

TI 2026-072/VII
Tinbergen Institute Discussion Paper

Truth, Lies, and Social Ties: When Image Concerns Fuel Fake News

*Dana Sisak*¹

*Philipp Denter*²

¹ Erasmus University Rotterdam

² Universidad Carlos III de Madrid

Truth, Lies, and Social Ties: When Image Concerns Fuel Fake News*

Dana Sisak (r) Philipp Denter

September 16, 2026

Abstract

We study how social image concerns shape information sharing among peers. Individuals receive a signal about a binary state with an observable direction and a veracity that can be learned only through costly verification, which depends on type. We analyze two motives: a desire to appear competent and a desire to signal one's worldview. For each, we characterize equilibrium sharing, derive implications for the relative diffusion of true versus false news, and compare welfare. The two motives can both generate greater sharing of false than factual news but imply distinct, testable patterns which we relate to existing empirical evidence.

Keywords: P2P Information Sharing, Social Image, Signaling, Fake News

JEL Codes: D83, D72

*Sisak: Erasmus University Rotterdam and Tinbergen Institute. E-Mail: sisak@ese.eur.nl. Denter: Universidad Carlos III de Madrid. E-Mail: pdenter@eco.uc3m.es. (r) indicates randomized order of author names. We thank Paul Bose, Micael Castanheira, Zuheir Desai, Paul Heidhues, Patricia Himml, Vladimir Karamychev, Jori Korpershoek, David Levine, Agustina Martínez, Kirill Pogorelskiy, Johannes Schneider, Marco Schwarz, Lotte Swank, Otto Swank and the participants of MAPE 2021, PET 2022, KVS New Paper Sessions 2022, the CEPR and Dortmund Political Economy Workshop 2023, the 2023 UC3M Micro Retreat, the 2024 Leuphana Microeconomics Workshop, EPCS 2024, the 2024 ETH/CEPR Workshop on Democracy, SAEe 2024, Political Economy: Theory Meets Empirics 2025, as well as the seminar audiences at Basel, Rotterdam, and King's College London for helpful suggestions. Philipp Denter gratefully acknowledges financial support by the Agencia Estatal de Investigación through grants CEX2021-001181-M and PID2022-141823NA-I00 as well as by the Consejería de Educación, Juventud y Deportes de la Comunidad de Madrid through grant CAM - EPUC3M11 (V PRICIT).

1 Introduction

Billions of people around the world share information with their friends, families, and other peers through social media platforms. The rise of these platforms has significantly reduced the cost of information sharing, making it easier for individuals to connect and interact with others.¹ However, concerns about the quality of shared information and its societal consequences have been mounting in recent years. This is due to the increasingly widespread phenomenon of disinformation, which has been shown to undermine trust in democratic institutions² and to influence electoral outcomes (Zimmermann and Kohring, 2020; Cantarella et al., 2023).³ Disinformation has also contributed to public mistrust in vaccines during the COVID-19 pandemic, leading to suboptimal vaccine uptake in many countries (Montagni et al., 2021).

Unfortunately, the problem of disinformation is not confined to those who create it. Peer-to-peer sharing can dramatically amplify the reach and impact of both factual and false information. Recent research shows that disinformation spreads “*farther, faster, deeper, and more broadly*” than factual content on platforms like Twitter/X (Vosoughi et al., 2018), suggesting that the diffusion of fake news is driven to a great extent by everyday users.

To develop effective remedies against the spread of disinformation and its consequences, it is essential to understand the underlying factors that lead individuals to share information. One important reason why individuals share information is to influence their social image, as shown by Lee et al. (2011), Lee and Ma (2012), and Kümpel et al. (2015). By sharing exclusively high-quality information, individuals can signal that they are *able*—that is, capable of discerning valuable information from dubious content. Conversely, sharing information that later turns out to be false can damage that image and lead to a loss of status. However, information sharing is not only a signal of ability; it can also serve to communicate one’s *worldview*. Sharing content with a particular slant or direction can signal ideological

¹Research has documented the importance of information sharing in various contexts, including the adoption of microfinance (e.g., Banerjee et al., 2013), vaccination campaigns (e.g., Banerjee et al., 2019), and electoral behavior (e.g., Allcott and Gentzkow, 2017; Pogorelskiy and Shum, 2019).

²See, for example, <https://www.brookings.edu/blog/fixgov/2022/07/26/misinformation-is-eroding-the-publics-confidence-in-democracy>.

³See also <https://www.theguardian.com/world/2019/oct/30/whatsapp-fake-news-brazil-election-favoured-jair-bolsonaro-analysis-suggests>.

alignment and indicate support for the perspective it promotes.

In this paper, we build a theoretical model to analyze the conditions under which individuals share information and the resulting consequences when they have social image concerns. We assume an unknown binary state of the world, ω , which could represent, for example, whether human activity drives climate change or whether a vaccine is effective. A sender (S) receives a binary signal σ about ω and can choose whether to share it with a receiver (R). The signal has two dimensions. First, it contains a headline, for instance, one that supports or opposes the idea of human-induced climate change. This headline may be surprising—contradicting prior beliefs—or unsurprising. Only surprising signals, provided they are sufficiently informative, influence the receiver’s decision-making. We refer to this first dimension as the signal’s *relevance*, which is immediately observable to both sender and receiver at no cost.

Second, the signal may be either proper, meaning it is fact-based, or fake, in which case it is fabricated and therefore false. We call this dimension *veracity*. Proper signals are correlated with the true state of the world and thus informative, whereas fake signals convey no information. Determining a signal’s veracity is possible but costly, and only individuals with high ability find it worthwhile to do so.

The sender is motivated by social image concerns: she cares about how she is perceived by her peers. We examine the two types of social image concerns mentioned earlier. First, the desire to be seen as *able*, which can be demonstrated by accurately distinguishing between proper and fake signals. Second, the desire to be perceived as holding a particular *worldview*.

The model is designed to capture a common scenario: an individual deciding whether to share, for example, a newspaper article with another person or a broader audience on social media. The article’s relevance can typically be inferred from its headline alone. Verifying its veracity, however, requires reading the article and engaging with its content—an effort that demands time, attention, and the ability to evaluate arguments critically.

We characterize equilibrium information-sharing behavior driven by each of the two social image motives. Under the ability motive, high-ability senders filter out disinformation and share only signals that are both relevant and based on facts. In contrast, low-ability senders cannot reliably identify fake signals and therefore randomize between sharing rele-

vant content and choosing not to share at all. When the motive is to signal one’s worldview, the sharing pattern changes: only individuals with strong or extreme beliefs are willing to share corresponding signals, while moderates refrain from sharing any signal.

Our main results concern the conditions under which the quality of information deteriorates through the sharing process.

- With an *ability* motive, fake news may be shared disproportionately when improper signals are predominantly *surprising* to receivers. They are shared especially by *low-ability* senders who hold *different worldviews* than their audience. In such cases, senders tend to share only signals that are surprising to receivers and withhold unsurprising ones. Lower sharing costs tend to exacerbate this pattern, making the disproportionate spread of fake news more likely. This may especially hurt low-ability receivers. At the same time, the quality of news *conditional on relevance* always improves under this motive.
- With a *worldview* motive, fake news may be shared disproportionately when improper signals predominantly *align* with the receivers’ prior beliefs. They are shared especially by senders who share *a similar worldview* as their audience. In such cases, senders are more likely to share signals that confirm the receivers’ expectations and withhold those that contradict them. Lower sharing costs encourage sharing by senders with moderate worldview with no clear effect on fake news propagation. Since, conditional on relevance there is no filtering of signals, sharing costs tend to decrease decision quality of both types of receivers by restricting access to information.

A deterioration of information quality after sharing may arise regardless of the underlying social image motive and reveals a fundamental vulnerability in peer-to-peer information transmission. At the same time, when fake news is shared disproportionately, the two social image motives give rise to clearly distinct sharing patterns. As a result, our model generates novel, testable predictions that allow researchers to empirically distinguish between different types of social image concerns, which is important to find effective remedies against the spread of disinformation. The patterns we uncover theoretically align with empirical findings (see Section 6).

