Bayesian Doublespeak

Ing-Haw Cheng, University of Toronto

Alice Hsiaw, Brandeis University

Working Paper Series

2023 | 55

Bayesian Doublespeak*

Ing-Haw Cheng† Alice Hsiaw‡

July 2023

Abstract

We show that misinformation distorts long-run beliefs in "doublespeak'' equilibria of a cheap talk game where receivers are uncertain of a state and the sender's type. A sender type who prefers receivers take wrong actions sends messages that plausibly come from a good type under a different state. Even after observing infinite messages, receivers disagree about the state and take different ex-post actions. A policymaker who believes that doublespeak would mislead receivers may restrict the sender to finite messages. An option for receivers to fact-check messages does not limit doublespeak, but sender concerns about reputation can.

Keywords: Misinformation, Disinformation, Disagreement, Polarization, Fake News

*The authors thank Avidit Acharya, S. Nageeb Ali, Dana Foarta, Robert Gibbons, Francesco Giovannoni, Navin Kartik, Andrew Little, Stephen Morris, Thomas Palfrey, Daniel Quigley, Jesse Shapiro, Adam Szeidl, Catherine Thomas, Dustin Tingley, seminar participants at Brandeis University, Pennsylvania State University, Virtual Seminars in Economic Theory, and conference participants at Stanford Institute for Theoretical Economics, National Bureau of Economic Research Organizational Economics meeting, and North American Summer Meeting of the Econometric Society for comments.

†Rotman School of Management, University of Toronto, 105 St. George St., Toronto, ON M5S3E6 (email: inghaw.cheng@rotman.utoronto.ca)

‡International Business School, Brandeis University, 415 South St., Waltham, MA 02453 (email: ahsiaw@brandeis.edu).

Can misinformation distort the long-run beliefs and actions of rational agents? The usual presumption is that rational agents should learn the truth about an unknown state of the world in the long run and can pierce through misinformation in equilibrium (Savage, 1954; Blackwell and Dubins, 1962; Stein, 1989; Fudenberg and Tirole, 1986; Holmström, 1999). However, growing amounts of misinformation on topics such as election fraud, vaccine safety, and climate change have raised fresh interest in its effect on behavior and heightened concerns about polarization and distrust in institutions (Myers and Sullivan, 2022).

This paper characterizes when and how misinformation distorts long-run beliefs and actions in "doublespeak" equilibria. Our foundation is the observation from Acemoglu et al. (2016) that rational agents may fail to learn the truth in the long run if they are exogenously uncertain about the distribution of signals they receive. We endogenize the uncertainty about signal distributions within a cheap talk game where receivers are uncertain of both the state of the world and the sender's preferences. Our core insight is that, even after observing an infinite number of messages, receivers only partially learn the state ex-post whenever both good and non-good sender types are possible ex-ante. Intuitively, the messages themselves do not resolve uncertainty over the meaning of messages since non-good types play strategies we call "doublespeak" that confound long-run learning.

Section 1 introduces our model. A continuum of receivers with possibly heterogeneous priors take an action "in the long run" after observing an infinite sequence of messages about an unknown state of the world and updating beliefs using Bayes' rule. A strategic sender reports a message after seeing a private signal about the state for an infinite number of such private signals. Each signal is i.i.d. with an accuracy that is known among all players. The sender is uncertain about the state but knows her type, which determines her preferences over receivers' actions. We allow for a generalized set of Crawford and Sobel (1982) preferences, including an unconditional preference for a specific action and preferences for actions positively or negatively correlated with the true state. Receivers prefer to take an action consistent with the true state, but are uncertain about the sender's type and the state. The state, messages, and actions are binary.

A sender type *doublespeaks* if her messages do not match her private information and produce a long-run distribution of messages that could have plausibly been generated by

another sender type in a different state. Our first result is that, regardless of the true sender type, receivers learn the true state in the long run if and only if no sender type doublespeaks and no type sends pure noise. As an example of how doublespeak confounds long-run learning, suppose a sender observes a sequence of private signals about a binary unknown state that is 75% accurate and sends a message after each signal. One sender type truthfully reports her signals and the other sender type doublespeaks by flipping their information content. After many messages, receivers observe that 25% of messages are "0" and 75% are "1." If both strategies are possible in equilibrium, receivers learn that either the state is "1" and the sender is truthful or the state is "0" and the sender is doublespeaking, but cannot disentangle these two cases.

In Section 2, we characterize when doublespeak occurs in equilibrium. Doublespeak equilibria exist when receivers are initially unsure of whether the sender is a good type or not, and take one of two forms that we call mimicking and mirroring equilibria. A mimicking equilibrium occurs when one sender type is good in that she prefers that receivers take actions that are consistent with the true state, but the other type is single-minded in that she prefers that receivers take a specific action irrespective of the true state. The single-minded type sends messages consistent with her desired action that are in fact noise but which mimic a distribution of messages that the good type could generate. For example, if the sender's private signal is 75% accurate and the single-minded type wants receivers to take action consistent with state "1," she sends noise that is "1" with 75% probability and "0" with 25% probability for each private signal received. This distribution of messages is indistinguishable from what a truthful good sender would produce if the true state were "1."

A "mirroring" equilibrium occurs when one sender type is good but the other type is malevolent in that she prefers receivers take actions inconsistent with the true state. The malevolent type sends messages that are perfectly negatively correlated with those of the good type, for example by reporting "0" whenever her private signal is "1," and vice versa.

In doublespeak equilibria, receivers only partially learn the true state even after observing infinite messages (even if the true sender type is good), leading some receivers to take actions that depend on their priors. Some of these actions may be incorrect ex post. For example, Figure 1 shows how receiver actions vary across regions of prior beliefs in a mirror-

ing equilibrium. The horizontal axis, λ, indicates receivers' prior probability that the sender type is good, while the vertical axis, ω, indicates receivers' prior probability that the state is 1. Receivers with priors in the UNSURE regions take actions corresponding with their prior beliefs about the state as they are sufficiently unsure of sender's type ex post. Receivers in the TRUST region take an action consistent with the state indicated by the sender's messages, and those in the DISTRUST region take an action inconsistent with the indicated state. If, for example, the sender were actually a good type who delivered messages indicating state "1" ex post, any receivers in the DISTRUST and bottom-UNSURE region would take action consistent with state "0." Partial learning distinguishes doublespeak equilibria from babbling equilibria, where all receivers learn nothing and take actions based only on priors.

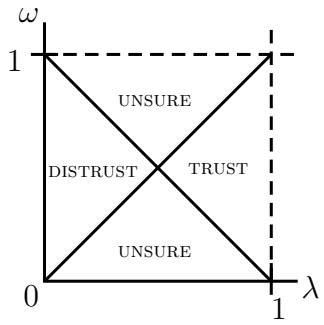


Figure 1: Receivers' actions in mirroring equilibrium

Section 3 shows when a policymaker whose objective is to maximize receivers' welfare chooses to allow some, but not unlimited, communication from the sender if she is also uncertain about both the state and sender's type. From the policymaker's perspective, messages would benefit receivers when receivers' priors are, on average, incorrect about the state or when the sender is sufficiently likely to be good. However, unlimited messages may not be best for receivers since a non-good type could more easily convince receivers to take the wrong action. Thus, she may restrict the sender to limited messages rather than unlimited. Our results characterize whether a policymaker would optimally choose, infinite, one, or no messages, depending on her own beliefs about the state and the sender type.

Section 4 shows that doublespeak equilibria exist even when receivers have access to a technology to verify the truthfulness of sender's messages before they choose actions. We allow receivers to pay a cost to "fact-check" and reveal the state after they have observed the

sender's messages. We characterize who fact-checks in equilibrium and show that, because fact-checking is endogenously limited in equilibrium, it does not affect the sender types required to sustain informative and doublespeak equilibria and thus does not induce more information transmission by senders. The overall welfare effect of giving receivers the option to fact-check is ambiguous since some receivers may needlessly fact check.

Section 5 shows that reputation concerns expand the set of sender types supporting fully informative equilibria, shrinks the set of sender types supporting mirroring equilibria, and expands the set of sender types for which only babbling equilibria exist. In this extension, senders' preferences depend on both receivers' actions and receivers' posterior beliefs that she is a good type. Only babbling equilibria may exist for certain sender types because reputation concerns deter doublespeak but are not strong enough to induce full pooling with a good type. Thus, reputation can either increase or decrease the amount of information transmission, depending on the possible sender types and the degree of reputation concern.

Section 6 discusses alternative assumptions and empirical implications. Doublespeak can occur even if receivers share a common prior, have a continuous action space, can fact-check with a second sender, when there are multiple senders (not just sender types), or when the sender knows the state. We provide empirical predictions and highlight evidence in different contexts to motivate future research.

Our theory extends the theoretical literature on long-run Bayesian learning and disagreement by incorporating strategic information transmission. Acemoglu et al. (2016) show that Bayesians can disagree in the long run when they are uncertain about the exogenous message distribution, in contrast to the longstanding literature showing that beliefs will converge to the truth when the information structure of messages is commonly known (Blackwell and Dubins, 1962; Kartik et al., 2021). We endogenize message distributions and long-run disagreement in terms of sender preferences and receivers' prior beliefs. Thus, we connect the extensive literatures on long-run disagreement, heterogeneous priors (Morris, 1995), and cheap talk (Crawford and Sobel, 1982).

Our work also relates to the applied literature on misinformation. Mullainathan and Shleifer (2005) show how sources can bias reports when consumers prefer information that confirms their beliefs. Gentzkow and Shapiro (2006) finds that a firm that wants to build

a reputation as a provider of accurate information tends to distort information toward a consumer's priors, and ex post verification weakens this incentive to distort. The mechanism and resulting information distortion in our model differs because the sender has preferences over receivers' actions rather than reputation alone. Cisternas and Vásquez (2023), Kranton and McAdams (2023), and Bowen et al. (2023) study the effect of selective news sharing on misinformation and belief polarization. Allcott and Gentzkow (2017) and Nyhan (2020) review the large body of empirical work in political science and economics on misinformation and misperceptions.

Finally, our work complements the literature explaining disagreement due to cognitive errors (Rabin and Schrag, 1999; Cheng and Hsiaw, 2022; Fryer et al., 2019; Ortoleva and Snowberg, 2015), misspecified models (Bohren and Hauser, 2021; Gentzkow et al., 2021; Szeidl and Szucs, 2022), and non-standard preferences (Baliga et al., 2013). Our work differs in that all agents are rational. Distinguishing doublespeak equilibria from other theories of disagreement and misinformation in different contexts is a fruitful area for future research.

1 Model

There is a continuum of receivers of mass 1 indexed by $i \in (0,1)$ and one sender. There are two principal dates, $\tau \in \{0,1\}$. At $\tau = 0$, nature chooses the state of the world $\theta \in \{0,1\}$ and sender type $j \in \{u,v\}$. Between dates 0 and 1, there are an infinite number of (sub-) periods indexed by n where the sender sends messages to receivers. Receivers take action $a_i \in \{0,1\}$ in the long run at $\tau = 1$, after which payoffs realize. This timing of messages and actions aligns with the long-run learning framework of Acemoglu et al. (2016).

Each receiver i has utility $-(a_i - \theta)^2$ and thus prefers to choose an action that corresponds to the state θ. Receivers are uncertain of the state θ and sender type j and learn about them from the sender's messages using Bayes' Rule. Receiver i has prior belief at $\tau = 0$ given by $(\lambda_i, \omega_i) \in (0,1) \times (0,1)$, where λ_i is the prior probability that $j = u$ and ω_i is the prior probability that $\theta = 1$. Prior beliefs over the state and sender type are independent: $P_i(j = u, \theta = 1) = \lambda_i \omega_i$. Receivers have possibly heterogeneous priors at $\tau = 0$ over the sender's type and the state. We let $f(\lambda, \omega)$ denote the density of receivers with prior (λ, ω),

and assume full support.[1]

Sender derives utility from receivers' actions corresponding with her type j. Type j's preferences over receivers' actions are given by: $-\int_0^1 [a_i - (c_j\theta + b_j)]^2 \, di$, which generalizes the parameterization from Crawford and Sobel (1982). The parameter b_j reflects the sender's desired receiver action when $\theta = 0$, and the sum $c_j + b_j$ reflects the sender's desired receiver action when $\theta = 1$.

We partition sender types into three regions based on what actions they desire from receivers. We say a sender type's preferences are "good" if the sender type prefers receivers take $a_i = 1$ if and only if $\theta = 1$, "single-minded" if the sender type prefers receivers always take $a_i = 1$, and "malevolent" if the sender type prefers receivers take $a_i = 1$ if and only if $\theta = 0$. We omit the case where sender types prefer receivers always take $a_i = 0$ since this case re-labels single-minded preferences.

Definition 1 (Sender preferences)**.** *A sender type j's preferences are* ***good*** *if $b_j \leq 1/2$ and $c_j + b_j \geq 1/2$,* ***single-minded*** *if $b_v \geq 1/2$ and $c_v + b_v \geq \frac{1}{2}$, and* ***malevolent*** *if $b_v \geq 1/2$ and $c_v + b_v \leq \frac{1}{2}$.*

Sender knows her own type but is uncertain about the state. At $\tau = 0$, she has prior belief that $\theta = 1$ with probability $\omega^S \in (0, 1)$. In each period n, sender observes a noisy private signal $s_n \in \{0, 1\}$ about θ with accuracy $\gamma \in (1/2, 1)$, so $P(s_n = \theta \mid \theta) = \gamma$. The accuracy γ is common knowledge, and signals are independently and identically distributed across periods. After observing s_n, the sender updates her beliefs using Bayes' Rule and costlessly announces a public message $m_n \in \{0, 1\}$. Let $\mathbf{m}_n$ denote the history of messages sent, and $\mathbf{s}_n$ denote the history of private signals, from subperiods 1 through n. Each sender type j chooses a strategy that specifies a probability of reporting $m_n = 1$ in each subperiod n as a function of her history of private signals and previous messages: $P_j(m_n = 1 | \mathbf{s}_n, \mathbf{m}_{n-1})$.

1.1 Receivers' long-run learning

We first establish several definitions and a key proposition about receivers' learning. Define the frequency of a history of messages $\mathbf{m}_n$ as the proportion of messages that are 1's:

[1]The assumption that $f(\lambda, \omega)$ has full support is not required for any of the main results. It is made purely for clarity of exposition because it rules out knife-edge cases that are not economically meaningful. We discuss this in detail in the Appendix.

Definition 2 (Frequency)**.** *The* ***frequency*** *of any finite history of messages* $\mathbf{m}_n$ *is* $p(\mathbf{m}_n) \equiv \frac{n_1}{n}$*, where* n_1 *is the number of ones reported in* $\mathbf{m}_n$*. The* ***long-run frequency*** *for an infinite history of messages* $\mathbf{m}_\infty$ *is* $p(\mathbf{m}_\infty) \equiv \lim_{n\to\infty} p(\mathbf{m}_n)$*, if such a limit exists.*

Suppose that sender strategies produce well-defined long-run frequencies almost surely. Let $p_{\theta j}$ denote the long-run frequency that sender type j's strategy produces conditional on the true state $\theta \in \{0, 1\}$. Note that we write $p_{\theta j}$ in terms of θ even though the sender conditions her message m_n on her signals $\mathbf{s}_n$ and does not know the state. For example, a sender who always reports $m_n = s_n$ produces $(p_{1j}, p_{0j}) = (\gamma, 1-\gamma)$ almost surely due to the strong law of large numbers.

We define misinformation as follows:

Definition 3 (Misinformation)**.** *Sender type* j*'s strategy is truthful (in the long run) if it produces long-run frequencies equal to* $p_{1j} = \gamma$ *and* $p_{0j} = 1-\gamma$*. A strategy that is not truthful delivers* ***misinformation****.*

We define doublespeak as a particular form of misinformation that produces long-run frequencies that are identical to those produced by another sender type in a different state:

Definition 4 (Doublespeak)**.** *Given the strategy of sender type* j'*, sender type* j ***double-speaks*** *if it plays a strategy that produces misinformation and long-run frequencies with* $p_{1j} = p_{0j'}$ *or* $p_{0j} = p_{1j'}$*, where* $j \neq j'$*.*

Proposition 1 says that receivers learn the true state in the long run regardless of sender type if and only if 1) no types send pure noise, and 2) no sender type doublespeaks. The intuition behind Proposition 1 is that long-run frequencies must identify the state in order for receivers to learn θ for certain. Doublespeak confounds long-run learning by garbling this identification. The implication of Proposition 1 is that receivers can indeed pierce through all misinformation in the long run, *except for* noise and doublespeak.

Proposition 1 (Long-run learning)**.** *When strategies produce message histories with well-defined long-run frequencies* $p(\mathbf{m}_\infty)$*, a receiver learns* θ *almost surely regardless of the true sender type and state if and only if:*

1. $p_{1j} \neq p_{0j}$ *for all* j*, and*
2. $p_{1j} \neq p_{0j'}$ *for all* j *and* $j' \neq j$*.*

A sketch of the proof illustrates the forwards direction. Suppose receivers learn θ for certain after observing $\mathbf{m}_\infty$ for all sender types, and proceed to prove statements (1) and (2) by contradiction. For (1), if $p_{1j} = p_{0j}$ for either $j = u$ or $j = v$, then one sender type's strategy is uninformative about the state, and receivers would not learn θ when the true sender is of that type. For (2), if $p_{1j} = p_{0j'}$ for some j and j', then when $p(\mathbf{m}_\infty) = p_{1j}$, receivers learn that either the true state is 1 and sender type was j, or that the true state was 0 and the sender type was j'. However, receivers cannot distinguish between these two.

We formulate our definitions and results in terms of strategies that produce long-run frequencies from a message space of $m_n \in \{0, 1\}$ for simplicity and concreteness. Neither long-run frequencies nor the specific message space we consider are required for our results. We can recharacterize our analysis for a completely unrestricted message space and obtain qualitatively identical results by redefining $p_{\theta j}$ to represent the history of messages that sender j's strategy produces (whatever their content) given the true state $\theta \in \{0, 1\}$. As Farrell and Rabin (1996) note, the sender's message space in any cheap talk game is very large because she could potentially "say anything," and what matters is the receivers' inference given a messaging strategy, not the specific messages per se.

2 Doublespeak Equilibria

Proposition 2 characterizes the set of equilibria that can exist over the space of sender types; note that babbling equilibria always exist (Crawford and Sobel, 1982). We employ the solution concept of perfect Bayesian Nash equilibrium. Our main result is that doublespeak equilibria, in which at least one sender type doublespeaks, exist whenever one sender type is good and the other sender type is not good. Thus, some receivers may fail to learn the state in the long run.

Proposition 2. *Non-babbling equilibria are characterized as follows:*

1. *Fully informative equilibria, in which senders play strategies satisfying Proposition 1, exist if and only if types u and v are good.*

2. *If both sender types are not good, only babbling equilibria exist.*

3. *Doublespeak equilibria, in which at least one sender type doublespeaks (Definition 4), exist if and only if u is good and v is not good. There are only two forms of doublespeak equilibria:*

   (a) ***Mimicking*** *equilibria, in which senders play strategies such that $p_{1u} \neq p_{0u}$ and $p_{1v} = p_{0v} = p_{1u}$, exist if and only if type u is good and type v is single-minded.*

   (b) ***Mirroring*** *equilibria, in which senders play strategies such that $p_{1j} \neq p_{0j}$ for all j, and $p_{1j} = p_{0j'}$ for all $j \neq j'$, exist if and only if type j is good and type j' is malevolent.*

Part 1 says that fully informative equilibria exist only when both sender types are good. Within fully informative equilibria, all receivers learn θ almost surely regardless of the true sender type and state. If a non-good type were a part of a fully informative equilibrium, they would have an incentive to deviate to a strategy that misleads receivers. For example, a non-good sender type that prefers receivers take $a_i = 0$ when $\theta = 1$ could profitably deviate to a messaging strategy that produces a corresponding long-run frequency that maps, under receivers' reasoning, to $\theta = 0$. Part 2 says that only babbling equilibria exist when neither sender type is good. If receivers believed messages conveyed information, then any non-good type would again have an incentive to deviate to a strategy that misleads receivers.

Part 3 says that doublespeak equilibria exist when one sender type is good and the other is not good. Within doublespeak equilibria, receivers partially learn about the state. Intuitively, receivers believe the messages convey information because a good sender type wants to communicate in a way that reveals the state. But this means that a non-good type can profitably doublespeak, obfuscating the meaning of the messages. There are only two forms of doublespeak equilibria, "mimicking" and "mirroring," which we describe below.[2]

2.1 Mimicking equilibrium

In any mimicking equilibrium, type u uses a messaging strategy that results in different long-run frequencies in each state. Type v "mimics" type u by using a messaging strategy that always generates the long-run frequency that type u would have produced in one state (e.g.,

[2]In the Appendix, we show that other forms of doublespeak equilibria only exist in knife-edge cases in which at least one sender has indifference about receivers' actions.

p_{1u}). The strategy induces a subset of receivers to take action 1 even when the true state is $\theta = 0$. The reason is that $p(\mathbf{m}_\infty) = p_{1u}$ is consistent with $(j, \theta) \in \{(u, 1), (v, 1), (v, 0)\}$. Thus, receivers cannot distinguish sender types, and $P_i(j, \theta | p(\mathbf{m}_\infty) = p_{1u}) < 1$ for all i.

An intuitive example of messaging strategies that fit this description is an equilibrium in which type u reports truthfully, and type v "mimics" u by randomizing so that v always generates a long-run frequency equal to γ, as in Table 1.

	$P(m_n = 1 \mid s_n = 0)$	$P(m_n = 1 \mid s_n = 1)$	p_{0j}	p_{1j}
Sender type u	0	1	$1 - \gamma$	γ
Sender type v	γ	γ	γ	γ

Table 1: Example of mimicking equilibrium.

