**FTG Working Paper Series**

Bayesian Doublespeak

by

Ing-Haw Cheng
Alice Hsiaw

Working Paper No. 00086-00

Finance Theory Group

www.financetheory.com

# Bayesian Doublespeak*

Ing-Haw Cheng[†] Alice Hsiaw[‡]

November 2022

**Abstract**

Why does misinformation persist, and how does it distort the long-run beliefs and actions of rational agents? Suppose receivers see an infinite stream of messages from a sender of unknown type who observes private signals about an unknown state of the world. We characterize the conditions for "doublespeak" equilibria where one sender type repeatedly reveals each private signal truthfully but another sender type repeatedly fabricates false values of her private signals. Receivers only partially learn the true state in the long run irrespective of the true sender type, resulting in long-run disagreement and ex post incorrect actions by some receivers. Equilibrium fact-checking by receivers does not induce more truth-telling among sender types but reputational concerns can. Our results cast doubt on the presumption that rational agents can pierce through persistent extreme lies in the long run and highlight the deleterious effects of such lies for receiver welfare.

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

*The authors thank Avidit Acharya, Dana Foarta, Andrew Little, Thomas Palfrey, Jesse Shapiro, Dustin Tingley, seminar participants at Brandeis University and conference participants at Stanford Institute for Theoretical Economics for comments.

[†]University of Toronto, Rotman School of Management. Email: inghaw.cheng@rotman.utoronto.ca.

[‡]Brandeis University, International Business School. Email: ahsiaw@brandeis.edu.

Why does misinformation persist, and how does it distort the long-run beliefs and actions of rational agents? The usual theoretical 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, the growing amount of misinformation on topics such as election fraud, vaccine safety, and climate change have raised fresh interest in its effect on the beliefs and behavior of agents and heightened concerns about its effect on polarization and public trust in institutions (Myers and Sullivan, 2022).

This paper characterizes when and how misinformation distorts long-run beliefs in equilibrium. We start with the observation from Acemoglu et al. (2016) that rational agents may fail to learn the truth in the long run if they are uncertain about the distribution of signals they receive. We endogenize the uncertainty about signal distributions within a cheap talk game and characterize possible long-run equilibrium beliefs as a function of sender incentives. Our core insight is that, in equilibrium, receivers only partially learn the state in the long run whenever the sender might be a bad type who does not want receivers to learn the truth. Bad-type senders send a form of misinformation that we call *doublespeak*, and the possibility of doublespeak confounds receivers' long-run learning.

Doublespeak confounds learning because it contains misinformation that plausibly comes from a truthful sender type even in the long run. For example, 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.

Specifically, our model features a continuum of receivers with possibly heterogeneous priors who 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, following Acemoglu et al. (2016). We depart by assuming that the infinite sequence of messages comes from a strategic sender who 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 are uncertain about the sender's type in addition to the state and prefer to take an action consistent with the true state. The state, messages, and actions are binary.

We define *misinformation* as any set of messages that are not truthful revelation of the sender's private signals. We say a sender type *doublespeaks* if she delivers misinformation that produces a long-run distribution of messages that could have plausibly been generated by a truthful messaging strategy.

Section 1 introduces the model and provides a key proposition: 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. The intuition follows the example above: If at least one sender type doublespeaks, receivers cannot infer whether messages have one meaning from one sender type or another meaning from the other sender type. Consistent with our usual theoretical intuitions, receivers can unwind other types of misinformation and learn the state for certain in the long run.

In Section 2, we characterize long-run equilibrium beliefs as a function of sender incentives and show when doublespeak occurs in equilibrium. As a benchmark, we show that receivers fully learn the state in equilibrium only when all sender types are good in that they sufficiently prefer that receivers take actions consistent with the true state. The reason is that this set of sender preferences eliminates doublespeak and pure noise strategies in equilibrium.

Our main result is that doublespeak equilibria can occur when receivers are unsure whether or not the sender is a good type. In these equilibria, the good type truthfully relays her private signals while a non-good type doublespeaks, and receivers only partially learn the true state in the long run. Partial learning distinguishes doublespeak equilibria from babbling equilibria, where receivers learn nothing.

Doublespeak equilibria take one of two forms that we call mimicking and mirroring equilibria. A mimicking equilibrium occurs when one sender type is good but the other type is single-minded in that she prefers that take a specific action irrespective of the true state. In equilibrium, 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 only 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 good sender would send if the true state were "1."

A "mirroring" equilibrium can occur when one sender type is good but the other type is malevolent in that she prefers receivers take actions inconsistent with the true state. In equilibrium, the malevolent type sends messages that are perfectly negatively correlated with those of the truthful type, for example by falsely reporting the opposite values of her private signals as in the simple example above.

Receivers may only partially learn the state in doublespeak equilibria, and partial learning leads them to take actions that depend on their priors, even in the long run. Some of these actions may be incorrect ex post. Figure 1 highlights how receiver actions vary across distinct regions of prior beliefs in the mirroring equilibrium. The horizontal axis, $\lambda$, indicates receivers' prior probability that the sender type is good, while the vertical axis, $\omega$, 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."

Doublespeak can be deleterious for receiver welfare. Receivers are unambiguously best off in a fully informative equilibrium because they all take the correct action when the sender's messages always identify the state. Despite partial learning, receivers may even be worse off in a doublespeak equilibrium than in a babbling equilibrium where receivers ignore all messages. This situation occurs when the sender is sufficiently likely to be a non-good type,

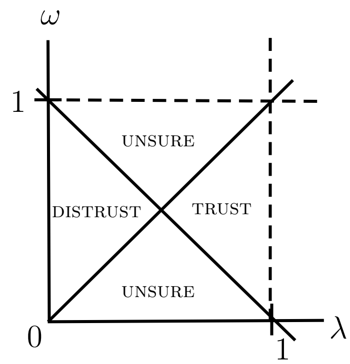



**Figure 1:** Receivers' actions in mirroring equilibrium

so receivers would be better off ignoring her.

Section 3 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. Thus, the possibility of being fact-checked does not induce more truth-telling by senders. Although there is more learning than in the base model in which no fact-checking is possible, the welfare effects of allowing for fact-checking are ambiguous and depend on the distribution of receivers' priors because some receivers needlessly fact check.

Section 4 shows that reputation concerns expand the set of sender types who report truthfully, shrinks the set of sender types who mirror, and expands the set of sender types for which neither informative nor doublespeak equilibria exist. In this extension, senders' preferences depend on both receivers' actions and receivers' posterior beliefs that she is a good type. The intuition for why neither informative and doublespeak equilibria may exist for certain sender type preferences is as follows. Reputation benefits may be insufficient to sustain truthful reporting, but sender types may also be deterred from doublespeak because reputation costs outweigh the intrinsic benefits from misleading receivers. Thus, reputation can either increase or decrease the amount of information revelation, depending on the possible sender types and the degree of reputation concern.

Section 5 discusses alternative assumptions and empirical implications. Doublespeak can occur even if receivers share a common prior or have a continuous action space. Furthermore, we show that doublespeak can still occur if receivers observe the signals of multiple senders before they choose actions or if 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). Indeed, we show that equilibrium outcomes of our cheap talk game with long-run disagreement are identical to those of a static one-message cheap talk game where senders know the true state. In that sense, our work shows that, absent additional assumptions, there is nothing special about receivers seeing many messages over the long run that enables them to ex post learn and agree on an ex ante unknown state.

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 (2020), Kranton and McAdams (2020), and Bowen et al. (2022) 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 also work complements the literature on 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 rational receivers only partially learn the true state in equilibrium because signals endogenously fail to identify the true state. Distinguishing our 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 $\theta$. Receivers are uncertain of the state $\theta$ 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 $\lambda_i$ is the prior probability that $j = u$ and $\omega_i$ 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 $(\lambda, \omega)$ and place no restrictions on its support or continuity.

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 $\theta$ with accuracy $\gamma \in (1/2, 1)$, so $P(s_n = \theta \mid \theta) = \gamma$. The accuracy $\gamma$ 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 $m_n \in \{0, 1\}$.[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 Examples of Model Settings

One model applies to settings with three key features: (1) receivers decide which action to take after seeing many messages from an informed sender, (2) receivers are uncertain about the state and the sender's motives, and (3) the sender's payoffs are linked to receivers' actions. Receivers cannot verify the truthfulness of sender's messages in our base model, but we allow receivers to engage in costly verification in Section 3.

**Lead example—Decisions involving medications.** 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

[1] The restriction of the message space to $m_n \in \{0, 1\}$ is without loss of generality. 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." However, what ultimately matters is the meaning that receivers infer from the messages. We can recharacterize our analysis for a completely unrestricted message space and obtain qualitatively identical equilibrium sender doublespeak and receiver beliefs.

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.

Decisions involving branded medications and vaccinations fit the above description. 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). Similarly, a common source of vaccine hesitancy is distrust in institutions and healthcare professionals due to financial motives (Larson et al., 2018; Jamison et al., 2019). Vaccine hesitancy is a longstanding and growing problem that has persisted despite accumulating evidence of safety and efficacy and repeated calls from medical professionals and public health officials to vaccinate for several diseases (Dubé et al., 2013; Gowda and Dempsey, 2013).

**Other examples.** *Decisions involving election fairness*—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.[2]

Controversy surrounding the 2020 U.S. election fits the above description. Among many voters, significant uncertainty swirled around the fairness of the election and over the credibility of election officials. Regarding the outcome, 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).[3] Regarding the credibility of officials, many voters doubted, and still doubt, officials' repeated insistence that Joseph Biden won fairly (Reinhard and Sanchez, 2022). For example, many voters alleged that Brad Raffensperger, a Republican election of-

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

[3]In contrast, those who believe that Joseph 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).

ficial in Georgia who repeatedly insisted that Biden fairly won in that state, was anti-Trump (Cillizza, 2020).

*Decisions involving investments*—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.

*Decisions involving supporting oversight*—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. This description fits a large literature in corporate finance concerned with whether shareholders persistently elect ineffective, captured corporate boards of directors based on management recommendations (Jensen, 1993; Hermalin and Weisbach, 1998).

The description above also fits political and regulatory contexts. Political leaders, especially authoritarians, may prefer that citizens support the leader's lackeys as candidates for roles that provide opposition or checks on the leader's power and make recommendations accordingly (Guriev and Treisman, 2022). Citizens may be worried whether the leader has authoritarian tendencies and whether a particular candidate for a role is a lackey or not. Similarly, regulated organizations may be a sender type that prefers citizens support candidates who are informed experts for oversight bodies or an alternate type that prefers candidates who are captured lobbyists. Citizens may be unsure of both the organization and candidate's type. For example, many individuals have raised concerns that the pharmaceutical industry has captured the U.S. Food and Drug Administration (Chen, 2018).

## 1.2 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:

**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 $\theta$ 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.

Although we provide Definitions 2, 3, and 4 and state Proposition 1 in terms of long-run frequencies for simplicity, we do not restrict sender strategies to produce well-defined long-run frequencies. Such strategies are on the equilibrium path for all equilibria we consider later. When we construct such equilibria, we allow for strategies that do not generate long-run frequencies. More generally, the results of Proposition 1 apply to strategies that do and do not generate long-run frequencies, because one can redefine $p_{\theta j}$ to represent the history of messages that sender $j$'s strategy produces given the true state $\theta \in \{0, 1\}$.

We classify strategies based on their truthfulness as follows:

**Definition 3** (Truthfulness)**.** *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****. Sender type* $j$*'s strategy is* ***quasi-truthful*** *if it produces a long-run frequency with* $p_{1j} \geq 1/2$ *and* $p_{0j} \leq 1/2$*, with strict inequality for at least one.*

We will focus on a particular form of misinformation:

**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* $j \neq j'$*.*

Intuitively, doublespeak produces long-run frequencies that are identical to those of another sender type under the opposite true state. The following two examples illustrate "mimicking" and "mirroring" strategies. Suppose that sender type $j = u$ truthfully reports $m_n = s_n$ each period. In the mimicking example, sender type $j' = v$ reports $m_n = 1$ with probability $\gamma$ independently of the signal $s_n$. Such a strategy produces $p_{0v} = p_{1u} = \gamma$ by sending messages that are noise. In the mirroring example, sender type $j' = v$ reports $m_n = 1$ if and only if $s_n = 0$. Such a strategy produces $p_{0v} = p_{1u} = \gamma$ and $p_{1v} = p_{0u} = 1 - \gamma$ by sending messages that flip the sender's private signal.

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.

**Proposition 1** (Long-run learning)**.** *When strategies produce message histories with well-defined long-run frequencies $p(\mathbf{m}_\infty)$, a receiver learns $\theta$ 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* $j' \neq j$.

The intuition behind Proposition 1 is that long-run frequencies must identify the state in order for receivers to learn $\theta$ for certain. Doublespeak confounds long-run learning by garbling this identification.

A sketch of the proof illustrates the forwards direction. Suppose receivers learn $\theta$ 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 $\theta$ 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.

The implication of Proposition 1 is that receivers can pierce through all misinformation in the long run *except for* doublespeak. For example, suppose that sender type $u$ reports truthfully as before, but that sender type $v$ plays a "slanting" strategy that tilts messages towards $\theta = 1$ by truthfully reporting $s_n = 1$ but obscuring $s_n = 0$ with noise. Specifically, she plays $m_n = 1$ when $s_n = 1$ but $m_n = X_n$ when $s_n = 0$, where $X_n$ is an i.i.d. Bernoulli random variable with $P(X_n = 1) = 1/2$. Such a strategy results in long-run frequencies of $p_{1v} = \gamma + (1/2)(1-\gamma) = 1/2(1+\gamma)$ and $p_{0v} = (1-\gamma) + 1/2\gamma = 1 - \gamma/2$ and is thus not doublespeak. The possible values of $p(\mathbf{m}_\infty)$ are then $\{p_{1u}, p_{0u}, p_{1v}, p_{0v}\} = \{\gamma, 1-\gamma, 1/2(1+\gamma), 1-\gamma/2\}$. Whenever these values are all distinct, receivers correctly deduce both the true state and sender type upon seeing the realized value of $p(\mathbf{m}_\infty)$. One can expand this intuition to show that receivers learn the true $\theta$ for certain even when no sender types play truthful strategies, so long as the strategies satisfy the conditions in Proposition 1.

# 2 Doublespeak Equilibria

We characterize the set of non-babbling equilibria that can exist over the space of sender types and investigate equilibrium conditions under which receivers fail to learn $\theta$ for certain in the long run even when one sender type optimally plays a quasi-truthful strategy.[4] Our main result is that equilibria with doublespeak may exist whenever one sender type is good but the other sender type is single-minded or malevolent. In these equilibria, receivers may fail to learn $\theta$ for certain in the long run even when one sender type optimally plays a quasi-truthful strategy. In contrast, a fully informative equilibrium where receivers always learn $\theta$ for certain only exists when both sender types are good. When neither sender type is good, only babbling equilibria exist.

We employ the solution concept of perfect Bayesian Nash equilibrium. We consider each type's incentive to deviate from the proposed messaging strategies, given receivers' beliefs within the equilibrium. First, each type must prefer her equilibrium strategy to any other strategy that generates plausible frequencies, meaning those that are possible in equilibrium. Second, each type must prefer her equilibrium strategy to any other strategy that receivers can clearly identify as an off-equilibrium strategy, i.e., those that generate frequencies that are not plausible or that do not generate a long-run frequency. We apply the refinement of neologism-proofness (Farrell, 1993), which restricts receivers' beliefs in cheap talk games to

[4]As is standard in cheap talk games, babbling equilibria where receivers infer no information from the sender's messages and senders thus have no incentive to convey informative messages always exist, for any combination of sender types (Crawford and Sobel, 1982).

credible off-equilibrium messages.[5]

The result that a fully informative equilibrium exists only when both sender types are good follows from standard equilibrium analysis (Online Appendix). The intuition is that, within any fully informative equilibrium, receivers can map long-run frequencies to the true $\theta$ under Proposition 1. Thus, a non-good sender type that (for example) prefers receivers take $a_i = 0$ when $\theta = 1$ can profitably deviate to a messaging strategy that produces a corresponding long-run frequency that maps, under receivers' reasoning, to $\theta = 0$. Thus, when at least one sender type is not good, no fully informative equilibrium exists.

Analogously, only babbling equilibria exist when neither sender type is good. In any non-babbling equilibrium, receivers believe the messages convey information about the state and take actions accordingly because at least one sender type is reporting quasi-truthfully. But if receivers believe the messages convey information, then any non-good type would have a strict incentive to deviate from reporting quasi-truthfully. Thus no information can be conveyed in equilibrium when neither sender type is good.

Instead, doublespeak equilibria may exist when one sender type is good and the other is not good. We show that there are two forms of such equilibria.[6] In each case, we assign type $u$ to report quasi-truthfully and show that $u$ must be a good type in equilibrium. The two cases differ in the preferences and strategies of type $v$.

To facilitate the analysis, we first define a partition of receivers based on how their actions depend on sender's messages:

**Definition 5** (Trust in messages)**.** *Consider* $\mathbf{m}_\infty \in \mathbb{M}$*, where* $\mathbb{M}$ *is the set of message histories that produce long-run frequencies almost surely. Receiver* $i$ ***trusts (distrusts)*** *the sender's messages if* $a_i(\mathbf{m}_\infty) = 1$ *whenever* $p(\mathbf{m}_\infty) > 1/2$ *(*$< 1/2$*) and* $a_i(\mathbf{m}_\infty) = 0$ *whenever* $p(\mathbf{m}_\infty) < 1/2$ *(*$> 1/2$*). Receiver* $i$ *is* ***unsure*** *of the meaning of sender's messages if* $a_i(\mathbf{m}_\infty) = a_i(\emptyset)$ *for all* $\mathbf{m}_\infty \in \mathbb{M}$*.*

[5] As an example within our game, receivers should not believe that a non-plausible frequency identifies a non-good type, since a non-good type has no incentive to reveal herself.

[6] In the Online Appendix, we show that a third equilibrium exists but only as a knife-edge case. The case is where one sender type is identifiable from the observed frequency of messages but is clearly uninformative.

## 2.1 Mimicking equilibrium

In a mimicking equilibrium, sender type $u$ plays a strategy that is quasi-truthful and sender type $v$ plays a strategy that mimics one of sender $u$'s long-run frequencies using noise. Such an equilibrium exists if and only if type $u$ is good and type $v$ is single-minded. We first state the result in Proposition 2 before discussing the intuition.

**Proposition 2** (Mimicking equilibrium)**.** *Consider a **mimicking equilibrium** in which:*

1. *Sender type $u$ plays a quasi-truthful strategy: $m_n = s_n$ with probability $\rho$ for $\rho > \frac{1}{2\gamma}$ and $1 - s_n$ otherwise;*
2. *Sender type $v$ plays a **mimicking strategy**: $m_n = X_n$ where $X_n$ is an i.i.d. Bernoulli random variable with $P(X_n = 1) = \gamma\rho$.*

*Such an equilibrium exists if and only if sender preferences are such that type $u$ is good and type $v$ is single-minded.*

Intuitively, sender type $v$ camouflages her type with that of $u$ by sending a message of 1 each period with i.i.d. probability $\gamma\rho$. The strategy induces a subset of receivers to take action 1 even when the true state turns out to be zero. The reason is that $p(\mathbf{m}_\infty) = \gamma\rho$ is consistent with $(j, \theta) \in \{(u, 1), (v, 1), (v, 0)\}$. Thus, receivers cannot distinguish sender types, and $P_i(j, \theta | p(\mathbf{m}_\infty) = \gamma\rho) < 1$ for all $i$.

For concreteness, suppose $\rho = 1$ so that sender type $u$ plays a truthful strategy and type $v$ reports $m_n = 1$ with probability $\gamma$ each period. Table 1 summarizes strategies and resulting long-run frequencies.

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

**Table 1:** Messaging strategy and long-run frequencies in mimicking equilibria.

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

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

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

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

Receiver $i$ thus chooses $a_i(\mathbf{m}_\infty \mid p(\mathbf{m}_\infty)=1-\gamma)=0$. Furthermore, she chooses $a_i(\mathbf{m}_\infty \mid p(\mathbf{m}_\infty)=\gamma)=1$ if $\omega_i>\frac{1-\lambda_i}{2-\lambda_i}$, $a_i(\mathbf{m}_\infty \mid p(\mathbf{m}_\infty)=\gamma)=0$ if $\omega_i<\frac{1-\lambda_i}{2-\lambda_i}$, and randomizes between actions with equal probability if $\omega_i=\frac{1-\lambda_i}{2-\lambda_i}$.[7]

Figure 2 illustrates which receivers take what actions in the long run. Receivers above the curve with $\omega_i>\frac{1-\lambda_i}{2-\lambda_i}$ trust the sender. They take $a_i=1$ when $p(\mathbf{m}_\infty)=\gamma$ in Panel (a) and $a_i=0$ when $a_i=0$ when $p(\mathbf{m}_\infty)=1-\gamma$ in Panel (b). Receivers with $\omega_i<\frac{1-\lambda_i}{2-\lambda_i}$ are unsure and always choose $a_i=0$.

