma(a_1 \mid \hat{s}, t) s(\hat{s} \mid t). \tag{IS}$$

In light of Lemma 1, without loss of generality, we can focus on the case where the sender chooses a standard Bayesian persuasion experiment $\sigma : T \to \triangle(A)$, that is, as if the private information did not exist. We conclude this subsection by updating our example:

Running Example, Part 3 *To support the bill, the sponsor includes evidence in his report containing details about the proposal. As before, the sponsor's goal is to reveal with probability one when the bill is good, while engaging in fake news when the bill is bad by revealing that it is good with probability x. Such a goal is achieved in two steps. When the bill is good, independently of the evidence, the sponsor reveals that the bill is good, namely $\sigma^{DC}(a_1 \mid \hat{s}, t_g) = 1$ for all $\hat{s}$. When the bill is bad, the sponsor must* bend the facts. *One possibility is recommending a_1 with probability $\sigma^{DC}(a_1 \mid \hat{s} = t_g, t_b) = 1$ when the evidence suggests the bill is good even though it is not, and recommending a_1 with probability $\sigma^{DC}(a_1 \mid \hat{s} = t_b, t_b) = \frac{\epsilon_b - (1-x)}{\epsilon_b}$ when the evidence suggests the bill is bad.*[12] *In the process, the sponsor maximizes the probability of acceptance at the expense of the public's expected utility.*

Surprise Mediated Communication

In preparation for our equilibrium analysis found in the next section, we present a second preliminary benchmark. In light of Theorem 1, our current goal is to highlight the role and capabilities of the intermediary in particularly simple circumstances. To this end, we currently make the following assumption about the sender-intermediary interaction:

Assumption 3 (Surprise Intervention) *The sender believes that the intermediary will pass along his recommendation to the receiver intact, i.e., without any interference. Specifically, the sender does not strategize for the intermediary's presence, justifying the term* surprise intervention.

[12] For this particular solution, we must assume that ϵ_b is large enough so that $\frac{\epsilon_b - (1-x)}{\epsilon_b} \geq 0$.

Moreover, in the current subsection we make two technical simplifying assumptions. First, we assume that when the receiver is indifferent between two experiments, the intermediary chooses the one that favors the sender. Second, we assume the private information is imperfectly informative, namely $\epsilon_t \in (0,1)$ for all t. In our main section, which comes next, we show that the insights of the analysis under these assumptions can be properly extended to the case where: (i) sender and intermediary strategize for each other (equilibrium analysis), (ii) ties are not arbitrarily broken, and (iii) perfectly accurate or inaccurate private information is also allowed. The chronology of events that we currently analyze is the following:

1. The sender chooses an experiment $\sigma : T \to \triangle(A)$ under Assumption 3 and commits to it.

2. The intermediary knows σ, chooses an experiment $\mu : T \times T \to \triangle(A)$, and commits to it.

3. The state of the world is realized, the sender privately observes it, and makes a recommendation.

4. The intermediary observes the sender's recommendation and her private signal, and makes a recommendation.

5. The receiver observes the intermediary's recommendation and updates her beliefs to make a decision.

If the sender believes his recommendation will be passed intact to the receiver, as Assumption 3 implies, he insists on the optimal BPCP, namely, $\sigma_b^{BP} = x$ and $\sigma_g^{BP} = 1$. Substituting the BPCP in Equations 5 and 6, the receiver's expected utility is

$$\begin{aligned} V_\mu^S = \frac{1}{2}\big[&x(\epsilon_g + \epsilon_b - 1)\big(\mu(a_1 \mid t_g, a_1) - \mu(a_1 \mid t_b, a_1)\big) \\ &-(1-x)\big(\epsilon_b \mu(a_1 \mid t_b, a_2) + (1-\epsilon_b)\mu(a_1 \mid t_g, a_2)\big)\big]. \end{aligned} \tag{9}$$

The following proposition summarizes the optimal intermediary's experiment and its dependence on the accuracy of her private information. The result suggests that the receiver's expected utility is increasing in the amount of information that the

intermediary possesses and that any information better than completely combined-uninformative signals is enough to guarantee the receiver a positive payoff (this fact will cease to hold in the equilibrium model of the next section):

Proposition 1 *The intermediary's optimal experiment is given by* $\mu^*(a_1 \mid t_b, a_2) = \mu^*(a_1 \mid t_g, a_2) = 0$,

$$\mu^*(a_1 \mid t_g, a_1) = \begin{cases} 1, & \text{if } \epsilon_g + \epsilon_b \geq 1, \\ 0, & \text{otherwise}, \end{cases}$$

and

$$\mu^*(a_1 \mid t_b, a_1) = \begin{cases} 0, & \text{if } \epsilon_g + \epsilon_b > 1, \\ 1, & \text{otherwise}. \end{cases}$$

Moreover, the receiver's expected utility under the optimal experiment can be written as $V_\mu^S = \frac{x}{2} I(\epsilon_g, \epsilon_b) = I(\epsilon_g, \epsilon_b) V^{\max}$, *which is strictly increasing in the combined informativeness of the private information.*

When the intermediary observes a recommendation a_2 from the sender, she confirms that the state is t_b, and, regardless of her private information, she maximizes the receiver's payoff by recommending a_2 with probability one. When the intermediary observes a recommendation a_1 from the sender, however, she is uncertain of the state of the world. As a result, she must rely on her private research—her private information—to determine the best recommendation to make. If $\epsilon_g + \epsilon_b > 1$, the private information is combined-accurate (i.e., it is closer to being perfectly accurate than perfectly inaccurate). Therefore, a realization of the private information t confirms that the world's most likely true state is t. Then, the intermediary maximizes the receiver's payoff by recommending a_1 with probability one when the private information realization is $\hat{s} = t_g$, and with probability zero when the realized private information is $\hat{s} = t_b$. Likewise, the opposite is true when $\epsilon_g + \epsilon_b < 1$, as the private information is combined-inaccurate.

The combined informativeness of the intermediary's private information serves as an instrument to transfer surplus to the receiver with respect to the Bayesian persuasion outcome.[13] Indeed, for sufficiently high combined informativeness, almost the entire surplus can be transferred. As an example, if $\epsilon_g = 0.97$ and $\epsilon_b = 0.05$,

[13]Note well how the important variable is combined informativeness and not accuracy. Even in the very inaccurate case, the intermediary can greatly help improve the receiver's payoff.

a realization t_g is viewed like this: it is more likely than the state is good (where the private information is very accurate) than the state is bad (where the private information is very inaccurate), because 0.97 is closer to 1 than 0.05 is to 0; then, the recommended action is a_1 (note how the reverse conclusion would be reached if $\epsilon_g = 0.93$ and $\epsilon_b = 0.05$). Because the combined informativeness is the same in both cases, the same surplus amount is transferred to the receiver in the solution. But the way this is implemented is different. In the former case, upon receiving a good realization of the private information, being more informative in the good state, it is estimated that the good state is more likely, and a_1 is recommended. In the latter, receiving a bad realization of the private information is deemed more informative about a good state, and, despite the bad realization, the nondefault action a_1 is recommended.

Running Example, Part 4 *Coming back to the running example, the public may rely on the media (intermediary), who evaluates the bill based on its own research of the evidence* (s)*. If the sponsor does not know what evidence the media will find or how it will be used to evaluate the bill, the best he can hope for is that the media's opinion agrees with his (Surprise Intervention Assumption). The media learns about the bill only through the sponsor's report and its own evaluation (Information Assumption) and, conditional on the quality of the bill, its evaluation is independent of the sponsor's report (Independence Assumption). Finally, the media's research efforts are imperfect, so it might misinterpret the bill's scope and focus on irrelevant aspects: it can correctly confirm a bill of quality* t *only with probability* ϵ_t*.*

Proposition 1 establishes that, after accounting for the quality of its private information, the media will favor a_1 *only if its research confirms that the state is* t_g*. Moreover, the media's efforts to mitigate the persuasive effect of fake news will be increasingly successful as it becomes better informed (in the combined informativeness sense) about the quality of the bill. Since the media's goals are aligned with the public, the media should learn to compensate for its private information's shortcomings (as in the case* $(\epsilon_g, \epsilon_b) = (0.93, 0.05)$ *discussed above, to support a recommendation that goes against its realization of the evidence) and provide the socially desirable recommendation.*

4. Equilibrium Intermediation Model

Setup

This is our central section. In it, we allow the sender and intermediary to design their experiments in equilibrium, accounting for each other's strategic behavior. Thus, we currently eliminate Assumption 3, so the timing of events is modified as follows:

1. Simultaneously, the sender chooses an experiment $\sigma : T \to \triangle(A)$, the intermediary chooses an experiment $\mu : T \times T \to \triangle(A)$, and they commit to them.
2. The state of the world is realized, the sender privately observes it, and he makes a recommendation.
3. The intermediary observes the sender's recommendation and her private signal, and makes a recommendation.
4. The receiver observes the intermediary's recommendation and updates her beliefs to make a decision.

We stress that, by relaxing the surprise intervention assumption, we modify how the sender and intermediary strategize for each other, not the timing of the recommendation process (stages 2 through 5 above). Two remarks are on point. First, we no longer insist on the sender relying on the BPCP; rather, his strategy must hold as an equilibrium object if it is to be used. Second, commitment plays an important role for the intermediary: she must stick to her communication policy, regardless of the realization of her private information and the sender's recommendation. As is typically the case in the literature (Bergemann and Morris, 2019; Kamenica, 2019), the validity of commitment power is better addressed in the context of applications. We thus postpone our discussion until later in this section as we update our running example. We note, however, that our model is strategically equivalent to one without an intermediary, where the receiver ex-ante commits to a stochastic decision rule mapping the sender's realized message and the private information realization into actions. For the applications we have in mind, it is more natural to envision an intermediary, instead of a receiver's stochastic plan.[14]

[14] In Section 5, we provide further detail and differences with existing literature on information design with receiver's commitment power.

Equilibrium

We look for recommendation experiments μ^* and σ^* that constitute a Perfect Bayesian Equilibrium.[15] An important challenge is guaranteeing that agents' equilibrium experiments are obedient. A key simplifying observation is that the intermediary recommending the default action with probability one is always obedient, and so we can analyze the game where the obedience constraint (OC) is not actively considered.[16] As a result, any intermediary's best response to the sender's experiment cannot do worse and will also be obedient. The following lemma formalizes this logic:

Lemma 2 *The intermediary's best response to any σ is always obedient.*

Next, Proposition 2 establishes that an equilibrium exists for any parameters of the model:

Proposition 2 *For any $(\epsilon_g, \epsilon_b) \in [0,1]^2$ and $x \in (0,1)$, an equilibrium (μ^*, σ^*) exists.*

While this result gives us the existence of equilibrium, its uniqueness is not to be found. Yet our results will characterize the relevant properties of all equilibria.

The Sender's and Receiver's Expected Utility Revisited

It is helpful for our analysis to write the sender's and receiver's expected utility in terms of two objects of interest. The first one is *the contribution to the receiver's expected utility from the intermediary recommending a_1 when her private information is $\hat{s}$ and the sender's recommendation is $\hat{\sigma}$*, denoted $\overline{v}_{s,\sigma}(\hat{s}, \hat{\sigma})$. Denoting by $v(t)$ the receiver's payoff of choosing a_1 when the true state is t, then:[17]

$$\overline{v}_{s,\sigma}(\hat{s}, \hat{\sigma}) \equiv \mathbb{P}(\hat{s}, \hat{\sigma})\mathbb{E}(v(t) \mid \hat{s}, \hat{\sigma}) = \frac{1}{2}\big(x\sigma(\hat{\sigma} \mid t_g)s(\hat{s} \mid t_g) - \sigma(\hat{\sigma} \mid t_b)s(\hat{s} \mid t_b)\big). \quad (10)$$

[15] Our equilibrium notion is a special case of the correlated Bayes-Nash equilibrium in Bergemann and Morris (2016). The two players aim to maximize expected payoffs with their designed obedient experiments, committing to the plan after every state or private information realization. No correlation device features in our model. Our equilibrium notion is also a special case of the communication equilibrium in Forges (1986), reinterpreting the intermediary's experiment as a communication device that takes $(\hat{s}, \hat{\sigma})$ as inputs to produce $\hat{\mu}$ as output.

[16] In other words, there is no equilibrium of the unrestricted game that is not an equilibrium of the game restricted by obedience.