What do receivers infer in equilibrium? If $p(\mathbf{m}_\infty) = p_{0u}$, then *all* receivers are sure that $(j, \theta) = (u, 0)$. But if $p(\mathbf{m}_\infty) = p_{1u}$, then receivers only know that $(j, \theta) \neq (u, 0)$. Receivers' posterior beliefs are $P(u, 0 | p(\mathbf{m}_n) = p_{1u}) = 0$ and:

$$P(u, 1 | p(\mathbf{m}_\infty) = p_{1u}) = \frac{\omega_i \lambda_i}{\omega_i \lambda_i + \omega_i (1 - \lambda_i) + (1 - \omega_i)(1 - \lambda_i)} \tag{1}$$

$$P(v, 1 | p(\mathbf{m}_\infty) = p_{1u}) = \frac{\omega_i (1 - \lambda_i)}{\omega_i \lambda_i + \omega_i (1 - \lambda_i) + (1 - \omega_i)(1 - \lambda_i)} \tag{2}$$

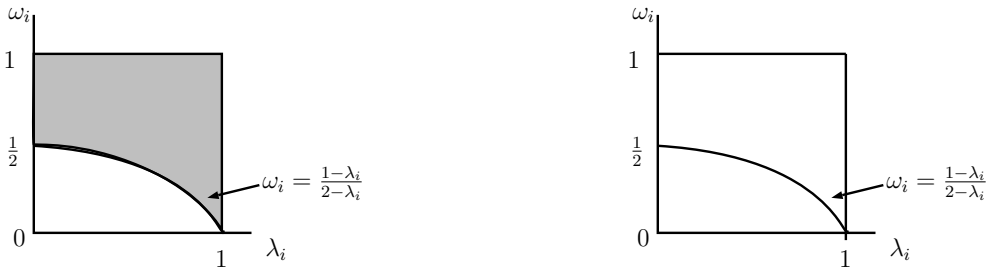
$$P(v, 0 | p(\mathbf{m}_\infty) = p_{1u}) = \frac{(1 - \omega_i)(1 - \lambda_i)}{\omega_i \lambda_i + \omega_i (1 - \lambda_i) + (1 - \omega_i)(1 - \lambda_i)}. \tag{3}$$

Receiver i thus chooses $a_i(\mathbf{m}_\infty \mid p(\mathbf{m}_\infty) = p_{0u}) = 0$. She chooses $a_i(\mathbf{m}_\infty \mid p(\mathbf{m}_\infty) = p_{1u}) = 1$ if $\omega_i > \frac{1-\lambda_i}{2-\lambda_i}$, $a_i(\mathbf{m}_\infty \mid p(\mathbf{m}_\infty) = p_{1u}) = 0$ if $\omega_i < \frac{1-\lambda_i}{2-\lambda_i}$, and randomizes between actions with equal probability if $\omega_i = \frac{1-\lambda_i}{2-\lambda_i}$.[3]

Figure 2 illustrates which receivers take what actions in the long run, based on the messages and their prior beliefs about the state and the sender's type. Receivers above the curve with $\omega_i > \frac{1-\lambda_i}{2-\lambda_i}$ trust the sender. They take $a_i = 1$ when $p(\mathbf{m}_\infty) = p_{1u}$ in panel (a) and $a_i = 0$ when $a_i = 0$ when $p(\mathbf{m}_\infty) = p_{0u}$ in panel (b). Receivers with $\omega_i < \frac{1-\lambda_i}{2-\lambda_i}$ are unsure and always choose $a_i = 0$.

Proposition 2 shows that a mimicking equilibrium exists if and only if type u is good and type v is single-minded. The intuition behind the equilibrium conditions is as follows.

[3]We assume that receivers who believe that $\theta = 1$ and $\theta = 0$ are equally likely tiebreak by randomizing between actions with equal probability.



(a) $p(\mathbf{m}_\infty) = p_{1u} = p_{1v} = p_{0v}$ **(b)** $p(\mathbf{m}_\infty) = p_{0u}$

Figure 2: Receivers' actions in mimicking equilibrium. Panel (a) shows behavior when receivers observe $p(\mathbf{m}_\infty) = p_{1u}$. Panel (b) shows behavior when receivers observe $p(\mathbf{m}_\infty) = p_{0u}$. In each panel: Only receivers whose priors lie in the gray areas choose $a_i = 1$ in response to the observed frequencies.

Consider type u's problem given a mimicking strategy by type v and receivers' beliefs within the mirroring equilibrium. Following her equilibrium strategy leads all receivers to choose $a_i = 0$ from Figure 2(b), and any other strategy would potentially induce some receivers to choose $a_i = 1$ when $\theta = 0$ instead. Likewise, her equilibrium strategy generates $p(\mathbf{m}_\infty) = p_{1u}$ when $\theta = 1$ and induces the most receivers to choose $a_i = 1$ when $\theta = 1$, even though mimicking by sender v means that u cannot induce all receivers to choose $a_i = 1$ when $\theta = 1$. Thus, her strategy is optimal if and only if u is a good type. Analogously, when considering type v's problem given u's strategy, mimicking is optimal if and only if v is a single-minded type who wants receivers to choose $a_i = 1$ in both states.

2.2 Mirroring equilibrium

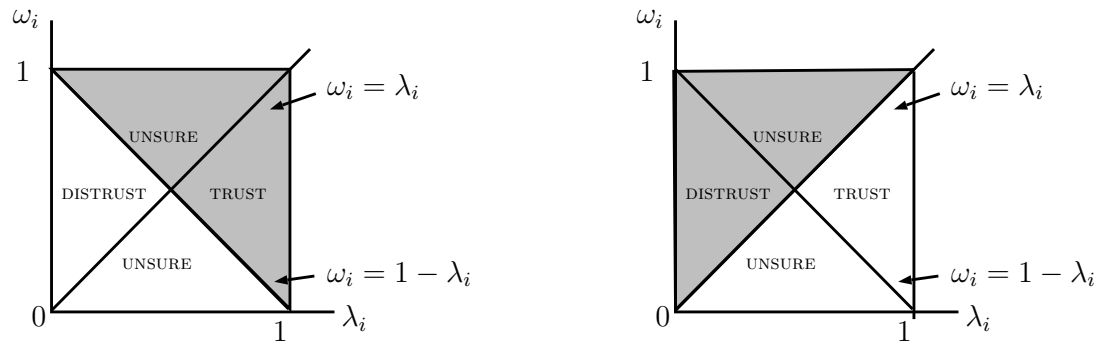
In a mirroring equilibrium, type u uses a messaging strategy that results in different long-run frequencies in each state. Type v "mirrors" type u by using a messaging strategy that generates the same long-run frequencies as type u but in opposite states. The strategy induces a subset of receivers to take the action opposite to their desired action had they known θ for certain. No receiver learns the sender's type and state for certain even in the long run: $P_i(j, \theta|\mathbf{m}_\infty) \neq 1$ for all i and $\mathbf{m}_\infty$.

An intuitive example of messaging strategies that fit this description is an equilibrium in which type u reports truthfully, and type v "mirrors" u by reporting a message of 1 whenever her signal is 0 and reporting a message of 0 whenever her signal is 1. Table 2 summarizes the strategies and long-run frequencies.

	$P(m_n = 1 \mid s_n = 0)$	$P(m_n = 1 \mid s_n = 1)$	p_{0j}	p_{1j}
Sender type u	0	1	$1-\gamma$	γ
Sender type v	1	0	γ	$1-\gamma$

Table 2: Example of mirroring equilibrium.

Figure 3 maps receivers' priors into receivers' actions within a mirroring equilibrium.[4] Receivers whose prior beliefs lie in the UNSURE regions are unsure of the meaning of sender's messages. Intuitively, given their priors, these receivers do not find the messages sufficiently conclusive about sender types to influence their action. Receivers whose prior beliefs lie in the TRUST regions always trust the sender's messages. They take $a_i = 1$ if and only if $p(\mathbf{m}_\infty) = p_{1u}$ since their priors indicate these messages map to state 1, even if their priors also suggest that the state is 0 as in region D. Finally, receivers whose priors lie in the DISTRUST regions distrust the sender's messages and take action $a_i = 0$ if and only if $p(\mathbf{m}_\infty) = p_{1u}$ since their priors indicate the sender is the mirroring type. These regions also underlie the UNSURE, TRUST, and DISTRUST regions in Figure 1.



(a) $p(\mathbf{m}_\infty) = p_{1u} = p_{0v}$ **(b)** $p(\mathbf{m}_\infty) = p_{0u} = p_{1v}$

Figure 3: Receivers' actions in mirroring equilibrium. Panel (a) shows behavior when receivers observe $p(\mathbf{m}_\infty) = p_{1u}$. Panel (b) shows behavior when receivers observe $p(\mathbf{m}_\infty) = p_{0u}$. In each panel: Only receivers whose priors lie in the gray areas choose $a_i = 1$ in response to the observed frequencies.

By implication, receivers whose priors lie in the TRUST region take the correct action ex post if and only if the sender is u because they trust the messages, while receivers whose priors in regions in the DISTRUST regions take the correct action ex post if and only if the sender is v because they distrust the messages. Receivers whose priors lie in the UNSURE regions take the correct action ex post if and only if their priors match the true θ. Thus,

[4]The derivation of receivers' beliefs and actions is analogous to that mimicking equilibrium, so we provide full details in the Appendix.

within mirroring equilibria, the beliefs of trusting and distrusting receivers move in opposite directions in response to common information so they also take opposing actions. One of either the trusting or distrusting group of receivers takes the incorrect action ex post.

Proposition 2 shows that a mirroring equilibrium exists if and only if one sender type is good and the other type is malevolent. Because trusting and distrusting receivers always take opposing actions in response to the same messages, the sender's incentives depend on the relative distributions of these two groups. If there are more receivers in the TRUST than in the DISTRUST region, then u must be a good type and v must be a malevolent type to sustain a mirroring equilibrium. Intuitively, a good type prefers a greater mass of trusting receivers because they will follow her messages and take the correct action on average. But a malevolent type shares this preference so that receivers will follow her messages and take the wrong action on average. Analogously, if there are more receivers in the DISTRUST than in the TRUST region, then receivers tend to do the opposite of the messages, so u must be a malevolent type and v must be a good type to sustain a mirroring equilibrium.

2.3 Real-world examples

Decisions involving medications. A common concern, among the public and within the medical profession, is that some doctors prescribe expensive drugs rather than cheaper alternatives because they have been "bought'' by the pharmaceutical industry (Richmond et al., 2017; Fiore, 2010; Dale, 2017; Groningen, 2017). In the model, consider individuals who decide whether or not to take a drug, which is beneficial for them or not. A doctor sends messages about medical evidence on the drug to patients. A doctor is either a good type who recommends the drug to the patient if and only if the doctor sincerely believes it is in the patient's interests or a single-minded type who has an unconditional interest in the patient taking the drug. For example, the single-minded type may receive a financial benefit from pharmaceutical companies if the patient takes the drug. Individuals' prior beliefs reflect uncertainty over the doctor's type and the drug's efficacy.

Larkin et al. (2017) and Ornstein et al. (2016) document behavior that is consistent with a mimicking equilibrium. Patients' concerns and uncertainty over doctor types are well-founded as doctors who receive gifts or payments from pharmaceutical companies are

significantly more likely than other doctors to persistently prescribe expensive branded medications instead of cheaper generics. These doctors try to camouflage as good types by often claiming to act in the best interest of patients. Finally, patients of such doctors take branded medications at an above-average rate but not exclusively, suggesting that some patients are sufficiently skeptical of the doctor's motives to obtain and take the cheaper, similarly effective generic version of the drug while others are convinced.

Decisions involving election fairness. Among many voters, significant uncertainty swirled around the fairness of the 2020 U.S. presidential election and over the credibility of election officials. Many voters support the idea that Joseph Biden did not fairly win the 2020 election and instead stole the election through "THE BIG LIE" (Trump, 2021).[5] Regarding the credibility of officials, many voters doubted, and still doubt, officials' repeated insistence that Biden won fairly (Reinhard and Sanchez, 2022). For example, many voters alleged that Brad Raffensperger, a Republican election official in Georgia who repeatedly insisted that Biden fairly won in that state, was anti-Trump (Cillizza, 2020).

In the model, consider citizens who decide whether to support the results of an election, which was either fair or unfair. Suppose that citizens prefer to support the results of fair elections. An election official investigates the extent of election fraud and sends messages about their findings to voters. The official is either a good type who wants to truthfully deliver the findings to the public or a single-minded type who has a personal agenda to sway voters toward the winning candidate. Citizens' priors reflect uncertainty over the extent of election fraud and the official's type.[6]

The behaviors of election officials and lingering uncertainty over the 2020 election in the United States are consistent with a mimicking equilibrium. Brad Raffensperger and other election officials failed to convince many skeptics of the election's legitimacy, despite repeated messages, due to concerns that he and others had single-minded anti-Trump leanings and were covering up an unfair election (Cillizza, 2020; Reinhard and Sanchez, 2022). Several years later, skeptics are still concerned about election fraud despite multiple recounts

[5]In contrast, those who believe that Biden fairly won the 2020 election refer to the claim itself that the election was stolen as "The Big Lie" (e.g., Block, 2021; Longwell, 2022).

[6]One can write an analogous setup where the election official has a personal agenda to sway voters toward the losing candidate.

affirming the results (King, 2022; Montellaro, 2022; Fowler, 2022). In a subsequent election, Raffensperger only narrowly defeated a competing candidate who supported claims of election fraud, suggesting continued voter disagreement over the issue (Fowler, 2022).

Decisions involving investments. There is long-running concern that financial firms trade against their clients and recommendations (e.g., Dealbook, 2010). In the model, consider investors who decide whether or not to buy a stock. They receive advice from a financial advisor who is better informed about its fundamental value. The advisor is either a good type who recommends the stock to the investor if and only if the advisor sincerely believes it is in the investor's interests or a malevolent type who wants to mislead investors to trade against them (as in Bénabou and Laroque, 1992). Investors' prior beliefs reflect uncertainty over the asset's value and the advisor's type.

The behavior of some advisors is consistent with mirroring strategies - saying that the stock is doing well when it is not ("pump and dump") and that the stock is doing poorly when it is not ("short and distort"). For example, Enron executives repeatedly reported inflated profits, watched the firm's stock rise, and then finally sold their stocks shortly before Enron's subsequent collapse into bankruptcy (Cramer, 2002). Some investors were skeptical of the firm's reported profits and shorted Enron, suggesting that not all investors were convinced by the company's claims (Bryan-Low and McGee, 2001).

Decisions involving oversight. A large literature in corporate finance studies whether shareholders elect ineffective, captured corporate boards of directors based on management recommendations (Jensen, 1993; Hermalin and Weisbach, 1998). In the model, consider shareholders who decide whether to vote for a candidate for a firm's board of directors. The candidate's type is the unknown state of the world: She is either an independent type who will effectively monitor management or a political lackey of the firm's manager who will not. Shareholders prefer to vote yes if and only if the candidate is the independent type. The firm's manager sends messages to shareholders by making recommendations about whether to vote for the candidate, and a good manager type shares shareholders' preferences. A malevolent manager type has preferences opposite to the shareholders in that she would like shareholders to vote yes to the lackey and no to the independent type. Shareholders are uncertain about the types of the candidate and the manager.

Evidence on management recommendations and boards of directors is consistent with a mirroring equilibrium. First, shareholders' concerns about director types are well-founded: Directors often have social ties to management (Hwang and Kim, 2009), and those appointed after a new CEO takes office are often friendlier to management and enable pet projects (Coles et al., 2014). Second, management typically claims that their recommended directors are in the best interests of shareholders, consistent with a desire to camouflage with good types. Finally, although most shareholders approve management-recommended slates of directors in elections, some shareholders withhold votes for those slates (Cai et al., 2009; Del Guercio et al., 2008), suggesting that some shareholders are skeptical of management recommendations. Moreover, some activist shareholders mount campaigns to install alternative slates of directors, and these slates are routinely opposed by management (Kang et al., 2022), consistent with a mirroring strategy.

3 Limiting Communication

Given that some receivers may take the wrong actions ex post in doublespeak equilibria, a natural question is whether limiting communication ex ante can benefit receivers. We show that a policymaker who is also unsure of the sender's type and state may choose to limit the number of messages the sender can transmit.

There is a policymaker whose objective is to maximize expected receiver welfare, $W = E[-\int_0^1 (a_i - \theta)^2 di]$. The policymaker is uncertain about the sender's type and state, and her prior belief is (λ^P, ω^P).[7] At the beginning of period $\tau = 0$, the policymaker chooses how many subperiods $N \in \{0, 1, \infty\}$ the sender can transmit messages to the receiver. Senders and receivers observe the policymaker's choice of N, and the sender receives private signals and transmits messages during the N subperiods. After subperiod $n = N$, receivers receive no further messages and choose actions in period $\tau = 1$. Let $n^* \in \{0, 1, \infty\}$ be the policymaker's optimal choice of N.

Proposition 3 characterizes the conditions under which each n^* is optimal when one sender type is good and the other is either malevolent or single-minded. Because the policymaker's

[7] This prior belief is public knowledge and is not based on private information.

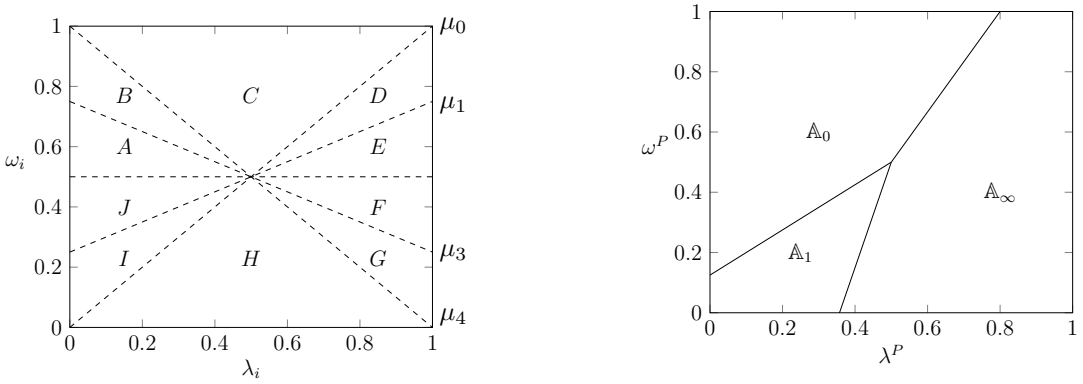
problem is of greatest interest when the $N = 0$, $N = 1$, and $N = \infty$ games each result in different outcomes, we suppose non-babbling equilibria, rather than babbling equilibria, are realized in the $N = 1$ and $N = \infty$ games. Given any distribution of receiver priors $f(\lambda, \omega)$, the proposition characterizes partitions of policymaker priors $\{\mathbb{A}_0, \mathbb{A}_1, \mathbb{A}_\infty \mid \cup_k \mathbb{A}_k = (0,1) \times (0,1)\}$ such that $n^* = k$ is the policymaker's optimal choice whenever $(\lambda^P, \omega^P) \in \mathbb{A}_k$. In general, $\mathbb{A}_\infty$ and $\mathbb{A}_0$ are non-empty. When the policymaker believes the sender is sufficiently likely to be good, she permits $n^* = \infty$ to maximize the number of receivers who will take the correct action. When she believes receivers would, on average, take the correct action based on their priors and would be misled by a non-good sender, she permits no messages.

More interestingly, Proposition 3 shows when $\mathbb{A}_1$ can be non-empty so a policymaker may choose limited communication. To illustrate the general intuition, suppose there is a reasonably large mass of receivers whose priors are that the sender is good and that the state is 1. In this case, the policymaker chooses limited communication when she thinks the state might be zero (so that some messages are necessary to induce receivers to take the correct action) but when she also believes the sender might not be the good type (making unlimited communication undesirable). Thus, $\mathbb{A}_1$ may be non-empty for a subset of low values of (λ^P, ω^P). In other cases of receiver priors, one can develop analogous intuitions for when $\mathbb{A}_1$ is non-empty and for its location. Below, we discuss detailed intuitions when the sender type can be good or malevolent and when the sender type can be good or single-minded.

When sender types are good or malevolent, a single message may be optimal because it can induce receivers, on average, to take the correct action even if it comes from the malevolent type when the sender's private signal was incorrect. Figure 4(a) partitions receivers' priors according to how their actions differ with message content in mirroring equilibria in the $N = \{0, 1, \infty\}$ games. To develop intuition for why $\mathbb{A}_1$ can be non-empty, suppose receivers are concentrated in a certain region, such as E. Receivers in E always take actions consistent with sender messages for both $N = 1$ and $N = \infty$ due to prior beliefs that the sender is good; their priors are that $\theta = 1$ is more likely.[8]

[8]More completely: Receivers in region C and H always take actions consistent with their priors irrespective of sender messages for all N. Those in regions D and G take actions consistent with their priors for $N = 1$ and with sender messages for $N = \infty$. Those in regions E and F take actions consistent with sender messages for both $N = 1$ and $N = \infty$. Those in regions B and I take actions consistent with their priors for $N = 1$ but opposite of sender messages for $N = \infty$. Finally, those in regions A and J take actions opposite

When receivers are concentrated in E, Figure 4(b) shows that $n^* = 1$ is uniquely optimal when λ^P and ω^P are sufficiently low in the lower-left area indicated by $\mathbb{A}_1$. If the policymaker permitted no messages, receivers in E would choose $a_i = 1$ given their priors, which conflicts with her view that θ is likely zero and $a_i = 0$ is the correct choice. If she permitted $N = \infty$, she believes these receivers would also likely choose $a_i = 1$ since 1) she thinks the messages will likely come from a malevolent type who will generate $p(\mathbf{m}_\infty) = p_{0v} = p_{1u}$ and 2) she knows receivers in E will follow these messages. Thus, $n^* = 1$ since there is a chance that $m_1 = 0$ when the sender's private signal is $s_1 = 1$, which would lead receivers to take $a_i = 0$ even though policymaker expects that a malevolent type plays a mirroring strategy.



(a) Partition of receivers' priors. Lines: $\mu_0 = \lambda_i, \mu_1 = \gamma\lambda_i + (1-\gamma)(1-\lambda_i)$, $\mu_3 = (1-\gamma)\lambda_i + \gamma(1-\lambda_i), \mu_4 = 1-\lambda_i$. Parameters: $\gamma = 0.75$.

(b) Optimal $n^* \in \{0, 1, \infty\}$ given $f(\lambda, \omega)$ such that $E = 0.3$, $F = G = 0.1$, $D = 0.2$, $A = B = J = I = C = H = 0.05$, where the values correspond to the mass of receivers in each region.

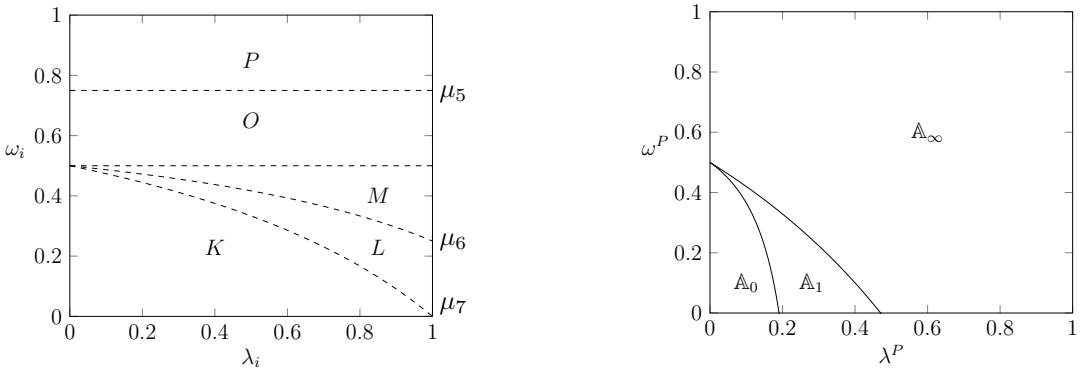
Figure 4: Mirroring Example.

When sender types are good or single-minded, a single message may be optimal because it can induce receivers to take the correct action when the message comes from a good type and her private signal is correct. Figure 5(a) partitions receivers' priors according to how their actions differ with message content in mimicking equilibria in the $N = \{0, 1, \infty\}$ games. To develop intuition, suppose receivers are concentrated in O. These receivers always take action consistent with sender messages in both $N = 1$ and $N = \infty$ games, and have priors that $\theta = 1$ is more likely.[9]

of sender messages for both $N = 1$ and $N = \infty$. Receivers in regions A, B, C, D, E believe $\theta = 1$ is more likely ex-ante, while those in regions F, G, H, I, J believe $\theta = 0$ is more likely ex-ante.