Our findings contribute to the ongoing debate on how to reduce fake news sharing on social media. Pennycook et al. (2021) and Guriev et al. (2026) show that nudging users to focus on accuracy is an effective way to lower the spread of disinformation while increasing the spread of proper information.⁴ Our model shows that this may not always be effective, at least in increasing the overall quality of shared news. While in contexts where fake news is shared disproportionately due to a worldview motive, nudging towards accuracy—and thus an ability motive—may work, in other contexts prompting ability signaling decreases the quality of information.⁵ In addition, our model cautions that to keep the accuracy motive salient, senders need to believe that receivers will consistently engage with the information shared.

Literature: This paper contributes to several literatures both in political economy and microeconomic theory. One related strand studies peer-to-peer sharing of non-falsifiable information. A closely related paper from this literature is due to Che and Kartik (2009), who analyze a setting in which a sender decides whether to disclose a piece of information about a binary state to a receiver. In their model, the purpose of information sharing is not to gain social image or status utility—which generates incentives to share information in the present paper—but rather to influence the receiver’s action. Pogorelskiy and Shum (2019) show that voters are more likely to share political news with their peers if the news aligns with their preferences. Levy and Razin (2018) and Bowen et al. (2023) examine the consequences of peer-to-peer information sharing for belief polarization in society. Levy and Razin (2018) show how information sharing can lead to polarization when individuals update according to what they call the “Bayesian Peer Influence heuristic,” while Bowen et al. (2023) demonstrate that even minor misperceptions about the signal-generating process can render collective learning inefficient and result in polarization of beliefs. While similar in spirit, in these papers information sharing is non-strategic. By contrast, our results are driven by individuals’ desire for social image or status utility, which shapes their information-sharing behavior.

⁴In fact, Guriev et al. (2026) show that all of their policy interventions (additional click, fact check available, first assess veracity, accuracy prime) predominantly work through increasing the salience of reputation.

⁵Moreover, the supply and characteristics of fake news are not fixed and may adapt to the prevailing social image motive.

This paper also fits into the growing theoretical literature on the propagation of fake news. Grossman and Helpman (2023) study a model of electoral competition in which political parties can spread fake news about policy positions. They abstract from the sharing of such information among peers as a main driver of the spread of fake news, and sharing serves primarily to exert influence. Acemoglu et al. (2010), Papanastasiou (2020), Denter et al. (2021), Denter and Ginzburg (2026), and Germano et al. (2026) study how misinformation and fake news spread through social networks, though they do not explicitly model the sharing decision.

In contrast, Kranton and McAdams (2024) assume that an individual shares information if and only if she believes, with sufficiently high probability, that the information is genuine. Similarly, Acemoglu et al. (2023) study the decision to dislike or share potentially fake information with one’s peers. Disliking a piece of information brings utility if it turns out to be fake, whereas sharing brings utility if the information turns out to be factual or if peers decide to reshare it. If peers decide to dislike a shared piece of information, one incurs disutility. Unlike in Kranton and McAdams (2024) and Acemoglu et al. (2023), in our paper agents are both horizontally (by ideology) and vertically (by ability) differentiated, and we consider two distinct social-image motives. Moreover, our analysis reveals that the veracity of a piece of information may be orthogonal to it being shared.

One focus of our paper is to determine the circumstances that lead to the sharing of fake news. There exists a wide empirical literature concerned with this questions. Guess et al. (2019) show that over 65 year old individuals and conservative-leaning individuals were most likely to share (mostly pro-Trump) fake news. Pennycook et al. (2021) argue that various motives matter for sharing information on Twitter. When signaling one’s partisanship is salient, copartisan fake news are spread knowingly, but if accuracy becomes more important, fake news sharing is reduced.⁶ Similarly, Osmundsen et al. (2021) find that fake news sharing is often driven by partisan sentiment, rather than by individuals’ ignorance regarding the veracity of a news item. Guriev et al. (2026) show that ability signaling, a desire to persuade receivers, and signaling of partisanship are all drivers of information sharing. Moreover, they

⁶This is in line with recent evidence that people are reasonably good at detecting real from fake news (Angelucci and Prat, 2024).

show that priming the circulation of fake news is an effective tool to reduce sharing of fake news while preserving the dissemination of genuine news. Just as Pennycook et al. (2021) and Guriev et al. (2026), we study the motives of ability and partisanship signaling, using a formal model. Unlike these papers, we model active receivers who may engage with the shared news to ascertain its quality (“fact checking”) and then form beliefs about the type of the sender in Bayesian fashion. Interestingly, we find that even with an accuracy motive, fake news may be spread disproportionately, as found by Vosoughi et al. (2018). We identify and characterize settings in which fake news sharing is most problematic with each motive and derive empirical predictions that can help to distinguish between the different motives.

2 Model

Players. There are two roles in the game: a sender, S , and one or more receivers, R . A sender S receives a signal σ about an unknown state of the world and decides whether to fact-check and share it. We use R to denote a generic receiver. Depending on the application, the sender communicates either with a single receiver, as in a private message on Twitter/ℳ, WhatsApp, or Telegram, or with a large audience, as in a public post or group message on these platforms. We approximate the latter by a continuum of receivers of mass one.

State of the World. At the beginning of the game, nature draws the state of the world $\omega \in \{0, 1\}$. The true probability that $\omega = 1$ is $\mathbb{P}[\omega = 1] = p_T \in (0, 1)$, while $\omega = 0$ occurs with complementary probability $1 - p_T$. Players need not share this objective probability and may hold heterogeneous prior beliefs about the state.

Signal Generating Process. Nature then generates a signal $\sigma \in \{0, 1\}$ that is received by the sender. Each signal has a veracity status $\nu \in \{0, 1\}$, where $\nu = 1$ denotes a *proper* signal and $\nu = 0$ a *fake* signal. With probability $1 - q$, the signal is proper, $\nu = 1$, meaning that it is based on facts and informative about the state. With probability $q \in (0, 1)$, the signal is fake, $\nu = 0$, and carries no information about ω .

If $\nu = 1$, the signal matches the state with probability

$$\mathbb{P}[\sigma = 1 \mid \omega = 1, \nu = 1] = \mathbb{P}[\sigma = 0 \mid \omega = 0, \nu = 1] = \eta \in \left(\frac{1}{2}, 1\right],$$

and is incorrect with complementary probability $1 - \eta$. We refer to η as the *precision* of a proper signal. If $\nu = 0$, the signal is independent of the state. In particular,

$$\mathbb{P}[\sigma = 1 \mid \omega = 1, \nu = 0] = \mathbb{P}[\sigma = 1 \mid \omega = 0, \nu = 0] = \beta \in [0, 1],$$

while $\sigma = 0$ occurs with probability $1 - \beta$. We interpret β as the *bias* of a fake signal. Such a signal may, for example, have been created by a biased agent attempting to influence players' beliefs. The signal-generating process is common knowledge.

Player Types. Before any player observes the signal, her private type $t_i = (p_i, \theta_i)$ is realized. The two components of type are drawn independently, and types are independent across players. Their distributions are common knowledge, while their realizations are private information.

Players differ along two dimensions. First, they may hold different prior beliefs about the state of the world. We assume that an underlying ideological or partisan dimension shapes these beliefs. For example, a player who is generally skeptical of climate change may approach a new environmental policy with skepticism and therefore attach a relatively low prior probability to a favorable state, while a more environmentally minded player may hold the opposite view.⁷ For simplicity, we identify this ideological or partisan dimension directly with player i 's prior $p_i \in [0, 1]$ and refer to it as her *worldview*.⁸ Player i 's worldview is drawn from distribution F_i with strictly positive density f_i on $[0, 1]$, where $i \in \{S, R\}$. Its realization p_i is private information, while F_i is common knowledge.