[17] Mathematically, if $\mathbb{I}$ is the indicator function, $\overline{v}_{s,\sigma}(\hat{s}, \hat{\sigma}) = \mathbb{E}(v(t)\mathbb{I}_{(\hat{s},\hat{\sigma})})$, interpreted as the contribution of the realization $(\hat{s}, \hat{\sigma})$ to the unconditional (ex-ante) expected payoff $\mathbb{E}(v(t))$.

The second object is *the average increment in the probability of the intermediary recommending a_1, when the sender recommends a_1 instead of a_2 in state t*, denoted by $\overline{\triangle}_{\mu,s}(t)$. To formally define $\overline{\triangle}_{\mu,s}(t)$, we note that the change in the probability that the intermediary recommends a_1 when she has private information $\hat{s}$ and the sender changes his recommendation from a_2 to a_1 is $\triangle\mu(\hat{s}) = \mu(a_1 \mid \hat{s}, a_1) - \mu(a_1 \mid \hat{s}, a_2)$, so that

$$\overline{\triangle}_{\mu,s}(t) \equiv \mathbb{E}(\triangle\mu(\hat{s}) \mid t) = \sum_{\hat{s}} \triangle\mu(\hat{s}) s(\hat{s} \mid t). \tag{11}$$

On the one hand, $\overline{v}_{s,\sigma}$ determines how the sender's experiment affects the receiver's expected utility, so it determines the intermediary's best-response behavior to σ. On the other hand, $\overline{\triangle}_{\mu,s}$ determines how the intermediary's experiment changes when the sender changes her recommendation, so it determines the sender's best-response behavior to μ. The following lemma characterizes the intermediary's and sender's expected payoff and best responses in terms of $\overline{v}_{s,\sigma}$ and $\overline{\triangle}_{\mu,s}$:

Lemma 3 *The intermediary's expected payoff V_μ^E and (unrestricted) best-response correspondence μ^* are*

$$V_\mu^E(\mu, \sigma) = \sum_{\hat{\sigma}, \hat{s}} \mu(a_1 \mid \hat{s}, \hat{\sigma}) \overline{v}_{s,\sigma}(\hat{s}, \hat{\sigma}), \tag{12}$$

and

$$\mu^*(a_1 \mid \hat{s}, \hat{\sigma}) = \begin{cases} 1 & \text{if } \overline{v}_{s,\sigma}(\hat{s}, \hat{\sigma}) > 0, \\ \in [0,1] & \text{if } \overline{v}_{s,\sigma}(\hat{s}, \hat{\sigma}) = 0, \\ 0 & \text{if } \overline{v}_{s,\sigma}(\hat{s}, \hat{\sigma}) < 0. \end{cases} \tag{13}$$

The sender's expected payoff U_μ^E and (unrestricted) best-response correspondence σ^* are

$$U_\mu^E(\mu, \sigma) = \frac{1}{2}\left(\sum_t \sigma(a_1 \mid t) \overline{\triangle}_{\mu,s}(t) + \sum_{t, \hat{s}} s(\hat{s} \mid t) \mu(a_1 \mid \hat{s}, a_2) \right), \tag{14}$$

and

$$\sigma^*(a_1 \mid t) = \begin{cases} 1 & \text{if } \overline{\triangle}_{\mu,s}(t) > 0, \\ \in [0,1] & \text{if } \overline{\triangle}_{\mu,s}(t) = 0, \\ 0 & \text{if } \overline{\triangle}_{\mu,s}(t) < 0. \end{cases} \tag{15}$$

Lemma 3 reveals key insights behind the sender-intermediary strategic interaction. On the one hand, the intermediary is only willing to recommend a_1 with positive

probability when the contribution to the receiver's expected utility is strictly positive. In other words, the intermediary's optimal strategy is to recommend a_2 when the sender's recommendation and the private information suggest that recommending a_1 is expected to harm the receiver. On the other hand, the sender understands the intermediary's optimal strategy, so he recommends a_1 with positive probability only if it is expected to increase the average probability of the intermediary recommending a_1.

The Autarky Value of the Intermediary's Private Information

One of the main goals is to determine how the intermediary's optimal experiment depends on the characteristics of her private information. Toward our goal, we construct a novel information measure, the *Autarky Value of the Intermediary's Private Information (AVIPI)*. The AVIPI is the receiver's maximum expected utility gain over direct communication when the intermediary only trusts her private information.

When the intermediary recommends a_1 with probability one after observing $\hat{s} = t_g$, as if she fully trusted her private information, the receiver's expected utility is[18]

$$\underline{V}_g(\epsilon_g, \epsilon_b) \equiv \frac{\epsilon_g}{\epsilon_g + (1-\epsilon_b)} x - \frac{1-\epsilon_b}{\epsilon_g + (1-\epsilon_b)}. \tag{16}$$

Meanwhile, when the intermediary recommends a_1 with probability one after $\hat{s} = t_b$, as if she fully distrusted her private information, the receiver's expected utility is[19]

$$\underline{V}_b(\epsilon_g, \epsilon_b) \equiv \frac{1-\epsilon_g}{(1-\epsilon_g) + \epsilon_b} x - \frac{\epsilon_b}{(1-\epsilon_g) + \epsilon_b}. \tag{17}$$

Finally, if $\underline{V}_{\hat{s}}(\epsilon_g, \epsilon_b) < 0$ for both private information realizations $\hat{s}$, the intermediary could recommend the default action with probability 1.

Therefore, since these three strategies are always available to the intermediary in the model, and recalling that the receiver's expected utility under direct communication is zero, we define the following:

[18] If $\mu^*(a_1 \mid t_g, \hat{\sigma}) = 1$ and $\mu^*(a_1 \mid t_b, \hat{\sigma}) = 0$ for all $\hat{\sigma}$, then $\mu_g = \epsilon_g$ and $\mu_b = 1 - \epsilon_b$, and the posterior of t after a recommendation $\hat{\mu} = a_1$ is $\frac{\epsilon_g}{\epsilon_g + (1-\epsilon_b)}$.

[19] If $\mu^*(a_1 \mid t_b, \hat{\sigma}) = 1$ and $\mu^*(a_1 \mid t_g, \hat{\sigma}) = 0$ for all $\hat{\sigma}$, then $\mu_g = 1 - \epsilon_g$ and $\mu_b = \epsilon_b$, and the posterior of t after a recommendation $\hat{\mu} = a_1$ is $\frac{1-\epsilon_g}{(1-\epsilon_g) + \epsilon_b}$.

Definition 1 *The Autarky Value of the Intermediary's Private Information (AVIPI) is the receiver's maximum equilibrium expected utility when the intermediary recommends a_1 based only on her private information, namely,*

$$A(\epsilon_g, \epsilon_b) \equiv \max\left(\underline{V}_g(\epsilon_g, \epsilon_b), \underline{V}_b(\epsilon_g, \epsilon_b), 0\right). \tag{18}$$

The following lemma characterizes a strictly positive AVIPI as simple conditions over the accuracy of the intermediary's private information. As such, it provides a helpful tool to derive the insights behind our main results.

Lemma 4 *$A(\epsilon_g, \epsilon_b) > 0$ if and only if one of the following conditions holds*

$$\textit{Sufficient Accuracy: } \underline{v}_A(\epsilon_g, \epsilon_b) \equiv \epsilon_g x - (1 - \epsilon_b) > 0. \tag{SA}$$

$$\textit{Sufficient Inaccuracy: } \underline{v}_I(\epsilon_g, \epsilon_b) \equiv (1 - \epsilon_g)x - \epsilon_b > 0 \tag{SI}$$

The Receiver's Welfare and the AVIPI

Can the intermediary help the receiver improve her expected utility, relative to direct communication? Put differently, could the sender *persuade the receiver through the intermediary* and leave her indifferent between following a recommendation and the default action? The following Theorem, our main result, provides a complete answer to these questions: the receiver's expected utility when there is an intermediary is equal to the AVIPI. Specifically, the intermediary can help the receiver improve over direct communication if and only if the AVIPI is strictly positive.

Theorem 2 *If (μ^*, σ^*) is an equilibrium, then $V_\mu^E(\mu^*, \sigma^*) = A(\epsilon_g, \epsilon_b)$. Moreover, the following statements are equivalent:*

(i) $V_\mu^E(\mu^*, \sigma^*) > 0$*, i.e., the receiver has an expected utility gain over choosing the default action a_2.*

(ii) There is $(\hat{s}, \hat{\sigma})$ such that $\overline{v}(\hat{s}, \hat{\sigma}) > 0$.

The equivalence between (i) and (ii) in Theorem 2 is the building block behind the logic of the result, and stems from the intermediary's best response correspondence in Lemma 3. Specifically, it is optimal for the intermediary to recommend a_1 if and

only if her private information-recommendation pair $(\hat{s}, \hat{\sigma})$ positively contributes to the receiver's expected utility, namely, $\overline{v}_{s,\sigma^*}(\hat{s}, \hat{\sigma}) > 0$. If $(\hat{s}, \hat{\sigma})$ contributes negatively, the intermediary can neutralize such negative effect by "playing it safe" and recommending a_2 with probability one. The result is that the receiver's expected utility is positive if and only if there is an private information-recommendation pair with a strictly positive contribution.

The sender-intermediary interaction is a persuasion game: the sender tries to persuade the intermediary to recommend a_1 with the highest probability possible. As such, it is important to identify the intermediary's default action. In particular, the intermediary can always base a recommendation entirely on her private information, securing a payoff equal to the AVIPI for the receiver. Thus, any best response for the intermediary must yield an expected utility greater or equal than the AVIPI. However, in this case, as in the standard Bayesian persuasion model (Kamenica and Gentzkow, 2011), the sender can maximize the total probability of the intermediary recommending a_1 by leaving the intermediary indifferent between the equilibrium recommendation experiment and her default action. It turns out that no other outcome is compatible with equilibrium. The result is that the only possible equilibrium receiver's expected utility is the AVIPI. A crucial remark follows: a strictly positive AVIPI does not imply that the intermediary bases her recommendation *only* on her private information.[20]

A strictly positive AVIPI endows the intermediary with two important tools that guarantee the receiver a strictly positive expected utility. First, as argued above, an experiment that is beneficial for the receiver, which completely shuts the sender's persuasive efforts off. Second, a mechanism to identify cases where complementing the intermediary's private information with the sender's recommendation is strictly beneficial to the receiver, compared to direct communication.

Lemma 4 implies that when $A(\epsilon_g, \epsilon_b) > 0$, the accuracy of the intermediary's private information must satisfy one of two conditions, SA or SI. Suppose for simplicity that (SA) holds, so $\underline{v}_A(\epsilon_g, \epsilon_b) > 0$. Such a condition implies that the intermediary always can find $\hat{\sigma}$ such that $\overline{v}_{s,\sigma^*}(t_g, \hat{\sigma}) > 0$. If, on the one hand, $\overline{v}_{s,\sigma^*}(t_g, a_1) > 0$, the problem is trivial. If, on the other hand, $\overline{v}_{s,\sigma^*}(t_g, a_1) \leq 0$, the intermediary can compute $\overline{v}_{s,\sigma^*}(t_g, a_2) = \underline{v}_A - \overline{v}_{s,\sigma^*}(t_g, a_2) > 0$, thanks to $\underline{v}_A > 0$. As a result of $A(\epsilon_g, \epsilon_b) > 0$, the intermediary can identify the sender's recommendation-private information pairs

[20] Further detail is provided in Section 4.

that positively contribute to the receiver's expected utility, improving the receiver's situation over trusting only private information.

Theorem 2 in the Context of the Running Example

We revisit our bill sponsor example to discuss the interpretation of our commitment assumptions and highlight the mechanics behind Theorem 2. In our example, the sponsor designs a report σ that, with probability σ_t, recommends approval of the bill in state t. We interpret the realization of the sponsor's recommendation as the media's interpretation of the sponsor's report. Simultaneously, the media designs a communication experiment μ to make a recommendation to the public when the outcome of its private research is $\hat{s}$ and its interpretation of the sponsor's report is $\hat{\sigma}$. In this context, we justify our commitment assumption on the grounds of reputation. One possibility is that the media fears potential credibility issues in the future if it suitably changes its experiment only after realizing the sponsor's communication, as opposed to sticking with institutional guidelines.