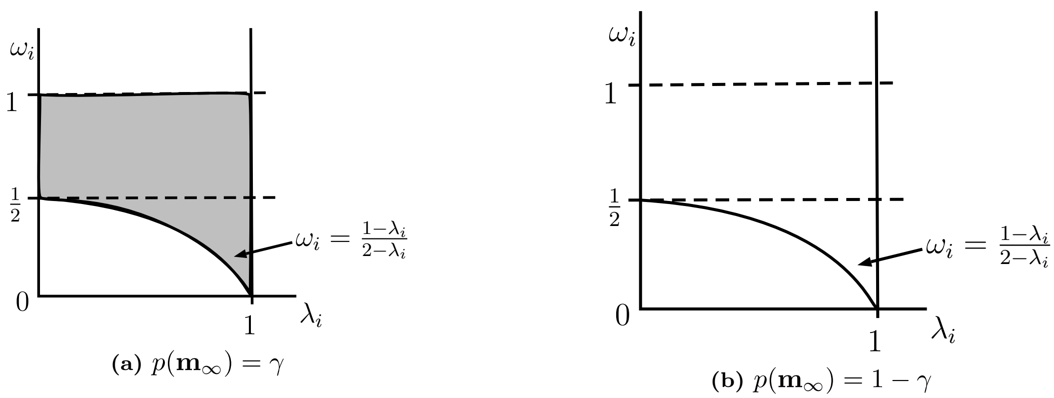



**Figure 2:** Receivers' actions in mimicking equilibrium. Panel (a) shows behavior when receivers observe $p(\mathbf{m}_\infty)=\gamma$. Panel (b) shows behavior when receivers observe $p(\mathbf{m}_\infty)=1-\gamma$. In each panel: Only receivers whose priors lie in the gray areas choose $a_i=1$ in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

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. Consider type $u$'s problem given a mimicking strategy by type $v$ and receivers' beliefs within the mirroring equilibrium. Truthful reporting generates $p(\mathbf{m}_\infty)=1-\gamma$ when $\theta=0$ and induces all receivers to choose $a_i=0$ from Figure 2 Panel (b). Any other strategy would

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

potentially induce some receivers to choose $a_i = 1$ when $\theta = 0$ instead.[8] Likewise, truthful reporting generates $p(\mathbf{m}_\infty) = \gamma$ 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, truthful reporting is optimal if and only if $u$ is a good type. Analogously, when considering type $v$'s problem given a truthful strategy by type $u$, mimicking is optimal if and only if $v$ is a single-minded type who wants receivers to choose $a_i = 1$ in both states. The Appendix contains details and rules out deviations to off-equilibrium strategies.

## 2.2 Mirroring equilibrium

In a mirroring equilibrium, sender type $u$ plays a strategy that is quasi-truthful and sender type $v$ plays a mirroring strategy that reports the opposite of whatever type $u$ would report. Such an equilibrium exists if and only if type $u$ is good, type $v$ is malevolent, and the distribution of receivers' priors satisfies a specific condition. We first state the result in Proposition 3 before discussing the intuition.

**Proposition 3** (Mirroring equilibrium). *Consider a* ***mirroring equilibrium*** *in which:*

1. *Sender type $u$ plays a quasi-truthful strategy: $m_n = s_n$ with probability $\rho$ for $\rho > \frac{1}{2\gamma}$ and $1 - s_n$ otherwise;*
2. *Sender type $v$ plays a* ***mirroring strategy****: $m_n = s_n$ with probability $1 - \rho$ and $1 - s_n$ otherwise.*

*Such an equilibrium exists if and only if: 1) sender preferences are such that type $u$ is good and type $v$ is malevolent, and 2) The distribution of receivers' priors satisfies:*

$$\int_0^{\frac{1}{2}} \int_{1-\omega}^{1} f(\lambda,\omega) d\lambda d\omega - \int_{\frac{1}{2}}^{1} \int_0^{1-\omega} f(\lambda,\omega) d\lambda d\omega \geq 0 \geq \int_0^{\frac{1}{2}} \int_0^{\omega} f(\lambda,\omega) d\lambda d\omega - \int_{\frac{1}{2}}^{1} \int_{1-\omega}^{1} f(\lambda,\omega) d\lambda d\omega. \quad (4)$$

Intuitively, sender type $v$ camouflages her type with that of $u$ by perfectly flipping her private signals. The strategy induces a subset of receivers to take the action opposite to their desired action had they known $\theta$ 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 $\mathbf{m}_\infty$.

[8] For example, deviating to a mimicking strategy would generate $p(\mathbf{m}_\infty) = \gamma$ which induces receiver actions according to Figure 2 Panel (a) instead of $a_i = 0\ \forall i$.

Our discussion below proceeds analogously to our preceding analysis of mimicking equilibria. As before, suppose that $\rho = 1$ so that sender type $u$ plays a truthful strategy and type $v$ reports the opposite of her private signals. 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 | $1-\gamma$ | $\gamma$ |
| Sender type $v$ | 1 | 0 | $\gamma$ | $1-\gamma$ |

**Table 2:** Messaging strategies and long-run frequencies in mirroring equilibria.

What do receivers infer in equilibrium? If $p(\mathbf{m}_\infty) = \gamma$, then all receivers learn that the truth is either $(u,1)$ or $(v,0)$ the other possibilities generate $p(\mathbf{m}_\infty) = 1-\gamma$. However, the receivers cannot distinguish between the two. Receiver $i$'s posterior beliefs are $P(u,0|p(\mathbf{m}_\infty) = \gamma) = P(v,1|p(\mathbf{m}_\infty) = \gamma) = 0$ and:

$$P(u,1|p(\mathbf{m}_\infty) = \gamma) = \frac{\omega_i \lambda_i}{\omega_i \lambda_i + (1-\omega_i)(1-\lambda_i)} \tag{5}$$

$$P(v,0|p(\mathbf{m}_\infty) = \gamma) = \frac{(1-\omega_i)(1-\lambda_i)}{\omega_i \lambda_i + (1-\omega_i)(1-\lambda_i)}. \tag{6}$$

Receiver $i$ chooses $a_i(\mathbf{m}_\infty \mid p(\mathbf{m}_\infty) = \gamma) = 1$ if $\omega_i > 1-\lambda_i$, $a_i(\mathbf{m}_\infty \mid p(\mathbf{m}_\infty) = \gamma) = 0$ if $\omega_i < 1-\lambda_i$, and randomizes between actions with equal probability if $\omega_i = 1-\lambda_i$. Likewise, if $p(\mathbf{m}_\infty) = 1-\gamma$, posterior beliefs are $P(u,1|p(\mathbf{m}_\infty) = 1-\gamma) = P(v,0|p(\mathbf{m}_\infty) = 1-\gamma) = 0$ and:

$$P(u,0|p(\mathbf{m}_\infty) = 1-\gamma) = \frac{(1-\omega_i)\lambda_i}{(1-\omega_i)\lambda_i + \omega_i(1-\lambda_i)} \tag{7}$$

$$P(v,1|p(\mathbf{m}_\infty) = 1-\gamma) = \frac{\omega_i(1-\lambda_i)}{(1-\omega_i)\lambda_i + \omega_i(1-\lambda_i)}. \tag{8}$$

Receiver $i$ chooses $a_i(\mathbf{m}_\infty \mid p(\mathbf{m}_\infty) = 1-\gamma) = 1$ if $\omega_i > \lambda_i$, $a_i(\mathbf{m}_\infty \mid p(\mathbf{m}_\infty) = 1-\gamma) = 0$ if $\omega_i < \lambda_i$, and randomizes between actions with equal probability if $\omega_i = \lambda_i$.

Figure 3 maps receivers' priors into receivers' actions. Receivers whose prior beliefs lie in regions $E$ and $F$ 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 regions $C$ and $D$ always trust the sender's messages. They take $a_i = 1$ if and only if $p(\mathbf{m}_\infty) = \gamma$ since their priors indicate the sender is the good type, even if their priors also suggest that the state is 0 as in region $D$. Finally, receivers whose priors lie in regions $A$ and $B$ distrust the sender's messages and take action $a_i = 0$ if and only if $p(\mathbf{m}_\infty) = \gamma$ since their priors indicate the sender is the malevolent type. These regions also underlie the unsure, trust, and distrust regions in Figure 1.

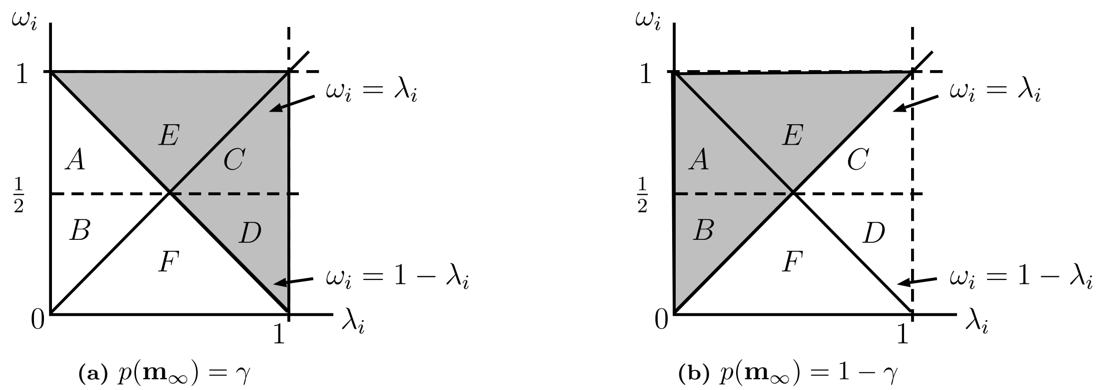



**Figure 3:** Receivers' actions in mirroring equilibrium. Panel (a) shows behavior when receivers observe $p(\mathbf{m}_\infty) = \gamma$. Panel (b) shows behavior when receivers observe $p(\mathbf{m}_\infty) = 1 - \gamma$. In each panel: Only receivers whose priors lie in the gray areas choose $a_i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

By implication, receivers whose priors lie in regions $C$ and $D$ take the correct action ex post if and only if the sender is good because they trust the messages, while receivers whose priors in regions in $A$ and $B$ take the correct action ex post if and only if the sender is malevolent because they distrust the messages. Receivers whose priors lie in regions $E$ and $F$ take the correct action ex post if and only if their priors match the true $\theta$.

Proposition 3 shows that a mirroring equilibrium exists if and only if type $u$ is good, type $v$ is single-minded, and the distribution of receivers' priors satisfies Equation 4. Intuitively, the existence of mirroring equilibria depends on the distribution of receivers' priors because trusting and distrusting receivers always take opposing actions in response to the same messages. Thus, the sender's incentives to deviate depend on the relative distributions of these two groups.

A necessary condition for the truthful strategy to be optimal for a good type and for the mirroring strategy to be optimal for a malevolent type is that there are more trusting receivers than distrusting receivers ($C + D \geq A + B$). Intuitively, a good sender prefers a greater mass of trusting receivers because they will follow her truthful messages. But a malevolent sender shares this preference so that receivers will follow her untruthful messages.

Equation 4 is a stronger condition: $D - A \geq 0 \geq B - C$. The requirement of $D \geq A$ says that the mass of trusting receivers whose priors favor $\theta = 0$ but take action $a = 1$ is greater than the mass of distrusting receivers whose priors favor $\theta = 1$ but take action $a = 0$ when all receivers see $p(\mathbf{m}_\infty) = \gamma$. An analogous statement applies for the condition that $C \geq B$ when all receivers see $p(\mathbf{m}_\infty) = 1 - \gamma$. These conditions rule out any profitable deviation for either sender type that induces receivers to take actions consistent with their priors about the state.

## 2.3 Comparing Mimicking and Mirroring Equilibria

The common feature of both forms of doublespeak equilibria is that receivers must be unsure whether the sender is a good type. The possibility that the sender is good lends credibility to the messages, but a non-good type camouflages her type through doublespeak. The main implication is that receivers only partially learn the truth, so priors influence posterior beliefs and actions even in the long run. This partial learning distinguishes doublespeak equilibria from babbling equilibria, where receivers learn nothing, and fully informative equilibria, where receivers learn everything.

Mimicking and mirroring equilibria differ from each other in three key ways. First, they differ in the set of supported sender preferences: Mimicking equilibria exist when sender type $v$ is single-minded while mirroring equilibria can exist when sender type $v$ is malevolent.

Second, receivers' beliefs respond to messages differently within each equilibrium. Within mirroring equilibria, the beliefs of trusting and distrusting receivers move in opposite directions in response to the same messages, so they also take opposing actions. Thus, one of either the trusting or distrusting group of receivers takes the incorrect action ex post. Within mimicking equilibria, all receivers' beliefs move in the same direction in response to the same messages, so receivers are only either trusting or unsure. Thus, some receivers may take the wrong action if they observe $p(\mathbf{m}_\infty) = \gamma$.[9] Nonetheless, if $p(\mathbf{m}_\infty) = 1 - \gamma$, all receivers agree on the interpretation of messages and take the correct action.

[9] If $\theta = 1$, unsure receivers incorrectly choose $a_i = 0$ because their beliefs do not move sufficiently toward $\theta = 1$. If $j = v$ and $\theta = 0$, trusting receivers incorrectly choose $a_i = 1$.

Third, the existence of mirroring equilibria requires a condition on the distribution of receivers' prior beliefs whereas the existence of mimicking equilibria does not. This additional condition occurs precisely because some receivers distrust the sender in the mirroring equilibrium. Since trusting and distrusting receivers always take opposing actions in response to the same messages, the incentives to deviate to an off-equilibrium strategy depend on the relative distributions of these two groups.

## 2.4 Receivers' Welfare

We define receiver welfare as equal to the expected sum of all receivers' utility given true ex-ante probabilities $\hat{\lambda}$ and $\hat{\omega}$ that nature selects the sender type as $u$ and the state as $\theta = 1$, respectively. Specifically, welfare $W$ equals $E\left[-\int_0^1 (a_i - \theta)^2 \, di\right]$ where the expectation is taken with respect to these probabilities. Proposition 4 compares receiver welfare across the equilibria we have studied.

**Proposition 4** (Receiver welfare)**.** *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. *Receivers are weakly better off in a fully informative equilibrium than in any doublespeak or babbling equilibria supported by any other $f(\lambda, \omega)$ and sender preferences $\{(b_j, c_j) : j \in \{u, v\}\}$.*
2. *For a fixed $f(\lambda, \omega)$ and set of sender preferences $\{(b_j, c_j) : j \in \{u, v\}\}$:*
    - *(a) Receivers are better off in mimicking than in babbling equilibrium:*
        - *i. for all $\hat{\lambda} \in (0, 1]$ if $\hat{\omega} \geq 1/2$, and*
        - *ii. for all $\hat{\lambda} \in \left( \frac{(1-2\hat{\omega})\left(\int_0^1 \int_{\frac{1-\lambda}{2-\lambda}}^{\frac{1}{2}} f(\lambda,\omega) d\omega d\lambda\right)}{(1-\hat{\omega})\left(\int_0^1 \int_{\frac{1-\lambda}{2-\lambda}}^{1} f(\lambda,\omega) d\omega d\lambda\right)}, 1 \right]$ if $\hat{\omega} < 1/2$;*
    - *(b) Receivers are better off in mirroring than in babbling equilibrium if and only if $\hat{\lambda} > 1/2$.*

Part 1 says that receivers are best off in a fully informative equilibrium compared to any other equilibria that may exist with other sender preferences or distributions of receivers' prior beliefs. Since receivers always take the correct action ex post in a fully informative equilibrium, no other equilibrium can produce strictly greater receiver welfare.

Perhaps more surprisingly, Part 2 shows that receiver welfare in a doublespeak equilibrium may be greater or less than welfare in a babbling equilibrium. Intuitively, when receivers tend to trust messages, they will be better off listening to messages than ignoring them when the messages are truthful but worse off when they are not.

Specifically, receivers are better off in a mimicking equilibrium than in a babbling equilibrium if the state is more likely to be $\theta = 1$ or the sender is sufficiently likely to be good. In a mimicking equilibrium, receivers believe the messages are positively correlated with the state. Thus they would only be better off ignoring the messages if it is sufficiently likely that the sender is the single-minded type and the state is $\theta = 0$.

Receivers are better off in a mirroring equilibrium than in a babbling equilibrium when the sender is more likely to be good than malevolent. In order for a mirroring equilibrium to exist, among those receivers for whom the sender's messages induce them to take an action contrary to their prior, more receivers trust than distrust each of the sender's messages (Equation 4). Thus if the sender is more likely to be good and therefore report truthfully, more receivers take the correct action based on the messages than if they ignored them.

## 2.5 Examples of Sender Strategies

We describe sender strategies and equilibria in the real-world examples of Section 1.1. In each of these examples, senders provide many messages to receivers, but uncertainty about sender types and the state among receivers persists.

**Lead example—Decisions involving medications.** If the doctor could be single-minded, a mimicking equilibrium exists where that type sends messages indicating the drug is beneficial or not at the same frequencies that a good type would send were the drug truly beneficial. Unlike the good type, the single-minded type ignores her private signals about the drug.

Evidence and anecdotes on the behavior of doctors who prescribe branded medications is consistent with a mimicking equilibrium (Larkin et al., 2017; Ornstein et al., 2016). First, 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.

Second, these doctors try to camouflage as good types by often claiming to act in the best interest of patients. Third, 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.

**Other examples.** *Decisions involving election fairness*—If the election official could be single-minded in favor of the winning candidate, a mimicking equilibrium exists where that type sends messages indicating the election was fair or unfair at the same frequencies that a good type would send were the election truly fair. Unlike the good type, the single-minded type ignores her private signals about election fairness.

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 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).

More broadly, the behavior of modern autocrats is also consistent with a mimicking equilibrium. Such autocrats repeatedly unfairly win elections by healthy but not overwhelming margins in order to claim that they were plausibly fair. As a result, some portion of the citizenry tend to believe these claims (Guriev and Treisman, 2022, Ch.5). Indeed, evidence that some citizens in both democracies and autocracies remain persistently uncertain about election fairness (Guriev and Treisman, 2022, p.133) is consistent with the possible existence of mimicking equilibria in both types of regimes.

*Decisions involving investments*—If the advisor could be malevolent, there can exist a mirroring equilibrium in which that type says that the stock is doing well when it is not, and that the stock is doing poorly when it is not, in order to trade against the receiver. The former strategy is known as "pump and dump," and the latter is known as "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). More broadly, there is long-running concern that financial firms trade against their clients and recommendations (Dealbook, 2010).

*Decisions involving supporting oversight*—If a firm manager can be malevolent, there exists a mirroring equilibrium where the good type recommends that shareholders vote for the director if and only if the information suggests the director is effective, but the malevolent type recommends that shareholders vote for the director if and only if the information suggests the director is a lackey. Trusting shareholders follow management recommendations, while distrusting shareholders vote against management recommendations.

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.

Evidence from political contexts also suggests the existence of mirroring equilibria. Authoritarian leaders have a history of staying in power through mirroring tactics (Guriev and Treisman, 2022). For example, in 1989 then-First Deputy Prime Minister Goh Chok Tong introduced the Nominated Member of Parliament (NMP) Scheme, ostensibly to represent opposition viewpoints and allow the government to accommodate constructive dissent (Baharudin, 2022). However, there is evidence that suggests that the NMP is a body of lackeys meant to strengthen authoritarian rule while maintaining the appearance of democracy (Guriev and Treisman, 2022). Its true function is still debated among Singaporean citizens (Lim, 2021), who decide whether or not to support the government in the future.

# 3 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 the $n = \infty$ 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 $\theta$. 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 $p(\mathbf{m}_\infty) = 1 - \gamma$, no receiver fact-checks. When $p(\mathbf{m}_\infty) = \gamma$, 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 $p(\mathbf{m}_\infty) = \gamma$, 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 $p(\mathbf{m}_\infty) = 1 - \gamma$, 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 $H_0^\rho(\lambda_i, \phi) \equiv \frac{(1-\phi)\lambda_i}{(1-\phi)\lambda_i+\phi(1-\lambda_i)}$.*

4. *Babbling: Receivers fact-check if their priors satisfy $\omega_i \in [\phi, 1 - \phi]$.*

Proposition 5 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.