[9]More completely: Receivers in region K always take actions consistent with their priors in both $N = 1$ and $N = \infty$. Those in regions P and L always take actions consistent with their priors for $N = 1$ and with sender messages for $N = \infty$. Those in regions O and M always take actions consistent with sender messages

When receivers are concentrated in O, Figure 5(b) shows that $n^* = 1$ is uniquely optimal when λ^P is intermediate and ω^P is sufficiently low in the lower-left area indicated by $\mathbb{A}_1$. If the policymaker permitted no messages, receivers in O would choose $a_i = 1$ given their priors, which conflicts with her view that $a_i = 0$ is the correct choice. If she permitted $N = \infty$, she believes these receivers would also likely choose $a_i = 1$ since 1) she thinks the messages will come from a single-minded type who will generate $p(\mathbf{m}_\infty) = p_{1v} = p_{1u}$ and 2) knows that receivers will follow these messages. Thus $n^* = 1$ because she believes there is some chance the sender is a good type who would transmit $m_1 = 0$ due to $s_1 = 0$ and induce receivers to choose $a_i = 0$.



(a) Partition of receivers' priors. Lines: $\mu_5 = \gamma$, $\mu_6 = \frac{1-\gamma\lambda_i}{2-\lambda_i}$, $\mu_7 = \frac{1-\lambda_i}{2-\lambda_i}$. Parameters: $\gamma = 0.75$.

(b) Optimal $n^* \in \{0, 1, \infty\}$ given $f(\lambda, \omega)$ such that $O = 0.6$, $M = 0.1$, $L = 0.2$, $K = P = 0.05$, where the values correspond to the mass of receivers in each region.

Figure 5: Mimicking Example.

Proposition 3 (Optimal $n^* = 1$)**.** *Let the distribution of receiver beliefs $f(\lambda, \omega)$ be given. Suppose non-babbling equilibria are realized in the $N = 1$ and $N = \infty$ games. There exist partitions $\{\mathbb{A}_0, \mathbb{A}_1, \mathbb{A}_\infty \mid \cup_k \mathbb{A}_k = (0,1) \times (0,1)\}$, whose boundaries are characterized in the Appendix, such that $n^* = k$ is the policymaker's optimal choice whenever $(\lambda^P, \omega^P) \in \mathbb{A}_k$. The optimum is unique if and only if (λ^P, ω^P) lies in the interior of $\mathbb{A}_k$.*

$\mathbb{A}_1$ is a set with non-empty interior if and only if:

1. *Mimicking: $(O + M)(\gamma L - (1-\gamma)M) - M(P + L) > 0$.*
2. *Mirroring: Neither of the following two conditions are satisfied:*
   (a) $E - J = F - A = 0$,

for $N = 1$ and $N = \infty$. Receivers in regions O, P believe $\theta = 1$ is more likely ex-ante, while those in regions M, L, K believe $\theta = 0$ is more likely ex-ante.

(b) $(E-J)\left((1-\gamma)(D+E-J-I)+\gamma G\right)=(F-A)\left((1-\gamma)(F-A+G-B)+\gamma D\right)$,

where, with a slight abuse of notation, the letters denote the mass of receivers in each region.

Proposition 3 does not depend on assumptions that receivers are concentrated in region E or O, which were made only for exposition. Generally, changes in $f(\lambda,\omega)$ change the boundaries for partitions $\mathbb{A}_0, \mathbb{A}_1$, and $\mathbb{A}_\infty$, and Figures 4(b) and 5(b) would change. In the mirroring case, $\mathbb{A}_1$ is a set with non-empty interior unless receivers are distributed such that the net effect of a single message on receivers' actions is zero (condition 2(a)), or if receivers are distributed such that $n^*=1$ is optimal only when the policymaker is indifferent among $n^*\in\{0,1,\infty\}$ (condition 2(b)). In the mimicking case, $\mathbb{A}_1$ is a set with non-empty interior when the mass of receivers who would have chosen $a_i=1$ without information and would respond to a single message (receivers in O) is large relative to two groups: those who would have chosen $a_i=0$ without information and could be misled by a single message (M), and those who do not respond to a single message (P and L) (condition 1).

In summary, Proposition 3 provides necessary and sufficient conditions for when policymakers adopt interior solutions for how much communication to permit. The key takeaway is that the policymaker's subjective beliefs and how they compare with receivers' matter a great deal for her optimal n^*. A choice of $n^*=1$ relies on the policymaker believing that a noisy message will benefit receivers (i.e., that she believes enough receivers' priors are incorrect about the state) despite her concern that the sender type is not good.

4 Endogenous fact-checking

Does fact-checking by receivers mitigate the sender's incentive to doublespeak? Suppose that receivers have the option to fact-check sender's messages at a fixed cost. Just after observing $n=\infty$ messages at the end of $\tau=0$ but before choosing their actions, all receivers have the option to incur a fixed cost $\phi\geq 0$ to learn the true state θ. We characterize which receivers fact-check, how fact-checking changes sender strategies, and whether the option to fact-check improves receiver welfare.

Lemma 1 describes who fact-checks in any fully informative, mimicking, and mirroring equilibria that exist. It shows that, in any equilibrium, a receiver only fact-checks if she is

sufficiently uncertain about the state after observing the sender's messages.

Lemma 1 (Who fact-checks)**.** *Let $\mu_i = P_i(\theta = 1|\mathbf{m}_\infty)$ be the receiver's posterior belief that $\theta = 1$ after observing messages $\mathbf{m}_\infty$.*

In any equilibrium, a receiver with $\mu_i \geq 1/2$ fact-checks when $\mu_i \leq 1 - \phi$, and a receiver with $\mu_i < 1/2$ fact-checks when $\mu_i \geq \phi$. Thus, in fully informative, mimicking, and mirroring equilibria, the following receivers fact-check on the equilibrium path:

1. *Fully informative: No receiver fact-checks.*

2. *Mimicking: When $p(\mathbf{m}_\infty) = p_{0u}$, no receiver fact-checks. When $p(\mathbf{m}_\infty) = p_{1u}$, receiver i fact-checks if $\omega_i \in [L^\kappa(\lambda_i, \phi), H^\kappa(\lambda_i, \phi)]$, where $L^\kappa(\lambda_i, \phi) \equiv \frac{\phi(1-\lambda_i)}{1-\lambda_i\phi}$ and $H^\kappa(\lambda_i, \phi) \equiv \frac{(1-\phi)(1-\lambda_i)}{(1-\phi)(1-\lambda_i)+\phi}$.*

3. *Mirroring: When $p(\mathbf{m}_\infty) = p_{1u}$, receiver i fact-checks if $\omega_i \in [L_1^\rho(\lambda_i, \phi), H_1^\rho(\lambda_i, \phi)]$, where $L_1^\rho(\lambda_i, \phi) \equiv \frac{\phi(1-\lambda_i)}{\phi(1-\lambda_i)+\lambda_i(1-\phi)}$, $H_1^\rho(\lambda_i, \phi) \equiv \frac{(1-\phi)(1-\lambda_i)}{(1-\phi)(1-\lambda_i)+\phi\lambda_i}$. When $p(\mathbf{m}_\infty) = p_{0u}$, receiver i fact-checks if $\omega_i \in [L_0^\rho(\lambda_i, \phi), H_0^\rho(\lambda_i, \phi)]$ where $L_0^\rho(\lambda_i, \phi) \equiv \frac{\phi\lambda_i}{\phi\lambda_i+(1-\lambda_i)(1-\phi)}$, and $H_0^\rho(\lambda_i, \phi) \equiv \frac{(1-\phi)\lambda_i}{(1-\phi)\lambda_i+\phi(1-\lambda_i)}$.*

Proposition 4 shows that fact-checking largely does not affect the conditions necessary to sustain fully informative and doublespeak equilibria from the base game without fact-checking. The only equilibrium condition that changes is the requirement on the distribution of receivers' priors for a mirroring equilibrium. Since no receivers fact-check in any equilibrium when it is too costly ($\phi > 1/2$), the proposition considers the case where $0 < \phi \leq 1/2$.

Proposition 4 (Equilibrium existence with endogenous fact-checking)**.** *Let $0 < \phi \leq 1/2$. Comparing the conditions for fully informative and doublespeak equilibria in the fact-checking game to the conditions in the base game:*

1. *The conditions on the sender types to support each non-babbling equilibrium (fully informative, mimicking, mirroring) are identical with or without the option to fact-check.*

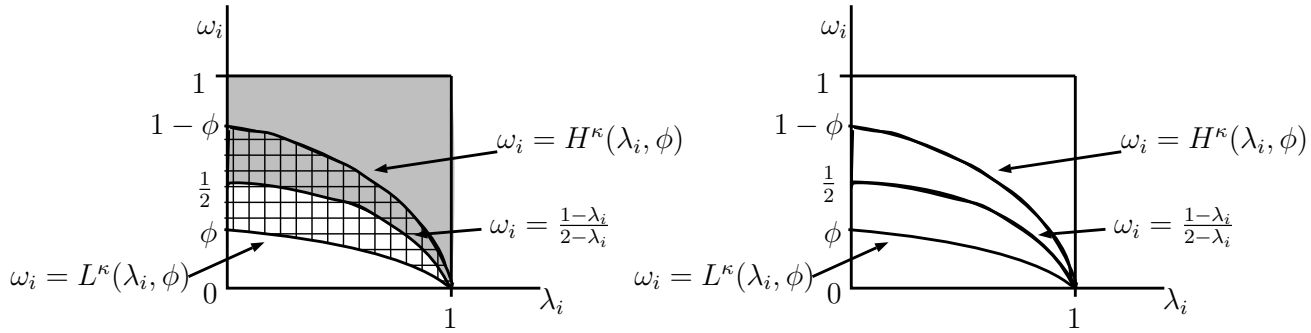
2. *The conditions on the distribution of receivers' priors to support mirroring equilibrium in the game with fact-checking are:*

$$\int_{1/2}^{1}\int_{L_1^\rho(\lambda_i,\phi)}^{L_0^\rho(\lambda_i,\phi)} f(\lambda,\omega)d\omega d\lambda \geq \int_{0}^{1/2}\int_{L_0^\rho(\lambda_i,\phi)}^{L_1^\rho(\lambda_i,\phi)} f(\lambda,\omega)d\omega d\lambda \tag{4}$$

$$\int_{1/2}^{1}\int_{H_1^\rho(\lambda_i,\phi)}^{H_0^\rho(\lambda_i,\phi)} f(\lambda,\omega)d\omega d\lambda \geq \int_{0}^{1/2}\int_{H_0^\rho(\lambda_i,\phi)}^{H_1^\rho(\lambda_i,\phi)} f(\lambda,\omega)d\omega d\lambda, \tag{5}$$

or these conditions with both inequalities reversed.

Part 1 shows that non-babbling equilibria can exist with the same sender types as in Proposition 2 when receivers have the option to fact-check. For mimicking equilibria, the conditions for existence are unchanged from Proposition 2. Some receivers fact-check in a mimicking equilibrium if $p(\mathbf{m}_\infty) \neq p_{0u}$, as Figure 6(a) depicts and Lemma 1 describes. Specifically, receivers fact-check $p(\mathbf{m}_\infty) = p_{1u}$ when they are ex ante: (1) quite uncertain about the state (ω near 1/2) and thought the sender type is probably not good (λ close to 0), (2) quite certain that $\theta = 0$ (ω near 0) even though they thought the sender type is probably good (λ near 1), or (3) in between these two cases along the $\omega_i = \frac{1-\lambda_i}{2-\lambda_i}$ curve. However, equilibrium fact-checking fails to induce the single-minded type to switch from a mimicking strategy to a strategy that maps to the true state. Intuitively, fact-checking reduces the absolute benefit of mimicking, but does not eliminate its relative benefit over a strategy that maps to the true state.[10]

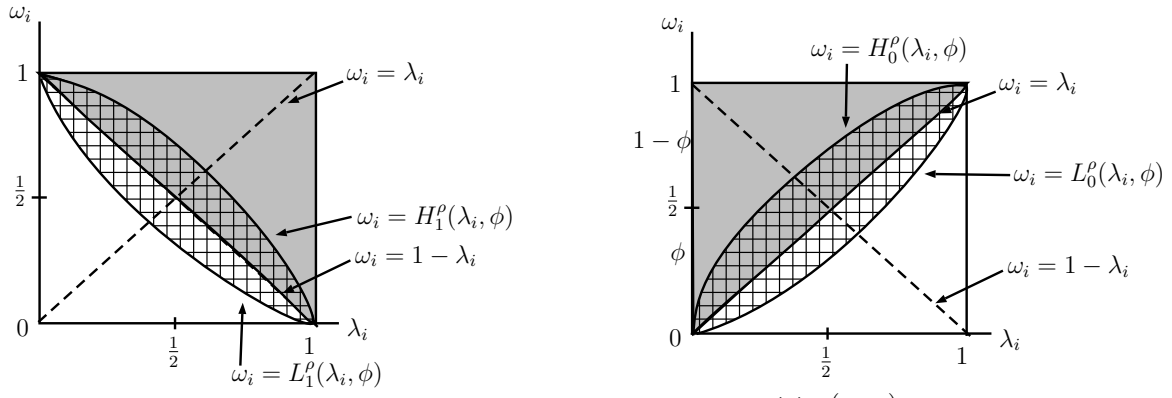


(a) $p(\mathbf{m}_\infty) = p_{1u} = p_{1v} = p_{0v}$ **(b)** $p(\mathbf{m}_\infty) = p_{0u}$

Figure 6: Fact-checking in mimicking equilibrium. Panel (a) shows behavior when receivers observe $p(\mathbf{m}_\infty) = p_{1u}$. Panel (b) shows behavior when receivers observe $p(\mathbf{m}_\infty) = p_{0u}$. In each panel: Only receivers whose priors lie in the hatched areas fact-check. Receivers whose priors lie in the gray areas would have chosen $a_i = 1$ in the base game.

For a mirroring equilibrium, Part 1 says that sender types need to be good or malevolent. Intuitively, receivers fact-check in equilibrium when (1) the messages are consistent with their priors on the state but they thought the sender was probably malevolent, or (2) the messages are inconsistent with their priors on the state but they thought the sender was probably good.

[10]Moreover, the good type does not deviate to an off-equilibrium strategy. Intuitively, an off-equilibrium strategy may be appealing to the good type if she can trigger enough fact-checking by receivers. However, the amount of fact-checking that the sender can trigger is limited since fact-checking is costly for receivers. The Appendix contains further details.



(a) $p(\mathbf{m}_\infty) = p_{1u} = p_{0v}$ **(b)** $p(\mathbf{m}_\infty) = p_{0u} = p_{1v}$

Figure 7: Fact-checking in mirroring equilibrium. Panel (a) shows behavior when receivers observe $p(\mathbf{m}_\infty) = p_{1u}$. Panel (b) shows behavior when receivers observe $p(\mathbf{m}_\infty) = p_{0u}$. In each panel: Only receivers whose priors lie in the hatched areas fact-check. Receivers whose priors lie in the gray areas would have chosen $a_i = 1$ in the base game.

Figure 7 shows receivers' fact-checking behavior given equilibrium frequencies in a mirroring equilibrium. In each panel, receivers in the hatched areas fact-check in response to the observed frequencies.

Part 2 describes an additional condition (relative to Proposition 2) on the distribution of receivers' priors needed to sustain mirroring equilibria. The additional condition appears in Proposition 4 because potential deviations must account for differential fact-checking behavior by receivers on the equilibrium paths. The Online Appendix discusses in detail.

The overall effect of introducing a costly option to fact-check on receivers' welfare is ambiguous and depends on how many receivers benefit from fact checking relative to those who needlessly do it. We provide a full characterization in the Online Appendix and discuss the intuition here. Intuitively, welfare depends on the relative mass of receivers whose prior beliefs lead them to needlessly fact-check (in that they would have taken the correct action anyway) versus those who benefit. For example, in the mimicking equilibrium, receivers whose priors lie in the hatched white area of Figure 6(a) fact-check when they observe $p(\mathbf{m}_\infty) = p_{1u}$. Since they would have chosen $a_i = 0$ in the base game, fact-checking enables them to choose the correct action and benefits them if $\theta = 1$, but is needlessly costly if the sender type is v and $\theta = 0$. Receivers whose priors lie in the hatched gray area also fact-check when they observe $p(\mathbf{m}_\infty) = p_{1u}$. Since they would have chosen $a_i = 1$ in the base game, fact-checking is needlessly costly if $\theta = 1$, but benefits them if the sender is v and $\theta = 0$.

5 Reputation

Can reputational concerns mitigate doublespeak? Suppose the sender cares about her reputation, which is the average receiver posterior belief that the sender is a good type after the receivers have received messages, chosen actions, and payoffs have been realized (and thus the state θ is revealed). Let $r \geq 0$ be the sender's preference weight on reputation, and let $\mathbb{G} \equiv \{(b,c) : b \leq 1/2, b+c \geq 1/2\}$ denote the set of possible sender preferences that are good. Sender type j's preferences are $-\int_0^1 [a_i - (c_j\theta + b_j)]^2 \, di + r \int_0^1 P_i((b_j, c_j) \in \mathbb{G} \mid \mathbf{m}_\infty, \theta) \, di$. To isolate the role of reputation, we do not allow fact-checking here.

Generally, reputation concerns strengthen a non-good sender type's incentive to pool with a good type because the revelation of the state allows receivers to learn whether the sender's messages are consistent with a good type. For example, in a mimicking equilibrium, if $p(\mathbf{m}_\infty) = p_{1u}$ and $\theta = 0$, then ex post receivers are sure that the sender is not good. But if $p(\mathbf{m}_\infty) = p_{1u}$ and $\theta = 1$, then ex post receivers still cannot identify the sender and their posteriors that the sender is good are λ_i.

Proposition 5 provides the existence conditions for fully informative and doublespeak equilibria. It characterizes how reputation can either increase or decrease the amount of information revelation and learning, depending on the possible sender types and the degree of reputation concern.

Proposition 5 (Equilibrium conditions with reputation)**.** *Reputation expands the set of sender types to support fully informative equilibria, shrinks the set of sender types to support mirroring equilibria, and may either expand or shrink the set of sender types to support mimicking equilibria.*

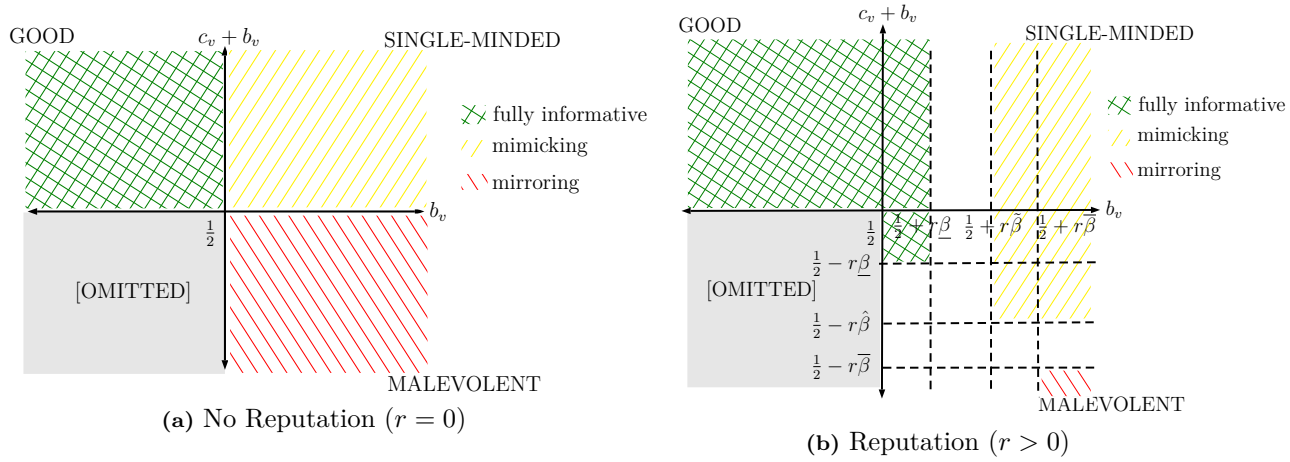
Define $\underline{\beta} = \frac{1}{2}\int_0^1 \int_0^1 \lambda f(\lambda,\omega) d\lambda d\omega$, $\tilde{\beta} = \frac{1}{2\int_0^1 \int_{\frac{1-\lambda}{2-\lambda}}^1 f(\lambda,\omega) d\lambda d\omega}$, $\hat{\beta} = \frac{\int_0^1 \int_0^1 \lambda f(\lambda,\omega) d\lambda d\omega}{2\int_0^1 \int_{\frac{1-\lambda}{2-\lambda}}^1 f(\lambda,\omega) d\lambda d\omega}$, *and* $\overline{\beta} = \frac{1}{2(\int_{\frac{1}{2}}^1 \int_{1-\lambda}^{\lambda} f(\lambda,\omega) d\omega d\lambda - \int_0^{\frac{1}{2}} \int_{\lambda}^{1-\lambda} f(\lambda,\omega) d\omega d\lambda)}$.

1. *A fully informative equilibrium exists if and only if sender type u is good and sender type v satisfies $b_v \leq \frac{1}{2} + r\underline{\beta}$ and $c_v + b_v \geq \frac{1}{2} - r\underline{\beta}$.*

2. *A mimicking equilibrium exists if and only if sender type u is good and sender type v satisfies $b_v \geq \frac{1}{2} + r\tilde{\beta}$ and $c_v + b_v \geq \frac{1}{2} - r\hat{\beta}$.*

3. *A mirroring equilibrium exists if and only if sender type u is good and sender type v satisfies $b_v \geq \frac{1}{2} + r\overline{\beta}$ and $c_v + b_v \leq \frac{1}{2} - r\overline{\beta}$.*

Figure 8 graphically depicts the requirements for sender v's type and v's behavior in non-babbling equilibria, given that sender u is a good type. Panel (a) shows which non-babbling equilibria can be sustained for each sender type v when there are no reputation concerns ($r = 0$). For clarity, we omit a single-minded type who always prefers $a_i = 0$ ($b_v < 1/2$ and $c_v + b_v < 1/2$). As Proposition 2 shows, a fully informative equilibrium exists when v is good, a mimicking equilibrium exists when v is single-minded, and a mirroring equilibrium exists when v is malevolent. Panel (b) shows which non-babbling equilibria can be sustained for each sender type v when there are reputation concerns ($r > 0$) from Proposition 5. Several observations follow from comparing the panels in the figure and thus the results of the propositions.[11]



(a) No Reputation ($r = 0$)

(b) Reputation ($r > 0$)

Figure 8: Reputation effects. This figure shows sender v's type and behavior in non-babbling equilibria. For clarity, we omit the single-minded type who prefers $a_i = 0$ in either state ($b_v < 1/2$ and $c_v + b_v > 1/2$). Panel (a) shows the base case of no reputation concerns ($r = 0$). Panel (b) shows the case of reputation concerns ($r > 0$), when $\hat{\beta} < \overline{\beta}$ and $\tilde{\beta} < \overline{\beta}$.

First, a comparison of Panels (a) and (b) of Figure 8 shows that reputation expands the set of sender types supporting fully informative equilibria, and shrinks the set of sender types supporting mirroring equilibria. Single-minded and malevolent types whose intrinsic preferences are sufficiently weak relative to reputation concern r will increase their reputations by pooling with the good type when they otherwise would have mimicked or mirrored, respectively. For single-minded types, this occurs when their preference for $a_i = 1$ when $\theta = 0$ is

[11] The key effects of reputation described in this section hold for all distributions of receiver priors for which the equilibria exist. Panel (b) of Figure 8 shows a case in which mimicking and mirroring equilibria are mutually exclusive. For some distributions of receiver priors, it is possible for mimicking and mirroring equilibria to both exist for a given set of malevolent types.

sufficiently weak, $b_v \leq 1/2 + r\underline{\beta}$. For malevolent types, this occurs when $b_v \leq 1/2 + r\underline{\beta}$ and their preference for $a_i = 0$ when $\theta = 1$ is also sufficiently weak, $c_v + b_v \geq 1/2 - r\underline{\beta}$.