Suppose $\epsilon_g = \frac{1}{2}$ and $\epsilon_b = \frac{1}{4}x$ so that $A(\epsilon_g, \epsilon_b) = \frac{1}{8}x > 0$. First, assume that the media is overly optimistic and expects the sponsor to favor the public by designing an experiment $\sigma_g = 1$ and $\sigma_b = 0$. Then, the contribution to the receiver's expected utility of each private information-recommendation pair is

$$\overline{v}_{s,\sigma}(\hat{s}, \hat{\sigma}) = \frac{1}{2} \begin{cases} \frac{x}{2}, & \text{if } (\hat{s}, \hat{\sigma}) = (t_g, a_1), \\ \frac{x}{2}, & \text{if } (\hat{s}, \hat{\sigma}) = (t_b, a_1), \\ -(1 - \frac{x}{4}), & \text{if } (\hat{s}, \hat{\sigma}) = (t_g, a_2), \\ -\frac{x}{4}, & \text{if } (\hat{s}, \hat{\sigma}) = (t_b, a_2). \end{cases}$$

The media expects a sender's recommendation a_1 to benefit the receiver and a recommendation a_2 to hurt her. As a result, the media best-responds by *passing the sponsor's recommendation intact to the public*, independently of the private information, namely,

$$\mu^*(a_1 \mid \hat{s}, \hat{\sigma}) = \begin{cases} 1, & \text{if } \hat{\sigma} = a_1, \\ 0, & \text{if } \hat{\sigma} = a_2. \end{cases}$$

Under (μ^*, σ), the receiver's expected utility is $V(\mu^*, \sigma) = x > A(\epsilon_g, \epsilon_b)$. However,

$\sigma_g = 1$ and $\sigma_b = 0$ are not optimal for the sponsor when he expects the media to rely on experiment μ^*. More specifically, under μ^*, the average increment in the probability of the intermediary recommending a_1, when the sender recommends a_1 instead of a_2 in state t_b is $\overline{\triangle}_{\mu}(t_b) = 1$. Thus, the sponsor expects an increase in σ_b to increase the total probability that the public is recommended a_1, leading him to deviate to $\sigma_b' = 1$.

If the media expects $\sigma_g^* = \sigma_b^* = 1$ instead, the contribution to the receiver's expected utility of each private information-recommendation pair is

$$\overline{v}_{s,\sigma^*}(\hat{s},\hat{\sigma}) = \frac{1}{2}\begin{cases} \frac{3}{4}x - 1, & \text{if } (\hat{s},\hat{\sigma}) = (t_g, a_1), \\ \frac{x}{4}, & \text{if } (\hat{s},\hat{\sigma}) = (t_b, a_1), \\ 0, & \text{if } (\hat{s},\hat{\sigma}) = (t_g, a_2), \\ 0, & \text{if } (\hat{s},\hat{\sigma}) = (t_b, a_2). \end{cases}$$

As outlined in the previous section, $A(\epsilon_g, \epsilon_b) > 0$ allows the media to identify at least a private information-recommendation pair that has a strictly positive contribution to the public's expected utility—(t_b, a_1) in this case. Then, one possible best response for the media, being part of an equilibrium along σ^*, is

$$\mu^*(a_1 \mid \hat{s},\hat{\sigma}) = \begin{cases} 1, & \text{if } (\hat{s},\hat{\sigma}) = (t_b, a_1), \\ 0, & \text{otherwise.} \end{cases}$$

Under (μ^*, σ^*), the public's expected utility is $V_{\mu}^{E} = \frac{1}{8}x = A(\epsilon_g, \epsilon_b)$. The example shows several features of the equilibrium interaction between the media and the sponsor. First, the public's equilibrium expected utility is only as good as the media's private information. Second, in equilibrium, the sponsor cannot rely on strategies that benefit the public over the AVIPI because the media's best response creates a channel for the sponsor to deviate profitably. Finally, even though the public's expected utility is the AVIPI, this does not mean that the media only relies on its private information. In the example, the media recommends a_1 only when the sponsor recommends a_1 and the private information confirms the state as t_b.

The Relationship of the AVIPI with Combined Informativeness and a Visual Representation of the Receiver's Equilibrium Expected utility

How does the AVIPI compare to combined informativeness? Corollary 1 highlights the difference between the two measures, in terms of the benefit of the receiver in expected utility over direct communication.

Corollary 1 *Suppose* (μ^*, σ^*) *is an equilibrium. Then, the following statements hold:*

(a) $I(\epsilon_g, \epsilon_b) > 1 - x$ *is sufficient (but not necessary) for* $V_\mu^E(\mu^*, \sigma^*) > 0$.

(b) $I(\epsilon_g, \epsilon_b) > 0$ *is necessary (but not sufficient) for* $V_\mu^E(\mu^*, \sigma^*) > 0$.

Corollary 1 distills one of the main economic insights of Theorem 2: the intermediary needs enough private information—not necessarily perfect—to secure a strictly positive expected utility for the receiver. Specifically, strictly positive combined informativeness is not enough, as was in the surprise intervention model.

To see that Corollary 1's sufficient condition in (a) is not necessary for nonpersuasive equilibrium outcomes,[21] it will be argued in a later subsection that the BPCP with any $\epsilon_g > 0$ and $\epsilon_b = 1$ are compatible with nonpersuasive equilibria. This leads to $I(\epsilon_g, \epsilon_b) = \epsilon_g$. As any ϵ_g works, we can just pick $\epsilon_g < 1 - x$. Meanwhile, to understand the lower bound $(1 - x)$ on combined informativeness for nonpersuasive equilibria, think of this intuition. If x is very close to 1 ($1 - x$ very close to 0), the expected utility under the prior for the receiver is approximately zero for the nondefault action. The smaller x is (the greater $(1 - x)$ is), the more attractive a_2 is as a default option. The more attractive a_2 is, the more combined informativeness the intermediary needs to overcome the harm of choosing a_1 in t_b.

Our model allows a visual representation of the receiver's equilibrium expected utility as a function of the accuracy pair (ϵ_g, ϵ_b), and provides insight on the comparison between the AVIPI and combined informativeness.

The red line connecting $(0, 1)$ to $(1, 1 - x)$ is the locus of accuracy levels such that $\underline{v}_A(\epsilon_g, \epsilon_b) = x\epsilon_g - (1 - \epsilon_b) = 0$. Lemma 4 and Theorem 2 imply that the accuracy levels in the blue region above $\underline{v}_A(\epsilon_g, \epsilon_b) = 0$ guarantee a strictly positive AVIPI (R_1 and R_2 in the figure). Being closer to perfect accuracy (ϵ_A) than to perfect inaccuracy

[21] By "nonpersuasive," we refer to those equilibrium outcomes where the sender fails to maximize the total probability of a_1 by extracting the whole surplus over the AVIPI, as under direct communication.

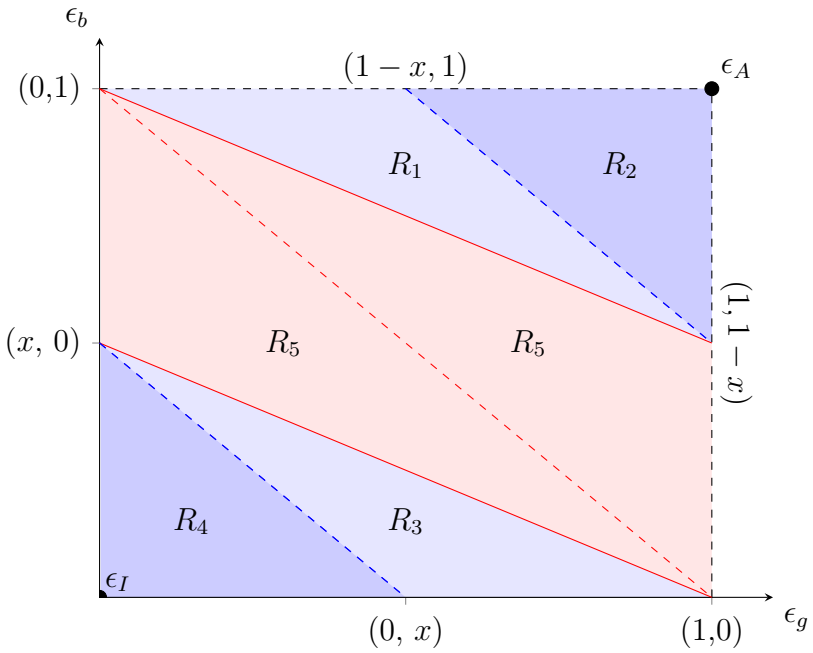

Figure 1: Informativeness Regions and Receiver's Equilibrium Expected Utility.

(ϵ_I), private information with accuracy levels in regions R_1 and R_2 can be interpreted as those that are *sufficiently accurate* to guarantee a strictly positive AVIPI.

The red line connecting $(1,0)$ to $(0,x)$ is the locus of accuracy levels such that $\underline{v}_I(\epsilon_g,\epsilon_b) = x(1-\epsilon_g) - \epsilon_b = 0$. Lemma 4 and Theorem 2 imply that the accuracy levels in the blue region below $\underline{v}_I(\epsilon_g,\epsilon_b) = 0$ guarantee a strictly positive AVIPI (R_3 and R_4 in the figure). Being closer to perfect inaccuracy than to perfect accuracy, private information with accuracy levels in regions R_3 and R_4 can be interpreted as those that are *sufficiently inaccurate* to guarantee a strictly positive AVIPI.

The red region between the loci of sufficient accuracy and inaccuracy ($\underline{v}_A = 0$ and $\underline{v}_I = 0$, respectively), including the boundary, is the region where the AVIPI is zero. Such a region is represented by R_5 in the figure. Private information with accuracy levels in region R_5 can be interpreted as those that are neither *sufficiently accurate* nor *sufficiently inaccurate* to guarantee a strictly positive AVIPI.

Lemma 4 implies that accuracy levels in the union of the blue regions are necessary and sufficient for a strictly positive AVIPI. Alternatively, accuracy levels in the red region, including the boundary, are necessary and sufficient for a zero AVIPI. Consequently, the identity $V_\mu(\mu^*,\sigma^*) = A(\epsilon_g,\epsilon_b)$ implies that *the receiver's expected utility is strictly positive if and only if the intermediary's private information is sufficiently*

accurate or sufficiently inaccurate.

Moreover, there are two additional regions of interest, pertaining to Corollary 1. On the one hand, the dashed blue line connecting $(1, 1-x)$ to $(1-x, 1)$ represents the locus of accuracy levels closer to perfect accuracy with a combined informativeness of $(1-x)$. The region above such a locus, depicted in darker blue as R_2, are those accuracy levels with enough combined informativeness, in the accuracy direction, to be sufficiently accurate in the AVIPI sense. Symmetrically, the dashed blue line connecting $(0, x)$ to $(x, 0)$ represents the locus of accuracy levels closer to perfect inaccuracy with a combined informativeness of $(1-x)$. The region below such a locus, depicted in darker blue as R_4, are those accuracy levels with enough combined informativeness, in the inaccuracy direction, to be sufficiently inaccurate in the AVIPI sense. The existence of regions R_1 and R_3 show that $I(\epsilon_g, \epsilon_b) > 1-x$ is sufficient but not necessary for a strictly positive AVIPI.

Finally, the counter-diagonal connecting $(0, 1)$ and $(1, 0)$, depicted as a dashed red line, represents the locus of accuracy levels with a combined informativeness of zero. Such a locus is contained in the region R_5, implying a zero AVIPI and a zero receiver's expected utility in equilibrium. The fact that such a locus is entirely contained in R_5 shows why $I(\epsilon_g, \epsilon_b) > 0$ is necessary but insufficient for a strictly positive AVIPI.

Classes of Equilibria as a Function of the Intermediary's Private Information

The final goal of our work is to understand how the sender and intermediary plan for each other in equilibrium. In this subsection, we analyze three different classes of equilibria, their consequences for the receiver's expected utility, and their dependence on the intermediary's private information.

Ignoring the Sender's Recommendation

Theorem 2 relies on the fact that when the AVIPI is strictly positive, the intermediary can leverage its private information to mitigate the sender's persuasive communication and avoid the receiver's whole surplus extraction. Two natural questions are whether the intermediary can trust her private information and ignore the sender in equilibrium and whether this is her only equilibrium possibility when she has a strictly positive AVIPI.

Formally, when the intermediary has private information $\hat{s}$ and recommends a_1 with the same probability, independently of the sender's recommendation, we say that *the intermediary ignores the sender's recommendation when her private information is* $\hat{s}$. Mathematically, $\mu(a_1 \mid \hat{s}, a_1) = \mu(a_1 \mid \hat{s}, a_2) \equiv \mu(\hat{s})$, where $\mu(\hat{s})$ is the probability of recommending a_1 upon realizing private information $\hat{s}$ when ignoring the sender. If the intermediary ignores the sender's message for every possible outcome of the private information, we say that *the intermediary always ignores the sender's recommendation.* The next result establishes the existence of equilibria where the intermediary ignores the sender and their relationship with the receiver's expected utility.