**Proposition 5** (Equilibrium existence with endogenous fact-checking)**.** *Comparing the conditions for fully informative and doublespeak equilibria in the game when receivers can fact-check to the conditions in the base game:*

1. *If $\phi > 1/2$, all conditions are identical because no receivers fact-check in any equilibrium.*

2. *If $0 < \phi \leq 1/2$:*

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

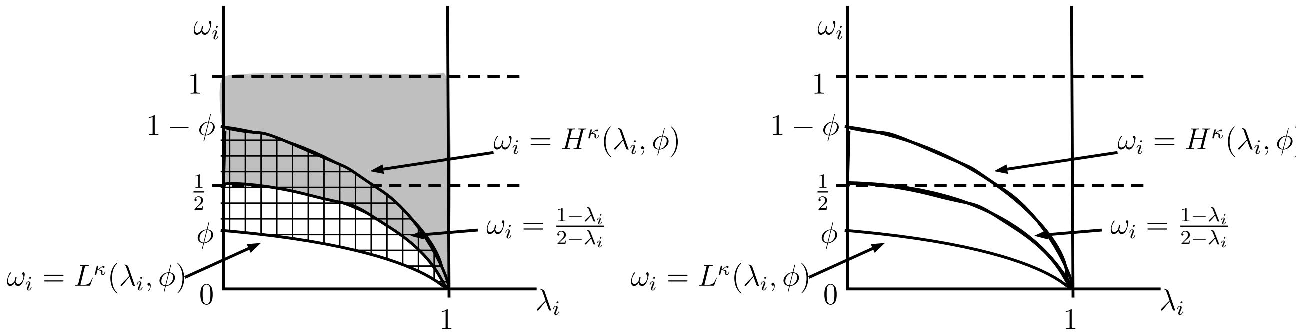
   (b) *The conditions on the distribution of receivers' priors to support mirroring equilibrium in the game with fact-checking are:*

$$\int_{1/2}^{1}\int_{L_1^\rho(\lambda_i,\phi)}^{\phi} f(\lambda,\omega)d\omega d\lambda - \int_{0}^{1/2}\int_{\phi}^{L_1^\rho(\lambda_i,\phi)} f(\lambda,\omega)d\omega d\lambda \geq 0 \geq \int_{0}^{1/2}\int_{L_0^\rho(\lambda_i,\phi)}^{\phi} f(\lambda,\omega)d\omega d\lambda - \int_{1/2}^{1}\int_{\phi}^{L_0^\rho(\lambda_i,\phi)} f(\lambda,\omega)d\omega d\lambda \tag{9}$$

$$\int_{1/2}^{1}\int_{H_1^\rho(\lambda_i,\phi)}^{1-\phi} f(\lambda,\omega)d\omega d\lambda - \int_{0}^{1/2}\int_{1-\phi}^{H_1^\rho(\lambda_i,\phi)} f(\lambda,\omega)d\omega d\lambda \geq 0 \geq \int_{0}^{1/2}\int_{H_0^\rho(\lambda_i,\phi)}^{1-\phi} f(\lambda,\omega)d\omega d\lambda - \int_{1/2}^{1}\int_{1-\phi}^{H_0^\rho(\lambda_i,\phi)} f(\lambda,\omega)d\omega d\lambda. \tag{10}$$

Part 1 says that, when fact-checking is too costly ($\phi > 1/2$), no receivers fact-check in any equilibrium and the set of equilibria is identical to the base game. Part 2 considers the more interesting case of $0 < \phi \leq 1/2$.

Part 2(a) says that, even with endogenous fact-checking, the fully informative equilibrium exists if and only if both sender types are good. Intuitively, sender types must be good precisely because the amount of fact-checking on the equilibrium path is endogenously equal to zero. The amount of fact-checking in equilibrium is zero because because all receivers believe the signals identify the state. With zero fact-checking, any non-good type would thus deviate from a strategy that leads receivers to correctly identify the state to a doublespeak strategy so that receivers mistakenly take the wrong action. Therefore, a non-good type cannot be part of a fully informative equilibrium.




**(a)** $p(\mathbf{m}_\infty) = \gamma$ **(b)** $p(\mathbf{m}_\infty) = 1 - \gamma$

**Figure 4:** Fact-checking in mimicking equilibrium. Panel (a) shows behavior when receivers observe $p(\mathbf{m}_\infty) = \gamma$. Panel (b) shows behavior when receivers observe $p(\mathbf{m}_\infty) = 1 - \gamma$. 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.

Part 2(a) also says that the mimicking equilibrium exists if and only if the possible sender types are good and single-minded, irrespective of whether receivers can fact-check. If $p(\mathbf{m}_\infty) = \gamma$, receivers whose prior beliefs with $(\omega_i, \lambda_i)$ in the area with $\omega_i \in (L^\kappa(\lambda_i, \phi), H^\kappa(\lambda_i, \phi))$ lying in the hatched areas of Figure 4 Panel (a) fact-check, per Lemma 1. Intuitively, receivers fact-check $p(\mathbf{m}_\infty) = \gamma$ when they are ex ante: (1) quite uncertain about the state ($\omega$ near 1/2) and thought the sender type is probably not good ($\lambda$ close to 0), (2) quite certain that $\theta = 0$ ($\omega$ near 0) even though they thought the sender type is probably good ($\lambda$ near 1), or (3) in between these two cases along the $\omega_i = \frac{1-\lambda_i}{2-\lambda_i}$ curve. If $p(\mathbf{m}_\infty) = 1 - \gamma$, all receivers are ex post sure that $\theta = 0$, so that no one fact-checks per Lemma 1 leading to no hatched areas in Figure 4 Panel (b).

The implication of the above analysis is that fact-checking fails to induce the single-minded type to report truthfully instead of playing the mimicking strategy. Intuitively, fact-checking reduces the absolute benefit of mimicking, but does not eliminate its relative benefit over being truthful. If $\theta = 1$ ex post, then receivers would see the same long-run frequency of $\gamma$ from both mimicking and truthful strategies and so take the same action. If $\theta = 0$ ex post, then mimicking induces some receivers to take $a_i = 1$ despite some fact-checking (the grey non-hatched area in Figure 4 Panel (a)), while a truthful strategy would have led to no receivers taking $a_i = 1$ (Figure 4 Panel (b)). Thus, for a single-minded type, the expected payoff from mimicking still exceeds that of a truthful strategy.

Moreover, the good type still prefers to report truthfully instead of deviating 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.

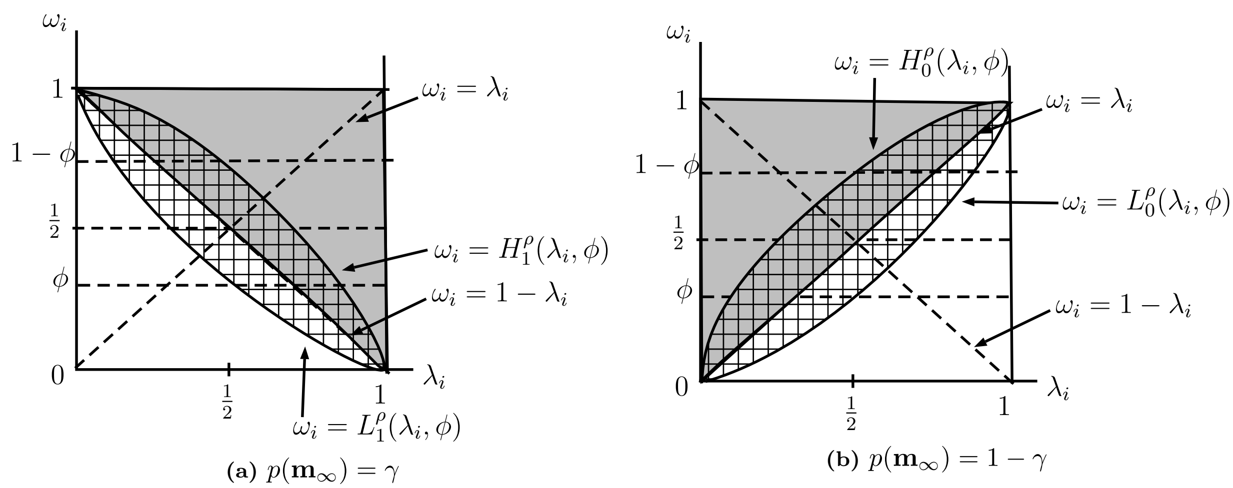



**Figure 5:** Fact-checking in mirroring equilibrium. Panel (a) shows behavior when receivers observe $p(\mathbf{m}_\infty) = \gamma$. Panel (b) shows behavior when receivers observe $p(\mathbf{m}_\infty) = 1 - \gamma$. In each panel: Only receivers whose priors lie in the hatched areas fact-check. Receivers whose priors lie in the gray areas would have chosen $a_i = 1$ in the base game.

Part 2(a) also says that the mirroring equilibrium can only exist if the sender is either 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. Figure 5 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(b) says that the mirroring equilibrium depends on the distribution of receivers' priors qualitatively similar manner to the base model in Proposition 3. Intuitively, if there are more receivers who trust the messages than distrust them, then only a good sender would prefer truthful reporting over other strategies that generate plausible frequencies. Likewise, only a malevolent sender would prefer mirroring. The restrictions on the distribution of receivers' preferences in Equations 9 and 10 only differ from Equation 4 in Proposition 3 because potential deviations must account for differential fact-checking behavior by receivers in response to each plausible and non-plausible frequency. The Appendix contains all details.

Proposition 6 characterizes welfare and shows that the overall effect of introducing the a costly option to fact-check depends on how many receivers benefit from fact checking relative to those who needlessly do it.

**Proposition 6** (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:*

$$\begin{aligned}&\left(\int_0^1 \int_{\frac{1-\lambda}{2-\lambda}}^{H^\kappa(\lambda,\phi)} f(\lambda,\omega)d\omega d\lambda\right)\left(-\hat{\omega}\phi + (1-\hat{\lambda})(1-\hat{\omega})(1-\phi)\right)\\ &+\left(\int_0^1 \int_{L^\kappa(\lambda,\phi)}^{\frac{1-\lambda}{2-\lambda}} f(\lambda,\omega)d\omega d\lambda\right)\left(\hat{\omega}(1-\phi) - (1-\hat{\lambda})(1-\hat{\omega})\phi\right) \geq 0. \end{aligned} \tag{11}$$

   *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:*

$$\begin{aligned}&\left(\int_0^1 \int_{L_1^\rho(\lambda,\phi)}^{1-\lambda} f(\lambda,\omega)d\omega d\lambda\right)\left(\hat{\lambda}\hat{\omega}(1-\phi) - (1-\hat{\lambda})(1-\hat{\omega})\phi\right)\\ &+\left(\int_0^1 \int_{1-\lambda}^{H_1^\rho(\lambda,\phi)} f(\lambda,\omega)d\omega d\lambda\right)\left(-\hat{\lambda}\hat{\omega}\phi + (1-\hat{\lambda})(1-\hat{\omega})(1-\phi)\right)\\ &+\left(\int_0^1 \int_{\lambda}^{H_0^\rho(\lambda,\phi)} f(\lambda,\omega)d\omega d\lambda\right)\left(\hat{\lambda}(1-\hat{\omega})(1-\phi) - (1-\hat{\lambda})\hat{\omega}c\right)\\ &+\left(\int_0^1 \int_{L_0^\rho(\lambda,\phi)}^{\lambda} f(\lambda,\omega)d\omega d\lambda\right)\left(-\hat{\lambda}(1-\hat{\omega})\phi + (1-\hat{\lambda})\hat{\omega}(1-\phi)\right) \geq 0. \end{aligned} \tag{12}$$

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

Part 1 shows that, for a fully informative equilibrium, the option to fact-check does not affect receiver welfare relative to the base game. Intuitively, no receivers fact-check in the fully informative equilibrium anyway.

Parts 2 and 3 show that, for doublespeak equilibria, whether the option to fact-check improves receiver welfare depends on the distribution of receivers' priors, specifically 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 4 Panel (a) fact-check when they observe $p(\mathbf{m}_\infty) = \gamma$. 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 $p(\mathbf{m}_\infty) = \gamma$. 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 $\theta = 0$.

# 4 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 $\theta$ 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 the sender's incentive to report truthfully because the revelation of the state may allow receivers to learn whether the sender's messages are consistent with a good sender type. For example, in a mimicking equilibrium, if $p(\mathbf{m}_\infty) = \gamma$ and $\theta = 0$, then ex post receivers are sure that the sender is not good. But if $p(\mathbf{m}_\infty) = \gamma$ and $\theta = 1$, then ex post receivers still cannot identify the sender so their posterior that the sender is good is $\lambda_i$.

Proposition 7 provides the equilibrium 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 7** (Equilibrium conditions with reputation)**.** *Reputation expands the set of sender types who report truthfully, shrinks the set of sender types who mirror, and may either expand or shrink the set of sender types who mimic.*

*Define* $\underline{\beta} = \frac{1}{2}\int_0^1\int_0^1 \lambda f(\lambda,\omega)d\lambda d\omega$, $\tilde{\beta} = \frac{1}{2\int_0^1\int_{\frac{1-\lambda}{2-\lambda}}^1 f(\lambda,\omega)d\lambda d\omega}$, $\hat{\beta} = \frac{\int_0^1\int_0^1 \lambda f(\lambda,\omega)d\lambda d\omega}{2\int_0^1\int_{\frac{1-\lambda}{2-\lambda}}^{\frac{1}{2}} f(\lambda,\omega)d\lambda d\omega}$, *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 $c_v + b_v \geq \frac{1}{2} - r\hat{\beta}$.*

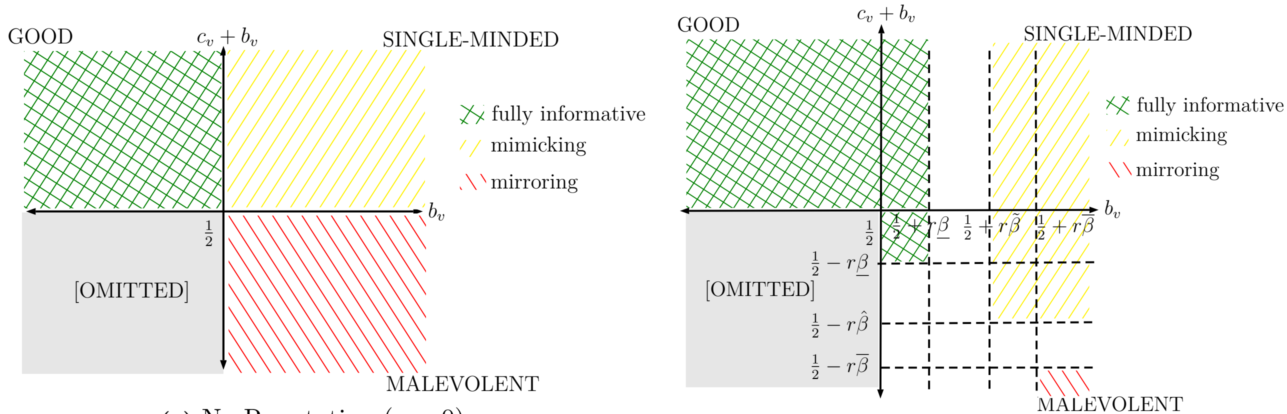
3. *A mirroring equilibrium exists if and only if sender type $u$ is good and sender type $v$ satisfies $b_v \geq \frac{1}{2} + r\overline{\beta}$ and $c_v + b_v \leq \frac{1}{2} - r\overline{\beta}$, and the distribution of receivers' priors satisfies:*

$$\int_0^{\frac{1}{2}}\int_{1-\omega}^1 f(\lambda,\omega)d\lambda d\omega - \int_{\frac{1}{2}}^1\int_0^{1-\omega} f(\lambda,\omega)d\lambda d\omega \geq 0 \geq \int_0^{\frac{1}{2}}\int_0^{\omega} f(\lambda,\omega)d\lambda d\omega - \int_{\frac{1}{2}}^1\int_{1-\omega}^1 f(\lambda,\omega)d\lambda d\omega. \tag{13}$$

Figure 6 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 who reports quasi-truthfully in equilibrium. 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 Propositions 2 and 3 show, a fully informative equilibrium exists when $v$ is good, a mimicking equilibrium exists when $v$ is single-minded, and a mirroring equilibrium can exist 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 7. Several observations follow from comparing the panels in the figure and thus the results of the propositions.[10]

First, a comparison of Panels (a) and (b) of Figure 6 shows that reputation expands the region of sender types who report truthfully, and shrinks the region of senders who mirror. 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

[10]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 6 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.




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

**(b)** Reputation ($r > 0$)

**Figure 6: 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 $\tilde{\beta} < \overline{\beta}$.

occurs when their preference for $a_i = 1$ when $\theta = 0$ is 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 region of sender types who mimic is ambiguous. On the one hand, the region shrinks because single-minded types with sufficiently strong reputation concerns relative to intrinsic preferences no longer mimic, $b_v < 1/2 + r\tilde{\beta}$. On the other hand, the region 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 6 Panel(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 lead sender $v$ to report truthfully. 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 report truthfully when she would not have done so when $r = 0$ (e.g., when a single-minded type reports truthfully instead of mimicking). But if receiver welfare is greater under doublespeak than babbling (per Proposition 4), 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.

# 5 Discussion

## 5.1 Robustness

### 5.1.1 Are heterogeneous priors necessary to support doublespeak equilibria?

Heterogeneous priors are not necessary to support doublespeak equilibria. A mimicking equilibrium is sustainable given any common prior, while a mirroring equilibrium is sustainable as long as the common prior lies in the region of trusting receivers.

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 $p(\mathbf{m}_\infty) = \gamma$. 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, disagreement occurs only if some receivers are unsure. If receivers' priors only differ about the sender's type, disagreement occurs because there are both trusting and distrusting receivers who interpret messages in opposing ways.

### 5.1.2 Is the discrete receiver action space necessary?

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 Propositions 2 and 3 hold.[11]

#### 5.1.3 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.

A mimicking equilibrium still exists in which a good sender type reports quasi-truthfully and a single-minded sender type mimics. Receivers learn that $\theta = 0$ whenever they observe a frequency of $1 - \gamma$ from at least one sender. But when they observe a frequency of $\gamma$ 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 good sender type reports quasi-truthfully and a malevolent sender type mirrors, 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.

#### 5.1.4 What if the sender knows the state?

In our model, the sender is never sure about the state when sending her messages even though the sender's beliefs about the state could become arbitrarily precise as the number of signals becomes arbitrarily large. In contrast, receivers can be sure of the state before

[11] 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 $c_u + b_u \geq 1$), sender $v$ prefers that the action mismatch the state ($b_u \geq 1$ and $c_u + b_u \leq 0$), and the distribution of receiver priors contains enough receivers who believe the sender is likely to be good relative to those who believe the sender is likely to be malevolent. For other combinations of sender types, doublespeak equilibria may also exist depending on the distribution of receivers.

they take actions if the long-run frequency of the $n = \infty$ messages identifies the state. This setting ostensibly gives receivers a favorable information environment.

Our results show that receivers are nevertheless no better off than they would be in an alternative game where sender possesses a significant informational advantage over receiver by knowing the state. In particular, the outcomes of our model where receivers see an infinite number of messages are equivalent to outcomes of a game where the sender knows the state and sends a single message $m \in [0, 1]$ when receivers are uncertain about sender's incentives in both games.[12] In both games, receivers would identify the true state if they were sure that the sender type is good, but a non-good sender type has the incentive to obfuscate the truth through doublespeak.[13] Thus, uncertainty about whether or not the sender is good can prevent receivers from full learning, regardless of how many messages they observe.

## 5.2 Empirical Implications

Propositions 2 and 3 describe sender strategies in mimicking and mirroring equilibria, and Section 2.5 discusses those strategies in real-world contexts. This section discusses empirical implications that do not condition on sender types as researchers may not be able to condition on such types in the data.

### 5.2.1 Who takes what action?

Suppose that an empiricist had cross-sectional data on which receivers took what action after a sender delivered messages that may contain information about the correct course of action. If receiver actions are heterogeneous, then the model suggests we are not in a

[12]For example, the base game in Sobel (1985) is the case in which there is one receiver, each state is equally likely, and a sender who knows the state is either good or malevolent. Sobel (1985) establishes conditions for a mirroring equilibrium that are a special case of Proposition 3. Morgan and Stocken (2003) consider a game where receivers do not know the extent to which a stock analyst favors pumping up a stock price or conveying its true fundamental value. Their setting is similar to the case in our model where the sender type could be single-minded.

[13]Note that our game does not converge to the alternative game in the following sense. Even though sender's knowledge about the state can become arbitrarily precise in our game, she is never completely sure of the state when sending her messages. The reason is that she expects an infinite number of future signals and knows those signals might indicate, however improbably, a state other than what she currently believes.

fully informative equilibrium.[14] Identifying the type of equilibrium then depends on how the receivers differ in their actions, with a particular focus on the behavior of receivers who ex ante believe that the sender is probably not the good type.

Suppose that, upon observing some heterogeneity in actions, an empiricist estimates a statistical 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. If $b_0 \neq 0$ but $b_1, b_2 = 0$, heterogeneity in receiver actions depends only on prior beliefs about the state and not on prior beliefs about sender type, suggesting a babbling equilibrium.

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 Panel (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).

### 5.2.2 Who fact-checks?

The model also makes predictions about who fact-checks in equilibrium, following the intuitions conveyed in Figures 4 and 5. In the cross-section, if some receivers fact-check, the model suggests we are not in a fully informative equilibrium.[15] The behavior of receivers who are fairly ex ante certain of the sender type can further distinguish the type of equilibrium.

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. 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

[14] If all receivers take the same action, the model suggests either a fully informative equilibrium or the outcome of a mimicking equilibrium as per Figure 2 Panel (b).

[15] If no receivers fact check, the model suggests either a fully informative equilibrium or the outcome of a mimicking equilibrium as per Figure 4 Panel (b).

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.[16] 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 more often than receivers who are fairly certain.

# 6 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, even if the sender is being truthful, can be fact-checked, and partially cares about reputation, due to suspicions about the sender's motives. Indeed, suspicions that the sender is malevolent may lead receivers to take the exact opposite action of what a truthful sender desires. Doublespeak is powerful because it may contain complete lies yet obfuscate the truth 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, D., V. Chernozhukov, and M. Yildiz (2016): "Fragility of Asymptotic Agreement under Bayesian Learning," *Theoretical Economics*, 11, 187–227.

Allcott, H. and M. Gentzkow (2017): "Social Media and Fake News in the 2016 Election," *Journal of Economic Perspectives*, 31, 211–236.

Baharudin, H. (2022): "'A baptism of fire': NMPs share their experiences in new book," *The Straits Times*, https://www.straitstimes.com/singapore/a-baptism-of-fire-nmps-share-their-experiences-in-new-book. (accessed September 21, 2022).