Second, players differ in their ability to verify information, $\theta_i \in \{L, H\}$. With probability

⁷A behavioral mechanism leading to such divergent priors based on an underlying ideology could be motivated reasoning as, for example, in Taber and Lodge (2006).

⁸This assumption helps us save on notation. Introducing instead a separate parameter capturing a player's worldview—for example a fixed ideology as in Buisseret and Van Weelden (2022), a bliss point as in the electoral competition literature (Downs, 1957), or a bias as in the cheap-talk literature (Crawford and Sobel, 1982)—and allowing it to be imperfectly correlated with p_i would not change our results.

$\lambda_i \in (0, 1)$, player i has high ability, $\theta_i = H$, and can verify a signal at zero cost, $c_F(H) = 0$. With complementary probability $1 - \lambda_i$, she has low ability, $\theta_i = L$, and the cost $c_F(L)$ is prohibitively high. We assume that $\lambda_S + \lambda_R < 1$. Ability is private information, while λ_i is common knowledge. Thus, player i 's private type is $t_i = (p_i, \theta_i) \in \mathcal{T}_i \equiv [0, 1] \times \{L, H\}$.

Fact-checking. After observing the signal and before taking any further action, player i may *fact-check* it. Denote this decision by $v_i(\sigma, t_i) \in \{0, 1\}$, where $v_i = 1$ means that player $i \in \{S, R\}$ checks the signal's veracity. Low-ability players cannot fact-check and therefore choose $v_i = 0$ for any signal. High-ability players, in contrast, always fact-check a signal that is surprising given their worldview p_i . We call a signal *surprising* if its realization runs counter to player i 's worldview p_i .

Fact-checking reveals the signal's veracity status ν . Denote player i 's information about veracity by $\nu_i \in \{0, 1, U\}$, where U indicates that veracity is unknown. Thus, $\nu_i = \nu$ if player i fact-checks and $\nu_i = U$ otherwise. Finally, let $\tilde{q}^i(\sigma, p_i, \nu_i)$ denote player i 's posterior belief that the signal is fake. If $\nu_i = U$, the signal realization may nevertheless be informative, yielding $\tilde{q}^i(\sigma, p_i, U) = \mathbb{P}[\nu = 0 \mid \sigma, p_i] \in [0, 1]$. If instead $\nu_i \in \{0, 1\}$, veracity is known and hence $\tilde{q}^i(\sigma, p_i, \nu_i) = 1 - \nu_i \in \{0, 1\}$.

Sharing and Receiver Actions. After observing the signal and any information obtained through fact-checking, the sender decides whether to share it. Let $\kappa(\sigma, t_S, \tilde{q}^S) \in [0, 1]$ denote the probability that sender S shares signal σ . Thus, sharing may depend on the signal realization, the sender's type, and any information she obtained about the signal's veracity.

If the signal is shared, receiver R observes its realization σ , but not the sender's fact-checking decision or the information obtained through fact-checking. She fact-checks according to the rule described above and then chooses an action $a \in \{0, 1\}$. The receiver wants her action to match the state of the world and obtains $u^R(a) = -|\omega - a| - v_R c_F(\theta_R)$. Let $\mathcal{I}_R$ denote the information available to receiver R after observing a shared signal: the realization σ , receiver R 's own type t_R , and any information ν_R obtained through fact-checking. Then, R chooses $a = 1$ whenever $\mathbb{P}[\omega = 1 \mid \mathcal{I}_R] \geq \frac{1}{2}$, and chooses $a = 0$ otherwise. We assume that she chooses $a = 1$ when indifferent.

Social Image. Sharing a signal entails a direct cost $c_S \geq 0$. The sender derives no direct utility from informing receivers about the state. Instead, the benefit from sharing comes from *social image*: by deciding which signals to share, the sender can influence receivers' beliefs about her type.

We consider two motives for social image. First, a sender may wish to be perceived as having high ability. In this case, her social image vis-à-vis receiver R is R 's posterior belief that $\theta_S = H$. Second, a sender may care about how accurately receiver R perceives her worldview. Denote by $\hat{p}_S = \mathbb{E}[p_S \mid \mathcal{I}_R]$ receiver R 's point estimate of the sender's worldview. A sender with a worldview motive obtains social image utility $-|p_S - \hat{p}_S|$ and therefore benefits when the receiver's point estimate is close to her true ideological or partisan type.

With a single receiver, the sender cares about expected social image, whereas with a large audience, she cares about aggregate social image across receivers. Since receiver types in the audience are distributed as a representative receiver and the audience has mass one, aggregate social image coincides with expected social image.

Equilibrium. Our focus is on Perfect Bayesian Equilibria of the information-sharing game. The sender first decides whether to fact-check the signal and then whether to share it, maximizing her social-image utility net of fact-checking and sharing costs. In doing so, she anticipates how each receiver type reacts to the shared information and what these reactions imply for beliefs about her type. Receivers, in turn, decide whether to fact-check and choose an action $a \in \{0, 1\}$ given the sender's strategy and all information available to them. Beliefs are updated according to Bayes' rule whenever possible and are consistent with equilibrium strategies.

We first analyze the case in which the sender uses information sharing to signal her ability. Section 4 then considers a sender whose sharing incentives are instead driven by her worldview.

3 Information Sharing to Signal Ability

3.1 Equilibrium

In this section we study situations where a sender wants to signal her ability to recognize improper signals. The sender thus chooses her sharing strategy to maximize the probability that she is perceived as a high ability type by R . To focus on the ability motive, we assume that sender types do not also differ in their beliefs p_S .⁹ Since worldviews are fixed throughout this section, we suppress them from the strategy notation. For simplicity, we assume that also receiver types do not differ in their beliefs, though we allow for $p_S \neq p_R$. To focus on situations where social image utility *can* be gained, we restrict attention to $\eta \geq \max\{p_R, 1 - p_R\}$, and thus a proper signal is sufficiently informative to influence the optimal decision of R . These signals can thus be used by S to gain social image utility. Without loss of generality, we assume that $p_R > \frac{1}{2}$, in which case only signal realizations $\sigma = 0$ are surprising and thus relevant for the decision of the receivers.¹⁰

As typical, there are multiple Perfect Bayesian Equilibria of the game. However, in our framework standard equilibrium refinements such as the Intuitive Criterion or D1 cannot help to narrow down the set of possible equilibria. Our interest here is in equilibria where social image utility is gained through signaling of ability. This can be done both by identifying and sharing proper signals, and by identifying and sharing fake signals.¹¹ We assume social image utility is gained through the sharing of proper signals. Then, a high ability sender has a strictly greater incentive to relay a proper and surprising signal than a fake and surprising signal, because the latter will be interpreted by high ability receivers as evidence that the sender has low ability. Therefore, we should expect the high ability sender to choose $\kappa(0, H, 1) = 0$ and never relay surprising, improper signals. The signal that should yield the greatest expected status gain is a surprising and proper signal, and hence we should expect

⁹Alternatively, we can assume that beliefs are common knowledge.

¹⁰While we assume fact-checking of surprising signals by high-ability players in the main analysis for simplicity, Supplementary Appendix A.1 offers a microfoundation of this assumption. The intuition is straightforward. If the signal is not surprising, or if the quality of a proper signal is too low, $\eta < p_R$, then independent of the signal's veracity, the optimal action of the receiver is $a = 1$. However, when the signal is surprising and $\eta \geq p_R$, it has the potential to change a receiver's optimal action. It follows that the player has an incentive to check the veracity, if the cost of checking is not too high.

¹¹In Section A.1 of the Supplementary Appendix we discuss which other types of equilibria cannot exist.

that these signals are always shared, $\kappa(0, H, 0) = 1$.