An important remark we made after Theorem 2 is that a strictly positive AVIPI does not imply that the intermediary simply ignores the sender's recommendation in *any* equilibrium. As a simple example, under perfect accuracy, $\sigma_g^* = 1$, $\sigma_b^* = x$, $\mu^*(a_1 \mid t_g, a_1) = 1$ and $\mu^*(a_1 \mid \hat{s}, \hat{\sigma}) = 0$ for all $(\hat{s}, \hat{\sigma}) \neq (t_g, a_1)$ is an equilibrium where the intermediary does not ignore the sender after private information shows $\hat{s} = t_g$, even though she has perfect information. Proposition 3 establishes under which conditions the intermediary can ignore the sender in equilibrium:

Proposition 3 *(i) There exists an equilibrium (μ^*, σ^*) where the intermediary always ignores the sender's recommendation if and only if $\overline{v}_{s,\sigma^*}(\hat{s}, a_1)\overline{v}_{s,\sigma^*}(\hat{s}, a_2) \geq 0$ for all $\hat{s}$.*

(ii) For any accuracy levels, there is an equilibrium where the intermediary always ignores the sender's recommendation.

(iii) In an equilibrium where the intermediary always ignores the sender's recommendation, the only possibilities are the following:

$$\mu^*(\hat{s}) = \begin{cases} 0, & \text{if } \underline{V}_{\hat{s}} < 0, \\ 1, & \text{if } \underline{V}_{\hat{s}} > 0, \\ \in [0,1], & \text{if } \underline{V}_{\hat{s}} = 0. \end{cases}$$

Part (i) follows from the fact that if $\overline{v}_{s,\sigma^*}(\hat{s}, a_1)$ and $\overline{v}_{s,\sigma^*}(\hat{s}, a_2)$ had opposite signs, the intermediary would benefit from being responsive to the sender's recommendation. For instance, if $\overline{v}_{s,\sigma^*}(\hat{s}, a_1) > 0$ and $\overline{v}_{s,\sigma^*}(\hat{s}, a_2) < 0$, then $\mu^*(a_1 \mid \hat{s}, a_1) = 1$ and

$\mu^*(a_1 \mid \hat{s}, a_2) = 0$, per Equation 13. Part (ii) explores whether the condition in part (i) imposes further restrictions on the accuracy levels that can support ignoring the sender in equilibrium. The answer is negative: such behavior can be sustained in equilibrium independently of the accuracy of the intermediary's private information.

However, part (iii) reveals that equilibria that feature the intermediary ignoring the sender can be qualitatively different, depending on the expected utility the intermediary creates for the receiver by ignoring the sender. For instance, when the expected utility the intermediary furnishes to the receiver by recommending a_1 after realizing private information $\hat{s}$ is negative ($\underline{V}_{\hat{s}} < 0$), the only equilibrium possibility is to recommend a_1 with probability zero. Otherwise, the intermediary would cause avoidable harm to the receiver. On the other hand, when the expected utility the intermediary awards to the receiver by recommending a_1 after realizing private information $\hat{s}$ is positive ($\underline{V}_{\hat{s}} > 0$), the only equilibrium possibility is to recommend a_1 with probability one. Otherwise, the intermediary would benefit the receiver from increasing such a probability. It follows that, in the nonpersuasive regions of Figure 1, equilibria of this sort exist, and in all of them, the receiver's expected utility is strictly positive.

On the other hand, in the region R_5—the persuasive region—when the intermediary's private information is not sufficiently accurate or inaccurate to grant a positive AVIPI, so $\underline{V}_g \leq 0$ and $\underline{V}_b \leq 0$. If both inequalities are strict, part (iii) implies that the only equilibrium possibility where the intermediary ignores the receiver is to recommend a_2 with probability one. As a result, the receiver's expected utility is zero. Meanwhile, if one of these inequalities binds—both binding is impossible—in any other equilibrium where the intermediary ignores the sender, she recommends with probability one a_2 for some realization of the private information but not necessarily for the other. The only case where the intermediary recommends a_2 with probability less than one regardless of her private information is on either the $\underline{V}_g = 0$ or the $\underline{V}_b = 0$ loci. In either of these two cases, the AVIPI is zero, and so is the receiver's expected utility.

Passing the Sender's Message Along

We say that *the intermediary passes the sender's message along intact* when she perfectly mimics the sender's recommendation, namely, $\mu(a_1 \mid \hat{s}, a_1) = 1$ and $\mu(a_1 \mid \hat{s}, a_2) = 0$ for all $\hat{s}$. As the example in Section 4 shows, passing the sender's message

along could open the door to profitable sender's deviations via persuasive communication. The following proposition determines whether passing the sender's message along is possible in equilibrium:

Proposition 4 *There exists an equilibrium (μ^*, σ^*) where the intermediary always passes the sender's message along if and only if the intermediary has no combined informativeness and the sender uses the BPCP, i.e. $I(\epsilon_g, \epsilon_b) = 0$, $\sigma_g^* = 1$ and $\sigma_b^* = x$.*

Proposition 4 generalizes the insights of the example in Section 4. Passing the sender's message to the receiver may create incentives for the sender to extract full surplus over the AVIPI and, consequently, for the intermediary to adjust its experiment. Indeed, when the intermediary passes the recommendation along to the receiver, the intermediary's unique obedient utility-maximizing experiment is the BPCP. This is because the problem is equivalent to direct communication. However, the only way the intermediary best-responds to the BPCP by passing along the sender's recommendation is when she cannot do better, *given the BPCP*. Such is the case only when her combined informativeness over the prior is zero.

Bayesian Persuasion Revisited

In light of Proposition 4, it is natural to investigate under what conditions the BPCP ($\sigma_g^{BP} = 1$ and $\sigma_b^{BP} = x$) is sustainable in equilibrium, namely, when does the sender's behavior is unaffected by the intermediary's presence. The next proposition complements the findings in Proposition 4 to provide an answer:

Proposition 5 *The BPCP is feasible in equilibrium only in the following cases:*

(a) If the intermediary's private information is combined-uninformative, namely, $I(\epsilon_g, \epsilon_b) = 0$.

(b) The intermediary's private information is combined-informative, namely, $I(\epsilon_g, \epsilon_b) > 0$ and the intermediary is perfectly informed about the good state of the world.[22]

In either case, the probability that the intermediary recommends a_1 is $\frac{1}{2} I(\epsilon_g, \epsilon_b)$. Moreover, the equilibrium payoffs can be written as

$$U_\mu^E = I(\epsilon_g, \epsilon_b)(U_{BP} - V^{\max}) \text{ and } V_\mu^E = I(\epsilon_g, \epsilon_b) V^{\max}.$$

[22] The latter case consists of two sub-cases. First, $\epsilon_b = 1$ and $\epsilon_g > 0$. Second, $\epsilon_b = 0$ and $\epsilon_g < 1$.

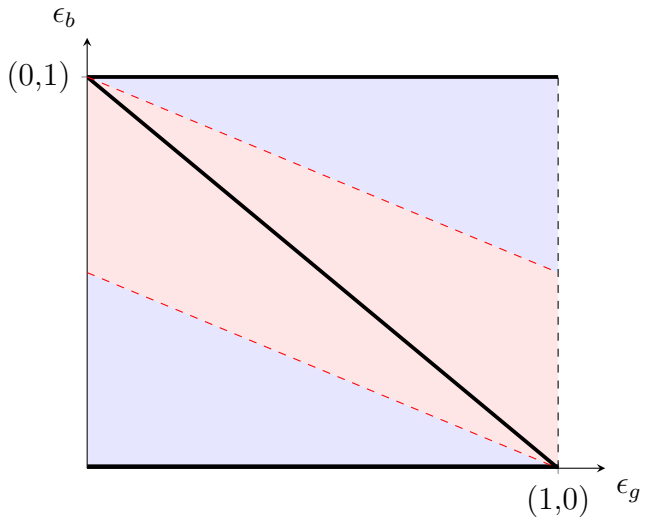

Figure 2: Locus of Accuracy Levels Supporting the BPCP as an Equilibrium.

Proposition 5 reveals an important insight into how the intermediary operates to protect the receiver in the context of her information. The first observation is that, in equilibrium, part of the social welfare is proportionally destroyed by the inaccuracy of the intermediary's private information, compared to the direct communication case. When the intermediary splits the remaining social surplus, he proportionally allocates the maximum possible expected utility to the receiver, leaving the rest to the sender.

In general, Proposition 5 establishes that the BPCP cannot be sustained in equilibrium under arbitrary conditions. In other words, such an experiment can be part of an equilibrium only under very specific circumstances: that the intermediary has no combined informativeness over the prior or can perfectly identify t_g with her private information. Consistent with Theorem 2, in the former case, the only attainable expected utility for the receiver is zero, whereas in the latter, the receiver's expected utility is strictly positive.

As the BPCP is designed to leave the receiver indifferent between a_1 and a_2, the only way the intermediary recommends a_1 is when he can be certain that the state is not t_b. Indeed, in such a case, the AVIPI is strictly positive and the intermediary can attain a strictly positive payoff for the receiver by simply ignoring the sender. This reveals that the BPCP is only vacuously sustainable in equilibrium when the intermediary has high-quality information. Alternatively, the only case in which the intermediary maximizes surplus extraction via the BPCP is when the intermediary

has no valuable information (null combined informativeness).

We can represent the locus of accuracy levels supporting the BPCP as an equilibrium in the (ϵ_g, ϵ_b) plane of Figure 1. Specifically, such a locus consists of three segments: $[(0,0),(1,0)]$, $[(1,0),(0,1)]$, and $[(0,1),(1,1)]$, forming an "inverted Z." We depict such a locus in Figure 2, as the black solid line. The top and bottom segments fall in the blue region representing a strictly positive AVIPI, hence, the strictly positive expected utility for the receiver. The counter-diagonal falls in the red region representing a zero AVIPI, yielding a zero expected utility for the receiver.

5. Discussion

Related Literature

We extend Kamenica and Gentzkow's (2011) Bayesian persuasion framework to accommodate a communication intermediary and private information. The two papers that are closest to our work are Kolotilin et al. (2017) and Perez-Richet and Skreta (2022). Kolotilin et al. (2017) studies persuasion mechanisms that condition the sender's experiment on a receiver's report of her private information. Our work differs mainly in two ways. First, the private information that our intermediary possesses is state-relevant, not payoff-relevant.[23] Second, consistent with our applications, we do not allow agents to communicate before the information structures are considered. Perez-Richet and Skreta (2022) studies optimal test design when the sender can falsify the state of the world after paying a cost. The main difference with our work is the existence of evidence available to both the sender and intermediary.

Our work also relates to the literature on information design with mediators. Ivanov (2010) studies whether a principal can improve over direct communication by choosing among multiple uninformed and self-interested mediators. Kosenko (2020) studies equilibrium information revelation under an uninformed, self-interested mediator. In contrast, we focus on the role of a privately-informed mediator representing the receiver's interests in mitigating persuasive communication and how her efficacy depends on the quality of her private information.

Several articles study sequential mediation by an arbitrary number of mediators. Ambrus et al. (2013), extends Crawford and Sobel's (1982) cheap-talk framework

[23] Equivalence is not guaranteed since Kolotilin et al. (2017) requires that the receiver's payoff is linear in the parameter over which she possesses private information.

to study whether mediated communication improves outcomes over direct communication. In the problem of hierarchical Bayesian persuasion by self-interested senders, Arieli et al. (2022) focuses on characterizing the sender's optimal value and extending concavification techniques.[24] Mahzoon (2022) focuses on the relationship between the receiver's welfare on the biases of uninformed intermediaries. Wu (2021) uses recursive concavification to study equilibrium information revelation. Lastly, Li and Norman (2021) study how the number of senders affects equilibrium information revelation.

More generally, our model can be considered one of Bayesian persuasion with multiple senders and receivers. In our model, however, the intermediary is not purely a sender or a receiver but both: she receives a persuasive message from a sender and uses it to design a new experiment to communicate with the receiver. As such, our framework presents substantial differences with the multi-receiver and multi-sender strategic communication literature (Kamenica, 2019; Milgrom and Roberts, 1986; Battaglini, 2002; Gentzkow and Kamenica, 2016a; Shin, 1998).[25] A critical difference with the multi-receiver literature is that, in our model, the intermediary does not choose a payoff-relevant action but an experiment. Our framework presents two crucial differences with the multi-sender literature. First, we consider a privately-informed intermediary. Second, the intermediary's preferences coincide with the receiver's.