Baliga, S., E. Hanany, and P. Klibanoff (2013): "Polarization and Ambiguity," *American Economic Review*, 103, 3071–3083.

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

[16] In this specification, receivers who are ex-ante certain the sender is a good type are the omitted category.

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

Block, M. (2021): "Can The Forces Unleashed By Trump's Big Election Lie Be Undone?" *National Public Radio News*, 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. A. and D. N. Hauser (2021): "Learning with Heterogeneous Misspecified Models: Characterization and Robustness," *Econometrica*, 89, 3025–3077.

Bowen, R., D. Dmitriev, and S. Galperti (2022): "Learning from Shared News: When Abundant Information Leads to Belief Polarization," Working paper.

Bryan-Low, C. and S. McGee (2001): "Enron Short Seller Detected Red Flags in Regulatory Filings," *The Wall Street Journal*.

Cai, J., J. L. Garner, and R. A. Walkling (2009): "Electing Directors," *The Journal of Finance*, 64, 2389–2421.

Chen, C. (2018): "FDA increasingly approves drugs without conclusive proof they work," *PBS NewsHour*, `https://www.pbs.org/newshour/health/fda-increasingly-approves-drugs-without-conclusive-proof-they-work`. (accessed September 29, 2022).

Cheng, I.-H. and A. Hsiaw (2022): "Distrust in Experts and the Origins of Disagreement," *Journal of Economic Theory*, 200, 105401.

Cillizza, C. (2020): "How this Republican official became the most hated man in his party," *CNN*.

Cisternas, G. and J. Vásquez (2020): "Fake News in Social Media: A Supply and Demand Approach," Working paper.

Coles, J. L., N. D. Daniel, and L. Naveen (2014): "Co-opted Boards," *The Review of Financial Studies*, 27, 1751–1796.

Cramer, J. J. (2002): "Pumping Enron," *New York Magazine*.

Crawford, V. P. and J. Sobel (1982): "Strategy Information Transmission," *Econometrica*, 50, 1431–1451.

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

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

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

Dubé, E., C. Laberge, M. Guay, P. Bramadat, R. Roy, and J. A. Bettinger (2013): "Vaccine hesitancy," *Human Vaccines & Immunotherapeutics*, 9, 1763–1773, pMID: 23584253.

Farrell, J. (1993): "Meaning and Credibility in Cheap-Talk Games," *Games and Economic Behavior*, 5, 514–531.

Farrell, J. and M. Rabin (1996): "Cheap Talk," *Journal of Economic Perspectives*, 10, 103–118.

Fiore, K. (2010): "Patients Wary of Doctors' Pharma Relationships," *ABC News*.

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

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

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

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

Gentzkow, M., M. B. Wong, and A. T. Zhang (2021): "Ideological Bias and Trust in Information Sources," Working paper.

Gowda, C. and A. F. Dempsey (2013): "The rise (and fall?) of parental vaccine hesitancy," *Human Vaccines & Immunotherapeutics*, 9, 1755–1762, pMID: 23744504.

Groningen, N. V. (2017): "Big Pharma gives your doctor gifts. Then your doctor gives you Big Pharma's drugs." *The Washington Post*.

Guriev, S. and D. Treisman (2022): *Spin Dictators: The Changing Face of Tyranny in the 21st Century*, Princeton and Oxford: Princeton University Press.

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

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

Hwang, B.-H. and S. Kim (2009): "It pays to have friends," *Journal of Financial Economics*, 93, 138–158.

Jamison, A. M., S. C. Quinn, and V. S. Freimuth (2019): ""You don't trust a government vaccine": Narratives of institutional trust and influenza vaccination among African American and white adults," *Social Science & Medicine*, 221, 87–94.

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

Kang, J.-K., H. Kim, J. Kim, and A. Low (2022): "Activist-Appointed Directors," *Journal of Financial and Quantitative Analysis*, 57, 1343–1376.

Kartik, N., F. X. Lee, and W. Suen (2021): "Information Validates the Prior: A Theorem on Bayesian Updating and Applications," *American Economic Review: Insights*, 3, 165–182.

King, M. (2022): "In Georgia, a G.O.P. Primary Tests the Power of a Trump Vendetta," *The New York Times*.

Kranton, R. and D. McAdams (2020): "Social Networks and the Market for News," Working paper.

Larkin, I., D. Ang, J. Steinhart, M. Chao, M. Patterson, S. Sah, T. Wu, M. Schoenbaum, D. Hutchins, T. Brennan, and G. Loewenstein (2017): "Association Between Academic Medical Center Pharmaceutical Detailing Policies and Physician Prescribing," *JAMA*, 317, 1785–1795.

Larson, H. J., R. M. Clarke, C. Jarrett, E. Eckersberger, Z. Levine, W. S. Schulz, and P. Paterson (2018): "Measuring trust in vaccination: A systematic review," *Human Vaccines & Immunotherapeutics*, 14, 1599–1609.

Lim, J. (2021): "The Big Read: 30 years of NMP scheme - are non-partisan, unelected voices still needed in Parliament?" *CNA*.

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

Montellaro, Z. (2022): "Georgia election official on Trump's enemies list takes his case to MAGA media," *Politico*.

Morgan, J. and P. C. Stocken (2003): "An analysis of stock recommendations," *RAND Journal of Economics*, 34, 183–203.

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

Mullainathan, S. and A. Shleifer (2005): "The Market for News," *The American Economic Review*, 95, 1031–1053.

Myers, S. L. and E. Sullivan (2022): "Disinformation Has Become Another Untouchable Problem in Washington," *New York Times*, available online: https://www.nytimes.com/2022/07/06/business/disinformation-board-dc.html, Last accessed July 2022.

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

Ornstein, C., M. Tigas, and R. G. Jones (2016): "Now There's Proof: Docs Who Get Company Cash Tend to Prescribe More Brand-Name Meds," *ProPublica*.

Ortoleva, P. and E. Snowberg (2015): "Overconfidence in Political Behavior," *American Economic Review*, 105, 504–535.

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

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

Richmond, J., W. Powell, M. Maurer, R. Mangrum, M. R. Gold, E. Pathak-Sen, M. Yang, and K. L. Carman (2017): "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*, 32, 1396–1402.

Savage, L. J. (1954): *The Foundations of Statistics*, New York: Wiley Publishing, reprinted in 1972 by Dover, New York.

Sobel, J. (1985): "A Theory of Credibility," *Review of Economic Studies*, 52, 557–573.

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

Szeidl, A. and F. Szucs (2022): "The Political Economy of Alternative Realities," Working paper.

Trump, D. J. (2021): "Statement by Donald J. Trump, 45th President of the United States of America," 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 $\lambda_i^j$ 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 $(j, \theta)$:

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

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

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

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 $j' \neq j$:

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

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 $\theta$ from the frequency of messages. Likewise, if the receiver can fully identify $\theta$ 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

After observing $p(\mathbf{m}_n)$, a receiver's posterior odds ratios are

$$\frac{P(v,1|p(\mathbf{m}_n))}{P(u,0|p(\mathbf{m}_n))} = \frac{(\rho\gamma)^{n_1}(1-\rho\gamma)^{n-n_1}\omega_i(1-\lambda_i)}{(1-\rho\gamma)^{n_1}(\rho\gamma)^{n-n_1}(1-\omega_i)\lambda_i} = \left(\left(\frac{1-\rho\gamma}{\rho\gamma}\right)^{1-\frac{2n_1}{n}}\right)^n\frac{\omega_i(1-\lambda_i)}{(1-\omega_i)\lambda_i} \tag{A.3}$$

$$\frac{P(v,0|p(\mathbf{m}_n))}{P(u,0|p(\mathbf{m}_n))} = \frac{(\rho\gamma)^{n_1}(1-\rho\gamma)^{n-n_1}(1-\omega_i)(1-\lambda_i)}{(1-\rho\gamma)^{n_1}(\rho\gamma)^{n-n_1}(1-\omega_i)\lambda_i} = \left(\left(\frac{1-\rho\gamma}{\rho\gamma}\right)^{1-\frac{2n_1}{n}}\right)^n\frac{1-\lambda_i}{\lambda_i} \tag{A.4}$$

$$\frac{P(u,1|p(\mathbf{m}_n))}{P(u,0|p(\mathbf{m}_n))} = \frac{(\rho\gamma)^{n_1}(1-\rho\gamma)^{n-n_1}\omega_i\lambda_i}{(1-\rho\gamma)^{n_1}(\rho\gamma)^{n-n_1}(1-\omega_i)\lambda_i} = \left(\left(\frac{1-\rho\gamma}{\rho\gamma}\right)^{1-\frac{2n_1}{n}}\right)^n\frac{\omega_i}{(1-\omega_i)}. \tag{A.5}$$

After observing $p(\mathbf{m}_\infty) = 1-\rho\gamma$, $P(u,0|p(\mathbf{m}_\infty) = 1-\rho\gamma) = 1$ so all receivers choose $a_i(\mathbf{m}_\infty|p(\mathbf{m}_\infty) = 1-\rho\gamma) = 0$.

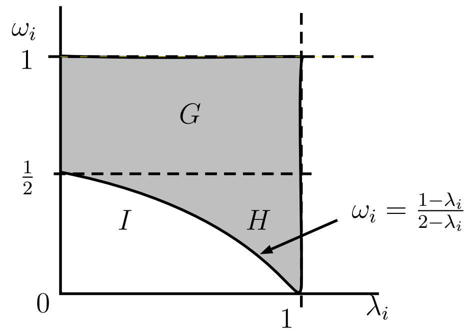



**Figure A1:** Partition of receivers' prior beliefs $(\lambda_i, \omega_i)$. Mimicking equilibrium: If $p(\mathbf{m}_\infty) = \rho\gamma$, receivers whose priors lie in the gray area choose $a_i = 1$ in response to the observed frequencies.

| Strategy | $P(m_n = 1 \vert s_n = 0)$ | $P(m_n = 1 \vert s_n = 1)$ | $p_{0j}$ | $p_{0j}$ |
|---|---|---|---|---|
| *Quasi-Truthful* | $0$ | $\rho$ | $1 - \rho\gamma$ | $\rho\gamma$ |
| *Mirror* | $\rho$ | $0$ | $\rho\gamma$ | $1 - \rho\gamma$ |
| *Ones* | $\rho\gamma$ | $\rho\gamma$ | $\rho\gamma$ | $\rho\gamma$ |
| *Zeros* | $1 - \rho\gamma$ | $1 - \rho\gamma$ | $1 - \rho\gamma$ | $1 - \rho\gamma$ |

**Table A1:** Strategies that generate plausible frequencies.

After observing $p(\mathbf{m}_\infty) = \rho\gamma$, receiver $i$'s posterior beliefs are $P(u, 0|p(\mathbf{m}_\infty) = \rho\gamma) = 0$ and

$$P(u, 1|p(\mathbf{m}_\infty) = \rho\gamma) = \frac{\omega_i\lambda_i}{\omega_i\lambda_i + \omega_i(1-\lambda_i) + (1-\omega_i)(1-\lambda_i)} \tag{A.6}$$

$$P(v, 1|p(\mathbf{m}_\infty) = \rho\gamma) = \frac{\omega_i(1-\lambda_i)}{\omega_i\lambda_i + \omega_i(1-\lambda_i) + (1-\omega_i)(1-\lambda_i)} \tag{A.7}$$

$$P(v, 0|p(\mathbf{m}_\infty) = \rho\gamma) = \frac{(1-\omega_i)(1-\lambda_i)}{\omega_i\lambda_i + \omega_i(1-\lambda_i) + (1-\omega_i)(1-\lambda_i)}, \tag{A.8}$$

so she chooses $a_i = 1$ if and only if $P(\theta = 1|p(\mathbf{m}_n) = \rho\gamma) > P(\theta = 0|p(\mathbf{m}_n) = \rho\gamma)$, which holds if and only if $\omega_i > \frac{1-\lambda_i}{2-\lambda_i}$. Figure A1 shows receivers' actions when they observe $p(\mathbf{m}_\infty) = \rho\gamma$. Receivers whose priors lie in area $G + H$ choose $a_i = 1$, and those whose priors lie in area $I$ choose $a_i = 0$. Let $G + H = \int_0^1 \int_{\frac{1-\lambda}{2-\lambda}}^1 f(\lambda, \omega) d\omega d\lambda$ and $I = \int_0^1 \int_0^{\frac{1-\lambda}{2-\lambda}} f(\lambda, \omega) d\omega d\lambda$.

Each sender type will play the proposed equilibrium strategy if there is no incentive to deviate to any other strategies. Although there is an arbitrarily large set of possible messaging strategies, we can partition message strategies into five classes based on their long-run frequencies, which are a sufficient statistic to determine receivers' posterior beliefs. Table A1 summarizes the four classes of strategies that generate plausible long-run frequencies, i.e. long-run frequencies that could be observed in equilibrium. The fifth class is strategies that receivers can identify as an off-equilibrium strategy, which are strategies that produce long-run frequencies that are not plausible or strategies that do not produce well-defined long-run frequencies. We first consider deviations to the four classes of strategies that generate plausible long-run frequencies, then consider deviations to the fifth class.

In general, a sender's payoff from a given strategy is

$$\omega^S[-\int_0^1\int_0^1(\mathbf{1}_{a_i=1}-c-b)^2 f(\lambda,\omega)d\lambda d\omega]+(1-\omega^S)[-\int_0^1\int_0^1(\mathbf{1}_{a_i=1}-0-b)^2 f(\lambda,\omega)d\lambda d\omega]. \tag{A.9}$$

Thus we can use Equation A.9 to determine a sender's payoff from a given messaging strategy by substituting in the relevant masses of receivers who choose $a_i = 1$ and $a_i = 0$ in response to the messages they observe. We use Table A1 and Equation B.55 to fill in Equation A.9 for each of the four classes of messaging strategies, since they generate plausible frequencies. Given the receivers' beliefs within a mirroring equilibrium, this yields the sender's payoffs for each of the first four classes:

1. *Quasi-Truthful*: $\omega^S[-(1-c-b)^2(G+H)-(0-c-b)^2(I)]+(1-\omega^S)[-(1-b)^2(0)-(0-b)^2(G+H+I)]$
2. *Mirror*: $\omega^S[-(1-c-b)^2(0)-(0-c-b)^2(G+H+I)]+(1-\omega^S)[-(1-b)^2(G+H)-(0-b)^2(I)]$
3. *Ones*: $\omega^S[-(1-c-b)^2(G+H)-(0-c-b)^2(I)]+(1-\omega^S)[-(1-b)^2(G+H)-(0-b)^2(I)]$
4. *Zeros*: $\omega^S[-(1-c-b)^2(0)-(0-c-b)^2(G+H+I)]+(1-\omega^S)[-(1-b)^2(0)-(0-b)^2(G+H+I)]$.

Using the sender's payoffs, sender $u$'s preferences must satisfy the following to avoid deviating from *Quasi-Truthful* to the other strategies that generate frequencies consistent with equilibrium:

$$(G+H)\left(\omega^S(-1+2c_u+2b_u)+(1-\omega^S)(1-2b_u)\right)\geq 0 \tag{A.10}$$
$$(G+H)\omega^S(-1+2c_u+2b_u)\geq 0 \tag{A.11}$$
$$(G+H)(1-\omega^S)(1-2b_u)\geq 0, \tag{A.12}$$

which together imply that $b_u \leq 1/2$ and $c_u + b_u \geq 1/2$. Sender $v$'s preferences must satisfy the following:

$$(G+H)(1-\omega^S)(-1+2b_v)\geq 0 \tag{A.13}$$
$$(G+H)\omega^S(-1+2c_v+2b_v)\geq 0 \tag{A.14}$$
$$(G+H)\left(\omega^S(-1+2c_v+2b_v)+(1-\omega^S)(-1+2b_v)\right)\geq 0, \tag{A.15}$$

which together imply that $b_v \geq 1/2$ and $c_v + b_v \geq 1/2$.

To consider each sender type's incentive to deviate to receivers can clearly identify as an off-equilibrium strategy, we apply the refinement of neologism-proofness (Farrell, 1993). First, sender $v$ has no incentive to reveal her type through an off-equilibrium deviation because receivers would best respond to such a revelation by taking actions according to their priors. This is strictly worse for sender $v$ than using her mimicking strategy. Second, suppose receivers believe the type is sender $u$ after an off-equilibrium deviation $p'(\mathbf{m}_\infty) \neq 1/2$, so they would all choose $a_i = 1$ if they observe $p' > 1/2$ and $a_i = 0$ if they observe $p' > 1/2$. Then sender $v$ would prefer to deviate by always sending $p' > 1/2$ to induce all receivers to choose $a_i = 1$. Thus such a deviation is not a credible neologism. Third, sender $u$ has no incentive to deviate if $\theta = 0$, so receivers should infer that an off-equilibrium deviation is intended to induce $a_i = 1$. If receivers believe $(j, \theta) \in \{(u, 1), (v, 0), (v, 1)\}$ after observing an out-of-equilibrium frequency or a sequence of messages that does not have a long-run frequency, then neither sender benefits from such a deviation. Finally, suppose receivers believe the messages convey no information about the state when they observe $p(\mathbf{m}_\infty) = 1/2$ or a sequence of messages that does not have a long-run frequency, so they take actions according to their priors. Sender $u$ weakly prefers reporting truthfully to this deviation when $H\omega^S(-1 + 2c_u + 2b_u) + G(1 - \omega^S)(1 - 2b_u) \geq 0$, and sender $v$ weakly prefers her mimicking strategy to this deviation when $H\left(\omega^S(-1 + 2c_v + 2b_v) + (1 - \omega^S)(-1 + 2b_v)\right) \geq 0$. These conditions are satisfied when $b_u \leq 1/2$, $c_u + b_u \geq 1/2$ and $b_v \geq 1/2$, $c_v + b_v \geq 1/2$, respectively. Thus the proposed mimicking equilibrium is neologism-proof if (1) sender $u$'s preferences satisfy $b_u \leq 1/2$ and $c_u + b_u \geq 1/2$ and (2) sender $v$'s preferences satisfy $b_v \geq 1/2$ and $c_v + b_v \geq 1/2$.

## A.3 Proof of Proposition 3

Given the sender's equilibrium strategy, a receiver $i$'s posterior odds ratios are

$$\frac{P(u,0|p(\mathbf{m}_n))}{P(u,1|p(\mathbf{m}_n))} = \frac{(1-\rho\gamma)^{n_1}(\rho\gamma)^{n-n_1}(1-\omega_i)\lambda_i}{(\rho\gamma)^{n_1}(1-\rho\gamma)^{n-n_1}\omega_i\lambda_i} = \left(\left(\frac{1-\rho\gamma}{\rho\gamma}\right)^{\frac{2n_1}{n}-1}\right)^n \left(\frac{1-\omega_i}{\omega_i}\right) \tag{A.16}$$

$$\frac{P(v,1|p(\mathbf{m}_n))}{P(u,1|p(\mathbf{m}_n))} = \frac{(1-\rho\gamma)^{n_1}(\rho\gamma)^{n-n_1}\omega_i(1-\lambda_i)}{(\rho\gamma)^{n_1}(1-\rho\gamma)^{n-n_1}\omega_i\lambda_i} = \left(\left(\frac{1-\rho\gamma}{\rho\gamma}\right)^{\frac{2n_1}{n}-1}\right)^n \left(\frac{1-\lambda_i}{\lambda_i}\right) \tag{A.17}$$

$$\frac{P(v,0|p(\mathbf{m}_n))}{P(u,1|p(\mathbf{m}_n))} = \frac{(\rho\gamma)^{n_1}(1-\rho\gamma)^{n-n_1}(1-\omega_i)(1-\lambda_i)}{(\rho\gamma)^{n_1}(1-\rho\gamma)^{n-n_1}\omega_i\lambda_i} = \frac{(1-\omega_i)(1-\lambda_i)}{\omega_i\lambda_i}. \tag{A.18}$$

Thus receiver $i$'s posterior beliefs after observing $p(\mathbf{m}_\infty) = \rho\gamma$ are $P(u,0|p(\mathbf{m}_\infty) = \rho\gamma) = P(v,1|p(\mathbf{m}_\infty) = \rho\gamma) = 0$ and:

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

$$P(v,0|p(\mathbf{m}_\infty) = \rho\gamma) = \frac{(1-\omega_i)(1-\lambda_i)}{\omega_i\lambda_i + (1-\omega_i)(1-\lambda_i)}. \tag{A.20}$$

Receiver $i$ chooses actions $a_i(\mathbf{m}_\infty|p(\mathbf{m}_\infty) = \rho\gamma) = 1$ if $\omega_i > 1 - \lambda_i$, $a_i(\mathbf{m}_\infty|p(\mathbf{m}_\infty) = \rho\gamma) = 0$ if $\omega_i < 1 - \lambda_i$, and randomizes between actions with equal probability if $\omega_i = 1 - \lambda_i$.

Likewise, receiver $i$'s posterior beliefs after observing $p(\mathbf{m}_n) = 1 - \rho\gamma$ are $P(u,1|p(\mathbf{m}_\infty) = 1 - \rho\gamma) = P(v,0|p(\mathbf{m}_\infty) = 1 - \rho\gamma) = 0$ and:

$$P(u,0|p(\mathbf{m}_\infty) = 1-\rho\gamma) = \frac{(1-\omega_i)\lambda_i}{(1-\omega_i)\lambda_i + \omega_i(1-\lambda_i)} \tag{A.21}$$

$$P(v,1|p(\mathbf{m}_\infty) = 1-\rho\gamma) = \frac{\omega_i(1-\lambda_i)}{(1-\omega_i)\lambda_i + \omega_i(1-\lambda_i)}, \tag{A.22}$$

and she chooses actions $a_i(\mathbf{m}_\infty|p(\mathbf{m}_\infty) = 1 - \rho\gamma) = 1$ if $\omega_i > \lambda_i$, $a_i(\mathbf{m}_\infty|p(\mathbf{m}_\infty) = 1 - \rho\gamma) = 0$ if $\omega_i < \lambda_i$, and randomizes between actions with equal probability if $\omega_i = \lambda_i$.