At the same time, it is unclear whether S should withhold or share a signal that is not surprising. Indeed, there are generally multiple equilibria of the game where social image utility can be gained from sharing proper signals, the major difference between them being if unsurprising signals are shared. In the following, we focus on the equilibrium in which S does not share *any* non-surprising signals.¹²

Proposition 1 (Ability Signaling). *There exist thresholds $0 < c_L < c_H < 1$, a strictly decreasing function $\bar{q} : [0, c_L] \rightarrow [0, 1]$ with $\bar{q}(0) = 1$ and $\bar{q}(c_L) = 0$, and an off-equilibrium belief threshold $\bar{\pi}_1 \in (0, 1)$ such that the following Perfect Bayesian Equilibrium exists for any $\pi_1 \leq \bar{\pi}_1$:*

- *No non-surprising signals are shared. Formally, $\kappa^*(1, H, \tilde{q}^S) = \kappa^*(1, L, \tilde{q}^S) = 0$.*
- *If $c_S \leq c_H$, a high ability sender checks every $\sigma = 0$ signal, shares every proper one, and never shares a fake one. Formally, $v_S^*(0, H) = 1$, $\kappa^*(0, H, 0) = 1$, and $\kappa^*(0, H, 1) = 0$. If $c_S > c_H$, a high ability sender shares no signal.*
- *If $c_S < c_L$ and $q \leq \bar{q}(c_S)$, a low ability sender shares a surprising signal with probability $\kappa^*(0, L, \tilde{q}^S(0, p_S, U)) \in (0, 1)$. If $c_S \geq c_L$ or $q > \bar{q}(c_S)$, a low ability sender shares no information.*

When the share of improper signals q and sharing costs c_S are not too large, sharing a surprising signal with positive probability is beneficial for a low ability sender in equilibrium. Moreover, as long as $c_S < c_L$, the probability that the low ability type shares a surprising signal, κ_0^* , decreases in c_S . Intuitively, a higher c_S makes sharing less attractive relative to remaining silent. A lower κ_0 restores the balance: it raises the status from sharing while lowering the status from silence, leaving the low ability sender indifferent between the two. Moreover, κ_0^* also decreases in p_S . What is the reason for this finding?

¹²This equilibrium is the ex-ante preferred one from the sender's perspective when $c_S > 0$. Bayes-plausibility implies that, in expectation, status remains unchanged. This means that the more signals are shared, the lower the expected utility of the players because expected sharing costs increase. Moreover, equilibria where non-surprising signals are shared with positive probability are less likely to feature a deterioration in the quality of information. Hence, our analysis identifies a worst-case scenario in line with our objective of identifying situations where fake news sharing is most problematic.

Fact 1. *The belief of a sender that a given signal σ is fake increases in $|p_S - \sigma|$.*

As p_S increases, the low ability sender's belief that her signal $\sigma = 0$ ($\sigma = 1$) is fake increases (decreases), which makes sharing less (more) attractive.¹³ All other comparative statics are not as clear cut.

The low ability sender's suspicion of her own signal also explains why she is the first to fall silent as sharing becomes more expensive. Once c_S reaches c_L , she stops sharing altogether while the high ability sender continues. The reason is that she cannot tell what she is passing on: some of the signals she relays are fake, and she loses social image whenever a receiver verifies one. Having checked before sharing, the high ability sender never bears this risk. Sharing is therefore worth less to the low type at any given cost.

3.2 The Quality of Shared Information

We now study the quality of shared information. We measure quality as the fraction of shared information that is fake. Defining the probability of proper news being shared as σ^P and the probability of fake news being shared as σ^F , implies that our measure of the expected quality of shared information equals

$$\gamma \equiv \frac{\sigma^F}{\sigma^F + \sigma^P}. \quad (1)$$

Naturally, when γ decreases, we interpret this as increasing quality, because the probability that a shared signal is fake is lower. Similarly, larger γ means the quality of shared information decreases.

We first study how the availability of social media platforms, which facilitate easy information sharing and therefore decrease the cost of sharing information with our peers, affects the quality of shared information. Many researchers have attributed increasing spread of misinformation to exactly those platforms. The next result shows that decreasing sharing cost indeed increases the share of disinformation of shared content. It also speaks to the effect of worldview polarization, as measured by the worldview of the sender p_S relative to the receiver(s) p_R (which by assumption $p_R > \frac{1}{2}$):

¹³The proofs of both comparative statics results are contained in the proof of Fact 1.

Proposition 2. *Consider the equilibrium described in Proposition 1. γ strictly decreases in both p_S and c_S for $c_S \in [0, c_L)$ and equals zero for $c_S \in [c_L, c_H]$.*

Sharing costs affect the spread of fake information through low-ability senders, since high-ability senders never share fake information. A low-ability sender randomizes between sharing surprising signals and keeping them to herself. When sharing becomes cheaper, it becomes, *ceteris paribus*, more attractive to share. Increasing κ_0 reduces the status gain from sharing and raises the status gain from not sharing, restoring the balance between relaying and withholding information. As a result, higher sharing costs reduce γ and improve the quality of shared information.

Also the sender's worldview p_S matters for the quality of information shared. The greater the sender's belief that $\omega = 1$, and thus the higher p_S , the lower κ_0^* , as explained above. Thus, low types are more conservative in their sharing decision, which improves the quality of information shared. Consequently, γ decreases in p_S . Recall that $p_R > \frac{1}{2}$. Then, γ is highest when $p_S \rightarrow 0$, and thus the sender and the receiver(s) maximally disagree in their belief about the state ω .

But how low can the quality of information after sharing become? Is the fraction of shared information that is fake always smaller than the expected fraction of fake information received by the sender, $\gamma < q$? This would imply that the quality of shared information increases due to filtering by senders. This is not necessarily the case. Figure 1 shows how $\gamma - q$ changes as a function of $q \in [\frac{1}{10}, \frac{9}{10}]$, $\beta \in [\frac{1}{10}, \frac{9}{10}]$, and $\lambda_S \in \{\frac{1}{10}, \frac{1}{5}\}$ when $c_S = 0$, $\eta = \frac{2}{3}$, $\lambda_R = \frac{1}{5}$, and $p_S = p_R = p_T = \frac{2}{3}$.¹⁴ We can see that in the example information quality deteriorates after sharing, $\gamma > q$, when q and β are small and thus fake news are not prevalent and biased towards the surprising state. Moreover, when the share of high ability senders λ_S decreases, the parameter range such that $\gamma > q$ increases. To shed further light on which situations are problematic, the next proposition offers a sufficient condition for the quality of information to improve after sharing.

Proposition 3. *Consider the equilibrium identified in Proposition 1. A sufficient condition*

¹⁴For these parameters, $0 < \kappa_0^* < 1$ and thus the equilibrium characterized in Proposition 1 exists.

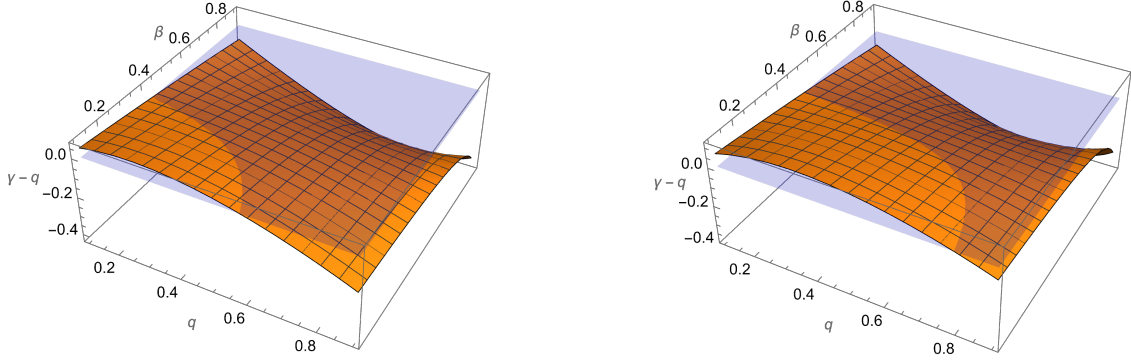


Figure 1: $\gamma - q$ as a function of q and β when $\lambda_S = \frac{1}{5}$ (left panel) and $\lambda_S = \frac{1}{10}$ (right panel), as well as $c_S = 0$, $\eta = \frac{2}{3}$, $\lambda_R = \frac{1}{5}$, and $p_S = p_R = p_T = \frac{2}{3}$. The light-shaded blue plane divides the positive and negative halfspace and marks $\gamma - q = 0$.

for $\gamma < q$ is $c_S < c_L$ and

$$\beta > 1 - \frac{p_T(1 - \eta) + (1 - p_T)\eta}{1 - \lambda_S},$$

or $c_S \in [c_L, c_H]$.