Finally, our model can be recast into one with receiver commitment without transfers, where the most notable application is mechanism design for optimal delegation, in the vein of Holmström (1978, 1984); Dessein (2002); Melumad and Shibano (1991); Alonso and Matouschek (2008). This is possible after noting that the intermediary and the receiver share the same preferences. As such, our model is strategically equivalent to one without an intermediary, where the receiver commits to a stochastic mapping from realizations of the sender's message and the private information into actions before such message and private information are realized. Similarly, the optimal delegation literature studies the case where a sender has private information that can be communicated through messages while the receiver commits to a deterministic or stochastic mapping from messages into actions. In contrast, we analyze how private information affects the efficacy of persuasive communication.

[24] Although our method of proof is different, concavification techniques are key in Kamenica and Gentzkow's (2011) and much of the literature following it; they were first found in Aumann and Maschler's analysis of repeated games with incomplete information (Aumann et al., 1995).

[25] The literature about Bayesian persuasion with multiple receivers typically evolves around specific applications. Kamenica (2019) provides a good survey.

Concluding Remarks

In this paper, we have constructed a simple model of persuasive communication with one sender (he), one receiver (she), one privately-informed intermediary (she) representing the receiver's interests, and private information about the state of the world. A salient feature of our model is that the private information is unverifiable by the receiver.

With our model, we answered several relevant questions. Can private information or an intermediary improve the receiver's payoff over the payoff of her default action? Does private information's effectiveness depend on who communicates it to the receiver? Theorem 1 implies that private information alone cannot mitigate persuasive communication; an intermediary is needed. Moreover, Theorem 2 complements the finding by the observation that private information must be accurate enough. Can private information or an intermediary change the sender's communication strategy? Propositions 3 through 5 imply that except for very specific vacuous conditions, the sender modifies his behavior to account for the intermediary. However, the extent to which it benefits the receiver depends on the accuracy of the private information.

We considered a simple binary-action and two states model to gain tractability and generate our main insights. However, this fact does not hinder the applicability or extension of our results. With our simplifying assumptions, we can interpret our model as one reflecting situations where agents face simple accept-reject choices in a world where the state variable is either "good enough to grant acceptance" or not. Thus, we have focused on simple "direct" communication mechanisms instead of considering general mechanisms. Two remaining relevant research questions to be explored in future work are whether the revelation principle applies to our framework and, if not, how our results are affected by more general mechanisms.

Two additional examples of our framework that the reader might consider are the following. First, a pharmaceutical company (sender) designs a clinical trial (communication) to persuade the FDA (receiver) to approve a new drug, which can be dangerous (bad) or safe (good) for the general population. The FDA typically designates a technical committee (intermediary) to review related scientific evidence (private information) to make a recommendation. Second, a Ph.D. advisor (sender) writes a recommendation letter (communication) to persuade a college (receiver) to hire his student, who can be a good or a bad teacher. The college typically appoints a search committee (intermediary) to evaluate the candidate independently and make

a final recommendation.

Speaking to the generality of our model, the insights in Theorem 2 extend to an arbitrary finite number of states and actions. First, because our model belongs to a highly-tractable class of rich-state-space models used in several applications (Gentzkow and Kamenica, 2016b; Perez-Richet and Skreta, 2022; Kolotilin et al., 2017). Second, at a technical level, because our proofs do not hinge on a binary action or the existence of only two states. On the one hand, the AVIPI definition naturally extends to an arbitrary number of actions and states. On the other hand, the two tools that a strictly positive AVIPI provides to the intermediary remain unchanged and are present also in such a case.

First, the intermediary can still resort to profitably ignoring the sender. Second, the intermediary can still identify the sender's recommendations that benefit the receiver. Finally, a zero AVIPI still allows the sender to extract the receiver's surplus over the AVIPI. The latter fact holds true because the existence of persuasion channels when the AVIPI is zero only depends on the continuity of the receiver's expected utility on the sender's experiment, a fact that is independent of the number of actions and states.[26]

[26]Lemma 5 in the Mathematical Appendix proves this fact.

References

Alonso, R. and Matouschek, N. . Optimal delegation. *The Review of Economic Studies*, 75(1):259–293, 2008.

Ambrus, A. , Azevedo, E. M. , and Kamada, Y. . Hierarchical cheap talk. *Theoretical Economics*, 8(1):233–261, 2013.

Arieli, I. , Babichenko, Y. , and Sandomirskiy, F. . Bayesian persuasion with mediators. *arXiv preprint arXiv:2203.04285*, 2022.

Aumann, R. J. , Maschler, M. , and Stearns, R. E. . *Repeated games with incomplete information*. MIT press, 1995.

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

Ben-Porath, E. , Dekel, E. , Lipman, B. L. , and Draft, F. . Mechanism design for acquisition of/stochastic evidence. *Hebrew University, Northwestern University and Boston University*, 2021.

Bergemann, D. and Morris, S. . Bayes correlated equilibrium and the comparison of information structures in games. *Theoretical Economics*, 11(2):487–522, 2016.

Bergemann, D. and Morris, S. . Information design: A unified perspective. *Journal of Economic Literature*, 57(1):44–95, 2019.

Crawford, V. P. and Sobel, J. . Strategic information transmission. *Econometrica: Journal of the Econometric Society*, pages 1431–1451, 1982.

Dessein, W. . Authority and communication in organizations. *The Review of Economic Studies*, 69(4):811–838, 2002.

FDA. Thimerosal and vaccines, Jan 2018. URL https://www.fda.gov/vaccines-blood-biologics/safety-availability-biologics/thimerosal-and-vaccines.

Forges, F. . An approach to communication equilibria. *Econometrica: Journal of the Econometric Society*, pages 1375–1385, 1986.

Gentzkow, M. and Kamenica, E. . Competition in persuasion. *The Review of Economic Studies*, 84(1):300–322, 2016a.

Gentzkow, M. and Kamenica, E. . A rothschild-stiglitz approach to bayesian persuasion. *American Economic Review*, 106(5):597–601, 2016b.

Holmström, B. . On the theory of delegation," in: Bayesian models in economic theory. ed. by m. boyer, and r. kihlstrom. north-holland, new york. 1984.

Holmström, B. R. . *On Incentives and Control in Organizations*. Stanford University, 1978.

Ivanov, M. . Communication via a strategic mediator. *Journal of Economic Theory*, 145(2):869–884, 2010.

Kamenica, E. . Bayesian persuasion and information design. *Annual Review of Economics*, 11:249–272, 2019.

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

Kolotilin, A. , Mylovanov, T. , Zapechelnyuk, A. , and Li, M. . Persuasion of a privately informed receiver. *Econometrica*, 85(6):1949–1964, 2017.

Kosenko, A. . Mediated persuasion. *arXiv preprint arXiv:2012.00098*, 2020.

Li, F. and Norman, P. . Sequential persuasion. *Theoretical Economics*, 16(2):639–675, 2021.

Mahzoon, M. . Hierarchical bayesian persuasion: Importance of vice presidents. *arXiv preprint arXiv:2204.05304*, 2022.

Melumad, N. D. and Shibano, T. . Communication in settings with no transfers. *The RAND Journal of Economics*, pages 173–198, 1991.

Milgrom, P. and Roberts, J. . Relying on the information of interested parties. *The RAND Journal of Economics*, pages 18–32, 1986.

Mnookin, S. . *The panic virus: a true story of medicine, science, and fear*. Simon and Schuster, 2011.

Perez-Richet, E. and Skreta, V. . Test design under falsification. *Econometrica*, 90 (3):1109–1142, 2022.

Shin, H. S. . Adversarial and inquisitorial procedures in arbitration. *The RAND Journal of Economics*, pages 378–405, 1998.

Wu, W. . Sequential bayesian persuasion. 2021.

A. Mathematical Appendix

Proof of Lemma 1 *Let* $\Gamma(\sigma_1, \sigma_2) = \sigma_1 s(a_1 \mid t) + \sigma_2 s(a_2 \mid t)$. $\Gamma(0,0) = 0$, $\Gamma(1,1) = 1$, *and* Γ *is continuous. The intermediate value theorem implies that for any* $y \in [0,1]$ *there exists* σ_1^y *and* σ_2^y *such that* $\Gamma(\sigma_1^y, \sigma_2^y) = y$.■

Proof of Theorem 1 *The sender's expected utility and the obedience constraint depend only on* $\sigma(a \mid t)$. *However, we can write* $\sigma_t = \sigma(a \mid t) = \sum_{\hat{\sigma}} \sigma(a \mid \hat{s}, t) s(\hat{s} \mid t)$. *Therefore, we can write the sender's optimization problem as*

$$\max_{\sigma} \frac{1}{2}(\sigma_{t_g} + \sigma_{t_b})$$

$$s.t.\ OC,$$

$$s.t.\ \sigma_t = \sum_{\hat{\sigma}} \sigma(a \mid \hat{s}, t) s(\hat{s} \mid t) \text{ for all } t.$$

As for any σ_t *there exists* $\sigma(a \mid \hat{s} = a_1, t)$ *and* $\sigma(a \mid \hat{s} = a_1, t)$ *such that IS is met, IS represents no additional restriction on the sender's problem over OC. Therefore, the solution to the problem is* $\sigma_t = \sigma_t^{BP}$. *Therefore, any* $\sigma(a \mid \hat{s}, t)$ *meeting IS is sender-optimal. Player's expected utilities follow from direct computation.*

Proof of Proposition 1 *Under the assumptions* $1 > \epsilon_b > 0$, *the intermediary maximizes the receiver's payoff only if* $\mu(a_1 \mid t, a_2) = 0$ *for all* t. *If* $\epsilon_g + \epsilon_b > 1$, *the receiver's payoff is strictly increasing in* $\mu(a_1 \mid t_g, a_1) - \mu(a_1 \mid t_b, a_1)$, *so the maximizing solution is* $\mu(a_1 \mid t_g, a_1) = 1$ *and* $\mu(a_1 \mid t_b, a_1) = 0$. *The opposite is true if* $\epsilon_g + \epsilon_b < 1$. *Finally, if* $\epsilon_g + \epsilon_b = 1$, *the receiver is indifferent between any* $\mu(a_1 \mid t_g, a_1)$ *and* $\mu(a_1 \mid t_b, a_1)$, *so the tie-breaker implies that* $\mu(a_1 \mid t, a_1) = 1$ *for all* t.

As a result, we can compute

$$V_\mu^S = \begin{cases} \frac{1}{2}(\epsilon_g + \epsilon_b - 1), & \text{if } \epsilon_g + \epsilon_b - 1 \geq 0 \\ \frac{1}{2}(1 - \epsilon_g - \epsilon_b), & \text{if } 1 - \epsilon_g - \epsilon_b > 0 \end{cases}.$$

We conclude that $V_\mu^S = \frac{1}{2} I(\epsilon_g, \epsilon_b)$.

■

Proof of Lemma 2 *Let* $V(\mu, \sigma)$ *be the receiver's payoff from the experiment that follows from any given policies* (σ, μ). *The obedience constraint in equation OC can*

be written as $V(\mu, \sigma) \geq 0$. *Notice that* $\mu(a_1 \mid \hat{s}, \hat{\sigma}) = 0$ *for all* $(\hat{s}, \hat{\sigma})$ *is a feasible experiment, which we call* μ_0. *Under* μ_0, $\mu_t = 0$ *for all* $t \in \{t_g, t_b\}$, *and thus, for any* σ, $V(\mu_0, \sigma) = 0$. *Then, for a given* σ, $\max_\mu V(\mu, \sigma) \geq V(\mu_0, \sigma) = 0$ *and the intermediary's best response to any* σ *leads to an obedient experiment.*[27]

■

Proof of Proposition 2 *Fix* $(\epsilon_g, \epsilon_b) \in [0,1]^2$ *and* $x \in (0,1)$. *Consider the set* $\Omega = [0,1]^6 \subset \mathbb{R}^6$ *and the correspondence* $\phi : \Omega \rightrightarrows \Omega$ *defined through equations 13 and 14 as*[28]

$$\phi = \begin{pmatrix} \mu^*(a_1 \mid t_g, a_1) \\ \mu^*(a_1 \mid t_g, a_2) \\ \mu^*(a_1 \mid t_b, a_1) \\ \mu^*(a_1 \mid t_b, a_2) \\ \sigma^*(a_1 \mid t_g) \\ \sigma^*(a_1 \mid t_b) \end{pmatrix}.$$

Given that $\mu^*(a_2 \mid \hat{s}, \hat{\sigma}) = 1 - \mu^*(a_1 \mid \hat{s}, \hat{\sigma})$ *and* $\sigma^*(a_2 \mid t) = 1 - \sigma^*(a_1 \mid t)$, *a fixed point* $\phi(\omega) = \omega \in \Omega$ *is an equilibrium of the game. At least one such fixed point exists by Kakutani's Fixed Point Theorem because*

1. Ω *is nonempty, convex, and compact.*
2. $\phi(\omega)$ *is nonempty for every* $\omega \in \Omega$.
3. $\phi(\omega)$ *is convex-valued for every* $\omega \in \Omega$ *because each dimension is either single-valued or the convex set* $[0,1]$.
4. ϕ *is upper hemicontinuous because* Ω *is compact and* ϕ *is closed-graph. The latter follows from the fact that each dimension of* ϕ *is closed-graph itself. For*

[27] Notice that the maximum is attained because $0 \leq \mu_t \leq 1$ for all t and $V(\mu, \sigma)$ is continuous in μ for all σ.