Second, reputation's effect on the set of sender types supporting mimicking equilibria is ambiguous. On the one hand, the set shrinks because single-minded types with sufficiently strong reputation concerns relative to intrinsic preferences ($b_v < 1/2 + r\tilde{\beta}$) no longer mimic. On the other hand, the set expands because malevolent types whose intrinsic preferences are intermediate relative to reputation concerns ($b_v \geq 1/2 + r\tilde{\beta}$ and $c_v + b_v \geq 1/2 - r\hat{\beta}$) mimic rather than mirror due to the reputational gains from partially pooling with the good type. The total effect of reputation depends on the relative size of these two countervailing effects.

Third, in the presence of reputational concerns, there exists an interim region of intrinsic preferences such that fully informative and doublespeak equilibria no longer exist (i.e., the white areas in Figure 8(b)). Intuitively, a sender type v with such preferences is deterred from double-speaking because reputation costs outweigh the intrinsic benefits from leading receivers to take incorrect actions. However, pooling with the good type does not provide enough of a reputation benefit to induce sender v to do so. Thus, when sender type v's preferences are in this region, the only remaining equilibria are babbling.

Finally, we note that receiver welfare may or may not be greater when reputational concerns exist ($r > 0$) compared to the when they do not exist ($r = 0$). Reputation concerns certainly increase receiver welfare if they induce type v to pool with the good type when she would not have done so otherwise. But if receiver welfare is greater under doublespeak than babbling, reputation concerns decrease welfare when they induce type v to babble instead of doublespeak. These and other examples make the effect of reputation concerns on receiver welfare ambiguous and dependent on the distribution of receiver priors and the true ex-ante probabilities of the sender's type and the state.

6 Discussion

6.1 Robustness

Are heterogeneous priors necessary to support doublespeak equilibria? No. Each form of doublespeak equilibrium is sustainable given any common prior. However, heterogeneous priors are necessary to generate long-run disagreement. In a mimicking equilibrium, heterogeneous priors over either the sender's type or the state are necessary to generate disagreement when receivers observe $p(\mathbf{m}_\infty) = p_{1u}$. In a mirroring equilibrium, heterogeneous priors over the sender's type alone generates qualitatively different behavior than heterogeneous priors over the state alone. Figure 3 illustrates. If receivers' priors only differ about the state, receivers only take different actions if some of them are unsure. If receivers' priors only differ about the sender's type, they take different actions because there are both trusting and distrusting receivers who interpret messages in opposing ways.

Is the discrete receiver action space necessary? No. If receivers' action space is continuous ($a_i \in \mathbb{R}$) rather than binary, then the types of sender u whose preferences can be interpreted as "good" becomes $b_u \leq 0$ and $c_u + b_u \geq 1$, rather than $b_u \leq 1/2$ and $c_u + b_u \geq 1/2$ (and analogously for single-minded and malevolent types). Doublespeak equilibria still exist, and the qualitative results of Proposition 2 holds.[12]

What if there are multiple senders, not just multiple sender types? Doublespeak equilibria can still occur if receivers observe the signals of multiple senders before they choose actions. Suppose receivers observe messages from two senders of unknown type, drawn independently by nature. Each sender observes private signals with accuracy $\gamma \in (1/2, 1)$ and reports messages to the receiver in the $n = \infty$ subperiods of period $\tau = 0$, then receivers choose actions in period $\tau = 1$. We sketch the intuition of the equilibria here and save the detailed results for the Online Appendix.

[12] If $a_i \in \mathbb{R}$, a receiver's optimal action is to choose an action that equals her posterior belief: $a_i = P_i(\theta = 1|p(\mathbf{m}_\infty))$. The mimicking equilibrium exists for any distribution of receiver priors when sender u prefers that the action match the state ($b_u \leq 0$ and $c_u + b_u \geq 1$) and sender v single-mindedly prefers action $a_i = 1$ ($b_u \geq 1$ and $c_u + b_u \geq 1$). The mirroring equilibrium exists when sender u prefers that the action match the state ($b_u \leq 0$ and $c_u + b_u \geq 1$) and sender v prefers that the action mismatch the state ($b_u \geq 1$ and $c_u + b_u \leq 0$). For other combinations of sender types, doublespeak equilibria may also exist depending on the distribution of receivers.

A mimicking equilibrium still exists in which a single-minded sender type mimics a good type. Receivers learn that $\theta = 0$ whenever they observe a frequency of p_{0u} from at least one sender. But when they observe a frequency of p_{1u} from both senders, they cannot be sure of the true state. Neither sender type has an incentive to deviate from these equilibrium strategies, for the same reasons as in the single-sender game.

Likewise, a mirroring equilibrium still exists in which a malevolent sender type mirrors a good type, if there are sufficiently many receivers who trust rather than distrust each sender. If both senders generate the same long-run frequencies, receivers are sure the senders are the same type but cannot identify which they are. If the senders generate different long-run frequencies, receivers can only be sure that the senders are different types. Regardless, receivers cannot identify the state.

What if sender knows the state? If sender knows the state perfectly, all of her knowledge can be communicated with one message in equilibrium. This has two implications. First, there is little distinction between a game with one message and many messages when sender knows the state. Adding sender uncertainty about the state and provides additional richness that allows our model to address questions such as the optimality of limited communication. Second, equilibria in a game where a perfectly informed sender sends one message are equivalent to equilibria in our game where an imperfectly informed sender sends infinite messages. Thus, our results imply that receiving infinite signals in itself does not guarantee that receivers always learn the state when signals are endogenously rather exogenously generated.

What if receivers can fact-check with a second sender? The results of Proposition 4 still hold when receivers can fact-check with a second source whose incentives are also unknown. Consider an alternative model in which receivers have the option to fact-check with such a second source. Intuitively, the main difference between this alternative model and ours is that receivers who are most uncertain about the state after the first sender speaks may be unlikely to fact-check if they were unsure how much to trust either of the two senders ex ante. But the possibility of being fact-checked against a second source still does not mitigate a non-good sender type's incentive to doublespeak, because fact-checking is still endogenously limited on the equilibrium path. Our main model assumes that fact-checking

reveals the state with certainty because this provides the strongest incentive for receivers to fact-check the sender and potentially mitigate the sender's incentive to doublespeak.

6.2 Empirical Implications

Suppose that an empiricist had cross-sectional data on which receivers took what action, $action_i$, after a sender delivered messages that may contain information about the correct course of action. A regression model of the form $action_i = a + b_0\omega_i + b_1\mathbb{1}_{[\lambda_i \text{ near } 0]} + b_2(\omega_i \times \mathbb{1}_{[\lambda_i \text{ near } 0]}) + e_i$, where a is a constant, b_0, b_1, b_2 are slopes, and e_i is the unexplained error term, is informative about what type of equilibrium receivers are in.

Specifically, $b_0 \neq 0$ but $b_1, b_2 = 0$ suggests a babbling equilibrium since heterogeneity in receiver actions depends only on prior beliefs about the state and not on prior beliefs about sender type. If $b_0, b_1, b_2 \neq 0$, receiver actions depend on both beliefs about the state and sender type, suggesting a doublespeak equilibrium. Further, if b_2 has the same sign as b_0, this suggests a mimicking equilibrium because receivers who view the sender as not good ex ante take actions that are more strongly dependent on their prior beliefs about the state than other receivers (as in Figure 2(a)). If instead $b_2 \approx -b_0$, this suggests a mirroring equilibrium because receivers who view the sender as not good ex ante take actions that depend very little on their prior beliefs about the state (as in each panel of Figure 3).

The model also makes predictions about who fact-checks in equilibrium, following the intuitions conveyed in Figures 6 and 7. Suppose that, upon observing some heterogeneity in fact-checking, an empiricist estimates a statistical model of the form $factcheck_i = a + b_0\mathbb{1}_{[\lambda_i \text{ near } 1/2]} + b_1\mathbb{1}_{[\lambda_i \text{ near } 0]} + e_i$, where a is a constant, b_0, b_1 are slopes, and e_i is the unexplained error term. In this specification, receivers with λ_i near 1 are the omitted category.

Evidence of $b_0, b_1 = 0$ suggests a babbling equilibrium because prior beliefs about sender type do not predict who fact-checks. Evidence of $b_1 > b_0 > 0$ suggests a mimicking equilibrium because fact-checking is monotone in prior beliefs about the sender type: Receivers who ex ante are fairly certain the sender is not the good type fact-check more than those who are are uncertain of the sender's type, who fact-check more than those who are certain the sender is the good type. Evidence of $b_0 > 0$ and $b_1 = 0$ suggests a mirroring equilibrium because fact-checking is non-monotone in prior beliefs about the sender type, with receivers

who are ex ante fairly uncertain of the sender's type fact-checking the most.

7 Conclusion

Our work casts doubt on the presumption that rational agents can pierce through misinformation in the long run. Even given an infinite history of public messages, Bayesian receivers may fail to learn the state in equilibrium and persistently disagree due to suspicions about the sender's motives, even if the true sender type is good, can be fact-checked, and partially cares about reputation. A policymaker who believes that doublespeak would mislead receivers may restrict communication from the sender. Doublespeak is powerful because it confounds learning for rational receivers who are able to pierce through less extreme forms of misinformation. Further research into doublespeak is an area of fruitful research given the growing importance of misinformation.

References

Acemoglu, Daron, Victor Chernozhukov, and Muhamet Yildiz, "Fragility of Asymptotic Agreement under Bayesian Learning," *Theoretical Economics*, 2016, *11*, 187–227.

Allcott, Hunt and Matthew Gentzkow, "Social Media and Fake News in the 2016 Election," *Journal of Economic Perspectives*, Spring 2017, *31* (2), 211–236.

Baliga, Sandeep, Eran Hanany, and Peter Klibanoff, "Polarization and Ambiguity," *American Economic Review*, 2013, *103* (7), 3071–3083.

Bénabou, Roland and Guy Laroque, "Using Privileged Information to Manipulate Markets: Insiders, Gurus, and Credibility," *The Quarterly Journal of Economics*, 08 1992, *107* (3), 921–958.

Blackwell, David and Lester Dubins, "Merging of Opinions in Increasing Information," *The Annals of Mathematical Statistics*, 1962, *33* (3), 882–886.

Block, Melissa, "Can The Forces Unleashed By Trump's Big Election Lie Be Undone?," *National Public Radio News*, January 2021. Available online: https://www.npr.org/2021/01/16/957291939/can-the-forces-unleashed-by-trumps-big-election-lie-be-undone, Last accessed June 2022.

Bohren, J. Aislinn and Daniel N. Hauser, "Learning with Heterogeneous Misspecified Models: Characterization and Robustness," *Econometrica*, November 2021, *89* (6), 3025–3077.

Bowen, Renee, Danil Dmitriev, and Simone Galperti, "Learning from Shared News: When Abundant Information Leads to Belief Polarization," *Quarterly Journal of Economics*, May 2023, *132* (2), 955–1000.

Bryan-Low, Cassell and Suzanne McGee, "Enron Short Seller Detected Red Flags in Regulatory Filings," *The Wall Street Journal*, November 2001.

Cai, Jie, Jacqueline L. Garner, and Ralph A. Walkling, "Electing Directors," *The Journal of Finance*, 2009, *64* (5), 2389–2421.

Cheng, Ing-Haw and Alice Hsiaw, "Distrust in Experts and the Origins of Disagreement," *Journal of Economic Theory*, March 2022, *200*, 105401.

Cillizza, Chris, "How this Republican official became the most hated man in his party," *CNN*, November 2020.

Cisternas, Gonzalo and Jorge Vásquez, "Misinformation in Social Media: The Role of Verification Incentives," February 2023. Working paper.

Coles, Jeffrey L., Naveen D. Daniel, and Lalitha Naveen, "Co-opted Boards," *The Review of Financial Studies*, 04 2014, *27* (6), 1751–1796.

Cramer, James J., "Pumping Enron," *New York Magazine*, February 2002.

Crawford, Vincent P. and Joel Sobel, "Strategy Information Transmission," *Econometrica*, 1982, *50* (6), 1431–1451.

Dale, David C., "President's Address: Physicians and the Pharmaceutical Industry," *Transactions of the American Clinical and Climatological Association*, 2017, *128*, 1–3.

Dealbook, "Goldman E-Mail Lays Bare Trading Conflicts," *The New York Times*, January 2010.

Del Guercio, Diane, Laura Seery, and Tracie Woidtke, "Do boards pay attention when institutional investor activists "just vote no"?," *Journal of Financial Economics*, 2008, *90* (1), 84–103.

Farrell, Joseph and Matthew Rabin, "Cheap Talk," *Journal of Economic Perspectives*, Summer 1996, *10* (3), 103–118.

Fiore, Kristina, "Patients Wary of Doctors' Pharma Relationships," *ABC News*, August 2010.

Fowler, Stephen, "Raffensperger declares victory over election denialism in Georgia GOP secretary of state's race," *Georgia Public Broadcasting*, May 2022. `https://www.gpb.org/news/2022/05/25/raffensperger-declares-victory-over-election-denialism-in-georgia-gop-secretary-of`. (accessed October 11, 2022).

Fryer, Roland G., Philipp Harms, and Matthew O. Jackson, "Updating Beliefs when Evidence is Open to Interpretation: Implications for Bias and Polarization," *Journal of the European Economic Association*, October 2019, *17* (5), 1470–1501.

Fudenberg, Drew and Jean Tirole, "A 'Signal-Jamming' Theory of Predation," *The RAND Journal of Economics*, 1986, *17* (3), 366–376.

Gentzkow, Matthew and Jesse M. Shapiro, "Media Bias and Reputation," *Journal of Political Economy*, 2006, *114* (2), 280–316.

__ , Michael B. Wong, and Allen T. Zhang, "Ideological Bias and Trust in Information Sources," August 2021. Working paper.

Groningen, Nicole Van, "Big Pharma gives your doctor gifts. Then your doctor gives you Big Pharma's drugs.," *The Washington Post*, June 2017.

Hermalin, Benjamin E. and Michael S. Weisbach, "Endogenously Chosen Boards of Directors and Their Monitoring of the CEO," *The American Economic Review*, 1998, *88* (1), 96–118.

Holmström, Bengt, "Managerial Incentive Problems: A Dynamic Perspective," *The Review of Economic Studies*, 01 1999, *66* (1), 169–182.

Hwang, Byoung-Hyoun and Seoyoung Kim, "It pays to have friends," *Journal of Financial Economics*, 2009, *93* (1), 138–158.

Jensen, Michael C., "The Modern Industrial Revolution, Exit, and the Failure of Internal Control Systems," *The Journal of Finance*, 1993, *48* (3), 831–880.

Kang, Jun-Koo, Hyemin Kim, Jungmin Kim, and Angie Low, "Activist-Appointed Directors," *Journal of Financial and Quantitative Analysis*, 2022, *57* (4), 1343–1376.

Kartik, Navin, Frances Xu Lee, and Wing Suen, "Information Validates the Prior: A Theorem on Bayesian Updating and Applications," *American Economic Review: Insights*, 2021, *3* (2), 165–182.

King, Maya, "In Georgia, a G.O.P. Primary Tests the Power of a Trump Vendetta," *The New York Times*, May 2022.

Kranton, Rachel and David McAdams, "Social Connectedness and Information Markets," *American Economic Journal: Microeconomics*, 2023. forthcoming.

Larkin, Ian, Desmond Ang, Jonathan Steinhart, Matthew Chao, Mark Patterson, Sunita Sah, Tina Wu, Michael Schoenbaum, David Hutchins, Troyen Brennan, and George Loewenstein, "Association Between Academic Medical Center Pharmaceutical Detailing Policies and Physician Prescribing," *JAMA*, 05 2017, *317* (17), 1785–1795.

Longwell, Sarah, "Trump Supporters Explain Why They Believe the Big Lie," *The Atlantic*, April 2022. Available online: https://www.theatlantic.com/ideas/archive/2022/04/trump-voters-big-lie-stolen-election/629572/, Last accessed June 2022.

Montellaro, Zach, "Georgia election official on Trump's enemies list takes his case to MAGA media," *Politico*, February 2022.

Morris, Stephen, "The Common Prior Assumption in Economic Theory," *Economics and Philosophy*, 1995, *11*, 227–253.

Mullainathan, Sendhil and Andrei Shleifer, "The Market for News," *The American Economic Review*, 2005, *95* (4), 1031–1053.

Myers, Steven Lee and Eileen Sullivan, "Disinformation Has Become Another Untouchable Problem in Washington," *New York Times*, July 2022. Available online: https://www.nytimes.com/2022/07/06/business/disinformation-board-dc.html, Last accessed July 2022.

Nyhan, Brendan, "Facts and Myths about Misperceptions," *Journal of Economic Perspectives*, Summer 2020, *34* (3), 220–236.

Ornstein, Charles, Mike Tigas, and Ryann Grochowski Jones, "Now There's Proof: Docs Who Get Company Cash Tend to Prescribe More Brand-Name Meds," *ProPublica*, March 2016.

Ortoleva, Pietro and Erik Snowberg, "Overconfidence in Political Behavior," *American Economic Review*, 2015, *105* (2), 504–535.

Rabin, Matthew and Joel L. Schrag, "First Impressions Matter: A Model of Confirmatory Bias," *Quarterly Journal of Economics*, 1999, *114* (1), 37–82.

Reinhard, Beth and Yvonne Wingett Sanchez, "As more states create election integrity units, Arizona is a cautionary tale," *The Washington Post*, September 2022. `https://www.washingtonpost.com/investigations/2022/09/26/arizona-election-integrity-unit/`. (accessed June 22, 2022).

Richmond, Jennifer, Wizdom Powell, Maureen Maurer, Rikki Mangrum, Marthe R. Gold, Ela Pathak-Sen, Manshu Yang, and Kristin L. Carman, "Public Mistrust of the U.S. Health Care System's Profit Motives: Mixed-Methods Results from a Randomized Controlled Trial," *Journal of General Internal Medicine*, 2017, *32*, 1396–1402.

Savage, Leonard J., *The Foundations of Statistics*, New York: Wiley Publishing, 1954. Reprinted in 1972 by Dover, New York.

Stein, Jeremy C., "Efficient Capital Markets, Inefficient Firms: A Model of Myopic Corporate Behavior," *The Quarterly Journal of Economics*, 11 1989, *104* (4), 655–669.

Szeidl, Adam and Ferenc Szucs, "The Political Economy of Alternative Realities," July 2022. Working paper.

Trump, Donald J., "Statement by Donald J. Trump, 45th President of the United States of America," 05 2021. Available online: https://www.donaldjtrump.com/news/statement-by-donald-j-trump-45th-president-of-the-united-states-of-america-05.03.21, Last accessed May 2022.

Appendix A Proofs

A.1 Proof of Proposition 1

Let λ_i^j be receiver i's prior on sender j, $P_i(j)$. Let $q_{\theta j} = P(m = 1|\theta, j)$.

To show the first part, consider a receiver's posterior likelihood ratios regarding (j, θ):

$$\frac{P(j,0|\mathbf{m}_n)}{P(j,1|\mathbf{m}_n)} = \frac{(q_{0j})^{n_1}(1-q_{0j})^{n-n_1}\lambda_i^j(1-\omega_i)}{(q_{1j})^{n_1}(1-q_{1j})^{n-n_1}\lambda_i^j(\omega_i)} = \left(\left(\frac{q_{0j}}{q_{1j}}\right)^{\frac{n_1}{n}}\left(\frac{1-q_{0j}}{1-q_{1j}}\right)^{1-\frac{n_1}{n}}\right)^n\left(\frac{1-\omega_i}{\omega_i}\right),$$

where $n = \infty$ and note that $p_{1j} \equiv \lim_{n\to\infty}\frac{n_1}{n}$. We can write this as $\frac{P(j,0|\mathbf{m}_n)}{P(j,1|\mathbf{m}_n)} = X^n\left(\frac{1-\omega_i}{\omega_i}\right)$ where $X = \left(\frac{q_{0j}}{q_{1j}}\right)^{\frac{n_1}{n}}\left(\frac{1-q_{0j}}{1-q_{1j}}\right)^{1-\frac{n_1}{n}}$. When is $X = 1$? Without loss of generality, suppose the truth is $(j, \theta) = (u, 1)$ so $p_{1j} = q_{1j}$. Holding fixed q_{1j}, we have

$$\begin{aligned}\frac{\partial X}{\partial q_{0j}} &= q_{1j}(\frac{q_{0j}}{q_{1j}})^{q_{1j}-1}(\frac{1}{q_{1j}})(\frac{1-q_{0j}}{1-q_{1j}})^{1-q_{1j}} + (\frac{q_{0j}}{q_{1j}})^{q_{1j}}(1-q_{1j})(\frac{1-q_{0j}}{1-q_{1j}})^{1-q_{1j}-1}(-\frac{1}{1-q_{1j}}) \\ &= (\frac{q_{0j}}{q_{1j}})^{q_{1j}-1}(\frac{1-q_{0j}}{1-q_{1j}})^{-q_{1j}}\left(\frac{1-q_{0j}}{1-q_{1j}} - \frac{q_{0j}}{q_{1j}}\right). \qquad \text{(A.1)}\end{aligned}$$

Thus, $\frac{\partial X}{\partial q_{0j}} > 0$ if $q_{0j} < q_{1j}$, $\frac{\partial X}{\partial q_{0j}} = 0$ if $q_{0j} = q_{1j}$, and $\frac{\partial X}{\partial q_{0j}} < 0$ if $q_{0j} > q_{1j}$. Since we can easily verify that $X = 1$ when $q_{0j} = q_{1j}$, this implies that $X \neq 1$ when $q_{0j} \neq q_{1j}$. Thus, $X = 1$ if and only if $q_{1j} = q_{0j}$ when $\frac{n_1}{n} = q_{1j}$. This means that if the truth is $(j, 1)$, then in equilibrium the receiver will know that it is *not* $(j, 0)$ whenever $q_{0j} \neq q_{1j}$. Thus, given j, the receiver learns the truth whenever $p_{1j} \neq p_{0j}$.

To show the second part, consider two senders j and $j' \neq j$:

$$\frac{P(j',0|\mathbf{m}_n)}{P(j,1|\mathbf{m}_n)} = \frac{(q_{0j'})^{n_1}(1-q_{0j'})^{n-n_1}(1-\lambda_i^j)(1-\omega_i)}{(q_{1j})^{n_1}(1-q_{1j})^{n-n_1}\lambda_i^j(\omega_i)}$$
$$= \left(\left(\frac{(q_{0j'})(1-q_{1j})}{q_{1j}1-q_{0j'}}\right)^{\frac{n_1}{n}}\left(\frac{1-q_{0j'}}{1-q_{1j}}\right)\right)^n\left(\frac{(1-\lambda_i^j)(1-\omega_i)}{\lambda_i^j\omega_i}\right), \tag{A.2}$$

Note that the first term of Equation A.2 is the same as the first term of Equation A.1 except that we have $1-q_{0j'}$ in place of $1-q_{0j}$. Thus by the same argument as in Part 1, if the truth is $(j,1)$, then in equilibrium the receiver will know that it is *not* $(j',0)$ whenever $q_{0j'} \neq q_{1j}$.