Thus, only when fake news are biased towards the surprising state, i.e., if β sufficiently small, is there scope for information quality deterioration. The same holds for sufficiently many low-ability senders, that is, λ_S small, because it's them who share fake signals, as well as a high true probability that the state is 1 (p_T large) and thus surprising signals $\sigma = 0$ are likely fake.¹⁵ To conclude, the quality of information shared relative to information received tends to be worse when β , p_S , λ_S and c_S are low.

While we established that the quality of information after sharing can deteriorate under an ability motive, our model also highlights the limitations of this measure. Receivers in our setting only care about relevant signals, as only those are pivotal for their decision. The quality of relevant signals unambiguously increases under the ability motive, due to filtering by high-ability senders. The decrease in quality is thus an artifact of the filtering of signals by *relevance*. This is in line with the empirical evidence in Vosoughi et al. (2018) who show that fake news, especially in the political domain, spread “*farther, faster, deeper, and more*

¹⁵For example, when $\lambda_S = \frac{1}{10}$, $\eta = \frac{2}{3}$ and $p_T = \frac{2}{3}$, we get a sufficient condition of $\beta > \frac{41}{81} = 0.506$ (right graph) while when λ_S increases to $\frac{1}{5}$, we get $\beta > \frac{4}{9} = 0.444$ (left graph).

broadly” on Twitter/ℳ than factual news.¹⁶ They also find that false news are more novel than proper news— β is small relative to p_T in terms of our model—and that people were more likely to share novel information, which is in line with the equilibrium in Proposition 1.

3.3 Welfare

We now examine how higher sharing costs affect receiver welfare. By discouraging low-ability senders from sharing, these costs raise information quality. Also empirically such a policy has been shown to be effective in reducing the sharing of fake news (Guriev et al., 2026).¹⁷ Since fact-checking is costless for high ability R , our measure of receiver welfare is the probability of a correct choice. Because high ability receivers fact check relevant signals they are hurt by an increase in sharing costs, because such a policy also weakly decreases the amount of proper signals shared with them. The absolute number of proper and surprising signals shared with them is the sole determinant of their welfare.

To the contrary, for low ability receivers, an increase in sharing costs may be *welfare improving*. A necessary condition for this is that $c_S < c_L$ and that R finds surprising signals shared with her sufficiently informative, leading her to change her optimal decision. When signals shared by low ability S are of low quality, because fake news are relatively prevalent (q is high) and/or biased to be surprising to R (β is small), higher sharing costs may improve low ability R ’s decisions in expectation. Thus, a policy affecting sharing costs is most likely welfare improving if there are many low ability R and S , shared signals by low ability S are of bad quality but low ability R still sufficiently trust the information they obtain to use it in their decision making.

¹⁶While they consider the differential spread of false relative to true news, we consider the quality of information before and after sharing. If fake signals are reshared more often than proper ones, this is equivalent to the quality of information deteriorating after sharing.

¹⁷We provide a formal analysis to complement this discussion as well as additional results on sender welfare in our Supplementary Appendix A.2.

4 Information Sharing to Signal Worldview

4.1 Equilibrium

To study information sharing as a means of signaling worldview, we now relax the assumption that the sender’s worldview p_S is common knowledge. Instead, the sender’s prior belief is distributed according to F_S , as introduced in Section 2. We interpret S ’s desire to signal her worldview to R as the wish to induce in R an estimated worldview $\hat{p}_S$ that is as close as possible to her actual worldview. Her social image utility equals

$$u^S(p_S) = - \int_0^1 |p_S - \hat{p}_S(p_R)| dF_R(p_R),$$

where $\hat{p}_S(p_R)$ denotes the perceived worldview of S by a receiver with worldview p_R .

As in most games of information transmission, there exists an equilibrium in which receivers ignore any information sent by S and do not update their beliefs about the sender’s worldview. Hence, in turn, it is optimal for S not to send any signal. We instead focus on equilibria in which R responds to the sender’s sharing decision. We refer to such an equilibrium as a *Responsive Equilibrium*. We focus on equilibria with natural language: sharing $\sigma = 1$ should weakly increase R ’s estimate of the worldview of S , while sharing $\sigma = 0$ should weakly decrease it, relative to not sharing anything.

While high-ability senders may learn the veracity of surprising signals, veracity does not confer differential social image utility, as we argue below. Hence, although fact-checking may occur, S ’s sharing decision does not condition on veracity.¹⁸ Therefore, for forming beliefs about S ’s worldview, R notes only whether a signal is sent and what the “headline” of the signal is, $\sigma \in \{0, 1\}$.

We will focus on responsive equilibria of the following form:

¹⁸To illustrate why it is not possible to gain differential social image utility from proper versus improper signals, assume that all senders and receivers observe veracity at no cost. Equilibrium then requires two versions of C_l as defined below, one for proper signals with threshold p_{Sl}^p and one for improper ones with threshold p_{Sl}^i , and analogously for C_h . Since social image utility from not sharing, $\hat{p}_S(\emptyset)$, is identical across all four conditions, and as there are no differential costs of sharing proper versus improper signals, it follows that $p_{Sl}^p = p_{Sl}^i$ and $p_{Sh}^p = p_{Sh}^i$. Note that this would be different in a mixed-motive setting as discussed in Section 5, where both ability and worldview social image utility can be gained. As improper signals are costly on the ability dimension, they could act as stronger signals of worldview.

- A sender that receives a signal $\sigma = 0$ shares the signal if and only if her worldview p_S lies in $[0, p_{Sl}]$ for some $p_{Sl} \in (0, 1)$;
- A sender that receives a signal $\sigma = 1$ shares the signal if and only if her worldview p_S lies in $[p_{Sh}, 1]$ for some $p_{Sh} \in (0, 1)$.

Denote by $\hat{p}_R = (1-q)(\eta p_R + (1-\eta)(1-p_R)) + q\beta$ a receiver's belief about the probability that the sender receives a signal $\sigma = 1$. In a *Responsive Equilibrium*, R 's posterior beliefs about the worldview of the sender—after observing either $\sigma = 0$, $\sigma = 1$, or no signal ($\emptyset$)—are given by

$$\begin{aligned}\hat{p}_S(0) &= \frac{\int_0^{p_{Sl}} p_S dF_S(p_S)}{F_S(p_{Sl})}, & \hat{p}_S(1) &= \frac{\int_{p_{Sh}}^1 p_S dF_S(p_S)}{1 - F_S(p_{Sh})}, \\ \hat{p}_S(\emptyset) &= \frac{\hat{p}_R \int_0^{p_{Sh}} p_S dF_S(p_S) + (1 - \hat{p}_R) \int_{p_{Sl}}^1 p_S dF_S(p_S)}{\hat{p}_R F_S(p_{Sh}) + (1 - \hat{p}_R)(1 - F_S(p_{Sl}))}.\end{aligned}$$

If a Responsive Equilibrium exists, receiving $\sigma = 1$ leads R to update her belief about the sender's worldview upwards, while receiving $\sigma = 0$ leads her to adjust it downwards. Note the following important fact: $\hat{p}_S(\emptyset)$ depends on the worldview of the receiver, whereas $\hat{p}_S(0)$ and $\hat{p}_S(1)$ do not. The reason is that, in the latter cases, R learns the realization of the signal. In a Responsive Equilibrium, the following two indifference conditions must hold and determine p_{Sl} and p_{Sh} :

$$C_l \equiv -|p_{Sl} - \hat{p}_S(0)| - c_S + \int_0^1 |\hat{p}_S(\emptyset) - p_{Sl}| dF_R(p_R) = 0, \quad (2)$$

$$C_h \equiv -|p_{Sh} - \hat{p}_S(1)| - c_S + \int_0^1 |\hat{p}_S(\emptyset) - p_{Sh}| dF_R(p_R) = 0. \quad (3)$$