We use Table A1 and receivers' actions to fill in Equation A.9 for each of the first four classes of messaging strategies, since they generate frequencies that receivers would expect to see in equilibrium. Given the receivers' beliefs within a mirroring equilibrium, this yields the sender's payoffs for each of the first four classes:

1. *Quasi-Truthful*: $\omega^S[-(1-c-b)^2(E+C+D)-(-c-b)^2(A+B+F)]+(1-\omega^S)[-(1-b)^2(A+B+E)-(-b)^2(C+D+F)]$

2. *Mirror*: $\omega^S[-(1-c-b)^2(A+B+E)-(-c-b)^2(C+D+F)]+(1-\omega^S)[-(1-b)^2(C+D+E)-(-b)^2(A+B+F)]$

3. *Ones*: $\omega^S[-(1-c-b)^2(E+C+D)-(-c-b)^2(A+B+F)]+(1-\omega^S)[-(1-b)^2(E+C+D)-(-b)^2(A+B+F)]$

4. *Zeros*: $\omega^S[-(1-c-b)^2(A+B+E)-(-c-b)^2(C+D+F)]+(1-\omega^S)[-(1-b)^2(A+B+E)-(-b)^2(C+D+F)]$.

Thus, Equations A.23, A.24, and A.25 must hold for sender 1 to avoid deviating from Truthful to the other strategies that generate frequencies consistent with equilibrium:

$$(C+D-A-B)\left(\omega^S(-1+2c_u+2b_u)+(1-\omega^S)(1-2b_u)\right)\geq 0 \tag{A.23}$$
$$-(C+D-A-B)(1-\omega^S)(-1+2b_u)\geq 0 \tag{A.24}$$
$$-(C+D-A-B)\omega^S(1-2c_u-2b_u)\geq 0. \tag{A.25}$$

First, suppose $C+D\geq A+B$. Then Equation A.24 implies that $b_u\leq 1/2$. Equation A.25 implies that $c_u+b_u\geq\frac{1}{2}$. Since $b_u\leq 1/2$, then this implies $c_u\geq 0$. Equation A.23 is automatically satisfied when Equations A.24 and A.25 are satisfied.

Analogously, sender $v$ mirrors if there is no incentive to deviate to *Quasi-Truthful*, *Ones*, or *Zeros* (for brevity, details not shown). When $C+D>A+B$, then this implies that $b_v\leq 1/2$ and $c_v+b_v\geq 1/2$.

Suppose receivers believe the type is sender $u$ after an off-equilibrium deviation $p'(\mathbf{m}_\infty)\neq 1/2$, so they would all choose $a_i=1$ if they observe $p'>1/2$ and $a_i=0$ if they observe $p'<1/2$. Then sender $v$ would also prefer to deviate by mirroring sender $u$'s deviation. Thus such a deviation is not a credible neologism because receivers should not believe the type is sender $u$, and there is no neologism-proof equilibrium in which receivers believe the type is sender $u$ after such an off-equilibrium deviation. Because sender $v$ has no incentive to reveal her type, there is no neologism-proof equilibrium in which receivers are sure the type is sender $v$ after an off-equilibrium deviation. Neither sender benefits from such a deviation if receivers believe $(j,\theta)\in\{(u,1),(v,0)\}$ after observing $p'>1/2$ or a sequence of messages that does not have a long-run frequency. Likewise neither sender benefits from such a deviation if receivers believe $(j,\theta)\in\{(u,0),(v,1)\}$ after observing $p'<1/2$ or a sequence of messages that does not have a long-run frequency.

Suppose receivers believe that the messages convey no information about the state when they observe $p(\mathbf{m}_\infty)=1/2$ or a sequence of messages that does not have a long-run frequency. Then they take actions according to their priors: those whose priors lie in area $A+C+E$ choose $a_i=1$ and those in $B+F+D$ choose $a_i=0$. Senders $u$ and $v$ will not deviate to $p(\mathbf{m}_\infty)=1/2$ if and only if Equations A.26 and A.27 are satisfied:

$$\omega^S(D-A)(-1+2c_u+2b_u)+(1-\omega^S)(B-C)(-1+2b_u)\geq 0. \tag{A.26}$$
$$\omega^S(B-C)(-1+2c_v+2b_v)+(1-\omega^S)(D-A)(-1+2b_v)\geq 0. \tag{A.27}$$

Equations A.26 and A.27 are satisfied for all $\omega^S$ if and only if $D-A\geq 0\geq B-C$. Otherwise, there exist values of $\omega^S$ such that either sender $u$ or $v$ prefer to deviate (i.e., when their private information $\mathbf{s}_n$ sufficiently indicates $\theta$). Formally, $D - A \geq 0 \geq B - C$ is $\int_0^{\frac{1}{2}} \int_{1-\omega}^1 f(\lambda, \omega) d\lambda d \; - \int_{\frac{1}{2}}^1 \int_0^{1-\omega} f(\lambda, \omega) d\lambda d\omega \geq 0 \geq \int_0^{\frac{1}{2}} \int_0^{\omega} f(\lambda, \omega) d\lambda d\omega - \int_{\frac{1}{2}}^1 \int_{1-\omega}^1 f(\lambda, \omega) d\lambda d\omega$.

Thus the proposed mirroring equilibrium is neologism-proof if (1) sender $u$'s preferences satisfy $b_u \leq 1/2$ and $c_u + b_u \geq \frac{1}{2}$, (2) sender $v$'s preferences satisfy $b_v \geq 1/2$ and $c_v + b_v \leq \frac{1}{2}$, and (3) $\int_0^{\frac{1}{2}} \int_{1-\omega}^1 f(\lambda, \omega) d\lambda d \; - \int_{\frac{1}{2}}^1 \int_0^{1-\omega} f(\lambda, \omega) d\lambda d\omega \geq 0 \geq \int_0^{\frac{1}{2}} \int_0^{\omega} f(\lambda, \omega) d\lambda d\omega - \int_{\frac{1}{2}}^1 \int_{1-\omega}^1 f(\lambda, \omega) d\lambda d\omega$. Analogously, if $C + D < A + B$, then these conditions apply with all inequalities reversed.

## A.4 Proof of Proposition 4

In the fully informative equilibrium, all receivers learn the state and chooses the correct action. Thus they are better off in the fully informative equilibrium than in babbling equilibrium.

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 $\theta = 1$. Using Figure 2, Table A2 shows which receivers choose the correct action in mimicking versus babbling equilibria, . Thus the mimicking equilibrium is worse for receivers than the babbling equilibrium if and only if $\hat{\lambda}(1 - \hat{\omega})(G + H) + H(-1 + 2\hat{\omega}) < 0$, which requires $\hat{\omega} < 1/2$ and $\hat{\lambda} < \frac{H(1-2\hat{\omega})}{(1-\omega)(G+H)}$ where $H = \int_0^1 \int_{\frac{1-\lambda}{2-\lambda}}^{\frac{1}{2}} f(\lambda, \omega) d\omega d\lambda$ and $G + H = \int_0^1 \int_{\frac{1-\lambda}{2-\lambda}}^1 f(\lambda, \omega) d\omega d\lambda$. Otherwise, the mimicking equilibrium is better for receivers than the babbling equilibrium.

| Sender Type ($j$) | State ($\theta$) | Mimicking | Babbling |
|---|---|---|---|
| $u$ | 0 | $G + H + I$ | $H + I$ |
| | 1 | $G + H$ | $G$ |
| $v$ | 0 | $I$ | $H + I$ |
| | 1 | $G + H$ | $G$ |

**Table A2:** Receivers who choose the correct action given sender type and state. Comparing mimicking and babbling equilibria.

Table A3 shows which receivers choose the correct action in mirroring versus babbling equilibria, using Figure 3. Thus the mirroring equilibrium is worse for receivers than the babbling equilibrium if and only if $(D - A)(1 - \hat{\omega})(2\hat{\lambda} - 1) + (C - B)(\hat{\omega})(2\hat{\lambda} - 1) < 0$, which holds only and only if $\hat{\lambda} < 1/2$ since $D - A \geq 0 \geq B - C$.

| Sender Type ($j$) | State ($\theta$) | Mirroring | Babbling |
|---|---|---|---|
| $u$ | 0 | $C + D + F$ | $B + D + F$ |
| | 1 | $C + D + E$ | $A + C + E$ |
| $v$ | 0 | $A + B + F$ | $B + D + F$ |
| | 1 | $A + B + E$ | $A + C + E$ |

**Table A3:** Receivers who choose the correct action given sender type and state. Comparing mirroring and babbling equilibria, where $D - A \geq 0 \geq B - C$.

## A.5 Proof of Proposition 5

Obviously, all receivers fact-check in equilibrium and learn $\theta$ 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 $\theta$ 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 $-(1 - \mu_i)$:

$$\mu_i[-(1-1)^2] + (1-\mu_i)[-(1-0)^2] = -(1-\mu_i). \tag{A.28}$$

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

$$\mu_i[-(1-0)^2] + (1-\mu_i)[-(0-0)^2] = -\mu_i. \tag{A.29}$$

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 Propositions 2 and 3, we provide full details in the Online Appendix.

## A.6 Proof of Proposition 7

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 otherwise analogous to that of Propositions 2 and 3, we provide the sender's payoffs from each of the strategies that generate plausible frequencies and leave full details of equilibrium construction in the Online Appendix.

### A.6.1 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 truthful, but only $u$ is good. Using the partitions denoted in Figure 2, The sender's payoffs from each of the strategies that generate plausible frequencies are:

1. *Quasi-Truthful*: $\omega^S[-(1-c-b)^2(G+H+I) - (0-c-b)^2(0) + r\int_0^1\int_0^1 \lambda f(\lambda,\omega)d\lambda d\omega] + (1-\omega^S)[-(1-b)^2(0) - (0-b)^2(G+H+I) + r\int_0^1\int_0^1 \lambda f(\lambda,\omega)d\lambda d\omega]$.

   Given the equilibrium strategies, receivers learn nothing about the sender's type when they learn that the message content matches the state: $P_i(j = u|p(\mathbf{m}_\infty) = \rho\gamma, \theta = 1) = \lambda_i$ and $P_i(j = u|p(\mathbf{m}_\infty) = 1-\rho\gamma, \theta = 0) = \lambda_i$.

2. *Ones*: $\omega^S[-(1-c-b)^2(G+H+I) - (0-c-b)^2(0) + r\int_0^1\int_0^1 \lambda f(\lambda,\omega)d\lambda d\omega] + (1-\omega^S)[-(1-b)^2(G+H+I) - (0-b)^2(0) + r(0)]$

   Given the equilibrium strategies, receivers learn nothing about the sender's type when they learn that the message content matches the state. But since there is no incentive for the good type to deviate from reporting truthfully, receivers should believe that only a non-good type would report message content that does not match the state: $P_i(j = u|p(\mathbf{m}_\infty) = \rho\gamma, \theta = 0) = 0$.

3. *Zeros*: $\omega^S[-(1-c-b)^2(0) - (0-c-b)^2(G+H+I) + r(0)] + (1-\omega^S)[-(1-b)^2(0) - (0-b)^2(G+H+I) + r\int_0^1\int_0^1 \lambda f(\lambda,\omega)d\lambda d\omega]$

4. *Mirror*: $\omega^S[-(1-c-b)^2(0) - (0-c-b)^2(G+H+I) + r(0)] + (1-\omega^S)[-(1-b)^2(G+H+I) - (0-b)^2(0) + r(0)]$.

### A.6.2 Mimicking Equilibrium

Suppose sender $u$ is good and reports truthfully, and sender $v$ mimics. Using the partitions denoted in Figure 2, the sender's payoffs from each of the strategies that generate plausible frequencies are:

1. *Quasi-Truthful*: $\omega^S[-(1-c-b)^2(G+H) - (0-c-b)^2(I) + r\int_0^1\int_0^1 \lambda f(\lambda,\omega)d\lambda d\omega] + (1-\omega^S)[-(1-b)^2(0) - (0-b)^2(G+H+I) + r(1)]$

   Given the equilibrium strategies, receivers are sure that the sender is good when $p(\mathbf{m}_\infty) = 1-\rho\gamma$ and $\theta = 0$: $P_i(j = u|p(\mathbf{m}_\infty) = 1-\rho\gamma, \theta = 0) = 1$. When $p(\mathbf{m}_\infty) = \rho\gamma$ and $\theta = 1$, receivers cannot identify the sender's type and they are back to their priors on type, so $P_i(j = u|p(\mathbf{m}_\infty = \rho\gamma) = \lambda_i$.

2. *Ones*: $\omega^S[-(1-c-b)^2(G+H) - (0-c-b)^2(I) + r\int_0^1\int_0^1 \lambda f(\lambda,\omega)d\lambda d\omega] + (1-\omega^S)[-(1-b)^2(G+H) - (0-b)^2(I) + r(0)]$

   Given the equilibrium strategies, receivers learn nothing about the sender's type when they learn that the message content matches the state. Since there is no incentive for the good type to deviate from reporting truthfully, receivers should believe that only a non-good type would report message content that does not match the state: $P_i(j = u|p(\mathbf{m}_\infty) = \rho\gamma, \theta = 0) = 0$.

3. *Zeros*: $\omega^S[-(1-c-b)^2(0) - (0-c-b)^2(G+H+I) + r(0)] + (1-\omega^S)[-(1-b)^2(0) - (0-b)^2(G+H+I) + r(1)]$

4. *Mirror*: $\omega^S[-(1-c-b)^2(0) - (0-c-b)^2(G+H+I) + r(0)] + (1-\omega^S)[-(1-b)^2(G+H) - (0-b)^2(I) + r(0)]$.

### A.6.3 Mirroring Equilibrium

Suppose sender $u$ is good and reports truthfully, and sender $v$ mirrors. Given the equilibrium strategies, receivers learn both the state and the type after the state is realized: $P_i(j = u|p(\mathbf{m}_\infty) = 1-\rho\gamma, \theta = 0) = P_i(j = u|p(\mathbf{m}_\infty) = \rho\gamma, \theta = 1) = 1$ and $P_i(j = v|p(\mathbf{m}_\infty) = 1-\rho\gamma, \theta = 1) = P_i(j = v|p(\mathbf{m}_\infty) = \rho\gamma, \theta = 0) = 1$.

Using the partitions denoted in Figure 3, the sender's payoffs from each of the strategies that generate plausible frequencies are:

1. *Quasi-Truthful*: $\omega^S[-(1-c-b)^2(E+C+D) - (0-c-b)^2(A+B+F) + r(1)] + (1-\omega^S)[-(1-b)^2(A+B+E) - (0-b)^2(C+D+F) - r(1)] \geq 0$

2. *Ones*: $\omega^S[-(1-c-b)^2(E+C+D) - (0-c-b)^2(A+B+F) + r(1)] + (1-\omega^S)[-(1-b)^2(E+C+D) - (0-b)^2(A+B+F) - r(1)] \geq 0$

3. *Zeros*: $\omega^S[-(1-c-b)^2(A+B+E)-(0-c-b)^2(C+D+F)+r(1)]+(1-\omega^S)[-(1-b)^2(A+B+E)-(0-b)^2(C+D+F)-r(1)] \geq 0$

4. *Mirror*: $\omega^S[-(1-c-b)^2(A+B+E)-(0-c-b)^2(C+D+F)+r(0)]+(1-\omega^S)[-(1-b)^2(C+D+E)-(0-b)^2(A+B+F)-r(0)] \geq 0.$

# Appendix B Online Appendix, "Bayesian Doublespeak"

## B.1 Fully Informative Equilibrium

In a fully informative equilibrium, all receivers choose $a_i = 1$ when $\theta = 1$ and $a_i = 0$ when $\theta = 0$. Thus each sender does not deviate from her equilibrium messaging strategy to the other strategies that generate frequencies consistent with equilibrium if the following conditions hold:

$$(A+B+C+D+E+F)\left(\omega^S(-1+2c_j+2b_j)+(1-\omega^S)(1-2b_j)\right) \geq 0 \tag{B.1}$$

$$(A+B+C+D+E+F)(1-\omega^S)(1-2b_j) \geq 0 \tag{B.2}$$

$$(A+B+C+D+E+F)\omega^S(-1+2c_j+2b_j) \geq 0. \tag{B.3}$$

Equation B.2 becomes the requirement that $b_j \leq 1/2$. Equation B.3 becomes $c_j + b_j \geq \frac{1}{2}$, which implies $c_j \geq 0$ since $b_j \leq 1/2$. Equation B.1 is satisfied when Equations B.2 and B.3 hold. Sender $j$ does not deviate to an off-equilibrium strategy that produces $p(\mathbf{m}_\infty) = 1/2$ if $\omega^S(B+D+F)(-1+2c_j+2b_j)+(1-\omega^S)(A+E+C)(1-2b_j) \geq 0$, which is also satisfied when Equations B.2 and B.3 hold. Thus, both senders must have $b_j \leq 1/2$ and $c_j + b_j \geq \frac{1}{2}$ to maintain their equilibrium strategies. A fully informative equilibrium in which $b_j \leq 1/2$ and $c_j + b_j \geq \frac{1}{2}$ for $j = 1, 2$ is neologism-proof, since neither sender has an incentive to deviate to an off-equilibrium messaging strategy.

## B.2 Equilibrium with one uninformative sender

Consider an equilibrium in which sender $v$ is identifiable from the observed frequency of messages but is clearly uninformative. This implies that $p_{1v} = p_{0v} \neq p_{1u} \neq p_{0u}$. Thus, all receivers learn the state when they observe $p(\mathbf{m}_\infty) = p_{1u}$ or $p(\mathbf{m}_\infty) = p_{0u}$: $P(u, 1|p(\mathbf{m}_\infty) = p_{1u}) = 1$ and $a_i(\lambda_i, \omega_i|p(\mathbf{m}_\infty) = p_{1u}) = 1$ for all $i$, $P(u, 0|p(\mathbf{m}_\infty) = p_{0u}) = 1$ and $a_i(\lambda_i, \omega_i|p(\mathbf{m}_\infty) = p_{0u}) = 0$. If receivers observe $p(\mathbf{m}_\infty) = p_{1v}$, they know that $j = v$ but cannot identify the state. Thus $P(1|p(\mathbf{m}_\infty) = p_{1v}) = \omega_i$ and

$$a_i(\mathbf{m}_\infty|p(\mathbf{m}_\infty) = p_{1v}) = \begin{cases} 1 & \text{if } \omega_i > 1/2 \\ 0 & \text{if } \omega_i < 1/2 \\ (0 \text{ w.p. } \frac{1}{2}; 1 \text{ w.p. } \frac{1}{2}) & \text{if } \omega_i = 1/2. \end{cases} \tag{B.4}$$

Using the partition of receivers' priors from Figure 3, sender 2 would prefer to be clearly uninformative over being truthful, mirroring sender 1, mimicking sender 1's frequency in state $\theta = 1$, and mimicking sender 1's frequency in state $\theta = 0$ if

$$(B+F+D)\omega^S[(1-c_v-b_v)^2 - (-c_v-b_v)^2] + (A+E+C)(1-\omega^S)[-(1-b_v)^2 + (-b_v)^2] \geq 0 \tag{B.5}$$

$$(A+E+C)\omega^S[-(1-c_v-b_v)^2 + (-c_v-b_v)^2] + (B+F+D)(1-\omega^S)[(1-b_v)^2 - (-b_v)^2] \geq 0 \tag{B.6}$$

$$(B+F+D)\left(\omega^S[(1-c_v-b_v)^2 - (-c_v-b_v)^2] + (1-\omega^S)[(1-b_v)^2 - (-b_v)^2]\right) \geq 0 \tag{B.7}$$

$$(A+E+C)\left(\omega^S[-(1-c_v-b_v)^2 + (-c_v-b_v)^2] + (1-\omega^S)[-(1-b_v)^2 + (-b_v)^2]\right) \geq 0, \tag{B.8}$$

which only holds if $c_v = 0$ and $b_v = 1/2$. Thus such an equilibrium can only exist in the knife-edge case in which sender 2 is indifferent about receivers' actions.

## B.3 Details for Proof of Proposition 5

### B.3.1 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 $p(\mathbf{m}_\infty) \in \{1-\gamma, \gamma\}$. This implies that the conditions for sender types to report truthfully rather that deviate to strategies that still generate $p(\mathbf{m}_\infty) \in \{1-\gamma, \gamma\}$ 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 $j$.