Parts 1 and 2 imply that if $p_{1j} \neq p_{0j}$ for all j and if $p_{1j} \neq p_{0j'}$ for $j \neq j'$, then in equilibrium the receiver can fully identify θ from the frequency of messages. Likewise, if the receiver can fully identify θ from the frequency of messages, then $p_{1j} \neq p_{0j}$ for all j and $p_{1j} \neq p_{0j'}$ for $j \neq j'$.

A.2 Proof of Proposition 2

Let R_p be the mass of receivers who choose action $a_i = 1$ when they observe some long-run frequency p. By definition, in any non-babbling equilibrium: There exists some frequencies p_1 and p_1' such that $R_{p_1} > R_{p_1'}$ if $\theta = 1$, and some frequencies p_0 and p_0' such that $R_{p_0} < R_{p_0'}$ if $\theta = 0$, where $p_1 \neq p_1'$ and $p_0 \neq p_0'$.

If $\theta = 1$, sender type j's payoff from using a strategy that generates p_1 is

$$-(1-c_j-b_j)^2(R_{p_1}) - (0-c_j-b_j)^2(1-R_{p_1}). \tag{A.3}$$

Thus if $\theta = 1$, u prefers a strategy that generates p_1 over a strategy that generates p_1' if and only if

$$(R_{p_1} - R_{p_1'})(-(1-c_j-b_j)^2 + (0-c_j-b_j)^2) \geq 0$$
$$(R_{p_1} - R_{p_1'})(-1+2c_j+b_j) \geq 0. \tag{A.4}$$

If $\theta = 0$, the sender type u's payoff from using a strategy that generates p_0 is

$$-(1-b_j)^2(R_{p_0}) - (0-b_j)^2(1-R_{p_0}). \tag{A.5}$$

Thus if $\theta = 0$, u prefers a strategy that generates p_0 over a strategy that generates p_0' if and only if

$$(R_{p_0} - R_{p_0'})(-(1-b_j)^2 + (0-b_j)^2) \geq 0$$
$$(R_{p_0} - R_{p_0'})(-1+2b_j) \geq 0. \tag{A.6}$$

We consider what each form of equilibrium can be and which sender types must exist to sustain them.

A.2.1 Fully Informative Equilibria

In a fully informative equilibrium, the receiver learns the state given any equilibrium long-run frequency. By Proposition 1, in any fully informative equilibrium, senders u and v must be using strategies such that $p_{1j} \neq p_{0j}$ for all j, and $p_{1j} \neq p_{0j'}$ for all $j \neq j'$. This implies that $R_{p_{1u}} = 1, R_{p_{0u}} = 0, R_{p_{1v}} = 1, R_{p_{0v}} = 0$. Equations A.6 and A.4 imply that senders u and v must be good

in order to sustain a fully informative equilibrium. Since all receivers take the correct action in equilibrium, a good sender type has no incentive to deviate to any off-equilibrium strategies.

A.2.2 Doublespeak Equilibria

In a doublespeak equilibrium, by Proposition 1 there must be at least $p_{1u} = p_{0v}$ for some $u \neq v$ or $p_{1u} = p_{0u}$ for some u. All of the cases are:

1. Mimicking: Senders use strategies such that $p_{1u} \neq p_{0u}$ and $p_{1v} = p_{0v} = p_{1u}$.

   By Bayes Rule, receiver i's posterior beliefs are

$$P(u, 0|p(\mathbf{m}_\infty) = p_{0u}) = 1 \tag{A.7}$$
$$P(u, 0|p(\mathbf{m}_\infty) \in \{p_{1v}, p_{0v}, p_{1u}\}) = 0 \tag{A.8}$$
$$P(u, 1|p(\mathbf{m}_\infty) \in \{p_{1v}, p_{0v}, p_{1u}\}) = \frac{\omega_i \lambda_i}{\omega_i \lambda_i + 1 - \lambda_i} \tag{A.9}$$
$$P(v, 1|p(\mathbf{m}_\infty) \in \{p_{1v}, p_{0v}, p_{1u}\}) = \frac{\omega_i (1 - \lambda_i)}{\omega_i \lambda_i + 1 - \lambda_i} \tag{A.10}$$
$$P(v, 0|p(\mathbf{m}_\infty) \in \{p_{1v}, p_{0v}, p_{1u}\}) = \frac{(1 - \omega_i)(1 - \lambda_i)}{\omega_i \lambda_i + 1 - \lambda_i}. \tag{A.11}$$

   This implies that if $p(\mathbf{m}_\infty) = p_{0u}$, all receivers choose $a_i = 0$ so $R_{p_{0u}} = 0$. If $p(\mathbf{m}_\infty) \in \{p_{1v}, p_{0v}, p_{1u}\}$, only receivers with priors such that $\omega_i \geq \frac{1-\lambda_i}{2-\lambda_i}$ choose $a_i = 1$, so $R_{p_{1u}} = R_{p_{1v}} = R_{p_{0v}} = \int_0^1 \int_{\frac{1-\lambda}{2-\lambda}}^1 f(\lambda, \omega) d\omega d\lambda$. Thus, in such an equilibrium, $R_{p_{0u}} = 0$ and $0 < R_{p_{1u}} = R_{p_{1v}} = R_{p_{0v}} < 1$. By Equations A.4 and A.6, sender type u's preferences must satisfy the following to be unwilling to deviate to any other strategies that generate equilibrium frequencies:

$$R_{p_{1v}}(-1 + 2c_u + 2c_b) \geq 0 \tag{A.12}$$
$$R_{p_{1v}}(1 - 2b_u) \geq 0. \tag{A.13}$$

   Thus sender type u must be good ($b_u \leq 1/2$ and $c_u + b_u \geq 1/2$). By Equations A.4 and A.6, sender type v's preferences must must satisfy the following to be unwilling to deviate to any other strategies that generate equilibrium frequencies:

$$R_{p_{1u}}(-1 + 2c_u + 2c_b) \geq 0 \tag{A.14}$$
$$R_{p_{1u}}(-1 + 2b_u) \geq 0. \tag{A.15}$$

   Thus sender type v must be single-minded ($b_v \geq 1/2$ and $c_v + b_v \geq 1/2$).

   Suppose the sender deviates to a strategy that receivers can clearly identify as out-of-equilibrium (such as a strategy that generates a non-equilibrium frequency or does not produce well-defined long-run frequencies). Out-of-equilibrium messages have zero probability in mimicking equilibrium, so receivers' beliefs in this event can be arbitrary. Suppose, in the worst case for the single-minded sender, that receivers take action $a_i = 0$ in this event. Equations A.4 and A.6 imply that neither sender type would deviate to strategies that generate out-of-equilibrium messages.

2. Mirroring: Senders use strategies such that $p_{1j} \neq p_{0j}$ for all j, and $p_{1j} = p_{0j'}$ for all $j \neq j'$.

By Bayes Rule, receiver i's posterior beliefs are

$$P(u,1|p(\mathbf{m}_\infty)\in\{p_{1u},p_{0v}\}) = \frac{\omega_i\lambda_i}{\omega_i\lambda_i+(1-\omega_i)(1-\lambda_i)} \tag{A.16}$$

$$P(v,0|p(\mathbf{m}_\infty)\in\{p_{1u},p_{0v}\}) = \frac{(1-\omega_i)(1-\lambda_i)}{\omega_i\lambda_i+(1-\omega_i)(1-\lambda_i)} \tag{A.17}$$

$$P(u,0|p(\mathbf{m}_\infty)\in\{p_{0u},p_{1v}\}) = \frac{(1-\omega_i)\lambda_i}{(1-\omega_i)\lambda_i+\omega_i(1-\lambda_i)} \tag{A.18}$$

$$P(v,1|p(\mathbf{m}_\infty)\in\{p_{0u},p_{1v}\}) = \frac{\omega_i(1-\lambda_i)}{(1-\omega_i)\lambda_i+\omega_i(1-\lambda_i)}, \tag{A.19}$$

and $P(u,0|p(\mathbf{m}_\infty)\in\{p_{1u},p_{0v}\}) = P(v,1|p(\mathbf{m}_\infty)\in\{p_{1u},p_{0v}\}) = P(u,1|p(\mathbf{m}_\infty)\in\{p_{0u},p_{1v}\}) = P(v,0|p(\mathbf{m}_\infty)\in\{p_{0u},p_{1v}\}) = 0$. This implies that if $p(\mathbf{m}_\infty)\in\{p_{1u},p_{0v}\}$, only receivers with priors such that $\omega_i \geq 1-\lambda_i$ choose $a_i = 1$. If $p(\mathbf{m}_\infty)\in\{p_{1v},p_{0u}\}$, only receivers with priors such that $\omega_i \geq \lambda_i$ choose $a_i = 1$.

This implies that $R_{p_{1u}} = R_{p_{0v}} = \int_0^1\int_{1-\lambda}^1 f(\lambda,\omega)d\omega d\lambda$ and $R_{p_{0u}} = R_{p_{1v}} = \int_0^1\int_\lambda^1 f(\lambda,\omega)d\omega d\lambda$.

Suppose $R_{p_{1u}} \geq R_{p_{0u}}$. By Equations A.4 and A.6, sender type u's preferences must satisfy the following to be unwilling to deviate to any other strategies that generate equilibrium frequencies:

$$(R_{p_{1u}} - R_{p_{0u}})(-1+2c_u+2c_b) \geq 0 \tag{A.20}$$

$$(R_{p_{1u}} - R_{p_{0u}})(1-2b_u) \geq 0. \tag{A.21}$$

Thus sender type u must be good ($b_u \leq 1/2$ and $c_u + b_u \geq 1/2$). For sender type v to mirror v's strategy, the equalities in Equation A.20 and A.21 must be reversed. Thus sender type v must be malevolent ($b_v \geq 1/2$ and $c_v + b_v \leq 1/2$).

Suppose the sender deviates to a strategy that receivers can clearly identify as out-of-equilibrium (such as a strategy that generates a non-equilibrium frequency or does not produce well-defined long-run frequencies). Out-of-equilibrium messages have zero probability in mirroring equilibrium, so receivers' beliefs in this event can be arbitrary. Suppose that all receivers treat such messages as though they had seen p_{1u}. Equations A.4 and A.6 imply that neither sender type would deviate to strategies that generate out-of-equilibrium messages.

Suppose $R_{p_{1u}} < R_{p_{0u}}$. Analogously, Equations A.20 and A.21 imply that a mirroring equilibrium exists in which sender type u is malevolent and type v is good.

3. There cannot exist equilibria in which senders use strategies such that $p_{1u} = p_{0u}$ for some u, $p_{1v} \neq p_{0v}$ for $v \neq u$, and $p_{1v} \neq p_{0v} \neq p_{1u}$, unless $b_u = 1/2$ and $c_u = 0$.

   In such an equilibrium, $R_{p_1u} = 1, R_{p_{0u}} = 0$ and $0 < R_{p_{1v}} = R_{p_{0v}} < 1$. Equations A.4 and A.6 imply that u would only employ this strategy if $b_u = 1/2$ and $c_u = 0$. Otherwise, type u has a strict incentive to deviate. Thus such an equilibrium can only exist in the knife-edge case in which sender u is indifferent about receivers' actions in both states.

4. There cannot exist equilibria in which senders use strategies such that $p_{1u} = p_{0v}$ for $u \neq v$, $p_{0u} \neq p_{1u}$, $p_{1v} \neq p_{0v}$, and $p_{0u} \neq p_{1v}$, unless $-1+2b_u+2c_u = 0$, $b_u \leq 1/2$ and $b_v = 1/2$, $-1+2b_v+2c_v \geq 1/2$.

In such an equilibrium, $R_{p_{1v}} = 1, R_{p_{0u}} = 0$ and $0 < R_{p_{1u}} = R_{p_{0u}} < 1$. Equations A.4 and A.6 imply that u would only employ this strategy if $-1 + 2b_u + 2c_u = 0$, $b_u \leq 1/2$. Otherwise, type u has a strict incentive to deviate. Likewise, v would only employ strategy if $b_v = 1/2$, $-1+2b_v+2c_v \geq 1/2$. Thus such an equilibrium can only exist in the knife-edge case in which sender u is indifferent about receivers' actions when $\theta = 1$ but prefers $a_i = 0$ when $\theta = 0$, and sender v is indifferent about receivers' actions when $\theta = 0$ but prefers $a_i = 1$ when $\theta = 1$.

Thus, doublespeak can only exist if and only if one sender type good and the other type is not good. The only two forms of doublespeak equilibria are mimicking and mirroring equilibria, except for knife-edge cases in which at least one sender has indifference about receivers' actions. It follows that if both sender types are not good, then only babbling equilibria exist.

A.2.3 Comment about Full Support Assumption

The assumption that $f(\lambda, \omega)$ has full support is not required for any of the main results. It is made purely for clarity because it rules out knife-edge cases that are not economically meaningful. In particular, the fully informative, mimicking, and mirroring equilibria exist for *any* $f(\lambda, \omega)$.

If $f(\lambda, \omega)$ does not have full support, then there exist cases in which doublespeak equilibria can exist for any combination of sender types u and v. As evident from Equations A.4 and A.6, any senders' preferences can be supported whenever the distribution of receivers' priors is such that receivers' actions on net do not change in response to the sender's messages. Consequently, any sender type would be indifferent among any messaging strategies, including those proposed in doublespeak equilibrium. Mimicking equilibria can exist for any combination of sender types u and v if and only if $\int_0^1 \int_0^{\frac{1-\lambda}{2-\lambda}} f(\lambda, \omega) d\omega d\lambda = 1$. This follows from Equations A.12, A.13, A.14, and A.15: if $R_{p_{1u}} = 0$, then any sender type has no incentive to deviate due to indifference among all strategies. Likewise, mirroring equilibria can exist for any combination of sender types u and v if and only if $\int_0^1 \int_{1-\lambda}^1 f(\lambda, \omega) d\omega d\lambda = \int_0^1 \int_\lambda^1 f(\lambda, \omega) d\omega d\lambda$ because any sender type has no incentive to deviate due to indifference among all strategies.

Note that even though receivers' actions do not change in response to the sender's messages, these knife-edge cases are not babbling equilibria because receivers do update their beliefs in response to the message content.

A.3 Proof of Proposition 3

We first solve the $N = 1$ games, then characterize the conditions for $n^* \in \{0, 1, \infty\}$.

A.3.1 Mimicking

Suppose the sender is either good or single-minded. Consider the subgame in which the policymaker has selected $N = 1$. This is a standard one-period cheap talk game. Let R_m be the mass of receivers who choose action $a_i = 1$ when they observe some message $m_1 = m$. Without loss of generality, suppose $R_1 > R_0$. To construct a non-babbling equilibrium, note that the good sender type must have an incentive to report truthfully. Sender type j will report $m_1 = s_1 = 1$ if and only if

$$P_j(\theta = 1|s_1 = 1)(-1 + 2c_j + 2b_j) - P_j(\theta = 0|s_1 = 1)(1 - 2b_j) \geq 0, \tag{A.22}$$

and will report $m_1 = s_1 = 0$ if and only if

$$-P_j(\theta = 1|s_1 = 0)(-1 + 2c_j + 2b_j) + P_j(\theta = 0|s_1 = 0)(1 - 2b_j) \geq 0. \tag{A.23}$$

It follows immediately that a single-minded type ($-1+2c_j+2b_j \geq 0$ and $1-2b_j \leq 0$) reports $m_1 = 1$ regardless of s_1 and her prior ω^S. Further, a necessary condition for type j to report truthfully is that she is a good type, $-1+2c_j+2b_j \geq 0$ and $1-2b_j \geq 0$. Since $P_j(\theta=1|s_1=1) = \frac{\gamma\omega^S}{\gamma\omega^S+(1-\gamma)(1-\omega^S)}$ and $P_j(\theta=1|s_1=0) = \frac{(1-\gamma)\omega^S}{(1-\gamma)\omega^S+\gamma(1-\omega^S)}$, then a good type will report truthfully if and only if $\omega^S \in [\frac{(1-\gamma)(1-2b_u)}{\gamma(-1+2c_u+2b_u)+(1-\gamma)(1-2b_u)}, \frac{\gamma(1-2b_u)}{(1-\gamma)(-1+2c_u+2b_u)+\gamma(1-2b_u)}]$. Thus, the only form of non-babbling equilibrium in the $N=1$ game is mimicking and it is sustained if and only if $\omega^S \in [\frac{(1-\gamma)(1-2b_u)}{\gamma(-1+2c_u+2b_u)+(1-\gamma)(1-2b_u)}, \frac{\gamma(1-2b_u)}{(1-\gamma)(-1+2c_u+2b_u)+\gamma(1-2b_u)}]$. Given the sender types' equilibrium strategies, receiver i chooses $a_i(m_1|m_1=0)=1$ if and only if $\omega_i \geq \frac{1-\gamma\lambda}{2-\lambda}$ and $a_i(m_1|m_1=1)=1$ if and only if $\omega_i \geq \gamma$. As before, we assume a receiver randomizes between actions with equal probability if indifferent.

Figure A1 partitions receivers according to the relevant areas in the mimicking equilibria for the $N=0$, and $N=1$ and $N=\infty$ games.

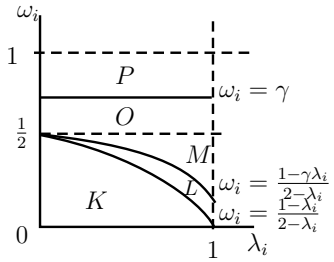


Figure A1: Mimicking: Who takes what actions given $\mathbf{m}_n$

Suppose the mimicking equilibria are realized in the $N=1$ and $N=\infty$ games.[13] Table A1 shows the expected welfare $W^N_{j\theta}$ for the $N=0,1,\infty$ games, conditional on (j,θ).

	$W^\infty_{j\theta}$	$W^0_{j\theta}$	$W^1_{j\theta}$
$(j,\theta)=(1,u)$	$-K$	$-(M+L+K)$	$-(L+K)-(1-\gamma)(O+M)$
$(j,\theta)=(0,u)$	0	$-(1-(M+L+K))$	$-1+(L+K)+\gamma(O+M)$
$(j,\theta)=(1,v)$	$-K$	$-(M+L+K)$	$-(L+K)$
$(j,\theta)=(0,v)$	$-(1-K)$	$-(1-(M+L+K))$	$-(1-(L+K))$

Table A1: Mimicking: Expected Welfare

Comparing the expected payoffs:

$$W^\infty \geq W^0 \iff \lambda^P(1-\omega^P)(1-K)+(2\omega^P-1)(M+L) \geq 0 \tag{A.24}$$

$$W^\infty \geq W^1 \iff \lambda^P\left((1-\omega^P)(1-K)+(\omega^P-\gamma)(O+M)\right)+(2\omega^P-1)L \geq 0 \tag{A.25}$$

$$W^0 \geq W^1 \iff \lambda^P(\gamma-\omega)(O+M)+(2\omega^P-1)M \geq 0. \tag{A.26}$$

[13] That is, suppose the sender's prior ω^S satisfies $\omega^S \in [\frac{(1-\gamma)(1-2b_u)}{\gamma(-1+2c_u+2b_u)+(1-\gamma)(1-2b_u)}, \frac{\gamma(1-2b_u)}{(1-\gamma)(-1+2c_u+2b_u)+\gamma(1-2b_u)}]$ and mimicking equilibria are realized in the $N=1$ and $N=\infty$ games. If the sender's prior does not satisfy this condition or a mimicking equilibrium is not realized in the $N=1$ game, then the $N=1$ game results in a babbling equilibrium and the outcome of the $N=1$ game is equivalent to the $N=0$ game. In that case, the policymaker's problem reduces to a comparison of the $N=\infty$ and $N=0$ games, and $n^*=1$ if and only if $n^*=0$ due to the policymaker's indifference between the $N=0$ and $N=1$ games. Because the policymaker's problem is of greatest interest when the $N=0$, $N=1$, and $N=\infty$ games each result in different outcomes, we suppose mimicking equilibria are realized in the $N=1$ and $N=\infty$ games.

Given (λ^P, ω^P) and $f(\lambda, \omega)$, these results imply that (1) $n^* = \infty$ if and only if Equations A.24 and A.25 are satisfied, (2) $n^* = 0$ if and only if the inequalities in Equation A.24 and A.26 hold in reverse, and (3) $n^* = 1$ if and only if Equation A.26 holds and Equation A.25 holds in reverse.

A.3.2 Mirroring

Suppose the sender is either good or malevolent. Consider the subgame in which the policymaker has selected $N = 1$. This is a standard one-period cheap talk game. Let R_m be the mass of receivers who choose action $a_i = 1$ when they observe some message $m_1 = m$. Without loss of generality, suppose $R_1 \geq R_0$. From the preceding case, we have already shown that sender u plays her equilibrium strategy if and only if $-1 + 2c_u + 2b_u \geq 0$ and $1 - 2b_u \geq 0$ and $\omega^S \in [\frac{(1-\gamma)(1-2b)}{\gamma(-1+2c_u+2b_u)+(1-\gamma)(1-2b_u)}, \frac{\gamma(1-2b_u)}{(1-\gamma)(-1+2c_u+2b_u)+\gamma(1-2b_u)}]$. Analogously, the type v mirrors given $\omega^S \in [\frac{(1-\gamma)(1-2b_v)}{\gamma(-1+2c_v+2b_v)+(1-\gamma)(1-2b_v)}, \frac{\gamma(1-2b_v)}{(1-\gamma)(-1+2c_v+2b_v)+\gamma(1-2b_v)}]$. Applying the same argument as in the $N = \infty$ game, this mirroring equilibrium is sustainable when $E + F \geq A + J$. If $E + F < A + J$, then type u is malevolent and type v is good. Thus, the only form of non-babbling equilibrium in the $N = 1$ game is mirroring and is sustained if and only if $\omega^S \in [\frac{(1-\gamma)(1-2b)}{\gamma(-1+2c_u+2b_u)+(1-\gamma)(1-2b_u)}, \frac{\gamma(1-2b_u)}{(1-\gamma)(-1+2c_u+2b_u)+\gamma(1-2b_u)}] \cap [\frac{(1-\gamma)(1-2b_v)}{\gamma(-1+2c_v+2b_v)+(1-\gamma)(1-2b_v)}, \frac{\gamma(1-2b_v)}{(1-\gamma)(-1+2c_v+2b_v)+\gamma(1-2b_v)}]$. Given the sender types' equilibrium strategies, receiver i chooses $a_i(m_1 = 0) = 1$ if and only if $\omega_i \geq \gamma\lambda_i + (1-\gamma)(1-\lambda_i)$ and $a_i(m_1 = 1) = 1$ if and only if $\omega_i \geq (1-\gamma)\lambda_i + \gamma(1-\lambda_i)$. As before, we assume a receiver randomizes between actions with equal probability if indifferent.

Figure A2 partitions receivers according to the relevant areas in the mirroring equilibria for the $N = 0$, and $N = 1$ and $N = \infty$ games.

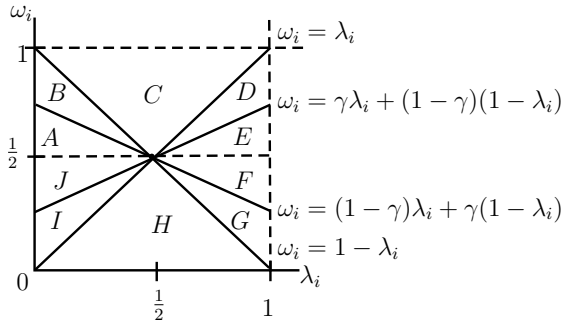


Figure A2: Mirroring: Who takes what actions given $\mathbf{m}_1$

Suppose the mirroring equilibria are realized in the $N = 1$ and $N = \infty$ games.[14] Table A2 shows the expected welfare $W^N_{j\theta}$ for the $N = 0, 1, \infty$ games, conditional on (j, θ).