The first condition implies that a sender with worldview p_{Sl} is indifferent between sharing $\sigma = 0$ and not sharing her signal, while the second condition implies the same for sharing $\sigma = 1$ and not sharing when the sender's worldview is p_{Sh} . A Responsive Equilibrium is a pair (p_{Sl}^*, p_{Sh}^*) such that both conditions are satisfied simultaneously. Our next result shows that, if c_S is not too large, then a Responsive Equilibrium always exists:

Proposition 4. *Define $\bar{c}_S \equiv \min\{1 - \xi, \mathbb{E}[p_S \mid p_S \leq \xi]\}$. There exists $\xi \in (0, 1)$, determined by F_S , such that for all $c_S \in [0, \bar{c}_S)$, an interior Responsive Equilibrium exists with*

$$(p_{Sl}^*, p_{Sh}^*) \in (0, \xi) \times (\xi, 1).$$

If the condition on c_S in the proposition is violated, then an interior equilibrium may not exist. In that case, either nobody shares $\sigma = 1$, nobody shares $\sigma = 0$, or no signals are shared at all. To facilitate the further analysis, we first assume that there is a *single receiver* with a *known worldview* p_R .¹⁹ This simplifies matters significantly, because we can show that in any interior equilibrium, it holds that $p_{Sh}^* > \hat{p}_S^*(\emptyset) > p_{Sl}^*$ —see Lemma 2 in Appendix A.6. Then, the equilibrium conditions (2) and (3) can be written as

$$\tilde{C}_l \equiv -(p_{Sl} - \hat{p}_S(0)) - c_S + (\hat{p}_S(\emptyset) - p_{Sl}) = 0, \quad (4)$$

$$\tilde{C}_h \equiv -(\hat{p}_S(1) - p_{Sh}) - c_S + (p_{Sh} - \hat{p}_S(\emptyset)) = 0. \quad (5)$$

Equipped with these simplified conditions, we now study how sharing costs c_S and the receiver's worldview p_R influence the equilibrium (p_{Sl}^*, p_{Sh}^*) . We henceforth assume that $c_S \leq \bar{c}_S$, implying that an interior Responsive Equilibrium exists.

To make statements about the comparative statics, we need to guarantee that a *unique* Responsive Equilibrium exists. Under Assumption 1 below, this is generally the case.

Assumption 1. $\frac{\partial \tilde{C}_l}{\partial p_{Sl}} < 0$, $\frac{\partial \tilde{C}_h}{\partial p_{Sh}} > 0$, and $\frac{\partial \tilde{C}_l}{\partial p_{Sl}} \frac{\partial \tilde{C}_h}{\partial p_{Sh}} - \frac{\partial \tilde{C}_l}{\partial p_{Sh}} \frac{\partial \tilde{C}_h}{\partial p_{Sl}} < 0$ for all $(p_{Sl}, p_{Sh}) \in [0, \xi] \times [\xi, 1]$.

These assumptions ensure that the Jacobian of the equilibrium conditions $(-\tilde{C}_l, \tilde{C}_h)$ is a P-matrix, in which case uniqueness of the Responsive Equilibrium follows from Gale and Nikaido (1965). Note that Assumption 1 is not very restrictive. Figure 2 illustrates several examples of f_S for which Assumption 1 holds.²⁰ For the remainder of the analysis, we assume that Assumption 1 is satisfied.

Under Assumption 1, the comparative statics of the equilibrium thresholds (p_{Sl}^*, p_{Sh}^*) with respect to the receiver's worldview p_R and sharing costs c_S are clear:

¹⁹Alternatively, we can assume multiple receivers with identical worldview p_R .

²⁰It is cumbersome to find general conditions on F_S that guarantee Assumption 1 without being overly restrictive. In Supplementary Appendix B.1, we show that it generally holds if F_S is uniform on $[\underline{p}_S, \bar{p}_S] \subseteq [0, 1]$. Moreover, in Supplementary Appendix B.2 we show that when $\frac{f_S(p_S)p_S}{F_S(p_S)} \in [1, 2]$ for $p_S \in [0, \frac{1}{2}]$, Assumption 1 holds in symmetric games, i.e., when f_S is symmetric around $\frac{1}{2}$ and $\hat{p}_R = \frac{1}{2}$.

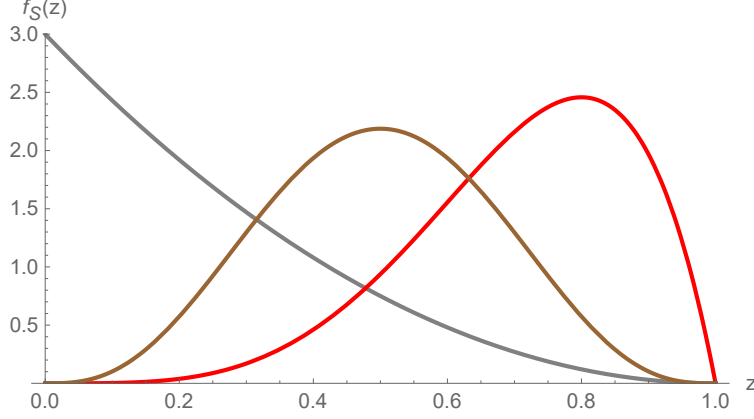


Figure 2: Possible shapes of f_S such that Assumption 1 is satisfied.

Proposition 5. *With a unique receiver with known worldview p_R , in any interior equilibrium the following holds: $\frac{\partial p_{Sl}^*}{\partial p_R} < 0$, $\frac{\partial p_{Sh}^*}{\partial p_R} < 0$, $\frac{\partial p_{Sl}^*}{\partial c_S} < 0$, and $\frac{\partial p_{Sh}^*}{\partial c_S} > 0$.*

The proposition reveals that an *echo chamber effect* emerges endogenously: when the receiver's prior belief about ω (p_R) increases, she receives more signals conforming with her prior, and fewer contradicting her prior. Therefore, to an external observer, the behavior of S may appear to stem from a conformity motive, because she adjusts her strategy toward the receiver's belief. However, it actually stems from moderate sender types, who are wary of being perceived as extreme when facing a receiver with extreme beliefs. Since these receivers expect signals to be aligned with their belief, not sharing is a strong indication of opposing beliefs.

Higher sharing costs c_S , on the other hand, imply that both $\sigma = 1$ and $\sigma = 0$ are shared less frequently. Intuitively, when c_S increases, the utility differential between sharing and not sharing needs to increase for sharing to remain worthwhile. A lower (higher) sharing threshold p_{Sl}^* (p_{Sh}^*) achieves exactly that: for a sender slightly below (above) the sharing threshold, the utility from sharing increases, while the utility from not sharing remains approximately constant.