Suppose the sender deviates to a strategy that generates off-equilibrium frequencies $p(\mathbf{m}_\infty) \notin \{1-\gamma, \gamma\}$. If both types are good ($b_j \leq 1/2$ and $c_j + b_j \geq 1/2$ for all $j$), such a deviation would not occur because they would have no incentive to deviate to an off-equilibrium frequency. Suppose there is an fully informative equilibrium with a good type and a non-good type. Receivers would infer that the deviation cannot come from the good type and take actions based on their priors because $\mu_i = P_i(\theta = 1 | p(\mathbf{m}_\infty) \notin \{1-\gamma, \gamma\}) = \omega_i$. By Lemma 1, for receivers with $\omega_i \geq 1/2$, they will fact-check an off-equilibrium deviation when $\omega_i \leq 1-\phi$ and not fact-check when $\omega_i > 1-\phi$. For receivers with $\omega_i < 1/2$, they will fact-check when $\omega_i \geq \phi$ and not fact-check when $\omega_i < \phi$.

Figure B2 shows receivers' actions in response to an off-equilibrium frequency in the fully informative equilibrium. Receivers whose priors lie in area $A$ do not fact-check and choose $a_i = 1$. Receivers whose priors lie in area $D$ do not fact-check and choose $a_i = 0$. Receivers in area $B+C$ fact-check and choose actions that match the state.

Anticipating this fact-checking, a sender's payoff from deviating to an off-equilibrium frequency is $\omega^S[-(1-c_j-b_j)^2(A+B+C) - (0-c_j-b_j)^2(D)] + (1-\omega^S)[-(1-b_j)^2(A) - (0-b_j)^2(B+C+D)] \geq 0$. Since all receivers choose the action that matches the state in equilibrium, a sender does not deviate to an off-equilibrium frequency if $D\omega^S(-1+2c_j+2b_j) + A(1-\omega^S)(1-2b_j) \geq 0$, which is satisfied when $b_j \leq 1/2$ and $c_j + b_j \geq 1/2$ for all $j$. In the fully informative equilibrium, no receivers fact-check because they know long-run frequency identifies the state. Intuitively, no receivers fact-check anything plausible in the equilibrium, and therefore there is always an incentive for a non-good type to deviate. Thus, endogenous fact-checking does not affect the sender types required to sustain fully informative equilibrium.

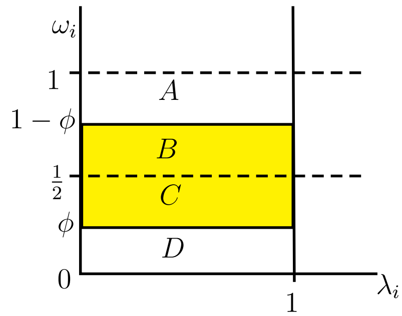



**Figure B2:** Fact-checking when receivers would otherwise use priors to take actions

### B.3.2 Mimicking Equilibrium

Within the mimicking equilibrium, all receivers are sure that $\theta = 0$ if $p(\mathbf{m}_\infty) = 1-\rho\gamma$: $\mu_i(p(\mathbf{m}_\infty) = 1-\rho\gamma) = 0$. Thus their expected payoff from not fact-checking is 0, and no receiver fact-checks if they observe $p(\mathbf{m}_\infty) = 1-\rho\gamma$.

If $p(\mathbf{m}_\infty) = \rho\gamma$, receivers are not sure of the state, and $\mu_i = P_i(\theta = 1|p(\mathbf{m}_\infty) = \rho\gamma) = \frac{\omega_i}{\omega_i+(1-\omega_i)(1-\lambda_i)}$. 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 areas $R+U+Y$ 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 areas $T+X$ in Figure B3. Figure B3 shows receivers' actions in response to $p(\mathbf{m}_\infty) = \rho\gamma$ in the mimicking equilibrium, where those whose priors lie in the yellow area fact-check. Receivers whose priors lie in area $Q+S+V+Z$ do not fact-check and choose $a_i = 1$. Receivers whose priors lie in area $W$ do not fact-check and choose $a_i = 0$. Receivers in area $R+T+U+X+Y$ fact-check and choose actions that match the state.

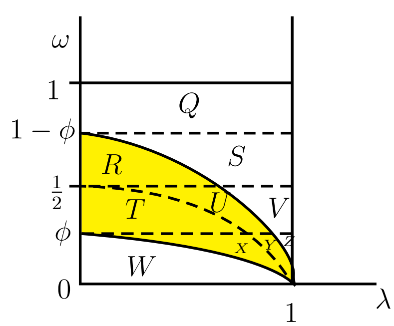



**Figure B3:** Fact-checking when receivers observe $p(\mathbf{m}_\infty) = \gamma$

Anticipating receivers' fact-checking after each observed frequency, the sender's payoffs from strategies that generate plausible frequencies are:

1. *Quasi-Truthful*: $\omega^S[-(1-c-b)^2(Q+R+S+T+U+V+X+Y+Z) - (0-c-b)^2(W)] + (1-\omega^S)[-(1-b)^2(0) - (0-b)^2(Q+R+S+T+U+V+W+X+Y+Z)]$
2. *Ones*: $\omega^S[-(1-c-b)^2(Q+R+S+T+U+V+X+Y+Z) - (0-c-b)^2(W)] + (1-\omega^S)[-(1-b)^2(Q+S+V+Z) - (0-b)^2(R+T+U+X+Y+W)]$

3. *Zeros*: $\omega^S[-(1-c-b)^2(0)-(0-c-b)^2(Q+R+S+T+U+V+W+X+Y+Z)]+(1-\omega^S)[-(1-b)^2(0)-(0-b)^2(Q+R+S+T+U+V+W+X+Y+Z)]$
4. *Mirror*: $\omega^S[-(1-c-b)^2(0)-(0-c-b)^2(Q+R+S+T+U+V+W+X+Y+Z)]+(1-\omega^S)[-(1-b)^2(Q+S+V+Z)-(0-b)^2(R+T+U+X+Y+W)]$

Sender $u$ does not deviate from reporting truthfully if

$$(Q+S+V+Z)(1-2b_u)\geq 0 \tag{B.9}$$

$$(Q+R+S+T+U+V+X+Y+Z)(-1+2c_u+2b_u)\geq 0 \tag{B.10}$$

$$(Q+R+S+T+U+V+X+Y+Z)[\omega^S(-1+2c_u+2b_u)+(1-\omega^S)(1-2b_u)]\geq 0, \tag{B.11}$$

which together imply that sender $u$ must be good: $b_u \leq 1/2$ and $c_u+b_u\geq 1/2$.

Does sender $u$ want to deviate to an off-equilibrium frequency? Receivers infer that an off-equilibrium frequency comes from the non-good type, so they fact-check according to Figure B2. Thus only those in $R+S+T+U+V$ fact-check, and sender $u$ does not deviate to an off-equilibrium frequency if $\omega^S(X+Y+Z)(-1+2c_u+2b_u)+(1-\omega^S)(Q)(1-2b_u)\geq 0$, which is satisfied when $b_u\leq 1/2$ and $c_u+b_u\geq 1/2$.

Sender $v$ does not deviate from the mimicking strategy of *Ones* if

$$(1-\omega^S)(Q+S+V+Z)(-1+2b_v)\geq 0 \tag{B.12}$$

$$\omega^S(Q+R+S+T+U+V+X+Y+Z)(-1+2c_v+2b_v)\geq 0 \tag{B.13}$$

$$\omega^S(Q+R+S+T+U+V+X+Y+Z)(-1+2c_v+2b_v)+(1-\omega^S)(Q+S+V+Z)(-1+2b_v)\geq 0, \tag{B.14}$$

which together imply that sender $v$ must single-mindedly prefer $a_i=1$: $b_v\geq 1/2$ and $c_v+b_v\geq 1/2$. Such a sender $v$ does not deviate to an off-equilibrium frequency if $\omega^S(X+Y+Z)(-1+2c_v+2b_v)+(1-\omega^S)(S+V+Z)(-1+2b_v)\geq 0$, which is satisfied. Thus, endogenous fact-checking has no effect on the space of sender types required to sustain a mimicking equilibrium.

### B.3.3 Mirroring Equilibrium

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

If $p(\mathbf{m}_\infty)=\rho\gamma$, then $\mu_i=P(\theta=1|p(\mathbf{m}_\infty)=\rho\gamma)=\frac{\omega_i\lambda_i}{\omega_i\lambda_i+(1-\omega_i)(1-\lambda_i)}$. 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 $C+J+K+O+U+V+EE$ 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 $B+I+N+S+T+Y+DD$ in Figure B4.

If $p(\mathbf{m}_\infty)=1-\rho\gamma$, then $\mu_i=P(\theta=1|p(\mathbf{m}_\infty)=1-\rho\gamma)=\frac{\omega_i(1-\lambda_i)}{\omega_i(1-\lambda_i)+(1-\omega_i)\lambda_i}$. 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 $E+L+K+N+S+R+AA$ 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 $F+M+O+U+T+X+BB$ 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: $D$, $A$, $G$, $H$, $P$, $Q$, $W$, $Z$, $CC$, and $FF$. Receivers in the blue area $K+N+O+S+T+U$ always fact-check. Receivers in the yellow areas $B+C+I+J$ and $V+Y+DD+EE$ only fact-check when they observe $p(\mathbf{m}_\infty) = \rho\gamma$. Receivers in the green areas $E+F+L+M$ and $R+X+AA+BB$ only fact-check when they observe $p(\mathbf{m}_\infty) = 1-\rho\gamma$.

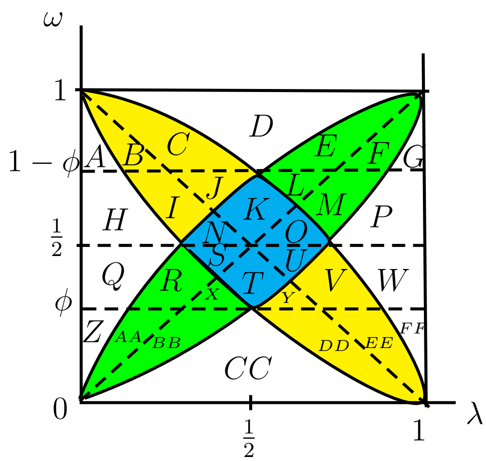



**Figure B4:** Fact-checking in mirroring equilibrium

Anticipating receivers' fact-checking after each observed frequency, the sender's payoffs from strategies that generate plausible frequencies are:

1. *Quasi-Truthful*: $\omega^S[-(1-c-b)^2(B+C+D+E+F+G+I+J+K+L+M+N+O+P+S+T+U+V+W+Y+DD+EE+FF)-(0-c-b)^2(A+H+Q+R+X+Z+AA+BB+CC)]+(1-\omega^S)[-(1-b)^2(A+B+C+D+H+I+J+Q+Z)-(0-b)^2(E+F+G+K+L+M+N+O+P+R+S+T+U+V+W+X+Y+AA+BB+CC+DD+EE+FF)]$

2. *Ones*: $\omega^S[-(1-c-b)^2(B+C+D+E+F+G+I+J+K+L+M+N+O+P+S+T+U+V+W+Y+DD+EE+FF)-(0-c-b)^2(A+H+Q+R+X+Z+AA+BB+CC)]+(1-\omega^S)[-(1-b)^2(D+E+F+G+L+M+P+W+FF)-(0-b)^2(A+B+C+H+I+J+K+N+O+Q+R+S+T+U+V+X+Y+Z+AA+BB+CC+DD+EE)]$

3. *Zeros*: $\omega^S[-(1-c-b)^2(A+B+C+D+E+F+H+I+J+K+L+M+N+O+Q+R+S+T+U+X+Z+AA+BB)-(0-c-b)^2(G+P+W+V+Y+CC+DD+EE+FF)]+(1-\omega^S)[-(1-b)^2(A+B+C+D+H+I+J+Q+Z)-(0-b)^2(E+F+G+K+L+M+N+O+P+R+S+T+U+V+W+X+Y+AA+BB+CC+DD+EE+FF)]$

4. *Mirror*: $\omega^S[-(1-c-b)^2(A+B+C+D+E+F+H+I+J+K+L+M+N+O+Q+R+S+T+U+X+Z+AA+BB)-(0-c-b)^2(G+P+W+V+Y+CC+DD+EE+FF)]+(1-\omega^S)[-(1-b)^2(D+E+F+G+L+M+P+W+FF)-(0-b)^2(A+B+C+H+I+J+K+N+O+Q+R+S+T+U+V+X+Y+Z+AA+BB+CC+DD+EE)]$

5. *Off-Equilibrium Frequency*: $\omega^S[-(1-c-b)^2(A+B+C+D+E+F+G+H+I+J+K+L+M+N+O+P+Q+R+S+T+U+V+W+X+Y)-(0-c-b)^2(Z+AA+BB+CC+DD+EE+FF)]+(1-\omega^S)[-(1-b)^2(A+B+C+D+E+F+G)-(0-b)^2(H+I+J+K+L+M+N+O+P+Q+R+S+T+V+W+X+Y+Z+AA+BB+CC+DD+EE+FF)]$

Sender $u$ does not deviate from reporting truthfully if the following conditions hold:

$$[(E+F+G+L+M+P+W+FF)-(A+B+C+H+I+J+Q+Z)](1-\omega^S)(1-2b_u)\geq 0 \tag{B.15}$$

$$[(G+P+V+W+Y+DD+EE+FF)-(A+H+Q+R+X+Z+AA+BB)](\omega^S)(-1+2c_u+2b_u)\geq 0 \tag{B.16}$$

$$[(E+F+G+L+M+P+W+FF)-(A+B+C+H+I+J+Q+Z)](1-\omega^S)(1-2b_u) + [(G+P+V+W+Y+DD+EE+FF)-(A+H+Q+R+X+Z+AA+BB)](\omega^S)(-1+2c_u+2b_u)\geq 0 \tag{B.17}$$

$$\omega^S[(DD+EE+FF)-(A+H+Z+R+X)](-1+2c_u+2b_u) + (1-\omega^S)[(E+F+G)-(H+I+J+Q+Z)](1-2b_u)\geq 0. \tag{B.18}$$

Sender $v$ does not deviate from mirroring if the following conditions hold:

$$[(E+F+G+L+M+P+W+FF)-(A+B+C+H+I+J+Q+Z)](1-\omega^S)(-1+2b_v)\geq 0 \tag{B.19}$$

$$[(G+P+V+W+Y+DD+EE+FF)-(A+H+Q+R+X+Z+AA+BB)](\omega^S)(1-2c_v+2b_v)\geq 0 \tag{B.20}$$

$$[(E+F+G+L+M+P+W+FF)-(A+B+C+H+I+J+Q+Z)](1-\omega^S)(-1+2b_v) + [(G+P+V+W+Y+DD+EE+FF)-(A+H+Q+R+X+Z+AA+BB)](\omega^S)(1-2c_v+2b_v)\geq 0 \tag{B.21}$$

$$\omega^S[(G+P+W+V+Y)-(Z+AA+BB)](1-2c_v-2b_v) + (1-\omega^S)[(L+M+P+W+FF)-(A+B+C)](-1+2b_v)\geq 0. \tag{B.22}$$

Together, these conditions imply that the mirroring equilibrium exists when sender $u$ is good ($b_u \leq 1/2$ and $c_u + b_u \geq 1/2$), sender $v$ is malevolent ($b_v \leq 1/2$ and $c_v + b_v \geq 1/2$) and the following conditions hold for the distribution of receivers:

$$(DD+EE+FF)-(A+H+Q+R+X)\geq 0 \tag{B.23}$$

$$(E+F+G)-(H+I+J+Q+Z)\geq 0 \tag{B.24}$$

$$(G+P+W+V+Y)-(Z+AA+BB)\geq 0 \tag{B.25}$$

$$(L+M+P+W+FF)-(A+B+C)\geq 0, \tag{B.26}$$

which can be re-written as Equations 9 and 10. In essence, these distribution conditions imply that senders will not deviate to off-equilibrium frequencies if enough receivers trust than distrust in each state.

## B.4 Proof of Proposition 6

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

### B.4.1 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.

### B.4.2 Mimicking Equilibrium

We compare receivers' welfare by using the partitions in Figure B3. If $(j, \theta) = (u, 0)$, all receivers observe $p(\mathbf{m}_\infty) = 1 - \rho\gamma$ and 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 $R + U + Y$ fact-check needlessly and receivers whose priors lie in $T + W$ benefit by fact-checking. If $(j, \theta) = (v, 0)$, receivers whose priors lie in $T + X$ fact-check needlessly and receivers whose priors lie in $R + U + Y$ benefit by fact-checking. Thus receivers are better off with the option to fact-check than without it if:

$$\hat{\lambda}(1-\hat{\omega})(0) + \hat{\omega}((R+U+Y)(-\phi) + (T+X)(1-\phi)) + (1-\hat{\lambda})(1-\hat{\omega})((R+U+Y)(1-\phi) + (T+X)(-\phi)) \geq 0$$

$$(R+U+Y)\left(\hat{\omega}\phi + (1-\hat{\lambda})(1-\hat{\omega})(1-\phi)\right) + (T+X)\left(\hat{\omega}(1-\phi) - (1-\hat{\lambda})(1-\hat{\omega})\phi\right) \geq 0,$$

which is Equation 11. Otherwise, they are worse off with the option to fact-check than without it.

### B.4.3 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:

$$\begin{aligned}
&(B+I+N+S+T+Y+DD)\left(\hat{\lambda}(1-\hat{\omega})(1-\phi) - (1-\hat{\lambda})(1-\hat{\omega})\phi\right) \\
&+ (C+J+K+O+U+V+EE)\left(\hat{\lambda}\hat{\omega}\phi + (1-\hat{\lambda})(1-\hat{\omega})(1-\phi)\right) \\
&+ (E+L+K+N+S+R+AA)\left(\hat{\lambda}(1-\hat{\omega})(1-\phi) - (1-\hat{\lambda})\hat{\omega}\phi\right) \\
&+ (F+M+O+U+T+X+BB)\left(-\hat{\lambda}(1-\hat{\omega})\phi + (1-\hat{\lambda})\hat{\omega}(1-\phi)\right) \geq 0,
\end{aligned}$$

which is Equation 12. Otherwise, they are worse off with the option to fact-check than without it.

## B.5 Details for Proof of Proposition 7

### B.5.1 Fully Informative Equilibrium

It is straightforward to verify that sender $u$ will still report truthfully when she is good: $b_u \leq 1/2$ and $c_u + b_u \geq 1/2$. Intuitively, reputation concerns only reinforce a good type's incentive to report truthfully.

Comparing payoffs, sender $v$ reports truthfully if

$$(G+H+I)(1-2b_v) + r\int_0^1\int_0^1 \lambda f(\lambda,\omega)d\lambda d\omega \geq 0 \tag{B.27}$$

$$(G+H+I)(-1+2c_v+2b_v) + r\int_0^1\int_0^1 \lambda f(\lambda,\omega)d\lambda d\omega \geq 0 \tag{B.28}$$

$$\omega^S[-(1-c_v-b_v)^2(G+H+I) - (0-c_v-b_v)^2(-G-H-I) + r\int_0^1\int_0^1 \lambda f(\lambda,\omega)d\lambda d\omega] \tag{B.29}$$

$$+ (1-\omega^S)[-(1-b_v)^2(-G-H-I) - (0-b_v)^2(G+H+I) + r\int_0^1\int_0^1 \lambda f(\lambda,\omega)d\lambda d\omega] \geq 0. \tag{B.30}$$

Equations B.27 and B.28 imply that sender $v$'s preferences must satisfy the following:

$$(1-2b_v) + r\int_0^1\int_0^1 \lambda f(\lambda,\omega)d\lambda d\omega \geq 0 \tag{B.31}$$

$$(-1+2c_v+2b_v) + r\int_0^1\int_0^1 \lambda f(\lambda,\omega)d\lambda d\omega \geq 0. \tag{B.32}$$

Note that Equation B.30 is satisfied when Equations B.31 and B.32 are satisfied.

If receivers believe an off-equilibrium frequency comes from sender $v$ but is uninformative about $\theta$, then sender $v$ does not deviate to an off-equilibrium frequency if

$$\omega^S[(H+I)(-1+2c_v+2b_v) + r\int_0^1\int_0^1 \lambda f(\lambda,\omega)d\lambda d\omega] + (1-\omega^S)[1-2b_v + r\int_0^1\int_0^1 \lambda f(\lambda,\omega)d\lambda d\omega] \geq 0, \tag{B.33}$$

which is satisfied when Equations B.27 and B.28 are satisfied.

Thus for sender $v$ to report truthfully, her preferences must satisfy Equations B.31 and B.32.

Equations B.31 and B.32 imply that reputation expands the set of sender types who report truthfully. First, if sender $v$ is also good ($b_v \leq 1/2$ and $c_v + b_v \geq 1/2$), then she will still report truthfully. Second, if sender $v$ is single-minded ($b_v > 1/2$ and $c_v + b_v \geq 1/2$), then she will report truthfully 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 report truthfully 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 $c_v + b_v \in [1/2 - \frac{r}{2}\int_0^1\int_0^1 \lambda f(\lambda,\omega)d\lambda d\omega, 1/2)$.

### B.5.2 Mimicking Equilibrium

Intuitively, reputation concerns only reinforce a good type's incentive to report truthfully. It is straightforward to verify that sender $u$ will still report truthfully when $b_u \leq 1/2$ and $c_u + b_u \geq 1/2$.