	$W^\infty_{j\theta}$	$W^0_{j\theta}$	$W^1_{j\theta}$
$(j,\theta) = (1,u)$	$-(B+A+J+I+H)$	$-(J+I+H+G+F)$	$-(I+H+G) - \gamma(A+J) - (1-\gamma)(E+F)$
$(j,\theta) = (0,u)$	$-(C+B+A+J+I)$	$-(1-(J+I+H+G+F))$	$-1 + (I+H+G) + (1-\gamma)(A+J) + \gamma(E+F)$
$(j,\theta) = (1,v)$	$-(1-(C+B+A+J+I))$	$-(J+I+H+G+F)$	$-(I+H+G) - (1-\gamma)(A+J) - \gamma(E+F)$
$(j,\theta) = (0,v)$	$-(1-(B+A+J+I+H))$	$-(1-(J+I+H+G+F))$	$-1 + (I+H+G) + \gamma(A+J) + (1-\gamma)(E+F)$

Table A2: Mirroring: Expected Welfare

[14] That is, suppose the sender's prior ω^S satisfies $\omega^S \in [\frac{(1-\gamma)(1-2b)}{\gamma(-1+2c_u+2b_u)+(1-\gamma)(1-2b_u)}, \frac{\gamma(1-2b_u)}{(1-\gamma)(-1+2c_u+2b_u)+\gamma(1-2b_u)}] \cap [\frac{(1-\gamma)(1-2b_v)}{\gamma(-1+2c_v+2b_v)+(1-\gamma)(1-2b_v)}, \frac{\gamma(1-2b_v)}{(1-\gamma)(-1+2c_v+2b_v)+\gamma(1-2b_v)}]$ and mirroring equilibria are realized in the $N = 1$ and $N = \infty$ games, for the reasons discussed in Footnote 13.

Comparing the expected payoffs:

$$W^\infty \geq W^0 \iff (\lambda - \omega^P)(D + E - I - J) + (\lambda + \omega^P - 1)(F + G - A - B) \geq 0 \tag{A.27}$$

$$W^\infty \geq W^1 \iff (\lambda^P - 1 + \omega^P)(G - B) + (\lambda^P - \omega^P)(D - I) + (1 - \gamma)(2\lambda^P - 1)(E + F - A - J) \geq 0 \tag{A.28}$$

$$W^1 \geq W^0 \iff \left((2\gamma - 1)\lambda^P + 1 - \omega^P - \gamma\right)(E - J) + \left((2\gamma - 1)\lambda^P + \omega^P - \gamma\right)(F - A) \geq 0. \tag{A.29}$$

Given (λ^P, ω^P) and $f(\lambda, \omega)$, these results imply that (1) $n^* = \infty$ if and only if Equations A.27 and A.28 are satisfied, (2) $n^* = 0$ if and only if the inequalities in Equation A.27 and A.29 hold in reverse, and (3) $n^* = 1$ if and only if Equation A.29 holds and Equation A.28 holds in reverse.

Thus there exist partitions $\mathbb{A}_0, \mathbb{A}_1, \mathbb{A}_\infty \subset (0,1) \times (0,1)$ such that $n^* = k$ is the policymaker's optimal choice whenever $(\lambda^P, \omega^P) \in \mathbb{A}_k$. The boundaries of these partitions are:

1. Mimicking: If the sender is either good or single-minded,
   (a) $n^* = \infty$ if and only if $\lambda^P(1-\omega^P)(1-K)+(2\omega^P-1)(M+L) \geq 0$ and $\lambda^P\left((1-\omega^P)(1-K)+(\omega^P-\gamma)(O+\right.$ $(2\omega^P - 1)L \geq 0$.
   (b) $n^* = 0$ if and only if $\lambda^P(1-\omega^P)(1-K)+(2\omega^P-1)(M+L) \leq 0$ and $\lambda^P(\gamma-\omega^P)(O+M)+$ $(2\omega^P - 1)M \leq 0$.
   (c) $n^* = 1$ if and only if $\lambda^P(\gamma-\omega^P)(O+M)+(2\omega^P-1)M \geq 0$ and $\lambda^P\left((1-\omega^P)(1-K)+(\omega^P-\gamma)(O+M\right.$ $(2\omega^P - 1)L \leq 0$.
2. Mirroring: If the sender is either good or malevolent,
   (a) $n^* = \infty$ if and only if $(\lambda^P - \omega^P)(D + E - J - I) + (\lambda^P + \omega^P - 1)(F + G - A - B) \geq 0$ and $(\lambda^P - 1 + \omega^P)(G - B) + (\lambda^P - \omega^P)(D - I) + (1-\gamma)(2\lambda^P - 1)(E + F - A - J) \geq 0$.
   (b) $n^* = 0$ if and only if $(\lambda^P - \omega^P)(D + E - J - I) + (\lambda^P + \omega^P - 1)(F + G - A - B) \leq 0$ and $\left((2\gamma - 1)\lambda^P + 1 - \omega^P - \gamma\right)(E - J) + \left((2\gamma - 1)\lambda^P + \omega^P - \gamma\right)(F - A) \leq 0$.
   (c) $n^* = 1$ if and only if $\left((2\gamma - 1)\lambda^P + 1 - \omega^P - \gamma\right)(E - J) + \left((2\gamma - 1)\lambda^P + \omega^P - \gamma\right)(F - A) \geq 0$ and $(\lambda^P - 1 + \omega^P)(G - B) + (\lambda^P - \omega^P)(D - I) + (1 - \gamma)(2\lambda^P - 1)(E + F - A - J) \leq 0$.

We now characterize the conditions under which $n^* = 1$ is a unique optimum.

A.3.3 Mimicking

Let $f_{0,\infty}$ denote the line when Equation A.24 holds with equality, so $f_{0,\infty} = \omega^P = \frac{L+M-\lambda^P(1-K)}{2(L+M)-\lambda^P(1-K)}$. Let $f_{1,\infty}$ denote the line when Equation A.25 holds with equality, so $f_{1,\infty} = \omega^P = \frac{\lambda^P(K+\gamma(O+M)-1)+L}{\lambda^P(K+O+M-1)+2L}$. Let $f_{0,1}$ denote the line when Equation A.26 holds with equality, so $f_{0,1} = \omega^P = \frac{\gamma\lambda^P(O+M)-M}{\lambda^P(O+M)-2M}$. Define $\lambda^P_{0,\infty}$ such that $f_{0,\infty}(\lambda^P_{0,\infty}) = 0$, $\lambda^P_{1,\infty}$ such that $f_{1,\infty}(\lambda^P_{1,\infty}) = 0$, and $\lambda^P_{0,1}$ such that $f_{0,1}(\lambda^P_{0,1}) = 0$.

We can show that $f_{0,\infty}$ is positive, decreasing and concave for all $\lambda^P \in [0, \lambda^P_{0,\infty}]$, and analogously for $f_{1,\infty}$ and $f_{0,1}$. First, note that $f_{0,\infty}(0) = f_{1,\infty}(0) = f_{0,1}(0) = \frac{1}{2}$. By direct differentiation:

$$\frac{\partial f_{0,\infty}}{\partial \lambda^P} = -\frac{(1-K)(L+M)}{(2(L+M)-\lambda^P(1-K))^2} < 0, \tag{A.30}$$

$$\frac{\partial^2 f_{0,\infty}}{\partial (\lambda^P)^2} = \frac{-2(1-K)(L+M)}{(2(L+M)-\lambda^P(1-K))^3}. \tag{A.31}$$

The denominator term $2(L+M)-\lambda^P(1-K)$ is decreasing in λ^P and positive at $\lambda^P=0$. Consider $\overline{\lambda}$ such that $2(L+M)-\overline{\lambda}(1-K)=0$, so $\overline{\lambda}=\frac{2(L+M)}{1-K}$. Since $\lambda^P_{0,\infty}=\frac{L+M}{1-K}<1$ and $\lambda^P_{0,\infty}<\overline{\lambda}$, then $2(L+M)-\lambda^P(1-K)$ is positive for all $\lambda\in[0,\lambda^P_{0,\infty}]$. Thus, $f_{0,\infty}$ is positive, decreasing and concave for all $\lambda^P\in[0,\lambda^P_{0,\infty}]$. The same exercise establishes the analogous properties for $f_{1,\infty}$ and $f_{0,1}$ (details omitted for brevity).

Recall that $n^*=1$ if and only if

$$\lambda^P(\gamma-\omega^P)(O+M)+(2\omega^P-1)M\geq 0$$
$$\lambda^P\left((1-\omega^P)(1-K)+(\omega^P-\gamma)(O+M)\right)+(2\omega^P-1)L\leq 0.$$

First, we show that if $\gamma(O+M)(M+L)-M(1-K)>0$, there exists a range of (λ^P,ω^P) such that $n^*=1$ is the unique optimum. It follows from the characterization of $\mathbb{A}_1$ and the above properties of $f_{1,\infty}$ and $f_{0,1}$ that there exists a range of (λ^P,ω^P) such that $n^*=1$ is the unique optimum if $\lambda^P_{1,\infty}>\lambda^P_{0,1}$, which holds if and only if $\gamma(O+M)(M+L)-M(1-K)>0$:

$$\lambda^P_{1,\infty}>\lambda^P_{0,1} \tag{A.32}$$
$$\frac{L}{(1-K)-\gamma(O+M)}>\frac{M}{\gamma(O+M)} \tag{A.33}$$
$$\gamma(O+M)(M+L)-M(1-K)>0. \tag{A.34}$$

Second, we show that there exists a range of (λ^P,ω^P) such that $n^*=1$ is the unique optimum only if $\gamma(O+M)(M+L)-M(1-K)>0$. Note that $f^{-1}_{1,\infty}=\lambda^P=\frac{(1-2\omega^P)L}{(1-\omega^P)(1-K)+(\omega^P-\gamma)(O+M)}$ and $f^{-1}_{0,1}=\lambda^P=\frac{(1-2\omega^P)M}{(\gamma=\omega^P)(O+M)}$. Moreover, $f^{-1}_{1,\infty}\geq f^{-1}_{0,1}$ if and only if:

$$f^{-1}_{1,\infty}\geq f^{-1}_{0,1} \tag{A.35}$$
$$\frac{(1-2\omega^P)L}{(1-\omega^P)(1-K)+(\omega^P-\gamma)(O+M)}\geq\frac{(1-2\omega^P)M}{(\gamma-\omega^P)(O+M)} \tag{A.36}$$
$$(1-2\omega^P)\left((\gamma-\omega^P)(L+M)(O+M)-(1-\omega^P)M(1-K)\right)\geq 0. \tag{A.37}$$

Suppose $\gamma(O+M)(M+L)-M(1-K)\leq 0$. Then

$$(\gamma-\omega^P)(L+M)(O+M)-(1-\omega^P)M(1-K)\leq(\gamma-\omega^P)(L+M)(O+M)-(1-\omega^P)\gamma(O+M)(M+L) \tag{A.38}$$
$$=-(O+M)(L+M)(1-\gamma)<0. \tag{A.39}$$

Thus, $f^{-1}_{1,\infty}\leq f^{-1}_{0,1}$ for all $\omega^P\in[0,\frac{1}{2}]$, with equality only at $\omega^P=\frac{1}{2}$. Thus there exists a range of (λ^P,ω^P) such that $n^*=1$ is the unique optimum if and only if $\gamma(O+M)(M+L)-M(1-K)>0$. Since $1-K=P+O+M+L$, this can be re-written as $(O+M)(\gamma L-(1-\gamma)M)-M(P+L)>0$.

A.3.4 Mirroring

Let $g_{0,\infty}$ denote the line when Equation A.27 holds with equality, so $g_{0,\infty}=\lambda^P=\frac{\omega^P(D-I+E-J)+(1-\omega^P)(F-A+G-B)}{(D+E+F+G)-(A+B+J+I)}$.
Let $g_{1,\infty}$ denote the line when Equation A.28 holds with equality, so $g_{1,\infty}=\lambda^P=\frac{(1-\omega^P)(G-B)+\omega^P(D-I)+(1-\gamma)(E+F-A-J)}{G-B+D-I+2(1-\gamma)(E+F-A-J)}$.
Let $g_{0,1}$ denote the line when Equation A.29 holds with equality, so $g_{0,1}=\lambda^P=\frac{(\omega^P+\gamma-1)(E-J)+(\gamma-\omega^P)(F-A)}{(2\gamma-1)(E-J+F-A)}$.

First, note that $g_{0,\infty}(\frac{1}{2})=g_{1,\infty}(\frac{1}{2})=g_{0,1}(\frac{1}{2})=\frac{1}{2}$. Second, $g_{0,\infty}$, $g_{1,\infty}$, and $g_{0,1}$ are linear. Thus, there exists a range of (λ^P,ω^P) such that $n^*=1$ is the unique optimum as long as $g_{0,1}$ exists and

$g_{0,1} \neq g_{1,\infty}$. The line $g_{0,1}$ does not exist if and only if Equation A.29 holds with equality. It is clear that Equation A.29 holds with equality for all (λ^P, ω^P) if and only if $E - J = F - A = 0$. By direct calculation, $g_{0,1} = g_{1,\infty}$ if and only if:

$$\frac{\omega^P(D-I+E-J)+(1-\omega^P)(F-A+G-B)}{(D+E+F+G)-(A+B+J+I)} = \frac{(1-\omega^P)(G-B)+\omega^P(D-I)+(1-\gamma)(E+F-A-J)}{G-B+D-I+2(1-\gamma)(E+F-A-J)} \tag{A.40}$$

$$(2\omega-1)\left((E-J)\left((1-\gamma)(D+E-J-I)+\gamma G\right)-(F-A)\left((1-\gamma)(F-A+G-B)+\gamma D\right)\right)=0. \tag{A.41}$$

Note that $g_{0,1} = g_{1,\infty}$ also implies $g_{0,1} = g_{1,\infty} = g_{0,\infty}$. Thus, there exists a range of (λ^P, ω^P) such that $n^* = 1$ is the unique optimum unless $f(\lambda, \omega)$ such that (1) $E - J = F - A = 0$, or (2) $(E-J)\left((1-\gamma)(D+E-J-I)+\gamma G\right) = (F-A)\left((1-\gamma)(F-A+G-B)+\gamma D\right)$.

A.4 Proof of Proposition 4

Obviously, all receivers fact-check in equilibrium and learn θ if fact-checking is free ($\phi = 0$). Thus any messaging strategy is sustainable because messages are irrelevant.

Let $\phi > 0$. Lemma 1 applies to any equilibrium.

Proof. A receiver who fact-checks learns θ and takes the action that matches the state. Thus a receiver's payoff from fact-checking is $0 - \phi = -\phi$. A receiver who does not fact-check chooses action $a_i(\mathbf{m}_\infty) = 1$ if $\mu_i > 1/2$, $a_i(\mathbf{m}_\infty) = 0$ if $\mu_i < 1/2$, and randomizes between actions with equal probability if $\mu_i = 1/2$.

Thus if $\mu_i \geq 1/2$, receiver i's expected payoff from not fact-checking is $-(1-\mu_i)$:

$$\mu_i[-(1-1)^2] + (1-\mu_i)[-(1-0)^2] = -(1-\mu_i). \tag{A.42}$$

If $\mu_i < 1/2$, receiver i's expected payoff from not fact-checking is $-\mu_i$:

$$\mu_i[-(1-0)^2] + (1-\mu_i)[-(0-0)^2] = -\mu_i. \tag{A.43}$$

Thus, a receiver with $\mu_i \geq 1/2$ fact-checks when $\mu_i \leq 1 - \phi$. A receiver with $\mu_i < 1/2$ fact-checks when $\mu_i \geq \phi$. ■

An implication of Lemma 1 is that if $\phi > 1/2$, then no receivers fact-check in any equilibrium.

We can construct each non-babbling equilibrium by accounting for receivers' fact-checking best responses (Lemma 1). Because the construction method is otherwise analogous to that of Proposition 2, we provide full details in the Online Appendix.

A.5 Proof of Proposition 5

We can construct each non-babbling equilibrium by accounting for receivers' beliefs that the sender type is good after they compare messages to the realized state. Because the construction method is analogous for each type of equilibrium, we provide a detailed analysis of the first type and provide the key elements for the remainder.

A.5.1 Fully Informative Equilibrium

If both sender types are good, then reputation has no effect on the equilibrium because receivers are always sure that the sender is good. Thus the fully informative equilibrium still exists when both senders are good.

Suppose both senders are use strategies that reveal the true state in equilibrium, but only u is good. Given the equilibrium strategies, receivers learn nothing about the sender's type when they learn that the message content reveals to the true state: $P_i(j = u|p(\mathbf{m}_\infty) = p_{1u}, \theta = 1) = P_i(j = u|p(\mathbf{m}_\infty) = p_{0u}, \theta = 0) = \lambda_i$. In equilibrium, the event in which the message content does not reveal to the true state occurs with zero probability. Suppose receivers believe that only a non-good type would report off-equilibrium message content: $P_i(j = u|p(\mathbf{m}_\infty) = p_{1u}, \theta = 0) = P_i(j = u|p(\mathbf{m}_\infty) = p_{0u}, \theta = 1) = 0$.

Since $R_{p_{1u}} = 1, R_{p_{0u}} = 0, R_{p_{1v}} = 1, R_{p_{0v}} = 0$, the sender's payoff from a strategy that results in p_{1u} if $\theta = 1$ is

$$-(1 - c_j - b_j)^2(1) - (0 - c_j - b_j)^2(1 - 1) + r \int_0^1 \int_0^1 \lambda f(\lambda, \omega) d\lambda d\omega, \tag{A.44}$$

and her payoff a strategy that results in p_{0u} if $\theta = 0$ is

$$-(1 - b_j)^2(0) - (0 - c_j - b_j)^2(0) + r \int_0^1 \int_0^1 \lambda f(\lambda, \omega) d\lambda d\omega. \tag{A.45}$$

Deviating to any strategy that results in message content that does not match the state results in a reputation payoff of 0 in that state. Thus a sender does not deviate to any strategy that generates p_1' in state 1 when

$$(1 - R_{p_1'})(-1 + 2b_j + 2c_j) + r \int_0^1 \int_0^1 \lambda f(\lambda, \omega) d\lambda d\omega \geq 0. \tag{A.46}$$

Equation A.46 is satisfied if $b_j + c_j \geq 1/2$. If $b_j + c_j < 1/2$, then the most profitable deviation is $p_1' = p_{0u}$. Thus j would not deviate to any strategy that generates p_1' in state 1 if

$$(-1 + 2c_v + 2b_v) + r \int_0^1 \int_0^1 \lambda f(\lambda, \omega) d\lambda d\omega \geq 0. \tag{A.47}$$

Likewise, a sender does not deviate to any strategy that generates p_0' in state 0 when

$$(R_{p_0'})(1 - 2b_j) + r \int_0^1 \int_0^1 \lambda f(\lambda, \omega) d\lambda d\omega \geq 0. \tag{A.48}$$

Equation A.48 is satisfied if $b_j \leq 1/2$. If $b_j > 1/2$, then the most profitable deviation is $p_0' = p_{1u}$. Thus j would not deviate to any strategy that generates p_0' in state 0 if

$$(1 - 2b_v) + r \int_0^1 \int_0^1 \lambda f(\lambda, \omega) d\lambda d\omega \geq 0. \tag{A.49}$$

Equations A.47 and A.49 imply that reputation expands the set of sender types supporting fully informative equilibria. First, if sender v is also good ($b_v \leq 1/2$ and $c_v + b_v \geq 1/2$), then she will still use her equilibrium strategy. Second, if sender v is single-minded ($b_v > 1/2$ and $c_v + b_v \geq 1/2$), then she will pool with u if reputation is sufficiently strong relative to her desire for receivers to choose

$a_i = 1$ in state $\theta = 0$: $b_v \in (1/2, 1/2 + \frac{r}{2}\int_0^1\int_0^1 \lambda f(\lambda,\omega)d\lambda d\omega]$. Third, if sender v is malevolent ($b_v > 1/2$ and $c_v + b_v > 1/2$), then she will pool with u if reputation is sufficiently strong relative to her desire for receivers to choose the wrong action in each state: $b_v \in (1/2, 1/2 + \frac{r}{2}\int_0^1\int_0^1 \lambda f(\lambda,\omega)d\lambda d\omega]$ and $c_v + b_v \in [1/2 - \frac{r}{2}\int_0^1\int_0^1 \lambda f(\lambda,\omega)d\lambda d\omega, 1/2)$.

A.5.2 Mimicking Equilibrium

Suppose sender u is good and uses a strategy that maps to the true state, and sender v mimics. Given equilibrium strategies, receivers are sure that the sender is good when $p(\mathbf{m}_\infty) = p_{0u}$ and $\theta = 0$: $P_i(j = u|p(\mathbf{m}_\infty) = p_{0u}, \theta = 0) = 1$. When $p(\mathbf{m}_\infty) = p_{1u}$ and $\theta = 1$, receivers cannot identify the sender's type so $P_i(j = u|p(\mathbf{m}_\infty) = p_{1u}) = \lambda_i$. Suppose receivers believe that only a non-good type would report message content that does not map to the true state: $P_i(j = u|p(\mathbf{m}_\infty) = p_{1u}, \theta = 0) = 0$. Likewise, suppose receivers believe out-of equilibrium frequencies come from the non-good type and choose $a_i = 0$. Intuitively, reputation concerns only reinforce a good type's incentive to reveal the state. We can easily verify that sender u will not deviate when $b_u \leq 1/2$ and $c_u + b_u \geq 1/2$. Sender v does not deviate from the mimicking strategy if

$$\left(\int_0^1\int_{\frac{1-\lambda}{2-\lambda}}^1 f(\lambda,\omega)d\omega d\lambda\right)(-1 + 2b_v) - r \geq 0, \tag{A.50}$$

$$\left(\int_0^1\int_{\frac{1-\lambda}{2-\lambda}}^1 f(\lambda,\omega)d\omega d\lambda\right)(-1 + 2c_v + 2b_v) + r\int_0^1\int_0^1 \lambda f(\lambda,\omega)d\lambda d\omega \geq 0. \tag{A.51}$$

Equations A.50 and A.51 imply that reputation leads fewer single-minded types to mimic, but some malevolent types will mimic instead of mirror. Equation A.50 implies that v does not mimic if $b_v \leq 1/2$, so she will not mimic if she is a good type. If v is single-minded, then Equation A.51 is satisfied. Thus the single-minded type will still mimic if $c_v + b_v \geq 1/2$ and $b_v \geq 1/2 + \frac{r}{2\int_0^1\int_{\frac{1-\lambda}{2-\lambda}}^1 f(\lambda,\omega)d\omega d\lambda}$. A malevolent type will mimic if $b_v \geq 1/2 + \frac{r}{2\int_0^1\int_{\frac{1-\lambda}{2-\lambda}}^1 f(\lambda,\omega)d\omega d\lambda}$ and $c_v + b_v \in [1/2 - \frac{r\int_0^1\int_0^1 \lambda f(\lambda,\omega)d\lambda d\omega}{2\int_0^1\int_{\frac{1-\lambda}{2-\lambda}}^1 f(\lambda,\omega)d\omega d\lambda}, 1/2)$.