4.2 Multiple Heterogeneous Receivers

We now discuss what happens when there are multiple receivers with differing worldviews, or when the receiver's worldview is not precisely known. Intuitively, the comparative statics

with respect to c_S are qualitatively unchanged to those in Proposition 5. However, when the distribution of p_R changes, things are not as clear unless we make further assumptions. Our last result of this section is a direct corollary of Proposition 5 and shows that the described echo chamber effect remains valid if receivers are ideologically sufficiently homogeneous and if there is a population wide ideological drift:

Corollary 1. *Consider two distinct distributions of receiver ideologies, $F_R(p_R)$ with support $[\underline{\alpha}, \bar{\alpha}] \subseteq [0, 1]$ and $\hat{F}_R(p_R)$ with support $[\underline{\beta}, \bar{\beta}] \subseteq [0, 1]$. Assume we can rank the two distributions by first-order stochastic dominance: $\hat{F}_R(p_R) \leq F_R(p_R)$. If the receiver population under each distribution is sufficiently homogeneous—both $|\bar{\alpha} - \underline{\alpha}|$ and $|\bar{\beta} - \underline{\beta}|$ are sufficiently small—then $p_{Sl}^*[\hat{F}_R] < p_{Sl}^*[F_R]$ and $p_{Sh}^*[\hat{F}_R] < p_{Sh}^*[F_R]$ in any interior equilibrium.*

This result directly follows from Proposition 5. Intuitively, starting at F_R and moving towards $\hat{F}_R$, we successively increase the ideology p_R of some receivers. Because under both distributions audiences are sufficiently homogeneous, the ranking $p_{Sh}^* > \hat{p}_S^*(\emptyset) > p_{Sl}^*$ holds for any p_R , and thus each of these increases leads to lower equilibrium thresholds by the echo chamber effect discussed above. Hence, the sum of all of these shifts needs to lead to lower sharing thresholds p_{Sl} and p_{Sh} as well.

With a sufficiently heterogeneous audience, it will often be the case that $p_{Sh}^* > \hat{p}_S^*(\emptyset) > p_{Sl}^*$ does not hold anymore for some (extreme) receiver types, implying this logic is not valid anymore. However, this does not imply that our comparative static result does not hold anymore. In Supplementary Appendix B.3 we show in an example that for a maximally polarized audience where $\hat{p}_R \in \{0, 1\}$, $p_{Sh}^* > \hat{p}_S^*(\emptyset) > p_{Sl}^*$ is always violated for one of the receiver types, but the comparative static in Corollary 1 still holds.

4.3 The Quality of Shared Information

We now study which situations are conducive to the spread of fake news. As in Section 3, we use the share of improper signals relative to all signals shared as a measure of the quality of information. This fraction equals

$$\gamma = \frac{q(1 - \beta)F_S(p_{sl}) + q\beta(1 - F_S(p_{sh}))}{\mathbb{P}(\sigma = 0)F_S(p_{sl}) + \mathbb{P}(\sigma = 1)(1 - F_S(p_{sh}))}, \quad (6)$$

where $\mathbb{P}(\sigma = 1) = q\beta + (1 - q)[p_T\eta + (1 - p_T)(1 - \eta)]$ and $\mathbb{P}(\sigma = 0) = 1 - \mathbb{P}(\sigma = 1)$ are the true probabilities of receiving signals with the different realizations. As in Section 3, γ can be interpreted as an inverse measure of the quality of information after sharing. The higher is γ , the lower the quality of information. Furthermore, whenever $\gamma > q$, the quality of information deteriorates after sharing.

For simplicity, we maintain our assumption of one receiver with known worldview p_R .²¹ We now state our first result on the spread of fake news when social image concerns regarding one's worldview are relevant.

Proposition 6. *Define $\hat{\beta} \equiv p_T\eta + (1 - p_T)(1 - \eta)$. If $\beta = \hat{\beta}$, then the quality of information does not change after sharing, $\gamma = q$. Moreover, if $\beta > \hat{\beta}$ ($\beta < \hat{\beta}$), then γ increases (decreases) in p_R .*

$\hat{\beta}$ denotes the bias of improper signals towards 1 at which signal realization $\sigma = 1$ is equally likely among proper and improper signals. Since S does not condition on veracity when making her sharing decision, but only on signal realization σ , the quality of information conditional on σ does not change after sharing. If $\beta > \hat{\beta}$, $\sigma = 1$ is more likely among improper signals than among proper signals, while the opposite holds when $\beta < \hat{\beta}$. If a realization that is more likely to be fake than the other is relayed by the sender more frequently, this will deteriorate information quality.

Proposition 5 shows that there is a positive correlation between the worldview of the receiver and the amount of signals she receives that concur with her worldview. Thus, when social image concerns revolve around worldview, the problem of fake news sharing is especially relevant when the worldview of the receiver is aligned with the bias of the fake news shared. This is in contrast to the ability motive, where the problem of fake news sharing was especially relevant when fake news were biased against the belief of the receiver

²¹We can again generalize our results to receiver populations that are sufficiently homogeneous, using analogous arguments as for Corollary 1.

and surprising signals were shared disproportionately.

The next proposition states when sharing leads to a deterioration of the quality of information:

Proposition 7. *In any interior equilibrium, the following holds:*

$$\text{Sign}[\gamma - q] = \text{Sign}\left[(\hat{\beta} - \beta)[F_S(p_{Sl}^*) - (1 - F_S(p_{Sh}^*))]\right].$$

From Proposition 7 it is clear that information deteriorates through sharing ($\gamma > q$) if fake signals are biased towards $x \in \{0, 1\}$ and $\sigma = x$ is shared more often than $\sigma = 1 - x$.

While the receiver's worldview p_R is an important determinant of the sign of $F_S(p_{Sl}^*) - (1 - F_S(p_{Sh}^*))$, as also shown in Proposition 5, the sender's strategy depends on it only through $\hat{p}_R$, the receiver's belief that the sender received a signal $\sigma = 1$. If fake news are prevalent and strongly biased, $\hat{p}_R \approx q\beta$. For example, if R expects $\sigma = 1$ because $q\beta$ is very high, moderate types are reluctant not to share such signals, as not sharing is a strong signal of having an opposing worldview. Because of this echo chamber effect, for high q the quality of information after sharing is likely to be lower than before sharing.

To illustrate this effect, in Figure 3 we plot $\gamma - q$ assuming that p_S follows a uniform distribution on $[0, 1]$. For low values of q we see that the quality of information deteriorates for $\beta < \frac{1}{2} = \hat{\beta}$. For high values of q , on the other hand, information quality deteriorates for any β that is not too close to $\frac{1}{2}$. For high enough q , the belief of the receiver about ω is not very relevant for her belief about the signal realization. She will expect more $\sigma = 1$ when $\beta > \frac{1}{2}$ and more $\sigma = 0$ else. Social image concerns about one's worldview together with a high prevalence of fake news thus provide a fertile ground for excessive fake news sharing.

4.4 Receiver Welfare

Finally, we turn to receiver welfare. First of all, signals that align with her prior are not relevant to the receiver, and as we showed, relevant signals are shared relatively infrequently with her. At the same time, senders do not “filter” signals conditional on their headline, thus the quality of information before and after sharing *conditional on* σ is the same. This also implies that increasing sharing cost c_S , which discourages sharing on the margin, will weakly

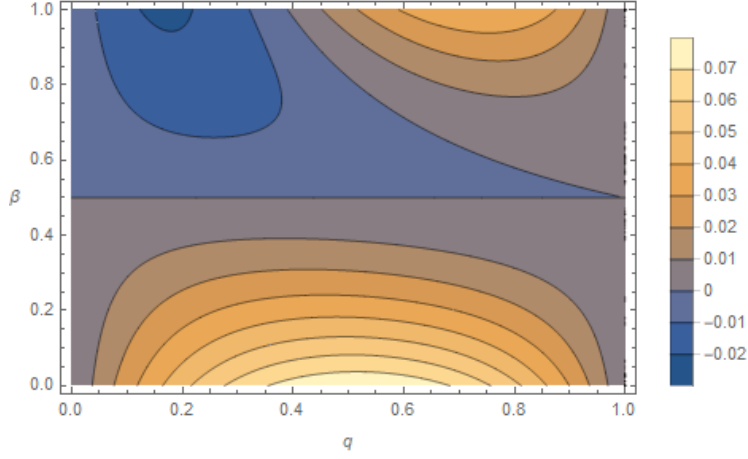


Figure 3: Quality of signals after relative to before sharing ($\gamma - q$) as a function of q and β , given $\eta = \frac{9}{10}$, $p_R = \frac{1}{10}$ and $p_T = \frac{1}{2}$. Positive values (yellow tones) denote a deterioration of information quality after sharing.

decrease receiver welfare as less signals are shared which are potentially informative. This is in contrast to ability signaling, where sharing costs differentially discourage low-ability senders from sharing and thus increase the quality of information shared conditional on σ , potentially increasing welfare of low-ability types.