[28] Note well the slight abuse of notation. In general, μ^* is used as a specific value of the intermediary's policy, whereas in this proof it may be a set-valued function that allows for [0,1] as image.

instance, consider $\mu^*(a_1 \mid t_g, a_1)$*, as the rest of dimensions are analogous. For* $(\sigma(a_1 \mid t_g), \sigma(a_1 \mid t_b))$ *such that* $\overline{v}_{s,\sigma}(t_g, a_1) \neq 0$*,* $\mu^*(a_1 \mid t_g, a_1)$ *is a continuous function, hence, upper hemicontinuous. For* $(\sigma(a_1 \mid t_g), \sigma(a_1 \mid t_b)$ *such that* $\overline{v}_{s,\sigma}(t_g, a_1) = 0$*,* $\mu^*(a_1 \mid t_g, a_1) \in \{0, 1\}$ *is part of the best-reply correspondence, hence, the image of any sequence in the domain converging to* $(\sigma(a_1 \mid t_g), \sigma(a_1 \mid t_b))$ *is part of and converges inside the image.*

■

Proof of Lemma 3 *Equations 6 and 3 imply that*

$$V = \frac{1}{2}(x\mu_g - \mu_b) = \frac{1}{2}\left(\sum_{\hat{s},\hat{\sigma}} \mu(a_1 \mid \hat{s}, \hat{\sigma}) s(\hat{s} \mid t_g)\sigma(\hat{\sigma} \mid t_g)x - \sum_{\hat{s},\hat{\sigma}} \mu(a_1 \mid \hat{s}, \hat{\sigma}) s(\hat{s} \mid t_b)\sigma(\hat{\sigma} \mid t_b)\right) =$$

$$\sum_{\hat{s},\hat{\sigma}} \mu(a_1 \mid \hat{s}, \hat{\sigma})\left(\frac{1}{2}\big(x\sigma(\hat{\sigma} \mid t_g)s(\hat{s} \mid t_g) - \sigma(\hat{\sigma} \mid t_b)s(\hat{s} \mid t_b)\big)\right) = \sum_{\hat{s},\hat{\sigma}} \mu(a_1 \mid \hat{s}, \hat{\sigma})\overline{v}_{s,\sigma}(\hat{s}, \hat{\sigma}).$$

When $\overline{v}_{s,\sigma}(\hat{s}, \hat{\sigma}) > 0$*,* $\mu(a_1 \mid \hat{s}, \hat{\sigma}) = 1$ *is maximal. Likewise, when* $\overline{v}_{s,\sigma}(\hat{s}, \hat{\sigma}) < 0$*,* $\mu(a_1 \mid \hat{s}, \hat{\sigma}) = 0$ *is maximal. Finally,* $\overline{v}_{s,\sigma}(\hat{s}, \hat{\sigma}) = 0$*, all* $\mu(a_1 \mid \hat{s}, \hat{\sigma}) \in [0, 1]$ *lead to the same payoff. As a result, the intermediary's best response to* σ *is*

$$\mu^*(a_1 \mid \hat{s}, \hat{\sigma}) = \begin{cases} 1 & \text{if } \overline{v}_{s,\sigma}(\hat{s}, \hat{\sigma}) > 0, \\ \in [0, 1] & \text{if } \overline{v}_{s,\sigma}(\hat{s}, \hat{\sigma}) = 0, \\ 0 & \text{if } \overline{v}_{s,\sigma}(\hat{s}, \hat{\sigma}) < 0. \end{cases}$$

Similarly, Equations 5 and 3 imply that

$$U = \frac{1}{2}(\mu_g + \mu_b) = \frac{1}{2}\sum_{\hat{\sigma},\hat{s},t} \mu(a_1 \mid \hat{s}, \hat{\sigma})\sigma(\hat{\sigma} \mid t)s(\hat{s} \mid t) =$$

$$\frac{1}{2}\sum_{\hat{s},t}\big(\mu(a_1 \mid \hat{s}, a_1)\sigma(a_1 \mid t)s(\hat{s} \mid t) + \mu(a_1 \mid \hat{s}, a_2)(1 - \sigma(a_1 \mid t))s(\hat{s} \mid t)\big) =$$

$$\frac{1}{2}\sum_{\hat{s},t}\big(\sigma(a_1 \mid t)s(\hat{s} \mid t)(\mu(a_1 \mid \hat{s}, a_1) - \mu(a_1 \mid \hat{s}, a_2)) + \mu(a_1 \mid \hat{s}, a_2)s(\hat{s} \mid t)\big) =$$

$$\frac{1}{2}\left(\sum_{t} \sigma(a_1 \mid t)\overline{\triangle}_{\mu,s}(t) + \sum_{\hat{s},t} \mu(a_1 \mid \hat{s}, a_2)s(\hat{s} \mid t)\right).$$

When $\overline{\triangle}_{\mu,s}(t) > 0$, $\sigma(a_1 \mid t) = 1$ *is maximal. Likewise, when* $\overline{\triangle}_{\mu,s}(t) < 0$, $\sigma(a_1 \mid t) = 0$ *is maximal. Finally, when* $\overline{\triangle}_{\mu,s}(t) = 0$, *all* $\sigma(a_1 \mid t) \in [0,1]$ *lead to the same payoff. As a result, the sender's best response to* μ *is*

$$\sigma^*(a_1 \mid t) = \begin{cases} 1 & \text{if } \overline{\triangle}_{\mu,s}(t) > 0, \\ \in [0,1] & \text{if } \overline{\triangle}_{\mu,s}(t) = 0, \\ 0 & \text{if } \overline{\triangle}_{\mu,s}(t) < 0. \end{cases}$$

■

Proof of Theorem 2 *To prove the first statement, we establish the following two steps:*

Step 1: *In any equilibrium* (μ^*, σ^*), $V(\mu^*, \sigma^*) \geq A(\epsilon_g, \epsilon_b)$.

Proof of Step 1: Fix any equilibrium experiment σ^* *for the sender. In equilibrium, the intermediary's experiment is such that* $\mu^* \in \arg\max_\mu V(\mu, \sigma^*)$. *Therefore,* $V(\mu^*, \sigma^*) \geq V(\mu, \sigma^*)$ *for all* μ. *Note that the strategy where the intermediary ignores the sender, namely,* $\mu(a_1 \mid \hat{s}, a_1) = \mu(a_1 \mid \hat{s}, a_2)$ *for all* $\hat{s}$, *is always feasible. Denote such a strategy by* μ^A. *We have, then, that* $V(\mu^A, \sigma^*) = A(\epsilon_g, \epsilon_b) \leq V(\mu^*, \sigma^*)$.□

Step 2: *In any equilibrium* (μ^*, σ^*), $V(\mu^*, \sigma^*) > A(\epsilon_g, \epsilon_b)$ *is impossible.*

To establish Step 2, we prove a series of lemmata, deriving the necessary logical implications and detailing all possible cases.

Lemma 5 *Let* (μ, σ) *be any experiment profile with* $V(\mu, \sigma) > A(\epsilon_g, \epsilon_b)$. *Then, there exists a neighborhood with radius* $\gamma^* > 0$ *around* σ, *denoted* $N_{\gamma^*}(\sigma)$, *such that for all* $\sigma' \in N_{\gamma^*}(\sigma)$, $V(\mu, \sigma') > A(\epsilon_g, \epsilon_b)$.[29] *Thus, if* (μ^*, σ^*) *is an equilibrium such that* $V(\mu^*, \sigma^*) > A(\epsilon_g, \epsilon_b)$, *the following are true:*

(a) $\overline{\triangle}_{\mu^*}(t) > 0$ *implies that* $\sigma_t^* = 1$.

(b) $\overline{\triangle}_{\mu^*}(t) < 0$ *implies that* $\sigma_t^* = 0$.

[29] Formally, constrained to the probability simplex, $N_\gamma(\sigma) = \{\sigma' \in [0,1]^2 : ||\sigma - \sigma'|| < \gamma\}$, where $||\cdot||$ is the Euclidean distance.

Proof of Lemma 5 *Consider an experiment profile (μ, σ). Equation 10 implies that $\overline{v}_{s,\sigma}(\hat{s}, \hat{\sigma})$ is continuous in σ for all $(\hat{s}, \hat{\sigma})$, given s and x. Thus, equation 12 implies that, given any fixed μ, $V(\mu, \sigma)$ is continuous in σ. Fixing μ, then, for any $\theta > 0$, there is $\gamma_\theta > 0$ such that $|V(\mu, \sigma') - V(\mu, \sigma)| < \theta$ whenever $\sigma' \in N_{\gamma_\theta}(\sigma)$. This implies, in particular, that $V(\mu, \sigma) - \theta < V(\mu, \sigma')$. As $V(\mu, \sigma) > A(\epsilon_g, \epsilon_b)$ by assumption, it is possible to set $\theta^* < V(\mu, \sigma)$ so that $V(\mu, \sigma') > A(\epsilon_g, \epsilon_b)$. Thus, there exists $\gamma^* = \gamma_{\theta^*}$, such that for all $\sigma' \in N_{\gamma^*}$, $V(\mu, \sigma') > A(\epsilon_g, \epsilon_b)$.*

Let (μ^*, σ^*) be an equilibrium such that $V(\mu^*, \sigma^*) > A(\epsilon_g, \epsilon_b)$. Suppose that $\overline{\triangle}_{\mu^*}(t) > 0$ and $\sigma_t^* < 1$. Denote by $-t$ the state that is not t. Step 1 and the first part of this lemma imply that it is possible to find $\sigma_t' > \sigma_t^*$ such that $\sigma' = (\sigma_t', \sigma_{-t}^*)$ is obedient given μ^*. However, as $\overline{\triangle}_{\mu^*}(t) > 0$, $U(\mu^*, \sigma') > U(\mu^*, \sigma^*)$, a contradiction to (μ^*, σ^*) being an equilibrium. The case of $\overline{\triangle}_{\mu^*}(t) < 0$ is completely symmetric.□

Lemma 6 *$\overline{\triangle}_{\mu^*}(t_g)\overline{\triangle}_{\mu^*}(t_b) < 0$ is not an equilibrium possibility.*

Proof of Lemma 6 *Suppose (μ^*, σ^*) is an equilibrium such that $\overline{\triangle}_{\mu^*}(t_g) > 0$ and $\overline{\triangle}_{\mu^*}(t_b) < 0$. Then, Lemma 5 implies that $\sigma_g^* = 1$ and $\sigma_b^* = 0$. Therefore, we can write*

$$\overline{v}(\hat{s}, \hat{\sigma}) = \begin{cases} \epsilon_g x, & \text{if } (\hat{s}, \hat{\sigma}) = (t_g, a_1), \\ (1 - \epsilon_g)x, & \text{if } (\hat{s}, \hat{\sigma}) = (t_b, a_1), \\ -(1 - \epsilon_b), & \text{if } (\hat{s}, \hat{\sigma}) = (t_g, a_2), \\ -\epsilon_b, & \text{if } (\hat{s}, \hat{\sigma}) = (t_b, a_2). \end{cases}$$

Consider the case where $(\epsilon_g, \epsilon_b) \in (0, 1)^2$. Therefore, $\mu^*(a_1 \mid t_g, a_1) = \mu^*(a_1 \mid t_b, a_1) = 1$ and $\mu^*(a_1 \mid t_g, a_2) = \mu^*(a_1 \mid t_b, a_2) = 0$. Therefore, $\triangle\mu(\hat{s}) = 1$ for all $\hat{s}$, and so $\overline{\triangle}_{\mu^*}(t_b) > 0$, a contradiction. Any remaining combination of ϵ_g and ϵ_b leads to a similar contradiction.□

Let (μ^*, σ^*) be an equilibrium. Lemma 6 implies that we can have the following cases:

- *Case 1: $\overline{\triangle}_{\mu^*}(t_g)\overline{\triangle}_{\mu^*}(t_b) > 0$. Suppose that $\overline{\triangle}_{\mu^*}(t_g) > 0$ and $\overline{\triangle}_{\mu^*}(t_b) > 0$. Then, Lemma 5 implies that $\sigma_g^* = \sigma_b^* = 1$. As a result, we can write*

$$\overline{v}(\hat{s}, \hat{\sigma}) = \begin{cases} \underline{V}_A, & \text{if } (\hat{s}, \hat{\sigma}) = (t_g, a_1), \\ \underline{V}_I, & \text{if } (\hat{s}, \hat{\sigma}) = (t_b, a_1), \\ 0, & \text{otherwise.} \end{cases}$$

Therefore, the receiver's expected utility is

$$V(\mu^*, \sigma^*) = \mu^*(a_1 \mid t_g, a_1)\underline{V}_A + \mu^*(a_1 \mid t_b, a_1)\underline{V}_I.$$

There are only three possibilities: $\underline{V}_A \geq 0$ *and* $\underline{V}_I < 0$, $\underline{V}_A < 0$ *and* $\underline{V}_I \geq 0$, *or* $\underline{V}_A < 0$ *and* $\underline{V}_I < 0$. *As a result, the intermediary best-responds to the sender only if* $V(\mu^*, \sigma^*) = \max\{0, \underline{V}_I, \underline{V}_A\} = A(\epsilon_g, \epsilon_b)$. *A symmetric argument implies that* $V(\mu^*, \sigma^*) = A(\epsilon_g, \epsilon_b)$ *when* $\overline{\triangle}_{\mu^*}(t_g) < 0$ *and* $\overline{\triangle}_{\mu^*}(t_b) < 0$. *Thus, it is impossible to have* $V(\mu^*, \sigma^*) > A(\epsilon_g, \epsilon_b)$ *in this first case.*

- *Case 2:* $\overline{\triangle}_{\mu^*}(t_g) = \overline{\triangle}_{\mu^*}(t_b) = 0$. *Then, we can write*

$$\overline{\triangle}_{\mu^*}(t_g) = \epsilon_g \triangle\mu(\hat{s} = t_g) + (1 - \epsilon_g)\triangle\mu(\hat{s} = t_b) = 0,$$

and

$$\overline{\triangle}_{\mu^*}(t_b) = (1 - \epsilon_b)\triangle\mu(\hat{s} = t_g) + \epsilon_b \triangle\mu(\hat{s} = t_b) = 0.$$

Notice that ϵ_g, $1 - \epsilon_g$, ϵ_b, *and* $1 - \epsilon_b$ *cannot be zero simultaneously. Suppose* $\epsilon_g \neq 0$, *as the remaining cases are symmetric. Then, we can write* $\triangle\mu(\hat{s} = t_g) = -\frac{1-\epsilon_g}{\epsilon_g}\triangle\mu(\hat{s} = t_b)$, *and, thus,* $0 = \frac{1-\epsilon_g-\epsilon_b}{\epsilon_g}\triangle\mu(\hat{s} = t_b)$.

Two sub-cases arise. If $1 - \epsilon_g + \epsilon_b = 0$, *so* $A(\epsilon_g, \epsilon_b) = 0$, *the model is equivalent to one where the intermediary has no private information, as* $\mathbb{P}(t = t_g \mid \hat{s} = t_g) = \mathbb{P}(t = t_g \mid \hat{s} = t_b) = \frac{1}{2}$. *As a result, the persuasion problem for the sender is equivalent to the standard Kamenica and Gentzkow's (2011) Bayesian persuasion problem, where* $V(\mu^*, \sigma^*) = 0 = A(\epsilon_g, \epsilon_b)$.

In the second sub-case, $1 - \epsilon_g + \epsilon_b \neq 0$, *so* $\triangle\mu(\hat{s} = t_b) = 0 = \triangle\mu(\hat{s} = t_g)$. *On the one hand,* $\triangle\mu(\hat{s} = t_g) = 0$ *implies that* $\mu^*(a_1 \mid t_g, a_1) = \mu^*(a_1 \mid t_g, a_2)$, *whereas* $\triangle\mu(\hat{s} = t_b) = 0$ *implies that* $\mu^*(a_1 \mid t_b, a_1) = \mu^*(a_1 \mid t_b, a_2)$. *In other words, the intermediary bases a recommendation exclusively on* $\hat{s}$. *As a result,* $V(\mu^*, \sigma^*) = A(\epsilon_g, \epsilon_b)$. *Thus, it is again impossible to have* $V(\mu^*, \sigma^*) > A(\epsilon_g, \epsilon_b)$ *in this second case.*

- *Case 3:* $\overline{\triangle}_{\mu^*}(t_b) \neq 0$ *and* $\overline{\triangle}_{\mu^*}(t_g) = 0$ *(as the opposite case is symmetric). We can write*

$$\overline{\triangle}_{\mu^*}(t_b) = (1 - \epsilon_b)\triangle\mu(\hat{s} = t_g) + \epsilon_b \triangle\mu(\hat{s} = t_b) = 0.$$

If $\epsilon_b \in (0,1)$, we must have $\triangle\mu(\hat{s}=t_g) = \triangle\mu(\hat{s}=t_b) = 0$, so, as in case 2, $\mu^*(a_1 \mid t_g, a_1) = \mu^*(a_1 \mid t_g, a_2)$ and $\mu^*(a_1 \mid t_b, a_1) = \mu^*(a_1 \mid t_b, a_2)$. In other words, the intermediary bases a recommendation exclusively on $\hat{s}$. As a result, $V(\mu^*, \sigma^*) = A(\epsilon_g, \epsilon_b)$.

Suppose that $\overline{\triangle}_\mu(t_g) > 0$, as $\overline{\triangle}_\mu(t_g) < 0$ is symmetric. The remaining sub-cases are $\epsilon_b \in \{0,1\}$. Suppose $\epsilon_b = 1$, as $\epsilon_b = 0$ is symmetric. Then, $\overline{\triangle}_{\mu^*}(t_b) = 0$ implies that $\triangle\mu^*(\hat{s}=t_b) = 0$. Consequently, $\mu^*(a_1 \mid t_b, a_1) = \mu^*(a_1 \mid t_b, a_2)$. Moreover, as $\overline{\triangle}_{\mu^*}(t_g) > 0$, then we have $\epsilon_g > 0$, and, from Lemma 5, $\sigma_g^* = 1$. Substituting, we compute

$$\overline{v}(\hat{s},\hat{\sigma}) = \begin{cases} x\epsilon_g, & \text{if } (\hat{s},\hat{\sigma}) = (a_1, t_g), \\ (1-\epsilon_g)x - \sigma_b^*, & \text{if } (\hat{s},\hat{\sigma}) = (a_1, t_b), \\ 0, & \text{if } (\hat{s},\hat{\sigma}) = (a_2, t_g), \\ -(1-\sigma_b^*), & \text{if } (\hat{s},\hat{\sigma}) = (a_2, t_b). \end{cases}$$

Finally, substituting μ^* and $\overline{v}$ in V^*, we have

$$V(\mu^*, \sigma^*) = \mu^*(a_1 \mid t_g, a_1)\epsilon_g x + \mu(a_1 \mid t_b, a_1)[(1-\epsilon_g)x - 1].$$

As we have $\epsilon_g > 0$ and $(1-\epsilon_g)x - 1 < 0$, the only possibility is $\mu^*(a_1 \mid t_g, a_1) = 1$ and $\mu^*(a_1 \mid t_b, a_1) = 0$, which leads to $V(\mu^*, \sigma^*) = \epsilon_g x$. Finally, $A(\epsilon_g, \epsilon_b) = \max\{0, \epsilon_g x, (1-\epsilon_g)x - 1\} = \epsilon_g x$, and so we conclude that $V(\mu^*, \sigma^*) = A(\epsilon_g, \epsilon_b)$, and thus, that it is also impossible to have $V(\mu^*, \sigma^*) > A(\epsilon_g, \epsilon_b)$ in this final third case.

This concludes the proof of Step 2, thereby establishing the first statement of the theorem.

To conclude the proof, we show the equivalence between (i) and (ii) next. Suppose that (ii) holds. Then for some $(\hat{s},\hat{\sigma})$, $\overline{v}(\hat{s},\hat{\sigma}) > 0$, and so $\mu^*(a_1 \mid \hat{s},\hat{\sigma}) = 1$. As $\mu^*(a \mid \hat{s},\hat{\sigma})\overline{v}(\hat{s},\hat{\mu}) \geq 0$ for all $(\hat{s},\hat{\sigma})$, Equation 12 implies that $V_\mu^E(\mu^*, \sigma^*) > 0$. We conclude that (i) holds.

We show that (i) implies (ii) by a contra-positive argument. Suppose that (ii) fails, so $\overline{v}(\hat{s},\hat{\sigma}) \leq 0$ for all $(\hat{s},\hat{\sigma})$. As $\mu^*(a_1 \mid \hat{s},\hat{\sigma})\overline{v}(\hat{s},\hat{\mu}) \geq 0$ and $\mu^*(a_1 \mid \hat{s},\hat{\sigma}) \geq 0$ for all $(\hat{s},\hat{\sigma})$, we conclude that $V_\mu^E(\mu^*, \sigma^*) = 0$, so (i) does not hold, and the proof is complete.

■

Proof of Corollary 1 *Suppose* $|\epsilon_g + \epsilon_b - 1| > 1 - x$. *Two cases arise. First, if* $\epsilon_g + \epsilon_b - 1 > 1 - x$, *then Theorem 2 implies that* $V_\mu^E(\mu^*, \sigma^*) > 0$ *because*

$$\underline{v}_A(\epsilon_g, \epsilon_b) = x\epsilon_g - (1 - \epsilon_b) > (1 - \epsilon_g)(1 - x) \geq 0.$$

Second, if $1 - \epsilon_g - \epsilon_b > 1 - x$, *a similar argument implies that* $\underline{v}_I(\epsilon_g, \epsilon_b) > 0$, *leading to* $V_\mu^E(\mu^*, \sigma^*) > 0$. *A counter-example to show the lack of necessity is provided in the main text, so part (i) is complete.*

For part (ii), we proceed by a contra-positive argument. Suppose that $I(\epsilon_g, \epsilon_b) = 0$. *Then,* $\epsilon_g - (1 - \epsilon_b) = 0$. *Therefore,*

$$\overline{v}_A(\epsilon_g, \epsilon_b) = x\epsilon_g - (1 - \epsilon_b) \leq 0,$$

and

$$\overline{v}_I(\epsilon_g, \epsilon_b) = x(1 - \epsilon_g) - \epsilon_b \leq 0.$$

This implies that (iii) of Theorem 2 does not hold, and hence, by the same result, that (i) does not either. We have shown that $I(\epsilon_g, \epsilon_b) = 0$ *implies that* $V_\mu^E(\epsilon_g, \epsilon_b) = 0$. *A counter-example for the lack of sufficiency is provided in the main text, so part (ii) of the corollary is complete.*

■

Proof of Lemma 4 *Note that* $\underline{V}_g > 0$ *if and only if (SA) holds. Meanwhile,* $\underline{V}_b > 0$ *if and only if (SI) holds. Furthermore,* $A(\epsilon_g, \epsilon_b) > 0$ *if and only if either* $\underline{V}_g > 0$ *or* $\underline{V}_b > 0$. *The result follows immediately.*

■

Proof of Proposition 3 *Part (i). To show the necessary condition, suppose* (μ^*, σ^*) *is an equilibrium such that* $\mu^*(a_1 \mid \hat{s}, a_1) = \mu^*(a_1 \mid \hat{s}, a_2)$ *for all* $\hat{s}$. *If* σ^* *leads to* $\overline{v}(\hat{s}, a_1)\overline{v}(\hat{s}, a_2) < 0$ *for some* $\hat{s}$, *it must be the case that* $\mu^*(a_1 \mid \hat{s}, a_1) = 0$ *and*

$\mu^*(a_1 \mid \hat{s}, a_2) = 1$ *or vice versa, a contradiction. Therefore, the only possibility is that* $\overline{v}(\hat{s}, a_1)\overline{v}(\hat{s}, a_2) \geq 0$ *for all* $\hat{s}$.