A.5.3 Mirroring Equilibrium

Without loss of generality, suppose sender u is good and sender v mirrors. Given the equilibrium strategies, receivers learn both the state and the type after the state is realized: $P_i(j = u|p(\mathbf{m}_\infty) = p_{0u}, \theta = 0) = P_i(j = u|p(\mathbf{m}_\infty) = p_{1u}, \theta = 1) = 1$ and $P_i(j = v|p(\mathbf{m}_\infty) = p_{0u}, \theta = 1) = P_i(j = v|p(\mathbf{m}_\infty) = p_{1u}, \theta = 0) = 1$. Suppose receivers believe that only a non-good type would report out-of equilibrium frequencies or message content that does not reveal to the true state and choose $a_i = 0$. Intuitively, reputation concerns only reinforce a good type's incentive to reveal the state. We can easily verify that sender u will not deviate when $b_u \leq 1/2$ and $c_u + b_u \geq 1/2$. Sender v does not deviate from mirroring if

$$\left(\int_{\frac{1}{2}}^1\int_{1-\lambda}^{\lambda} f(\lambda,\omega)d\omega d\lambda - \int_0^{\frac{1}{2}}\int_{\lambda}^{1-\lambda} f(\lambda,\omega)d\omega d\lambda\right)(1 - 2c_v - 2b_v) - r \geq 0 \tag{A.52}$$

$$\left(\int_{\frac{1}{2}}^1\int_{1-\lambda}^{\lambda} f(\lambda,\omega)d\omega d\lambda - \int_0^{\frac{1}{2}}\int_{\lambda}^{1-\lambda} f(\lambda,\omega)d\omega d\lambda\right)(-1 + 2b_v) - r \geq 0. \tag{A.53}$$

Equations A.52 and A.53 imply that reputation shrinks the set of sender types who mirror. Equation A.52 implies that the v does not mirror if $c_v + b_v \geq 1/2$, so she will not mirror if she is a good or single-minded type. Equations A.52 and A.53 imply that only malevolent types whose desire for receivers to choose the wrong action in each state is sufficiently strong relative to their weight on reputation continue to mirror.

A.5.4 Effect of reputation on equilibria

If sender v is a good type ($b_v \leq 1/2$, $c_v + b_v \geq 1/2$), then she reports truthfully regardless of reputation concerns and the fully informative equilibrium exists.

Suppose sender v is a single-minded type ($b_v > 1/2$, $c_v + b_v \geq 1/2$). We have shown that she pools with u if $b_v \in (1/2, 1/2 + \frac{r}{2}\int_0^1\int_0^1 \lambda f(\lambda,\omega)d\lambda d\omega]$, and she mimics if $b_v \geq 1/2 + \frac{r}{2\int_0^1\int_{\frac{1-\lambda}{2-\lambda}}^1 f(\lambda,\omega)d\omega d\lambda}$. Since $\frac{r}{2}\int_0^1\int_0^1 \lambda f(\lambda,\omega)d\lambda d\omega < \frac{r}{2\int_0^1\int_{\frac{1-\lambda}{2-\lambda}}^1 f(\lambda,\omega)d\omega d\lambda}$, the implication is that only a babbling equilibrium exists when $b_v \in (1/2 + \frac{r}{2}\int_0^1\int_0^1 \lambda f(\lambda,\omega)d\lambda d\omega, 1/2 + \frac{r}{2\int_0^1\int_{\frac{1-\lambda}{2-\lambda}}^1 f(\lambda,\omega)d\omega d\lambda})$.

Suppose sender v is a malevolent type ($b_v > 1/2$, $c_v + b_v < 1/2$). We have shown that she pools with u if $b_v \in (1/2, 1/2 + \frac{r}{2}\int_0^1\int_0^1 \lambda f(\lambda,\omega)d\lambda d\omega]$ and $c_v + b_v \in [1/2 - \frac{r}{2}\int_0^1\int_0^1 \lambda f(\lambda,\omega)d\lambda d\omega, 1/2)$. She mimics if $b_v \geq 1/2 + \frac{r}{2\int_0^1\int_{\frac{1-\lambda}{2-\lambda}}^1 f(\lambda,\omega)d\omega d\lambda}$ and $c_v + b_v \in [1/2 - \frac{r\int_0^1\int_0^1 \lambda f(\lambda,\omega)d\lambda d\omega)}{2\int_0^1\int_{\frac{1-\lambda}{2-\lambda}}^1 f(\lambda,\omega)d\omega d\lambda}, 1/2)$. She mirrors if $b_v \geq \frac{1}{2} + \frac{r}{2(\int_{\frac{1}{2}}^1\int_{1-\lambda}^{\lambda} f(\lambda,\omega)d\omega d\lambda - \int_0^{\frac{1}{2}}\int_{\lambda}^{1-\lambda} f(\lambda,\omega)d\omega d\lambda)}$ and $c_v + b_v \leq \frac{1}{2} - \frac{r}{2(\int_{\frac{1}{2}}^1\int_{1-\lambda}^{\lambda} f(\lambda,\omega)d\omega d\lambda - \int_0^{\frac{1}{2}}\int_{\lambda}^{1-\lambda} f(\lambda,\omega)d\omega d\lambda)}$. The implication is that only a babbling equilibrium exists if $b_v \in (1/2 + \frac{r}{2}\int_0^1\int_0^1 \lambda f(\lambda,\omega)d\lambda d\omega, \min\{1/2 + \frac{r}{2\int_0^1\int_{\frac{1-\lambda}{2-\lambda}}^1 f(\lambda,\omega)d\omega d\lambda}, \frac{1}{2} + \frac{r}{2(\int_{\frac{1}{2}}^1\int_{1-\lambda}^{\lambda} f(\lambda,\omega)d\omega d\lambda - \int_0^{\frac{1}{2}}\int_{\lambda}^{1-\lambda} f(\lambda,\omega)d\omega d\lambda)}$), or $c_v + b_v \in (\frac{1}{2} - \frac{r}{2(\int_{\frac{1}{2}}^1\int_{1-\lambda}^{\lambda} f(\lambda,\omega)d\omega d\lambda - \int_0^{\frac{1}{2}}\int_{\lambda}^{1-\lambda} f(\lambda,\omega)d\omega d\lambda)}$, $\frac{r\int_0^1\int_0^1 \lambda f(\lambda,\omega)d\lambda d\omega)}{2\int_0^1\int_{\frac{1-\lambda}{2-\lambda}}^1 f(\lambda,\omega)d\omega d\lambda}$) if $\frac{1}{(\int_{\frac{1}{2}}^1\int_{1-\lambda}^{\lambda} f(\lambda,\omega)d\omega d\lambda - \int_0^{\frac{1}{2}}\int_{\lambda}^{1-\lambda} f(\lambda,\omega)d\omega d\lambda)} > \frac{\int_0^1\int_0^1 \lambda f(\lambda,\omega)d\lambda d\omega)}{\int_0^1\int_{\frac{1-\lambda}{2-\lambda}}^1 f(\lambda,\omega)d\omega d\lambda}$.

Thus, if reputation concerns are sufficiently strong, they can induce some single-minded and malevolent types to pool with the good type when they otherwise would have mimicked or mirrored, respectively. If reputation concerns are sufficiently weak, single-minded types still mimic and malevolent types still mirror. But there also exists an interim range of reputation concerns for both single-minded and malevolent types in which babbling equilibria exist that otherwise would have been doublespeak equilibria. Figure 8 shows this graphically.

Appendix B For Online Publication: Online Appendix

B.1 Details for Proof of Proposition 4

B.1.1 Fully Informative Equilibrium

Within the fully informative equilibrium, a receiver's expected payoff from not fact-checking is 0 because she is sure that the long-run frequency identifies the state (i.e., the receiver's final belief is $\mu_i \in \{0, 1\}$). Thus no receivers fact check if they observe $p(\mathbf{m}_\infty) \in \{p_{1u}, p_{0u}\}$. This implies that the conditions for sender types to use equilibrium strategies apply as in the base model without fact-checking. Thus $b_j \leq 1/2$ and $c_j + b_j \geq 1/2$ for all j. Thus, endogenous fact-checking does not affect the sender types required to sustain fully informative equilibrium.

B.1.2 Mimicking Equilibrium

Within the mimicking equilibrium, all receivers are sure that $\theta = 0$ if $p(\mathbf{m}_\infty) = p_{0u}$: $\mu_i(p(\mathbf{m}_\infty) = p_{0u}) = 0$. Thus their expected payoff from not fact-checking is 0, and no receiver fact-checks if they observe $p(\mathbf{m}_\infty) = p_{0u}$.

If $p(\mathbf{m}_\infty) = p_{1u}$, receivers are not sure of the state, and $\mu_i = P_i(\theta = 1|p(\mathbf{m}_\infty) = p_{1u}) = \frac{\omega_i}{\omega_i + (1-\omega_i)(1-\lambda_i)}$. Receivers with $\mu_i \geq 1/2$ choose $a_i = 1$ if they do not fact-check. By Lemma 1, they will fact-check only when $\mu_i \leq 1 - \phi$, which implies that they will fact-check when $\omega_i \leq \frac{(1-\phi)(1-\lambda_i)}{(1-\phi)(1-\lambda_i)+\phi}$. Receivers whose priors satisfy this condition correspond to receivers in area R in Figure B3. Receivers with $\mu_i < 1/2$ choose $a_i = 0$ if they do not fact-check. By Lemma 1, they will fact-check only when $\mu_i \geq \phi$, which implies that they will fact-check when the following holds: $\omega_i \geq \frac{\phi(1-\lambda_i)}{1-\lambda_i\phi}$. Receivers whose priors satisfy this condition correspond to receivers in area S in Figure B3. Figure B3 shows receivers' actions in response to $p(\mathbf{m}_\infty) = p_{1u}$ in the mimicking equilibrium, where those whose priors lie in the yellow area fact-check. Receivers whose priors lie in area Q do not fact-check and choose $a_i = 1$. Receivers whose priors lie in area T do not fact-check and choose $a_i = 0$. Receivers in area $R + S$ fact-check and choose actions that match the state.

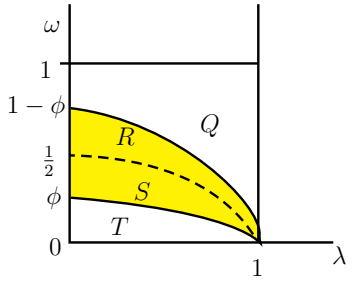


Figure B3: Fact-checking when receivers observe $p(\mathbf{m}_\infty) = p_{1u}$

Sender u's equilibrium strategy generates p_{1u} and leads to $R_{p_{1u}} = 1 - T$ in state 1. Deviating to a strategy that generates p_{0u} in state 1 would lead to $R_{p_{0u}} = 0$ in state 1. By Equation A.4, sender u does not deviate when $b_u + c_u \geq 1/2$. Likewise, sender u's equilibrium strategy generates p_{0u} and leads to $R_{p_{0u}} = 0$ in state 0. Deviating to a strategy that generates p_{1u} in state 0 would lead to $R_{p_{1u}} = Q$ in state 0. By Equation A.6, sender u does not deviate when $b_u \leq 1/2$. By an analogous argument, sender v does not deviate from mimicking when $b_v + c_v \geq 1/2$ and $b_v \geq 1/2$. Suppose sender $j \in \{u, v\}$ deviates to a strategy that generates off-equilibrium frequencies $p(\mathbf{m}_\infty) \notin \{p_{1u}, p_{0u}\}$. If receivers take $a_i = 0$ in response, then there is no incentive for either sender

type to deviate from their equilibrium strategies. Thus, endogenous fact-checking has no effect on the space of sender types required to sustain a mimicking equilibrium.

B.1.3 Mirroring Equilibrium

Within the mirroring equilibrium, no receivers are sure of the state when they observe the equilibrium frequencies.

If $p(\mathbf{m}_\infty) = p_{1u}$, then $\mu_i = P(\theta = 1|p(\mathbf{m}_\infty) = p_{0u}) = \frac{\omega_i\lambda_i}{\omega_i\lambda_i+(1-\omega_i)(1-\lambda_i)}$. Receivers with $\mu_i \geq 1/2$ choose $a_i = 1$ if they do not fact-check. By Lemma 1, they will fact-check only when $\mu_i \leq 1-\phi$, which implies that they will fact-check when $\omega_i \leq \frac{(1-\phi)(1-\lambda_i)}{(1-\phi)(1-\lambda_i)+\phi\lambda_i}$. Receivers whose priors satisfy this condition correspond to receivers in areas $C+I+J+M$ in Figure B4. Receivers with $\mu_i < 1/2$ choose $a_i = 0$ if they do not fact-check. By Lemma 1, they will fact-check only when $\mu_i \geq \phi$, which implies that they will fact-check when $\omega_i \geq \frac{\phi(1-\lambda_i)}{\phi(1-\lambda_i)+\lambda_i(1-\phi)}$. Receivers whose priors satisfy this condition correspond to receivers in areas $B+H+L+D$ in Figure B4.

If $p(\mathbf{m}_\infty) = p_{0u}$, then $\mu_i = P(\theta = 1|p(\mathbf{m}_\infty) = p_{0u}) = \frac{\omega_i(1-\lambda_i)}{\omega_i(1-\lambda_i)+(1-\omega_i)\lambda_i}$. Receivers with $\mu_i \geq 1/2$ choose $a_i = 1$ if they do not fact-check. By Lemma 1, they will fact-check only when $\mu_i \leq 1-\phi$, which implies that they will fact-check when $\omega_i \leq \frac{(1-\phi)\lambda_i}{(1-\phi)\lambda_i+\phi(1-\lambda_i)}$. Receivers whose priors satisfy this condition correspond to receivers in areas $E+I+H+K$ in Figure B4. Receivers with $\mu_i < 1/2$ choose $a_i = 0$ if they do not fact-check. By Lemma 1, they will fact-check only when $\mu_i \geq \phi$, which implies that they will fact-check when $\omega_i \geq \frac{\phi\lambda_i}{\phi\lambda_i+(1-\lambda_i)(1-\phi)}$. Receivers whose priors satisfy this condition correspond to receivers in areas $F+J+L+N$ in Figure B4.

Figure B4 shows receivers' actions in response to equilibrium frequencies in the mirroring equilibrium. To summarize, receivers in the white areas never fact-check: A, D, G, and P. Receivers in the blue area $H+I+J+L$ always fact-check. Receivers in the yellow areas $B+C$ and $M+D$ only fact-check when they observe $p(\mathbf{m}_\infty) = p_{1u}$. Receivers in the green areas $E+F$ and $K+N$ only fact-check when they observe $p(\mathbf{m}_\infty) = p_{0u}$.

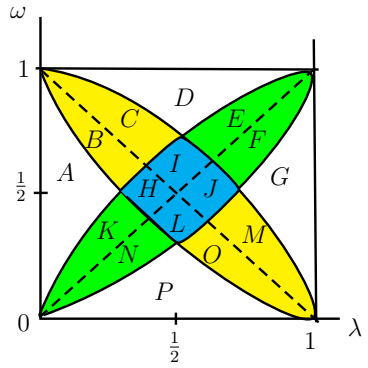


Figure B4: Fact-checking in mirroring equilibrium

Sender u's equilibrium strategy generates p_{1u} and leads to $R_{p_{1u}} = 1-(A+K+N+P)$ in state 1. Deviating to a strategy that generates p_{0u} in state 1 would lead to $R_{p_{0u}} = 1-(P+O+M+G)$ in state 1. By Equation A.4, sender u does not deviate when $b_u+c_u \geq 1/2$ and $O+M+G \geq A+K+N$. Likewise, sender u's equilibrium strategy generates p_{0u} and leads to $R_{p_{0u}} = A+B+C+D$ in state 0. Deviating to a strategy that generates p_{1u} in state 0 would lead to $R_{p_{1u}} = D+E+F+G$ in state 0. By Equation A.6, sender u does not deviate when $b_u \leq 1/2$ and $E+F+G \geq A+B+C$. By an analogous argument, sender v does not deviate from mirroring when $b_v+c_v \leq 1/2$ and $b_v \geq 1/2$. Suppose sender $j \in \{u, v\}$ deviates to a strategy that generates off-equilibrium

frequencies $p(\mathbf{m}_\infty) \notin \{p_{1u}, p_{0u}\}$. Suppose all receivers treat out-of-equilibrium messages as though they had seen p_{1u}. Then neither sender type has an incentive to deviate to such messages. Thus, if $O + M + G \geq A + K + N$ and $E + F + G \geq A + B + C$, then u is a good type and v is malevolent. If $O + M + G \leq A + K + N$ and $E + F + G \leq A + B + C$, then u is a malevolent type and v is good. Otherwise, it is straightforward to show that only babbling equilibria exist. Thus, endogenous fact-checking has no effect on the space of sender types required to sustain a mirroring equilibrium. In essence, when there are more trusting than distrusting receivers among those who do not fact-check on the equilibrium paths, u must be the good type and v must be the malevolent type to sustain the mirroring equilibrium; when there are more distrusting than trusting receivers among those who do not fact-check on the equilibrium paths, u must be malevolent and v must be good.

B.2 Comparison of receivers' welfare with and without fact-checking

We compare receivers' welfare in each non-babbling equilibrium when there is the option to fact-check to when there is not.

Proposition B1 (Receivers' welfare with fact-checking)**.** *Let $\hat{\lambda}$ be the true ex-ante probability that the sender is u, and $\hat{\omega}$ be the true ex-ante probability that the state is $\theta = 1$.*

1. *In fully informative equilibrium, receivers' welfare is unaffected by the option to fact-check.*
2. *In mimicking equilibrium, receivers are better off with the option to fact-check than without it if:*

$$\left(\int_0^1 \int_{\frac{1-\lambda}{2-\lambda}}^{H^\kappa(\lambda,\phi)} f(\lambda,\omega)d\omega d\lambda\right)\left(-\hat{\omega}\phi + (1-\hat{\lambda})(1-\hat{\omega})(1-\phi)\right) + \left(\int_0^1 \int_{L^\kappa(\lambda,\phi)}^{\frac{1-\lambda}{2-\lambda}} f(\lambda,\omega)d\omega d\lambda\right)\left(\hat{\omega}(1-\phi) - (1-\hat{\lambda})(1-\hat{\omega})\phi\right) \geq 0. \tag{B.1}$$

   Otherwise, they are worse off with the option to fact-check than without it.

3. *In mirroring equilibrium, receivers are better off with the option to fact-check than without it if:*

$$\begin{aligned}&\left(\int_0^1 \int_{L_1^\rho(\lambda,\phi)}^{1-\lambda} f(\lambda,\omega)d\omega d\lambda\right)\left(\hat{\lambda}\hat{\omega}(1-\phi) - (1-\hat{\lambda})(1-\hat{\omega})\phi\right)\\ &+ \left(\int_0^1 \int_{1-\lambda}^{H_1^\rho(\lambda,\phi)} f(\lambda,\omega)d\omega d\lambda\right)\left(-\hat{\lambda}\hat{\omega}\phi + (1-\hat{\lambda})(1-\hat{\omega})(1-\phi)\right)\\ &+ \left(\int_0^1 \int_{\lambda}^{H_0^\rho(\lambda,\phi)} f(\lambda,\omega)d\omega d\lambda\right)\left(\hat{\lambda}(1-\hat{\omega})(1-\phi) - (1-\hat{\lambda})\hat{\omega}c\right)\\ &+ \left(\int_0^1 \int_{L_0^\rho(\lambda,\phi)}^{\lambda} f(\lambda,\omega)d\omega d\lambda\right)\left(-\hat{\lambda}(1-\hat{\omega})\phi + (1-\hat{\lambda})\hat{\omega}(1-\phi)\right) \geq 0.\end{aligned} \tag{B.2}$$

   Otherwise, they are worse off with the option to fact-check than without it.

Proof.

1. Fully Informative Equilibrium

Within the fully informative equilibrium, no receivers fact-check on the equilibrium path even though there is the option to fact-check. Thus receivers' welfare is the same as in the base game, where there is no option to fact-check.

2. Mimicking Equilibrium

   We compare receivers' welfare by using the partitions in Figure B3. If $(j,\theta) = (u,0)$, no one fact-checks in either the fact-checking or base game. If $(j,\theta) \in \{(u,1),(v,1)\}$, receivers whose priors lie in R fact-check needlessly and receivers whose priors lie in S benefit by fact-checking. If $(j,\theta) = (v,0)$, receivers whose priors lie in S fact-check needlessly and receivers whose priors lie in R benefit by fact-checking. Thus receivers are better off with the option to fact-check than without it if:

$$\hat{\lambda}(1-\hat{\omega})(0) + \hat{\omega}((R)(-\phi) + (S)(1-\phi)) + (1-\hat{\lambda})(1-\hat{\omega})((R)(1-\phi) + (S)(-\phi)) \geq 0$$
$$(R)\left(\hat{\omega}\phi + (1-\hat{\lambda})(1-\hat{\omega})(1-\phi)\right) + (S)\left(\hat{\omega}(1-\phi) - (1-\hat{\lambda})(1-\hat{\omega})\phi\right) \geq 0,$$

   which is Equation B.1. Otherwise, they are worse off with the option to fact-check than without it.

3. Mirroring Equilibrium

   We compare receivers' welfare by using the partitions in Figure B4. Analogous to the analysis for the mimicking equilibrium, receivers are better off with the option to fact-check than without it if:

$$\begin{aligned}
&(B+H+L+O)\left(\hat{\lambda}(1-\hat{\omega})(1-\phi) - (1-\hat{\lambda})(1-\hat{\omega})\phi\right) \\
&+ (C+I+J+M)\left(\hat{\lambda}\hat{\omega}\phi + (1-\hat{\lambda})(1-\hat{\omega})(1-\phi)\right) \\
&+ (E+I+H+K)\left(\hat{\lambda}(1-\hat{\omega})(1-\phi) - (1-\hat{\lambda})\hat{\omega}\phi\right) \\
&+ (F+J+L+N)\left(-\hat{\lambda}(1-\hat{\omega})\phi + (1-\hat{\lambda})\hat{\omega}(1-\phi)\right) \geq 0,
\end{aligned}$$

   which is Equation B.2. Otherwise, they are worse off with the option to fact-check than without it.

■

B.3 Multiple Senders

Suppose there are two senders, 1 and 2, whose types are drawn independently by nature from $j \in \{u,v\}$. In each subperiod n of period $\tau = 0$, each sender observes independently drawn private signals with accuracy $\gamma \in (1/2,1)$ and reports messages m_{1n} and m_{2n}, respectively. The accuracy γ is common knowledge, and signals are independently and identically distributed across periods. As in the base game, there are $n = \infty$ subperiods and receivers take action $a_i \in \{0,1\}$ at $\tau = 1$, after which payoffs are realized.

Each receiver i's utility is still $-(a_i - \theta)^2$. Receivers are uncertain of the state θ and the types of senders 1 and 2. Receiver i has prior belief at $\tau = 0$ given by $(\lambda_{1i}, \lambda_{2i}, \omega_i) \in (0,1)\times(0,1)\times(0,1)$, where λ_{1i} is the prior probability that sender 1 is type u, λ_{2i} is the prior probability that sender 2 is type u, and ω_i is the prior probability that $\theta = 1$. Let $f(\lambda_1,\lambda_2,\omega)$ denote the density of receivers with prior $(\lambda_1,\lambda_2,\omega)$.