Sender $v$ does not deviate from *Ones* to *Quasi-Truthful*, *Zeros*, or *Mirror* if

$$(G+H)(-1+2b_v) - r \geq 0. \tag{B.34}$$

$$(G+H)(-1+2c_v+2b_v) + r\int_0^1\int_0^1 \lambda f(\lambda,\omega)d\lambda d\omega \geq 0. \tag{B.35}$$

where $G + H = \int_0^1 \int_{\frac{1-\lambda}{2-\lambda}}^1 f(\lambda,\omega)d\omega d\lambda$. Suppose receivers believe off-equilibrium frequencies come from the non-good type and are uninformative about the state. Sender $v$ does not deviate to an off-equilibrium frequency if

$$\omega^S[H(-1+2c_v+2b_v) + r\int_0^1\int_0^1 \lambda f(\lambda,\omega)d\lambda d\omega] + (1-\omega^S)[H(-1+2b_v)] \geq 0, \tag{B.36}$$

where $H = \int_0^1 \int_{\frac{1-\lambda}{2-\lambda}}^{\frac{1}{2}} f(\lambda,\omega)d\omega d\lambda$. Thus for sender $v$ to mimic, her preferences must satisfy Equations B.34, B.35, and B.36.

Equations B.34, B.35, and B.36 imply that reputation leads fewer single-minded types to mimic, but some malevolent types will mimic instead of mirror. Equation B.34 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 B.35 is satisfied and implies that Equation B.36 holds. Thus the single-minded type will still mimic if $c_v + b_v \geq 1/2$ and $b_v \geq 1/2 + \frac{r}{2\int_0^1\int_{\frac{1-\lambda}{2-\lambda}}^1 f(\lambda,\omega)d\omega d\lambda}$. If $v$ is malevolent, then Equation B.36 must hold for all $\omega^S$ and implies that Equation B.35 holds. Thus 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 $c_v + b_v \in [1/2 - \frac{r\int_0^1\int_0^1 \lambda f(\lambda,\omega)d\lambda d\omega}{2\int_0^1\int_{\frac{1-\lambda}{2-\lambda}}^{\frac{1}{2}} f(\lambda,\omega)d\omega d\lambda}, 1/2)$.

### B.5.3 Mirroring Equilibrium

Intuitively, reputation concerns only reinforce a good type's incentive to report truthfully. We can easily verify that sender $u$ will still report truthfully when $b_u \leq 1/2$ and $c_u + b_u \geq 1/2$.

Sender $v$ does not deviate from *Mirror* to *Ones*, *Zeros*, or *Quasi-Truthful* if

$$c_v + b_v \leq \frac{1}{2} - \frac{r}{2(C+D-A-B)} \tag{B.37}$$

$$b_v \geq \frac{1}{2} + \frac{r}{2(C+D-A-B)}, \tag{B.38}$$

where $C+D-A-B=\int_{\frac{1}{2}}^{1}\int_{1-\lambda}^{\lambda} f(\lambda,\omega)d\omega d\lambda-\int_{0}^{\frac{1}{2}}\int_{\lambda}^{1-\lambda} f(\lambda,\omega)d\omega d\lambda$.

Suppose receivers believe off-equilibrium frequencies come from the non-good type and are uninformative about the state. Sender $v$ does not deviate to an off-equilibrium frequency if $\omega^S[(B-C)(-1+2c_v+2b_v)]+(1-\omega^S)[(D-A)(-1+2b_v)]\geq 0$, which is satisfied when $D-A\geq 0\geq B-C$ and Equations B.37 and B.38 hold.

Thus for sender $v$ to mirror, her preferences must satisfy Equations B.37 and B.38.

Equations B.37 and B.38 imply that reputation shrinks the set of sender types who mirror. Equation B.37 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 B.37 and B.38 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.

### B.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 reports truthfully 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 reports truthfully 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 $c_v+b_v\in[1/2-\frac{r\int_0^1\int_0^1\lambda f(\lambda,\omega)d\lambda d\omega)}{2\int_0^1\int_{\frac{1-\lambda}{2-\lambda}}^{\frac{1}{2}} f(\lambda,\omega)d\omega d\lambda},1/2)$. 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 $c_v+b_v\in(\frac{1}{2}-\frac{r}{2(\int_{\frac{1}{2}}^1\int_{1-\lambda}^{\lambda} f(\lambda,\omega)d\omega d\lambda-\int_0^{\frac{1}{2}}\int_{\lambda}^{1-\lambda} f(\lambda,\omega)d\omega d\lambda)},1/2-\frac{r\int_0^1\int_0^1\lambda f(\lambda,\omega)d\lambda d\omega)}{2\int_0^1\int_{\frac{1-\lambda}{2-\lambda}}^{\frac{1}{2}} f(\lambda,\omega)d\omega d\lambda})$ if $\frac{1}{(\int_{\frac{1}{2}}^1\int_{1-\lambda}^{\lambda} f(\lambda,\omega)d\omega d\lambda-\int_0^{\frac{1}{2}}\int_{\lambda}^{1-\lambda} f(\lambda,\omega)d\omega d\lambda)}>\frac{\int_0^1\int_0^1\lambda f(\lambda,\omega)d\lambda d\omega)}{\int_0^1\int_{\frac{1-\lambda}{2-\lambda}}^{\frac{1}{2}} f(\lambda,\omega)d\omega d\lambda}$.

Thus, if reputation concerns are sufficiently strong, they can induce some single-minded and malevolent types to report truthfully 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 6 shows this graphically.

## B.6 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 $\gamma$ 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 $\theta$ 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 $\lambda_{1i}$ is the prior probability that sender 1 is type $u$, $\lambda_{2i}$ is the prior probability that sender 2 is type $u$, and $\omega_i$ 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 $\omega^{S_1}$ be sender 1's prior belief at $\tau = 0$ that $\theta = 1$ and $\omega^{S_2}$ 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.7 Mimicking Equilibrium

Suppose type $u$ reports quasi-truthfully so that $m_{1n} = s_{1n}$ with probability $\rho$ and $v$ mimics so that $P(m_{2n}) = \gamma\rho$. A receiver's posterior beliefs given $(m_{1n}, m_{2n})$ are

$$\frac{P(u_1, u_2, 0|\mathbf{m}_{1n}, \mathbf{m}_{2n})}{P(u_1, u_2, 1|\mathbf{m}_{1n}, \mathbf{m}_{2n})} = \left(\left(\frac{1-\rho\gamma}{\rho\gamma}\right)^{\frac{2n_{11}}{n}-1}\right)^n \left(\left(\frac{1-\rho\gamma}{\rho\gamma}\right)^{\frac{2n_{21}}{n}-1}\right)^n \left(\frac{1-\omega_i}{\omega_i}\right) \tag{B.39}$$

$$\frac{P(u_1, v_2, 0|\mathbf{m}_{1n}, \mathbf{m}_{2n})}{P(u_1, u_2, 1|\mathbf{m}_{1n}, \mathbf{m}_{2n})} = \left(\left(\frac{1-\rho\gamma}{\rho\gamma}\right)^{\frac{2n_{11}}{n}-1}\right)^n \frac{(1-\omega_i)(1-\lambda_{2i})}{\omega_i\lambda_{2i}} \tag{B.40}$$

$$\frac{P(v_1, v_2, 0|\mathbf{m}_{1n}, \mathbf{m}_{2n})}{P(u_1, u_2, 1|\mathbf{m}_{1n}, \mathbf{m}_{2n})} = \frac{(1-\omega_i)(1-\lambda_{1i})(1-\lambda_{2i})}{\omega_i\lambda_{1i}\lambda_{2i}} \tag{B.41}$$

$$\frac{P(v_1, u_2, 0|\mathbf{m}_{1n}, \mathbf{m}_{2n})}{P(u_1, u_2, 1|\mathbf{m}_{1n}, \mathbf{m}_{2n})} = \left(\left(\frac{1-\rho\gamma}{\rho\gamma}\right)^{\frac{2n_{21}}{n}-1}\right)^n \frac{(1-\omega_i)(1-\lambda_{1i})}{\omega_i\lambda_{1i}} \tag{B.42}$$

$$\frac{P(u_1, v_2, 1|\mathbf{m}_{1n}, \mathbf{m}_{2n})}{P(u_1, u_2, 1|\mathbf{m}_{1n}, \mathbf{m}_{2n})} = \frac{1-\lambda_{2i}}{\lambda_{2i}} \tag{B.43}$$

$$\frac{P(v_1, v_2, 1|\mathbf{m}_{1n}, \mathbf{m}_{2n})}{P(u_1, u_2, 1|\mathbf{m}_{1n}, \mathbf{m}_{2n})} = \frac{(1-\lambda_{1i})(1-\lambda_{2i})}{\lambda_{1i}\lambda_{2i}} \tag{B.44}$$

$$\frac{P(v_1, u_2, 1|\mathbf{m}_{1n}, \mathbf{m}_{2n})}{P(u_1, u_2, 1|\mathbf{m}_{1n}, \mathbf{m}_{2n})} = \frac{1-\lambda_{1i}}{\lambda_{1i}}. \tag{B.45}$$

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

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

$$P(v_1, v_2, 0|(\gamma, \gamma)) = \frac{(1-\lambda_{1i})(1-\lambda_{2i})(1-\omega_i)}{(1-\lambda_{1i})(1-\lambda_{2i})(1-\omega_i)+\omega_i} \tag{B.46}$$

$$P(u_1, v_2, 1|(\gamma, \gamma)) = \frac{\lambda_{1i}(1-\lambda_{2i})\omega_i}{(1-\lambda_{1i})(1-\lambda_{2i})(1-\omega_i)+\omega_i} \tag{B.47}$$

$$P(u_1, u_2, 1|(\gamma, \gamma)) = \frac{\lambda_{1i}\lambda_{2i}\omega_i}{(1-\lambda_{1i})(1-\lambda_{2i})(1-\omega_i)+\omega_i} \tag{B.48}$$

$$P(v_1, v_2, 1|(\gamma, \gamma)) = \frac{(1-\lambda_{1i})(1-\lambda_{2i})\omega_i}{(1-\lambda_{1i})(1-\lambda_{2i})(1-\omega_i)+\omega_i} \tag{B.49}$$

$$P(v_1, u_2, 1|(\gamma, \gamma)) = \frac{(1-\lambda_{1i})\lambda_{2i}\omega_i}{(1-\lambda_{1i})(1-\lambda_{2i})(1-\omega_i)+\omega_i} \tag{B.50}$$

$$P(u_1, v_2, 0|(\gamma, \gamma)) = P(u_1, u_2, 0|(\gamma, \gamma)) = 0. \tag{B.51}$$

$$a_i(\mathbf{m}_\infty \mid (1-\gamma,\gamma)) = 0 \tag{B.52}$$
$$a_i(\mathbf{m}_\infty \mid (\gamma,1-\gamma)) = 0 \tag{B.53}$$
$$a_i(\mathbf{m}_\infty \mid (1-\gamma,1-\gamma)) = 0 \tag{B.54}$$
$$a_i(\mathbf{m}_\infty \mid (\gamma,\gamma)) = \begin{cases} 1 & \text{if } \omega_i > \frac{(1-\lambda_{1i})(1-\lambda_{2i})}{1+(1-\lambda_{1i})(1-\lambda_{2i})} \\ 0 & \text{if } \omega_i < \frac{(1-\lambda_{1i})(1-\lambda_{2i})}{1+(1-\lambda_{1i})(1-\lambda_{2i})} \\ (0 \text{ w.p. } \frac{1}{2}; 1 \text{ w.p. } \frac{1}{2}) & \text{if } \omega_i = \frac{(1-\lambda_{1i})(1-\lambda_{2i})}{1+(1-\lambda_{1i})(1-\lambda_{2i})}. \end{cases} \tag{B.55}$$

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 being quasi-truthful to any other messaging strategy that generates plausible frequencies if

$$(1-\omega^{S_1})(1-\lambda_2^{S_1})\left(1-\int_0^1\int_0^1\int_0^{\frac{(1-\lambda_1)(-\lambda_2)}{1+(1-\lambda_1)(-\lambda_2)}} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_2 d\lambda_1\right)(1-2b_1) \geq 0 \tag{B.56}$$
$$\omega^{S_1}\left(\int_0^1\int_0^1\int_{\frac{(1-\lambda_1)(-\lambda_2)}{1+(1-\lambda_1)(-\lambda_2)}}^1 f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_2 d\lambda_1\right)(-1+2c_1+2b_1) \geq 0, \tag{B.57}$$

which imply that $b_1 \leq 1/2$ and $c_1 + b_1 \geq 1/2$, respectively. Moreover, sender 1 does not deviate from being quasi-truthful to sending an off-equilibrium strategy that is uninformative to receivers if

$$\begin{aligned}(1-\omega^{S_1})(1-\lambda_2^{S_1})&\left(1-\int_0^1\int_0^1\int_0^{\frac{1-\lambda_2}{2-\lambda_2}} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_2 d\lambda_1\right)(1-2b_1)\\ +\omega^{S_1}&\left(\int_0^1\int_0^1\int_{\frac{(1-\lambda_1)(1-\lambda_2)}{1+(1-\lambda_1)(1-\lambda_2)}}^1 - \int_0^1\int_0^1\int_{\frac{1-\lambda_2}{2-\lambda_2}}^1 f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_2 d\lambda_1\right)(-1+2c_1+2b_1) \geq 0,\end{aligned} \tag{B.58}$$

which is satisfied because $b_1 \leq 1/2$ and $c_1 + b_1 \geq 1/2$.

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

$$(1-\omega^{S_1})(1-\lambda_2^{S_1})\left(1-\int_0^1\int_0^1\int_{\frac{(1-\lambda_1)(1-\lambda_2)}{1+(1-\lambda_1)(1-\lambda_2)}}^1 f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_2 d\lambda_1\right)(-1+2b_1) \geq 0 \tag{B.59}$$
$$\omega^{S_1}\left(\int_0^1\int_0^1\int_{\frac{(1-\lambda_1)(1-\lambda_2)}{1+(1-\lambda_1)(1-\lambda_2)}}^1 f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_2 d\lambda_1\right)(-1+2c_1+2b_1) \geq 0, \tag{B.60}$$

which imply that $b_1 \geq 1/2$ and $c_1 + b_1 \geq 1/2$, respectively. Moreover, sender 1 does not deviate from mimicking to sending an off-equilibrium strategy that is uninformative to receivers if

$$(1-\omega^{S_1})(1-\lambda_2^{S_1})\left(\int_0^1\int_0^1\int_{\frac{(1-\lambda_1)(1-\lambda_2)}{1+(1-\lambda_1)(1-\lambda_2)}}^1 - \int_0^1\int_0^1\int_{\frac{1-\lambda_2}{2-\lambda_2}}^1 f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_2 d\lambda_1\right)(-1+2b_1)$$
$$+\omega^{S_1}\left(\int_0^1\int_0^1\int_{\frac{(1-\lambda_1)(1-\lambda_2)}{1+(1-\lambda_1)(1-\lambda_2)}}^1 f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_2 d\lambda_1\right)(-1+2c_1+2b_1) \geq 0, \tag{B.61}$$

which is satisfied because $b_1 \geq 1/2$ and $c_1 + b_1 \geq 1/2$.

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.8 Mirroring Equilibrium

Suppose type $u$ reports quasi-truthfully so that $m_{1n} = s_{1n}$ with probability $\rho$ and $v$ mirrors by so that $m_{1n} = s_{1n}$ with probability $1-\rho$. A receiver's posterior beliefs given $(m_{1n}, m_{2n})$ are

$$\frac{P(u_1,u_2,0|\mathbf{m}_{1n},\mathbf{m}_{2n})}{P(u_1,u_2,1|\mathbf{m}_{1n},\mathbf{m}_{2n})} = \left(\left(\frac{1-\rho\gamma}{\rho\gamma}\right)^{\frac{2n_{11}}{n}-1}\right)^n \left(\left(\frac{1-\rho\gamma}{\rho\gamma}\right)^{\frac{2n_{21}}{n}-1}\right)^n \left(\frac{1-\omega_i}{\omega_i}\right) \tag{B.62}$$

$$\frac{P(u_1,v_2,0|\mathbf{m}_{1n},\mathbf{m}_{2n})}{P(u_1,u_2,1|\mathbf{m}_{1n},\mathbf{m}_{2n})} = \left(\left(\frac{1-\rho\gamma}{\rho\gamma}\right)^{\frac{2n_{11}}{n}-1}\right)^n \frac{(1-\omega_i)(1-\lambda_{2i})}{\omega_i\lambda_{2i}} \tag{B.63}$$

$$\frac{P(v_1,v_2,0|\mathbf{m}_{1n},\mathbf{m}_{2n})}{P(u_1,u_2,1|\mathbf{m}_{1n},\mathbf{m}_{2n})} = \frac{(1-\omega_i)(1-\lambda_{1i})(1-\lambda_{2i})}{\omega_i\lambda_{1i}\lambda_{2i}} \tag{B.64}$$

$$\frac{P(v_1,u_2,0|\mathbf{m}_{1n},\mathbf{m}_{2n})}{P(u_1,u_2,1|\mathbf{m}_{1n},\mathbf{m}_{2n})} = \left(\left(\frac{1-\rho\gamma}{\rho\gamma}\right)^{\frac{2n_{21}}{n}-1}\right)^n \frac{(1-\omega_i)(1-\lambda_{1i})}{\omega_i\lambda_{1i}} \tag{B.65}$$

$$\frac{P(u_1,v_2,1|\mathbf{m}_{1n},\mathbf{m}_{2n})}{P(u_1,u_2,1|\mathbf{m}_{1n},\mathbf{m}_{2n})} = \left(\left(\frac{1-\rho\gamma}{\rho\gamma}\right)^{\frac{2n_{21}}{n}-1}\right)^n \frac{1-\lambda_{2i}}{\lambda_{2i}} \tag{B.66}$$

$$\frac{P(v_1,v_2,1|\mathbf{m}_{1n},\mathbf{m}_{2n})}{P(u_1,u_2,1|\mathbf{m}_{1n},\mathbf{m}_{2n})} = \left(\left(\frac{1-\rho\gamma}{\rho\gamma}\right)^{\frac{2n_{11}}{n}-1}\right)^n \left(\left(\frac{1-\rho\gamma}{\rho\gamma}\right)^{\frac{2n_{21}}{n}-1}\right)^n \frac{(1-\lambda_{1i})(1-\lambda_{2i})}{\lambda_{1i}\lambda_{2i}} \tag{B.67}$$

$$\frac{P(v_1,u_2,1|\mathbf{m}_{1n},\mathbf{m}_{2n})}{P(u_1,u_2,1|\mathbf{m}_{1n},\mathbf{m}_{2n})} = \left(\left(\frac{1-\rho\gamma}{\rho\gamma}\right)^{\frac{2n_{11}}{n}-1}\right)^n \frac{1-\lambda_{1i}}{\lambda_{1i}}. \tag{B.68}$$

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

$$P(v_1, v_2, 0|(\gamma, \gamma)) = \frac{(1-\lambda_{1i})(1-\lambda_{2i})(1-\omega_i)}{(1-\lambda_{1i})(1-\lambda_{2i})(1-\omega_i) + \lambda_{1i}\lambda_{2i}\omega_i} \tag{B.69}$$

$$P(u_1, u_2, 1|(\gamma, \gamma)) = \frac{\lambda_{1i}\lambda_{2i}\omega_i}{(1-\lambda_{1i})(1-\lambda_{2i})(1-\omega_i) + \lambda_{1i}\lambda_{2i}\omega_i} \tag{B.70}$$

$$\begin{aligned} P(u_1, u_2, 0|(\gamma, \gamma)) &= P(u_1, v_2, 0|(\gamma, \gamma)) = P(v_1, u_2, 0|(\gamma, \gamma)) = P(u_1, v_2, 0|(\gamma, \gamma)) \\ &= P(v_1, v_2, 1|(\gamma, \gamma)) = P(v_1, u_2, 1|(\gamma, \gamma)) = 0. \end{aligned} \tag{B.71}$$

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

$$P(v_1, v_2, 1|(1-\gamma, 1-\gamma)) = \frac{(1-\lambda_{1i})(1-\lambda_{2i})\omega_i}{(1-\lambda_{1i})(1-\lambda_{2i})\omega_i + \lambda_{1i}\lambda_{2i}(1-\omega_i)} \tag{B.72}$$

$$P(u_1, u_2, 0|(1-\gamma, 1-\gamma)) = \frac{\lambda_{1i}\lambda_{2i}(1-\omega_i)}{(1-\lambda_{1i})(1-\lambda_{2i})\omega_i + \lambda_{1i}\lambda_{2i}(1-\omega_i)} \tag{B.73}$$

$$\begin{aligned} P(u_1, u_2, 1|(1-\gamma, 1-\gamma)) &= P(u_1, v_2, 0|(1-\gamma, 1-\gamma)) = P(v_1, u_2, 0|(1-\gamma, 1-\gamma)) \\ &= P(u_1, v_2, 0|(1-\gamma, 1-\gamma)) = P(v_1, v_2, 0|(1-\gamma, 1-\gamma)) \\ &= P(v_1, u_2, 1|(1-\gamma, 1-\gamma)) = 0. \end{aligned} \tag{B.74}$$

Intuitively: If the senders' message content agree, then receivers know that they are the same type but are not sure which type.