To show the sufficient condition, suppose that σ^* *is an experiment such that* $\overline{v}(\hat{s}, a_1)\overline{v}(\hat{s}, a_2) \geq 0$ *for all* $\hat{s}$. *Only four possibilities arise, for a given* $\hat{s}$:

- $\overline{v}(\hat{s}, a_1) = 0$. *In this case, any* $\mu(a_1 \mid \hat{s}, a_1) \in [0, 1]$ *is optimal for the intermediary. Let* $\mu^*(a_1 \mid \hat{s}, a_2)$ *be any intermediary's best response when* $(\hat{s}, a_2)$ *is observed. Then, we can set* $\mu^*(a_1 \mid \hat{s}, a_1) = \mu^*(a_1 \mid \hat{s}, a_2)$, *which leads to* $\triangle\mu(\hat{s}) = 0$.

- $\overline{v}(\hat{s}, a_2) = 0$. *In this case, any* $\mu(a_1 \mid \hat{s}, a_2) \in [0, 1]$ *is optimal for the intermediary. Let* $\mu^*(a_1 \mid \hat{s}, a_1)$ *be any intermediary's best response when* $(\hat{s}, a_1)$ *is observed. Then, we can set* $\mu^*(a_1 \mid \hat{s}, a_2) = \mu^*(a_1 \mid \hat{s}, a_1)$, *which leads to* $\triangle\mu(\hat{s}) = 0$.

- $\overline{v}(\hat{s}, a_1) < 0$ *and* $\overline{v}(\hat{s}, a_2) < 0$. *In this case, the intermediary's only best response is* $\mu(a_1 \mid \hat{s}, a_1) = \mu(a_1 \mid \hat{s}, a_2) = 0$, *leading to* $\triangle\mu(\hat{s}) = 0$.

- $\overline{v}(\hat{s}, a_1) > 0$ *and* $\overline{v}(\hat{s}, a_2) > 0$. *In this case, the intermediary's only best response is* $\mu(a_1 \mid \hat{s}, a_1) = \mu(a_1 \mid \hat{s}, a_2) = 1$, *leading to* $\triangle\mu(\hat{s}) = 0$.

We conclude that when σ^* *is such that* $\overline{v}(\hat{s}, a_1)\overline{v}(\hat{s}, a_2) \geq 0$, $\mu^*(a_1 \mid \hat{s}, a_1) = \mu^*(a_1 \mid \hat{s}, a_2)$ *for all* $\hat{s}$ *is a best response for the intermediary, which automatically makes* σ^* *obedient per Lemma 2. It is left to check that* σ^* *is a best response to* μ^*. *In the four cases above, the only possible, it is implied that* $\triangle\mu(\hat{s}) = 0$ *for all* $\hat{s}$. *Thus,* $\overline{\triangle}_\mu(t) = 0$ *for all* t. *Therefore, the sender is indifferent for all* $\sigma_t \in [0, 1]$ *for all* t. *As a consequence,* σ^* *is a best response to* μ^*.

Part (ii). Trivially, if $\sigma_g^* = \sigma_b^* = 0$, *the necessary and sufficient condition in part (i) is satisfied, independently of* ϵ_g *and* ϵ_b. *This is not the only type of equilibria where the intermediary always ignores the sender, however.* $\sigma_g^* = 1$ *and* $\sigma_b^* = x$ *when* $\epsilon_g = 1 - \epsilon_b$ *is a different possibility.*

Part (iii). Suppose in equilibrium $\mu^*(a_1 \mid \hat{s}, a_1) = \mu^*(a_1 \mid \hat{s}, a_2) = \mu^*(\hat{s})$ *for all* $\hat{s}$. *As a result, we can write from equation 3,*

$$\mu_g = \mu^*(t_g)\epsilon_g + \mu^*(t_b)(1 - \epsilon_g),$$

and

$$\mu_b = \mu^*(t_g)(1-\epsilon_b) + \mu^*(t_b)\epsilon_b.$$

Using equation 6, we can write the expected utility of following the recommendation a_1 *as*

$$V = \frac{1}{2}(\mu_g x - \mu_b) = \frac{1}{2}(\mu^*(t_g)[x\epsilon_g - (1-\epsilon_b)] + \mu^*(t_b)[x(1-\epsilon_g) - \epsilon_b]) \geq 0,$$

where the inequality follows from the fact that the equilibrium must be obedient, via Lemma 2. Notice that the sign of $x\epsilon_g - (1-\epsilon_b) < 0$ *is the sign of* $\underline{V}_g < 0$*, whereas the sign of* $x(1-\epsilon_g) - \epsilon_b < 0$ *is the sign of* $\underline{V}_b < 0$*. Then, when* $\underline{V}_{\hat{s}} < 0$*, the intermediary maximizes the receiver's expected utility by setting* $\mu^*(\hat{s}) = 0$*. Meanwhile, when* $\underline{V}_{\hat{s}} > 0$*, the intermediary maximizes the receiver's expected utility by setting* $\mu^*(\hat{s}) = 1$*. Finally, when* $\underline{V}_{\hat{s}} > 0$*, any* $\mu^*(\hat{s})$ *is optimal for the intermediary.*

■

Proof of Proposition 4 *For the necessary condition, suppose that* (μ^*, σ^*) *is an equilibrium where the intermediary always passes the sender's message along. Then,* $\mu(a_1 \mid \hat{s}, a_1) = 1$ *and* $\mu(a_1 \mid \hat{s}, a_2) = 0$ *for all* $\hat{s}$*. Consequently,* $\triangle\mu(\hat{s}) = 1$ *for all* $\hat{s}$*, and so* $\overline{\triangle}_\mu(t) = 1$ *for all* t*.*

The sender's best response to μ^* *is found through the following problem:*

$$\max_{\sigma} U_\mu(\mu^*, \sigma)$$

$$s.t.\ V_\mu^E(\mu^*, \sigma) \geq 0.$$

Substituting the values of μ^* *and* $\overline{v}(\hat{s}, \hat{\sigma})$*, Equation 15 implies that such a problem is equivalent to*

$$\begin{aligned} &\max_{\sigma_g, \sigma_b} \sigma_g + \sigma_b \\ &s.t.\ x\sigma_g - \sigma_b = 0. \end{aligned} \tag{19}$$

The only solution to the problem is $\sigma_g^* = 1$ *and* $\sigma_b^* = x$*. Now, as* (μ^*, σ^*) *is an equilibrium, it must be the case that* μ^* *is a best response to* σ^**. So it must be the case that for all* $\hat{s}$*,* $\overline{v}(\hat{s}, a_1) \geq 0$ *and* $\overline{v}(\hat{s}, a_2) \leq 0$*. Under* σ^**,* $\overline{v}(\hat{s}, a_2) \leq 0$ *holds. However,*

$\overline{v}(\hat{s}, a_1) \geq 0$ *for all* $\hat{s}$ *holds if and only if* $I(\epsilon_g, \epsilon_b) \geq 0$ *and* $I(\epsilon_g, \epsilon_b) \leq 0$. *We conclude that* $I(\epsilon_g, \epsilon_b) = 0$.

For the sufficient condition, suppose that $I(\epsilon_g, \epsilon_b) = 0$, *and consider BPCP* $\sigma_g^{BP} = 1$ *and* $\sigma_b^{BP} = x$. *Under these conditions, for all* $\hat{s}$, $\overline{v}(\hat{s}, a_1) = 0$ *and* $\overline{v}(\hat{s}, a_2) \leq 0$. *Thus, a possible best response for the intermediary is* $\mu^*(a_1 \mid \hat{s}, a_1) = 1$ *and* $\mu^*(a_1 \mid \hat{s}, a_2) = 0$. *As outlined before, this implies that* $\overline{\triangle}_\mu(t) = 1$ *for all* t. *Consequently,* (μ^*, σ^{BP}) *constitutes an equilibrium only if* σ^{BP} *is a solution to the problem in Equation 19, which it does.*

■

Proof of Proposition 5 *Suppose that* $\sigma_g^* = 1$ *and* $\sigma_b^* = x$. *Part (a) follows from Proposition 4, which implies that* $V = 0$ *and, as the sender recommends* a_2 *with probability one, that* $U = 0$.

For part (b), we compute the following:

$$\overline{v}(\hat{s}, \hat{\sigma}) = \begin{cases} \frac{x}{2}(\epsilon_g + \epsilon_b - 1), & \text{if } (\hat{s}, \hat{\sigma}) = (t_g, a_1), \\ \frac{x}{2}(1 - \epsilon_g - \epsilon_b), & \text{if } (\hat{s}, \hat{\sigma}) = (t_b, a_1), \\ -\frac{(1-x)}{2}(1 - \epsilon_b), & \text{if } (\hat{s}, \hat{\sigma}) = (t_g, a_2), \\ -\frac{(1-x)}{2}\epsilon_b, & \text{if } (\hat{s}, \hat{\sigma}) = (t_b, a_2). \end{cases}$$

- *Case 1:* $\epsilon_g + \epsilon_b < 1$. *In this case,* $\epsilon_t < 1$. *We have that* $\overline{v}(t_g, a_1) < 0$ *and* $\overline{v}(t_b, a_1) > 0$, *so* $\mu^*(a_1 \mid t_g, a_1) = 0$ *and* $\mu^*(a_1 \mid t_b, a_1) = 1$. *If* $\epsilon_b \in (0, 1)$, $\overline{v}(t_g, a_2) < 0$ *and* $\overline{v}(t_b, a_2) < 0$, *so* $\mu^*(a_1 \mid \hat{s}, a_2) = 0$ *for all* $\hat{s}$. *Then,* $\triangle\mu(\hat{s} = t_g) = 0$ *and* $\triangle\mu(\hat{s} = t_b) = 1$, *leading to* $\overline{\triangle}_\mu(t = t_b) = \epsilon_b > 0$. *As* $\overline{v}(t_b, a_1) > 0$, *Theorem 2 implies that* $V_\mu(\mu^*, \sigma^*) > 0 = A(\epsilon_g, \epsilon_b)$. *Then, Lemma 5 implies that* $\triangle\mu(\hat{s} = t_b) = 1$ *implies that* $\sigma_t^* = 1$, *a contradiction.*

  The only case left to consider is $\epsilon_b = 0$. *As* $I(\epsilon_g, \epsilon_b) > 0$, *we must have* $\epsilon_g < 1$. *In this case,* $\overline{v}(t_g, a_2) < 0$ *and* $\overline{v}(t_b, a_2) = 0$, *leading to* $\mu^*(a_1 \mid t_g, a_2) = 0$ *and* $\mu^*(a_1 \mid t_b, a_2) \in [0, 1]$. *Thus,* $\triangle\mu(\hat{s} = t_g) = 0$ *and* $\triangle\mu(\hat{s} = t_b) = 1 - \mu^*(a_1 \mid t_b, a_2) \geq 0$. *Therefore,* $\overline{\triangle}_\mu(t_g) = (1 - \epsilon_g)\triangle\mu(\hat{s} = t_b)$ *and* $\overline{\triangle}_\mu(t_b) = 0$. *It is left to check that the BPCP is optimal under* μ^*, *namely, that the sender does not want to deviate. Indeed, as* $\overline{\triangle}_\mu(t_b) = 0$, $\sigma_b^* = x$ *is part of a best response for the sender.*

Moreover, as $\overline{\triangle}_\mu(t_g) \geq 0$, $\sigma_g^* = 1$ *is part of a best response for the sender, as well. Finally, as* $\overline{v}(t_b, a_1) > 0$, *Theorem 2 implies that* $V(\mu^*, \sigma^*) > 0 = A(\epsilon_g, \epsilon_b)$, *so* σ^* *is in fact obedient. We conclude that* (μ^*, σ^*) *is valid as an equilibrium.*

Finally, algebra implies that we can write

$$U_\mu^E = \frac{1}{2}\Big(\overline{\triangle}(t_g)+x\overline{\triangle}(t_b)+\big(\epsilon_g+(1-\epsilon_b)\big)\mu^*(a_1 \mid t_g, a_2)+\big((1-\epsilon_g)+\epsilon_b\big)\mu^*(a_1 \mid t_b, a_2)\Big) =$$

$$\frac{1}{2}(1-\epsilon_g) = \frac{1}{2}|\epsilon_g + \epsilon_b - 1| = |\epsilon_g + \epsilon_b - 1|(\frac{1}{2}(x+1) - \frac{1}{2}x),$$

where the last equality follows from the facts that $\epsilon_b = 0$ *and* $\epsilon_g + \epsilon_b < 1$.

As only $\overline{v}(t_b, a_1) > 0$ *when* $\mu^*(a_1 \mid t_b, a_1) = 1$, *then*

$$V_\mu^E = \overline{v}(t_b, a_1) = \frac{x}{2}|\epsilon_g + \epsilon_b - 1|.$$

- *Case 2:* $\epsilon_g + \epsilon_b > 1$. *The proof is completely symmetric to Case 1.*

■