Each sender's preference is still $-\int_0^1 [a_i - (c_j\theta + b_j)]^2 di$. Each sender knows her own type, but is uncertain about the state and the other sender's type. Let ω^{S_1} be sender 1's prior belief at $\tau = 0$ that $\theta = 1$ and ω^{S_2} be sender 2's prior belief at $\tau = 0$ that $\theta = 1$. Let $\lambda_2^{S_1}$ be sender 1's prior that sender 2 is type u and let $\lambda_1^{S_2}$ be sender 2's prior that sender 1 is type u. Let $\mathbf{m}_{1n}$ denote the history of messages sent by sender 1 and $\mathbf{s}_{1n}$ denote the history of private signals observed by sender 1, from subperiods 1 through n. Let $\mathbf{m}_{2n}$ denote the history of messages sent by sender 2 and $\mathbf{s}_{2n}$ denote the history of private signals observed by sender 2, from subperiods 1 through n.

Let n_{11} be the number of ones reported in $\mathbf{m}_{1n}$ and n_{21} be the number of ones reported in $\mathbf{m}_{2n}$.

B.3.1 Mimicking Equilibrium

Suppose type u uses a messaging strategy that results in different long-run frequencies in each state. Type v "mimics" type u by using a messaging strategy that always generates the long-run frequency that type u would have produced in one state (e.g., p_{1u}).

If receivers observe a long-run frequency of p_{0u} from either sender, they are sure that $\theta = 0$: $P(u_1, u_2, 1|(p(\mathbf{m}_{1n}), p(\mathbf{m}_{2n})) \in \{(p_{0u}, p_{1u}), (p_{1u}, p_{0u}), (p_{0u}, p_{0u})\}) = 1$.

If receivers observe $(p(\mathbf{m}_{1n}), p(\mathbf{m}_{2n})) = (p_{1u}, p_{1u})$, their posterior beliefs are

$$P(v_1, v_2, 0|(p_{1u}, p_{1u})) = \frac{(1-\lambda_{1i})(1-\lambda_{2i})(1-\omega_i)}{(1-\lambda_{1i})(1-\lambda_{2i})(1-\omega_i)+\omega_i} \tag{B.3}$$

$$P(u_1, v_2, 1|(p_{1u}, p_{1u})) = \frac{\lambda_{1i}(1-\lambda_{2i})\omega_i}{(1-\lambda_{1i})(1-\lambda_{2i})(1-\omega_i)+\omega_i} \tag{B.4}$$

$$P(u_1, u_2, 1|(p_{1u}, p_{1u})) = \frac{\lambda_{1i}\lambda_{2i}\omega_i}{(1-\lambda_{1i})(1-\lambda_{2i})(1-\omega_i)+\omega_i} \tag{B.5}$$

$$P(v_1, v_2, 1|(p_{1u}, p_{1u})) = \frac{(1-\lambda_{1i})(1-\lambda_{2i})\omega_i}{(1-\lambda_{1i})(1-\lambda_{2i})(1-\omega_i)+\omega_i} \tag{B.6}$$

$$P(v_1, u_2, 1|(p_{1u}, p_{1u})) = \frac{(1-\lambda_{1i})\lambda_{2i}\omega_i}{(1-\lambda_{1i})(1-\lambda_{2i})(1-\omega_i)+\omega_i} \tag{B.7}$$

$$P(u_1, v_2, 0|(p_{1u}, p_{1u})) = P(u_1, u_2, 0|(p_{1u}, p_{1u})) = 0, \tag{B.8}$$

and their actions are

$$a_i(\mathbf{m}_\infty \mid (p_{0u}, p_{1u})) = 0 \tag{B.9}$$

$$a_i(\mathbf{m}_\infty \mid (p_{1u}, p_{0u})) = 0 \tag{B.10}$$

$$a_i(\mathbf{m}_\infty \mid (p_{0u}, p_{0u})) = 0 \tag{B.11}$$

$$a_i(\mathbf{m}_\infty \mid (p_{1u}, p_{1u})) = \begin{cases} 1 & \text{if } \omega_i > \frac{(1-\lambda_{1i})(1-\lambda_{2i})}{1+(1-\lambda_{1i})(1-\lambda_{2i})} \\ 0 & \text{if } \omega_i < \frac{(1-\lambda_{1i})(1-\lambda_{2i})}{1+(1-\lambda_{1i})(1-\lambda_{2i})} \\ (0 \text{ w.p. } \frac{1}{2}; 1 \text{ w.p. } \frac{1}{2}) & \text{if } \omega_i = \frac{(1-\lambda_{1i})(1-\lambda_{2i})}{1+(1-\lambda_{1i})(1-\lambda_{2i})}. \end{cases} \tag{B.12}$$

Omitting details because the method generally follows the single-sender case, except that sender 1 must account for what sender 2's type might be:

Without loss of generality, suppose sender 1 is type u. Sender 1 will not deviate from her

equilibrium strategy if

$$\left(1-\int_0^1\int_0^1\int_0^{\frac{(1-\lambda_1)(1-\lambda_2)}{1+(1-\lambda_1)(1-\lambda_2)}} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_2 d\lambda_1\right)(1-2b_1)\geq 0 \tag{B.13}$$

$$\left(\int_0^1\int_0^1\int_{\frac{(1-\lambda_1)(1-\lambda_2)}{1+(1-\lambda_1)(1-\lambda_2)}}^1 f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_2 d\lambda_1\right)(-1+2c_1+2b_1)\geq 0, \tag{B.14}$$

which imply that $b_1 \leq 1/2$ and $c_1 + b_1 \geq 1/2$, respectively. Without loss of generality, suppose sender 1 is type v. Sender 1 will not deviate from mimicking to any other messaging strategy that generates plausible frequencies if

$$\left(1-\int_0^1\int_0^1\int_{\frac{(1-\lambda_1)(1-\lambda_2)}{1+(1-\lambda_1)(1-\lambda_2)}}^1 f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_2 d\lambda_1\right)(-1+2b_1)\geq 0 \tag{B.15}$$

$$\left(\int_0^1\int_0^1\int_{\frac{(1-\lambda_1)(1-\lambda_2)}{1+(1-\lambda_1)(1-\lambda_2)}}^1 f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_2 d\lambda_1\right)(-1+2c_1+2b_1)\geq 0, \tag{B.16}$$

which imply that $b_1 \geq 1/2$ and $c_1+b_1 \geq 1/2$, respectively. Suppose receivers treat out-of-equilibrium messages as if they had seen p_{1u}. Then neither sender type will deviate to any out-of-equilibrium messages. Thus the mimicking equilibrium exists in the two-sender game if and only if $b_u \leq 1/2$ and $c_u + b_u \geq 1/2$ and $b_v \geq 1/2$ and $c_v + b_v \geq 1/2$.

B.3.2 Mirroring Equilibrium

Suppose type u uses a messaging strategy that results in different long-run frequencies in each state, and type v mirrors u.

If receivers observe $(p(\mathbf{m}_{1n}), p(\mathbf{m}_{2n})) = (p_{1u}, p_{1u})$, their posterior beliefs are

$$P(v_1,v_2,0|(p_{1u},p_{1u})) = \frac{(1-\lambda_{1i})(1-\lambda_{2i})(1-\omega_i)}{(1-\lambda_{1i})(1-\lambda_{2i})(1-\omega_i)+\lambda_{1i}\lambda_{2i}\omega_i} \tag{B.17}$$

$$P(u_1,u_2,1|(p_{1u},p_{1u})) = \frac{\lambda_{1i}\lambda_{2i}\omega_i}{(1-\lambda_{1i})(1-\lambda_{2i})(1-\omega_i)+\lambda_{1i}\lambda_{2i}\omega_i} \tag{B.18}$$

$$\begin{aligned} P(u_1,u_2,0|(p_{1u},p_{1u})) &= P(u_1,v_2,0|(p_{1u},p_{1u})) = P(v_1,u_2,0|(p_{1u},p_{1u})) = P(u_1,v_2,0|(p_{1u},p_{1u})) \\ &= P(v_1,v_2,1|(p_{1u},p_{1u})) = P(v_1,u_2,1|(p_{1u},p_{1u})) = 0. \end{aligned} \tag{B.19}$$

Likewise if receivers observe $(p(\mathbf{m}_{1n}), p(\mathbf{m}_{2n})) = (p_{0u}, p_{0u})$, their posterior beliefs are

$$P(v_1,v_2,1|(p_{0u},p_{0u})) = \frac{(1-\lambda_{1i})(1-\lambda_{2i})\omega_i}{(1-\lambda_{1i})(1-\lambda_{2i})\omega_i+\lambda_{1i}\lambda_{2i}(1-\omega_i)} \tag{B.20}$$

$$P(u_1,u_2,0|(p_{0u},p_{0u})) = \frac{\lambda_{1i}\lambda_{2i}(1-\omega_i)}{(1-\lambda_{1i})(1-\lambda_{2i})\omega_i+\lambda_{1i}\lambda_{2i}(1-\omega_i)} \tag{B.21}$$

$$\begin{aligned} P(u_1,u_2,1|(p_{0u},p_{0u})) &= P(u_1,v_2,0|(p_{0u},p_{0u})) = P(v_1,u_2,0|(p_{0u},p_{0u})) \\ &= P(u_1,v_2,0|(p_{0u},p_{0u})) = P(v_1,v_2,0|(p_{0u},p_{0u})) \\ &= P(v_1,u_2,1|(p_{0u},p_{0u})) = 0. \end{aligned} \tag{B.22}$$

If receivers observe $(p(\mathbf{m}_{1n}), p(\mathbf{m}_{2n})) = (p_{1u}, p_{0u})$, their posterior beliefs are

$$P(v_1, u_2, 0|(p_{1u}, p_{0u})) = \frac{(1-\lambda_{1i})\lambda_{2i}(1-\omega_i)}{(1-\lambda_{1i})\lambda_{2i}(1-\omega_i) + \lambda_{1i}(1-\lambda_{2i})\omega_i} \tag{B.23}$$

$$P(u_1, v_2, 1|(p_{1u}, p_{0u})) = \frac{\lambda_{1i}(1-\lambda_{2i})\omega_i}{(1-\lambda_{1i})\lambda_{2i}(1-\omega_i) + \lambda_{1i}(1-\lambda_{2i})\omega_i} \tag{B.24}$$

$$P(u_1, u_2, 1|(p_{1u}, p_{0u})) = P(v_1, v_2, 1|(p_{1u}, p_{0u})) = P(u_1, u_2, 0|(p_{1u}, p_{0u})) \tag{B.25}$$

$$= P(u_1, v_2, 0|(p_{1u}, p_{0u})) = P(v_1, v_2, 0|(p_{1u}, p_{0u})) = P(v_1, u_2, 1|(p_{1u}, p_{0u})) = 0, \tag{B.26}$$

and analogously for $(p(\mathbf{m}_{1n}), p(\mathbf{m}_{2n})) = (p_{0u}, p_{1u})$ with the types reversed. Intuitively: If the senders' message content agree, then receivers know that they are the same type but are not sure which type. If the senders' message content disagree, then receivers know that they are different types but are not sure who is which type.

The receiver's optimal actions are

$$a_i(\mathbf{m}_\infty \mid (p_{1u}, p_{1u})) = \begin{cases} 1 & \text{if } \omega_i > \frac{(1-\lambda_{1i})(1-\lambda_{2i})}{\lambda_{1i}\lambda_{2i}+(1-\lambda_{1i})(1-\lambda_{2i})} \\ 0 & \text{if } \omega_i < \frac{(1-\lambda_{1i})(1-\lambda_{2i})}{\lambda_{1i}\lambda_{2i}+(1-\lambda_{1i})(1-\lambda_{2i})} \\ (0 \text{ w.p. } \frac{1}{2}; 1 \text{ w.p. } \frac{1}{2}) & \text{if } \omega_i = \frac{(1-\lambda_{1i})(1-\lambda_{2i})}{\lambda_{1i}\lambda_{2i}+(1-\lambda_{1i})(1-\lambda_{2i})}. \end{cases} \tag{B.27}$$

$$a_i(\mathbf{m}_\infty \mid (p_{0u}, p_{0u})) = \begin{cases} 1 & \text{if } \omega_i > \frac{\lambda_{1i}\lambda_{2i}}{\lambda_{1i}\lambda_{2i}+(1-\lambda_{1i})(1-\lambda_{2i})} \\ 0 & \text{if } \omega_i < \frac{\lambda_{1i}\lambda_{2i}}{\lambda_{1i}\lambda_{2i}+(1-\lambda_{1i})(1-\lambda_{2i})} \\ (0 \text{ w.p. } \frac{1}{2}; 1 \text{ w.p. } \frac{1}{2}) & \text{if } \omega_i = \frac{\lambda_{1i}\lambda_{2i}}{\lambda_{1i}\lambda_{2i}+(1-\lambda_{1i})(1-\lambda_{2i})}. \end{cases} \tag{B.28}$$

$$a_i(\mathbf{m}_\infty \mid (p_{1u}, p_{0u})) = \begin{cases} 1 & \text{if } \omega_i > \frac{(1-\lambda_{1i})\lambda_{2i}}{\lambda_{1i}(1-\lambda_{2i})+(1-\lambda_{1i})\lambda_{2i}} \\ 0 & \text{if } \omega_i < \frac{(1-\lambda_{1i})\lambda_{2i}}{\lambda_{1i}(1-\lambda_{2i})+(1-\lambda_{1i})\lambda_{2i}} \\ (0 \text{ w.p. } \frac{1}{2}; 1 \text{ w.p. } \frac{1}{2}) & \text{if } \omega_i = \frac{(1-\lambda_{1i})\lambda_{2i}}{\lambda_{1i}(1-\lambda_{2i})+(1-\lambda_{1i})\lambda_{2i}}. \end{cases} \tag{B.29}$$

$$a_i(\mathbf{m}_\infty \mid (p_{0u}, p_{1u})) = \begin{cases} 1 & \text{if } \omega_i > \frac{\lambda_{1i}(1-\lambda_{2i})}{\lambda_{1i}(1-\lambda_{2i})+(1-\lambda_{1i})\lambda_{2i}} \\ 0 & \text{if } \omega_i < \frac{\lambda_{1i}(1-\lambda_{2i})}{\lambda_{1i}(1-\lambda_{2i})+(1-\lambda_{1i})\lambda_{2i}} \\ (0 \text{ w.p. } \frac{1}{2}; 1 \text{ w.p. } \frac{1}{2}) & \text{if } \omega_i = \frac{\lambda_{1i}(1-\lambda_{2i})}{\lambda_{1i}(1-\lambda_{2i})+(1-\lambda_{1i})\lambda_{2i}}. \end{cases} \tag{B.30}$$

Omitting details because the method generally follows the single-sender case, except that sender 1 must account for what sender 2's type might be: Sender 1 will not deviate from her equilibrium strategy if

$$(1-2b_1)\left(\lambda_2^{S_1}\left(\int_0^1\int_0^1\int_0^{\frac{\lambda_1\lambda_2}{(1-\lambda_1)(1-\lambda_2)+\lambda_1\lambda_2}} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2 - \int_0^1\int_0^1\int_0^{\frac{(1-\lambda_1)\lambda_2}{\lambda_1(1-\lambda_2)+(1-\lambda_1)\lambda_2}} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2\right)\right.$$
$$\left.+(1-\lambda_2^{S_1})\left(\int_0^1\int_0^1\int_0^{\frac{\lambda_1(1-\lambda_2)}{(1-\lambda_1)\lambda_2+\lambda_1(1-\lambda_2)}} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2 - \int_0^1\int_0^1\int_0^{\frac{(1-\lambda_1)(1-\lambda_2)}{\lambda_1\lambda_2+(1-\lambda_1)(1-\lambda_2)}} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2\right)\right) \geq 0 \tag{B.31}$$

and

$$(-1+2b_1+2c_1)\left(\lambda_2^{S_1}\left(\int_0^1\int_0^1\int_{\frac{(1-\lambda_1)(1-\lambda_2)}{(1-\lambda_1)(1-\lambda_2)+\lambda_1\lambda_2}}^1 f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2-\int_0^1\int_0^1\int_{\frac{\lambda_1(1-\lambda_2)}{\lambda_1(1-\lambda_2)+(1-\lambda_1)\lambda_2}}^1 f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2\right)\right.$$
$$\left.+(1-\lambda_2^{S_1})\left(\int_0^1\int_0^1\int_{\frac{(1-\lambda_1)\lambda_2}{(1-\lambda_1)\lambda_2+\lambda_1(1-\lambda_2)}}^1 f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2-\int_0^1\int_0^1\int_{\frac{\lambda_1\lambda_2}{\lambda_1\lambda_2+(1-\lambda_1)(1-\lambda_2)}}^1 f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2\right)\right)\geq 0, \tag{B.32}$$

which can be re-written as

$$(1-2b_1)\left(\lambda_2^{S_1}\left(\int_0^1\int_{1/2}^1\int_{\frac{(1-\lambda_1)\lambda_2}{\lambda_1(1-\lambda_2)+(1-\lambda_1)\lambda_2}}^{\frac{\lambda_1\lambda_2}{(1-\lambda_1)(1-\lambda_2)+\lambda_1\lambda_2}} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2-\int_0^1\int_0^{1/2}\int_{\frac{\lambda_1\lambda_2}{(1-\lambda_1)(1-\lambda_2)+\lambda_1\lambda_2}}^{\frac{(1-\lambda_1)\lambda_2}{\lambda_1(1-\lambda_2)+(1-\lambda_1)\lambda_2}} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2\right)\right.$$
$$\left.+(1-\lambda_2^{S_1})\left(\int_0^1\int_{1/2}^1\int_{\frac{(1-\lambda_1)(1-\lambda_2)}{\lambda_1\lambda_2+(1-\lambda_1)(1-\lambda_2)}}^{\frac{\lambda_1(1-\lambda_2)}{(1-\lambda_1)\lambda_2+\lambda_1(1-\lambda_2)}} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2-\int_0^1\int_0^{1/2}\int_{\frac{\lambda_1(1-\lambda_2)}{(1-\lambda_1)\lambda_2+\lambda_1(1-\lambda_2)}}^{\frac{(1-\lambda_1)(1-\lambda_2)}{\lambda_1\lambda_2+(1-\lambda_1)(1-\lambda_2)}} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2\right)\right)\geq 0 \tag{B.33}$$

and

$$(-1+2b_1+2c_1)\left(\lambda_2^{S_1}\left(\int_0^1\int_{1/2}^1\int_{\frac{(1-\lambda_1)(1-\lambda_2)}{(1-\lambda_1)(1-\lambda_2)+\lambda_1\lambda_2}}^{\frac{\lambda_1(1-\lambda_2)}{\lambda_1(1-\lambda_2)+(1-\lambda_1)\lambda_2}} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2-\int_0^1\int_0^{1/2}\int_{\frac{\lambda_1(1-\lambda_2)}{\lambda_1(1-\lambda_2)+(1-\lambda_1)\lambda_2}}^{\frac{(1-\lambda_1)(1-\lambda_2)}{(1-\lambda_1)(1-\lambda_2)+\lambda_1\lambda_2}} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2\right)\right.$$
$$\left.+(1-\lambda_2^{S_1})\left(\int_0^1\int_{1/2}^1\int_{\frac{(1-\lambda_1)\lambda_2}{(1-\lambda_1)\lambda_2+\lambda_1(1-\lambda_2)}}^{\frac{\lambda_1\lambda_2}{\lambda_1\lambda_2+(1-\lambda_1)(1-\lambda_2)}} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2-\int_0^1\int_0^{1/2}\int_{\frac{\lambda_1\lambda_2}{\lambda_1\lambda_2+(1-\lambda_1)(1-\lambda_2)}}^{\frac{(1-\lambda_1)\lambda_2}{(1-\lambda_1)\lambda_2+\lambda_1(1-\lambda_2)}} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2\right)\right)\geq 0. \tag{B.34}$$

Since sender 1 must be unwilling to deviate for any belief about sender 2 and the state, then this implies $b_1 \leq 1/2$, $c_1+b_1 \geq 1/2$ and the following must hold:

$$\int_0^1\int_{1/2}^1\int_{\frac{(1-\lambda_1)\lambda_2}{\lambda_1(1-\lambda_2)+(1-\lambda_1)\lambda_2}}^{\frac{\lambda_1\lambda_2}{(1-\lambda_1)(1-\lambda_2)+\lambda_1\lambda_2}} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2-\int_0^1\int_0^{1/2}\int_{\frac{\lambda_1\lambda_2}{(1-\lambda_1)(1-\lambda_2)+\lambda_1\lambda_2}}^{\frac{(1-\lambda_1)\lambda_2}{\lambda_1(1-\lambda_2)+(1-\lambda_1)\lambda_2}} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2\geq 0 \tag{B.35}$$

$$\int_0^1\int_{1/2}^1\int_{\frac{(1-\lambda_1)(1-\lambda_2)}{\lambda_1\lambda_2+(1-\lambda_1)(1-\lambda_2)}}^{\frac{\lambda_1(1-\lambda_2)}{(1-\lambda_1)\lambda_2+\lambda_1(1-\lambda_2)}} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2-\int_0^1\int_0^{1/2}\int_{\frac{\lambda_1(1-\lambda_2)}{(1-\lambda_1)\lambda_2+\lambda_1(1-\lambda_2)}}^{\frac{(1-\lambda_1)(1-\lambda_2)}{\lambda_1\lambda_2+(1-\lambda_1)(1-\lambda_2)}} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2\geq 0 \tag{B.36}$$

$$\int_0^1\int_{1/2}^1\int_{\frac{(1-\lambda_1)(1-\lambda_2)}{(1-\lambda_1)(1-\lambda_2)+\lambda_1\lambda_2}}^{\frac{\lambda_1(1-\lambda_2)}{\lambda_1(1-\lambda_2)+(1-\lambda_1)\lambda_2}} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2-\int_0^1\int_0^{1/2}\int_{\frac{\lambda_1(1-\lambda_2)}{\lambda_1(1-\lambda_2)+(1-\lambda_1)\lambda_2}}^{\frac{(1-\lambda_1)(1-\lambda_2)}{(1-\lambda_1)(1-\lambda_2)+\lambda_1\lambda_2}} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2\geq 0 \tag{B.37}$$

$$\int_0^1\int_{1/2}^1\int_{\frac{(1-\lambda_1)\lambda_2}{(1-\lambda_1)\lambda_2+\lambda_1(1-\lambda_2)}}^{\frac{\lambda_1\lambda_2}{\lambda_1\lambda_2+(1-\lambda_1)(1-\lambda_2)}} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2-\int_0^1\int_0^{1/2}\int_{\frac{\lambda_1\lambda_2}{\lambda_1\lambda_2+(1-\lambda_1)(1-\lambda_2)}}^{\frac{(1-\lambda_1)\lambda_2}{(1-\lambda_1)\lambda_2+\lambda_1(1-\lambda_2)}} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2\geq 0. \tag{B.38}$$

Suppose receivers treat any out-of-equilibrium strategy as if they had seen p_{1u}. Then sender 1 does not want to deviate from her equilibrium strategy. We require the analogous conditions to hold for sender 2 not to deviate: $b_1 \geq 1/2$, $c_1+b_1 \leq 1/2$, and reverse the sender indices on Equations B.35, B.36, B.37, and B.38). The intuition is analogous to the single-sender case - in the two-sender game, the mirroring equilibrium requires that there are more receivers who trust than distrust

each sender in order for the good type to use a strategy that maps to the true state. Analogous conditions apply when there are more receivers who distrust than distrust each sender.