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

$$P(v_1, u_2, 0|(\gamma, 1-\gamma)) = \frac{(1-\lambda_{1i})\lambda_{2i}(1-\omega_i)}{(1-\lambda_{1i})\lambda_{2i}(1-\omega_i) + \lambda_{1i}(1-\lambda_{2i})\omega_i} \tag{B.75}$$

$$P(u_1, v_2, 1|(\gamma, 1-\gamma)) = \frac{\lambda_{1i}(1-\lambda_{2i})\omega_i}{(1-\lambda_{1i})\lambda_{2i}(1-\omega_i) + \lambda_{1i}(1-\lambda_{2i})\omega_i} \tag{B.76}$$

$$P(u_1, u_2, 1|(\gamma, 1-\gamma)) = P(v_1, v_2, 1|(\gamma, 1-\gamma)) = P(u_1, u_2, 0|(\gamma, 1-\gamma)) \tag{B.77}$$

$$= P(u_1, v_2, 0|(\gamma, 1-\gamma)) = P(v_1, v_2, 0|(\gamma, 1-\gamma)) = P(v_1, u_2, 1|(\gamma, 1-\gamma)) = 0. \tag{B.78}$$

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

$$P(v_1, u_2, 1|(1-\gamma, \gamma)) = \frac{(1-\lambda_{1i})\lambda_{2i}\omega_i}{(1-\lambda_{1i})\lambda_{2i}\omega_i + \lambda_{1i}(1-\lambda_{2i})(1-\omega_i)} \tag{B.79}$$

$$P(u_1, v_2, 0|(1-\gamma, \gamma)) = \frac{\lambda_{1i}(1-\lambda_{2i})(1-\omega_i)}{(1-\lambda_{1i})\lambda_{2i}(1-\omega_i) + \lambda_{1i}(1-\lambda_{2i})(1-\omega_i)} \tag{B.80}$$

$$P(u_1, u_2, 1|(\gamma, 1-\gamma)) = P(v_1, v_2, 1|(\gamma, 1-\gamma)) = P(u_1, u_2, 0|(\gamma, 1-\gamma)) \tag{B.81}$$

$$= P(u_1, v_2, 1|(\gamma, 1-\gamma)) = P(v_1, v_2, 0|(\gamma, 1-\gamma)) = P(v_1, u_2, 0|(\gamma, 1-\gamma)) = 0. \tag{B.82}$$

Intuitively: 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

$$a_i(\mathbf{m}_\infty \mid (\gamma,\gamma)) = \begin{cases} 1 & \text{if } \omega_i > \frac{(1-\lambda_{1i})(1-\lambda_{2i})}{\lambda_{1i}\lambda_{2i}+(1-\lambda_{1i})(1-\lambda_{2i})} \\ 0 & \text{if } \omega_i < \frac{(1-\lambda_{1i})(1-\lambda_{2i})}{\lambda_{1i}\lambda_{2i}+(1-\lambda_{1i})(1-\lambda_{2i})} \\ (0 \text{ w.p. } \frac{1}{2}; 1 \text{ w.p. } \frac{1}{2}) & \text{if } \omega_i = \frac{(1-\lambda_{1i})(1-\lambda_{2i})}{\lambda_{1i}\lambda_{2i}+(1-\lambda_{1i})(1-\lambda_{2i})}. \end{cases} \tag{B.83}$$

$$a_i(\mathbf{m}_\infty \mid (1-\gamma,1-\gamma)) = \begin{cases} 1 & \text{if } \omega_i > \frac{\lambda_{1i}\lambda_{2i}}{\lambda_{1i}\lambda_{2i}+(1-\lambda_{1i})(1-\lambda_{2i})} \\ 0 & \text{if } \omega_i < \frac{\lambda_{1i}\lambda_{2i}}{\lambda_{1i}\lambda_{2i}+(1-\lambda_{1i})(1-\lambda_{2i})} \\ (0 \text{ w.p. } \frac{1}{2}; 1 \text{ w.p. } \frac{1}{2}) & \text{if } \omega_i = \frac{\lambda_{1i}\lambda_{2i}}{\lambda_{1i}\lambda_{2i}+(1-\lambda_{1i})(1-\lambda_{2i})}. \end{cases} \tag{B.84}$$

$$a_i(\mathbf{m}_\infty \mid (\gamma,1-\gamma)) = \begin{cases} 1 & \text{if } \omega_i > \frac{(1-\lambda_{1i})\lambda_{2i}}{\lambda_{1i}(1-\lambda_{2i})+(1-\lambda_{1i})\lambda_{2i}} \\ 0 & \text{if } \omega_i < \frac{(1-\lambda_{1i})\lambda_{2i}}{\lambda_{1i}(1-\lambda_{2i})+(1-\lambda_{1i})\lambda_{2i}} \\ (0 \text{ w.p. } \frac{1}{2}; 1 \text{ w.p. } \frac{1}{2}) & \text{if } \omega_i = \frac{(1-\lambda_{1i})\lambda_{2i}}{\lambda_{1i}(1-\lambda_{2i})+(1-\lambda_{1i})\lambda_{2i}}. \end{cases} \tag{B.85}$$

$$a_i(\mathbf{m}_\infty \mid (1-\gamma,\gamma)) = \begin{cases} 1 & \text{if } \omega_i > \frac{\lambda_{1i}(1-\lambda_{2i})}{\lambda_{1i}(1-\lambda_{2i})+(1-\lambda_{1i})\lambda_{2i}} \\ 0 & \text{if } \omega_i < \frac{\lambda_{1i}(1-\lambda_{2i})}{\lambda_{1i}(1-\lambda_{2i})+(1-\lambda_{1i})\lambda_{2i}} \\ (0 \text{ w.p. } \frac{1}{2}; 1 \text{ w.p. } \frac{1}{2}) & \text{if } \omega_i = \frac{\lambda_{1i}(1-\lambda_{2i})}{\lambda_{1i}(1-\lambda_{2i})+(1-\lambda_{1i})\lambda_{2i}}. \end{cases} \tag{B.86}$$

Also, note that if receivers observe $(p(\mathbf{m}_{1n}), p(\mathbf{m}_{2n})) = (1/2, \gamma)$, their posterior beliefs are

$$P(u_1, v_2, 0|(1/2,\gamma)) = \frac{\lambda_{1i}(1-\lambda_{2i})(1-\omega_i)}{(1-\lambda_{2i})(1-\omega_i)+\lambda_{2i}\omega_i} \tag{B.87}$$

$$P(v_1, v_2, 0|(1/2,\gamma)) = \frac{(1-\lambda_{1i})(1-\lambda_{2i})(1-\omega_i)}{(1-\lambda_{2i})(1-\omega_i)+\lambda_{2i}\omega_i} \tag{B.88}$$

$$P(v_1, u_2, 1|(1/2,\gamma)) = \frac{(1-\lambda_{1i})\lambda_{2i}\omega_i}{(1-\lambda_{2i})(1-\omega_i)+\lambda_{2i}\omega_i} \tag{B.89}$$

$$P(u_1, u_2, 1|(1/2,\gamma)) = \frac{\lambda_{1i}\lambda_{2i}\omega_i}{(1-\lambda_{2i})(1-\omega_i)+\lambda_{2i}\omega_i} \tag{B.90}$$

$$P(u_1, v_2, 1|(1/2,\gamma)) = P(v_1, v_2, 1|(1/2,\gamma)) = P(v_1, u_2, 0|(1/2,\gamma)) = P(u_1, u_2, 0|(1/2,\gamma)) = 0. \tag{B.91}$$

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

$$P(u_1, v_2, 1|(1/2, 1-\gamma)) = \frac{\lambda_{1i}(1-\lambda_{2i})\omega_i}{(1-\lambda_{2i})\omega_i + \lambda_{2i}(1-\omega_i)} \tag{B.92}$$

$$P(v_1, v_2, 1|(1/2, \gamma)) = \frac{(1-\lambda_{1i})(1-\lambda_{2i})\omega_i}{(1-\lambda_{2i})\omega_i + \lambda_{2i}(1-\omega_i)} \tag{B.93}$$

$$P(v_1, u_2, 0|(1/2, \gamma)) = \frac{(1-\lambda_{1i})\lambda_{2i}(1-\omega_i)}{(1-\lambda_{2i})\omega_i + \lambda_{2i}(1-\omega_i)} \tag{B.94}$$

$$P(u_1, u_2, 0|(1/2, \gamma)) = \frac{\lambda_{1i}\lambda_{2i}(1-\omega_i)}{(1-\lambda_{2i})\omega_i + \lambda_{2i}(1-\omega_i)} \tag{B.95}$$

$$P(u_1, v_2, 0|(1/2, \gamma)) = P(v_1, v_2, 0|(1/2, \gamma)) = P(v_1, u_2, 1|(1/2, \gamma)) = P(u_1, u_2, 1|(1/2, \gamma)) = 0. \tag{B.96}$$

Thus

$$a_i(\mathbf{m}_\infty \mid (1/2, \gamma)) = \begin{cases} 1 & \text{if } \omega_i > 1-\lambda_{2i} \\ 0 & \text{if } \omega_i < 1-\lambda_{2i} \\ (0 \text{ w.p. } \frac{1}{2}; 1 \text{ w.p. } \frac{1}{2}) & \text{if } \omega_i = 1-\lambda_{2i}. \end{cases} \tag{B.97}$$

$$a_i(\mathbf{m}_\infty \mid (1/2, 1-\gamma)) = \begin{cases} 1 & \text{if } \omega_i > \lambda_{2i} \\ 0 & \text{if } \omega_i < \lambda_{2i} \\ (0 \text{ w.p. } \frac{1}{2}; 1 \text{ w.p. } \frac{1}{2}) & \text{if } \omega_i = \lambda_{2i}. \end{cases} \tag{B.98}$$

Intuitively: If sender 1 deviates to sending an off-equilibrium strategy that is uninformative, then receivers still have messages from sender 2 so their actions depend on their priors on sender 2's type.

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 being quasi-truthful to any other messaging strategy that generates plausible frequencies if

$$\begin{aligned}(1-2b_1)\Bigg(&\lambda_2^{S_1}\left(\int_0^1\int_0^1\int_0^{\frac{\lambda_1\lambda_2}{(1-\lambda_1)(1-\lambda_2)+\lambda_1\lambda_2}} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2 - \int_0^1\int_0^1\int_0^{\frac{(1-\lambda_1)\lambda_2}{\lambda_1(1-\lambda_2)+(1-\lambda_1)\lambda_2}} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2\right) \\ &+(1-\lambda_2^{S_1})\left(\int_0^1\int_0^1\int_0^{\frac{\lambda_1(1-\lambda_2)}{(1-\lambda_1)\lambda_2+\lambda_1(1-\lambda_2)}} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2 - \int_0^1\int_0^1\int_0^{\frac{(1-\lambda_1)(1-\lambda_2)}{\lambda_1\lambda_2+(1-\lambda_1)(1-\lambda_2)}} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2\right)\Bigg) \geq 0\end{aligned} \tag{B.99}$$

and

$$\begin{aligned}(-1+2b_1+2c_1)\Bigg(&\lambda_2^{S_1}\left(\int_0^1\int_0^1\int_{\frac{(1-\lambda_1)(1-\lambda_2)}{(1-\lambda_1)(1-\lambda_2)+\lambda_1\lambda_2}}^1 f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2 - \int_0^1\int_0^1\int_{\frac{\lambda_1(1-\lambda_2)}{\lambda_1(1-\lambda_2)+(1-\lambda_1)\lambda_2}}^1 f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2\right) \\ &+(1-\lambda_2^{S_1})\left(\int_0^1\int_0^1\int_{\frac{(1-\lambda_1)\lambda_2}{(1-\lambda_1)\lambda_2+\lambda_1(1-\lambda_2)}}^1 f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2 - \int_0^1\int_0^1\int_{\frac{\lambda_1\lambda_2}{\lambda_1\lambda_2+(1-\lambda_1)(1-\lambda_2)}} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2\right)\Bigg) \geq 0,\end{aligned} \tag{B.100}$$

which can be re-written as

$$(1-2b_1)\left(\lambda_2^{S_1}\left(\int_0^1\int_{1/2}^1\int_{\frac{(1-\lambda_1)\lambda_2}{\lambda_1(1-\lambda_2)+(1-\lambda_1)\lambda_2}}^{\frac{\lambda_1\lambda_2}{(1-\lambda_1)(1-\lambda_2)+\lambda_1\lambda_2}} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2 - \int_0^1\int_0^{1/2}\int_{\frac{\lambda_1\lambda_2}{(1-\lambda_1)(1-\lambda_2)+\lambda_1\lambda_2}}^{\frac{(1-\lambda_1)\lambda_2}{\lambda_1(1-\lambda_2)+(1-\lambda_1)\lambda_2}} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2\right)\right.$$
$$\left.+(1-\lambda_2^{S_1})\left(\int_0^1\int_{1/2}^1\int_{\frac{(1-\lambda_1)(1-\lambda_2)}{\lambda_1\lambda_2+(1-\lambda_1)(1-\lambda_2)}}^{\frac{\lambda_1(1-\lambda_2)}{(1-\lambda_1)\lambda_2+\lambda_1(1-\lambda_2)}} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2 - \int_0^1\int_0^{1/2}\int_{\frac{\lambda_1(1-\lambda_2)}{(1-\lambda_1)\lambda_2+\lambda_1(1-\lambda_2)}}^{\frac{(1-\lambda_1)(1-\lambda_2)}{\lambda_1\lambda_2+(1-\lambda_1)(1-\lambda_2)}} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2\right)\right)\geq 0 \tag{B.101}$$

and

$$(-1+2b_1+2c_1)\left(\lambda_2^{S_1}\left(\int_0^1\int_{1/2}^1\int_{\frac{(1-\lambda_1)(1-\lambda_2)}{(1-\lambda_1)(1-\lambda_2)+\lambda_1\lambda_2}}^{\frac{\lambda_1(1-\lambda_2)}{\lambda_1(1-\lambda_2)+(1-\lambda_1)\lambda_2}} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2 - \int_0^1\int_0^{1/2}\int_{\frac{\lambda_1(1-\lambda_2)}{\lambda_1(1-\lambda_2)+(1-\lambda_1)\lambda_2}}^{\frac{(1-\lambda_1)(1-\lambda_2)}{(1-\lambda_1)(1-\lambda_2)+\lambda_1\lambda_2}} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2\right)\right.$$
$$\left.+(1-\lambda_2^{S_1})\left(\int_0^1\int_{1/2}^1\int_{\frac{(1-\lambda_1)\lambda_2}{(1-\lambda_1)\lambda_2+\lambda_1(1-\lambda_2)}}^{\frac{\lambda_1\lambda_2}{\lambda_1\lambda_2+(1-\lambda_1)(1-\lambda_2)}} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2 - \int_0^1\int_0^{1/2}\int_{\frac{\lambda_1\lambda_2}{\lambda_1\lambda_2+(1-\lambda_1)(1-\lambda_2)}}^{\frac{(1-\lambda_1)\lambda_2}{(1-\lambda_1)\lambda_2+\lambda_1(1-\lambda_2)}} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2\right)\right)\geq 0. \tag{B.102}$$

Sender 1 will not deviate from being quasi-truthful to a messaging strategy that receivers interpret as uninformative if

$$\omega^S(-1+2b_1+2c_1)\left(\lambda_2^{S_1}\left(\int_0^1\int_0^1\int_{\frac{(1-\lambda_1)(1-\lambda_2)}{(1-\lambda_1)(1-\lambda_2)+\lambda_1\lambda_2}}^1 f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2 - \int_0^1\int_0^1\int_{1-\lambda_2}^1 f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2\right)\right.$$
$$\left.+(1-\lambda_2^{S_1})\left(\int_0^1\int_0^1\int_{\frac{(1-\lambda_1)\lambda_2}{(1-\lambda_1)\lambda_2+\lambda_1(1-\lambda_2)}}^1 f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2 - \int_0^1\int_0^1\int_{\lambda_2}^1 f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2\right)\right) \tag{B.103}$$

$$(1-\omega^S)(1-2b_1)\left(\lambda_2^{S_1}\left(\int_0^1\int_0^1\int_0^{\frac{\lambda_1\lambda_2}{(1-\lambda_1)(1-\lambda_2)+\lambda_1\lambda_2}} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2 - \int_0^1\int_0^1\int_0^{\lambda_2} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2\right)\right.$$
$$\left.+(1-\lambda_2^{S_1})\left(\int_0^1\int_0^1\int_0^{\frac{\lambda_1(1-\lambda_2)}{(1-\lambda_1)\lambda_2+\lambda_1(1-\lambda_2)}} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2 - \int_0^1\int_0^1\int_0^{1-\lambda_2} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2\right)\right)\geq 0, \tag{B.104}$$

which can be re-written as

$$\omega^S(-1+2b_1+2c_1)\left(\lambda_2^{S_1}\left(\int_0^1\int_{1/2}^1\int_{\frac{(1-\lambda_1)(1-\lambda_2)}{(1-\lambda_1)(1-\lambda_2)+\lambda_1\lambda_2}}^{1-\lambda_2} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2 - \int_0^1\int_0^{1/2}\int_{1-\lambda_2}^{\frac{(1-\lambda_1)(1-\lambda_2)}{(1-\lambda_1)(1-\lambda_2)+\lambda_1\lambda_2}} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2\right)\right.$$
$$\left.+(1-\lambda_2^{S_1})\left(\int_0^1\int_{1/2}^1\int_{\frac{(1-\lambda_1)\lambda_2}{(1-\lambda_1)\lambda_2+\lambda_1(1-\lambda_2)}}^{\lambda_2} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2 - \int_0^1\int_0^{1/2}\int_{\lambda_2}^{\frac{(1-\lambda_1)\lambda_2}{(1-\lambda_1)\lambda_2+\lambda_1(1-\lambda_2)}} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2\right)\right) \tag{B.105}$$

$$(1-\omega^S)(1-2b_1)\left(\lambda_2^{S_1}\left(\int_0^1\int_{1/2}^1\int_{\lambda_2}^{\frac{\lambda_1\lambda_2}{(1-\lambda_1)(1-\lambda_2)+\lambda_1\lambda_2}} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2 - \int_0^1\int_0^{1/2}\int_{\frac{\lambda_1\lambda_2}{(1-\lambda_1)(1-\lambda_2)+\lambda_1\lambda_2}}^{\lambda_2} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2\right)\right.$$
$$\left.+(1-\lambda_2^{S_1})\left(\int_0^1\int_{1/2}^1\int_{1-\lambda_2}^{\frac{\lambda_1(1-\lambda_2)}{(1-\lambda_1)\lambda_2+\lambda_1(1-\lambda_2)}} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2 - \int_0^1\int_0^{1/2}\int_{\frac{\lambda_1(1-\lambda_2)}{(1-\lambda_1)\lambda_2+\lambda_1(1-\lambda_2)}}^{1-\lambda_2} f(\lambda_1,\lambda_2,\omega)d\omega d\lambda_1 d\lambda_2\right)\right)\geq 0, \tag{B.106}$$

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:

$$\int_0^1 \int_{1/2}^1 \int_{\frac{(1-\lambda_1)(1-\lambda_2)}{(1-\lambda_1)(1-\lambda_2)+\lambda_1\lambda_2}}^{1-\lambda_2} f(\lambda_1, \lambda_2, \omega) d\omega d\lambda_1 d\lambda_2 - \int_0^1 \int_0^{1/2} \int_{1-\lambda_2}^{\frac{(1-\lambda_1)(1-\lambda_2)}{(1-\lambda_1)(1-\lambda_2)+\lambda_1\lambda_2}} f(\lambda_1, \lambda_2, \omega) d\omega d\lambda_1 d\lambda_2 \geq 0 \tag{B.107}$$

$$\int_0^1 \int_{1/2}^1 \int_{\frac{(1-\lambda_1)\lambda_2}{(1-\lambda_1)\lambda_2+\lambda_1(1-\lambda_2)}}^{\lambda_2} f(\lambda_1, \lambda_2, \omega) d\omega d\lambda_1 d\lambda_2 - \int_0^1 \int_0^{1/2} \int_{\lambda_2}^{\frac{(1-\lambda_1)\lambda_2}{(1-\lambda_1)\lambda_2+\lambda_1(1-\lambda_2)}} f(\lambda_1, \lambda_2, \omega) d\omega d\lambda_1 d\lambda_2 \geq 0 \tag{B.108}$$

$$\int_0^1 \int_{1/2}^1 \int_{\lambda_2}^{\frac{\lambda_1\lambda_2}{(1-\lambda_1)(1-\lambda_2)+\lambda_1\lambda_2}} f(\lambda_1, \lambda_2, \omega) d\omega d\lambda_1 d\lambda_2 - \int_0^1 \int_0^{1/2} \int_{\frac{\lambda_1\lambda_2}{(1-\lambda_1)(1-\lambda_2)+\lambda_1\lambda_2}}^{\lambda_2} f(\lambda_1, \lambda_2, \omega) d\omega d\lambda_1 d\lambda_2 \geq 0 \tag{B.109}$$

$$\int_0^1 \int_{1/2}^1 \int_{1-\lambda_2}^{\frac{\lambda_1(1-\lambda_2)}{(1-\lambda_1)\lambda_2+\lambda_1(1-\lambda_2)}} f(\lambda_1, \lambda_2, \omega) d\omega d\lambda_1 d\lambda_2 - \int_0^1 \int_0^{1/2} \int_{\frac{\lambda_1(1-\lambda_2)}{(1-\lambda_1)\lambda_2+\lambda_1(1-\lambda_2)}}^{1-\lambda_2} f(\lambda_1, \lambda_2, \omega) d\omega d\lambda_1 d\lambda_2 \geq 0. \tag{B.110}$$

Note that if Equations B.107, B.108, B.109, and B.110 are satisfied, then they imply that sender 1 also does not want to deviate to any other messaging strategy that generates plausible frequencies. 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.107, B.108, B.109, and B.110). 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.