While senders do not selectively share fake or proper signals, they do tailor their strategy to the receiver's worldview, such that receivers with extreme views receive more signals aligned with their worldview than moderates. In our model, rational receivers also update their belief from not receiving a signal. If this were not the case, for example because of selection neglect, we should see an increase in belief polarization as in Bowen et al. (2023) because of the worldview motive. Our worldview model thus provides a micro-foundation for selective sharing in low-attention information environments.

5 Combining Ability and Worldview Motives

The analyses in Sections 3 and 4 deliberately separated the two social image motives to provide a rich analysis of when fake news sharing is especially problematic under each motive. In this section we combine the two motives in a simplified framework to study how they interact. Assume that a fraction $\mu \in (0, 1)$ of high-ability senders cares about looking competent, as in Section 3 while the remaining high-ability senders and all low-ability

senders care about signaling their worldview, as in Section 4. We characterize equilibria that illustrate how different types of signals serve to signal different individual characteristics and how this can foster fake news sharing by high-ability types.

In contrast to the main model, we now assume that not just ability but also worldview is binary, $p_i \in \{p_L, p_H\}$, $i = \{R, S\}$. In addition, we assume symmetry in that a) both sender and receiver are equally likely to be of high or low worldview type and b) these types are equidistant from one half, $p_L = 1 - p_H$ and $0 < p_L < 1/2 < p_H < 1$. In addition, both state realizations are equally likely, $p_T = \frac{1}{2}$, and fake signals are unbiased, $\beta = \frac{1}{2}$. For simplicity, also assume that all receivers are of the high-ability type and verify *all* signals.²²

Under these assumptions, the model is fully symmetric: the population of senders and receivers is balanced across worldviews, and both proper and fake signals are equally likely to carry either content, $\sigma = 1$ or $\sigma = 0$. While this symmetry does not serve to study the severity of fake news sharing generally, which we learned in previous sections arises typically when either fake signals are sufficiently biased or sender and receiver worldviews are extreme and either conform or conflict, it serves to succinctly illustrate an additional channel through which fake news sharing becomes problematic. It highlights how receivers infer multiple characteristics from one sharing decision, and how different types of signals are thus useful for signaling different characteristics.

As in Section 4 we focus on equilibria with natural language and thus $\sigma = 1$ should weakly increase R 's estimate of S 's worldview (towards p_H) while $\sigma = 0$ should weakly decrease it (towards p_L). We thus define *congruent* signals as $\sigma = 0$ for $p_S = p_L$ types and $\sigma = 1$ for $p_S = p_H$ types. The following proposition states our result pertaining to this framework²³:

Proposition 8. *Consider the symmetric set-up described above. There exist sharing cost thresholds*

$$0 < c_S^{ii} < c_S^i$$

such that the following holds.

²²We do this to prevent the analysis from becoming overly long. Similar results can be obtained for λ_R and β as free parameters and under the original fact-checking assumption introduced in Section 2. Details are available upon request.

²³The proof can be found in Supplementary Appendix C.

(i) If $c_S \leq c_S^i$, there exists an equilibrium in which worldview-motivated high-ability senders do not check congruent signals and share all congruent signals, both proper and fake.

(ii) If $c_S \geq c_S^{ii}$ but sufficiently small, as well as $\lambda_S \geq \frac{1}{5}$, $10\mu \geq 5 + \frac{1}{\lambda_S}$ and $q \leq \frac{1+5\lambda_S-10\mu\lambda_S}{3\lambda_S-9\mu\lambda_S}$, there exists an equilibrium in which worldview-motivated high-ability senders check the veracity status of congruent signals and share only fake congruent signals and do not share proper congruent and all incongruent signals.

In both cases, low-ability (worldview-motivated) senders share congruent content only, ability-motivated high-ability senders check all signals and share all proper signals and no fake signals.

As in Section 3, ability-motivated high-ability types filter signals and thus improve the quality of information after sharing. As in Section 4, worldview-motivated low-ability types do not filter information by quality, only filtering on content. In addition, we now find that the presence of sufficiently many ability-motivated high-ability types using proper signals to signal ability increases the incentives of worldview-motivated high-ability types to share *improper* signals. In the extreme case, the proposition shows that high-ability types pursuing worldview image concerns might only share fake signals as sharing costs increase, thus deteriorating the quality of information after sharing.²⁴

6 Discussion

Our analysis shows that both social image motives can lead to disproportionate sharing of fake news. The motives, however, generate distinct sharing patterns, both in who shares fake news and in how its content relates to receivers' beliefs. Our empirical predictions build on the idea that different social image motives are more salient in different environments. We therefore focus on the benchmark models in Sections 3 and 4, which isolate the implications

²⁴Note that for low enough sharing costs also an equilibrium exists, where ability-motivated types never share, while worldview-motivated types only share congruent signals and perfectly reveal their type. Thus, ability-signaling through sharing is shut down in this equilibrium. The reverse is also possible for low enough sharing costs: ability-motivated senders share only proper signals while worldview-motivated senders do not share. Off-equilibrium beliefs after observing fake signals are equal to the prior.

of each motive. The simplified mixed-motive framework in Section 5 is primarily meant to illustrate how the two motives can interact rather than to deliver general empirical predictions. We now formulate the corresponding hypotheses and compare them with existing evidence.

We begin by formulating hypotheses about sharing behaviour conditional on the different motives to share information. With an *ability signaling motive*, the following hypotheses follow from our analysis:

Hypothesis A. *Consider a salient ability signaling motive.*

- (a) *Fake news are shared predominantly by users with a low ability.*
- (b) *Fake news are typically novel/surprising for the intended audience.*
- (c) *Fake news are mostly shared by users with opposing ideology as the intended receiver(s).*

With a *worldview signaling motive*, the hypotheses derived from our model are distinct:

Hypothesis W. *Consider a salient worldview signaling motive.*

- (a) *There is little to no engagement with the shared news, and the expected quality of the news is irrelevant for the sharing decision.*
- (b) *Moderates tend to share no information, whereas those on the extremes of the ideological spectrum share information in line with their belief, both fake and proper.*
- (c) *Fake news are mostly shared by users with similar ideology as the intended receiver(s).*
- (d) *The propensity to share fake news is orthogonal to ability.*

To distinguish between the two motives empirically, we consider how their salience varies across social networks. Closed networks, such as Facebook and Instagram, consist largely of pre-existing social ties, whereas audiences on open networks such as Twitter/X tend to be more heterogeneous. Audience homogeneity and familiarity facilitate identity-driven self-presentation and self-disclosure (Bazarova and Choi, 2014; van der Does et al., 2022; Altman and Taylor, 1973; Greene et al., 2006). We therefore expect worldview signaling

to be relatively more salient in closed networks. Consistent with this argument, Marwick and boyd (2011) document greater self-censorship on Twitter/X, while the platform is used disproportionately for disseminating scholarly work (Erdt et al., 2016; Enkhbayar et al., 2020), which is more consistent with ability signaling.

We first consider open or public networks, where we expect ability signaling to be relatively salient. Guriev et al. (2026) estimate a structural model of news sharing on Twitter/X and find ability signaling to be the empirically most important motive behind sharing. This is consistent with the premise underlying Hypothesis A. Further evidence comes from Vosoughi et al. (2018). Using Twitter/X data, they show that false news is more novel than factual news and that users who spread false news have significantly fewer followers. If follower counts are positively related to ability, both of these are consistent with Hypothesis A as well.²⁵

Turning now to evidence from closed networks, where we expect worldview signaling to be salient. Sundar et al. (2025) analyze over 35 million Facebook posts containing URLs shared between 2017 and 2020. They find that around 75% are “*shares without clicks*,” and thus most senders do not engage with the shared content themselves. Differences in the propensity of sharing false news without previously clicking on the URL between individuals of different groups are found to be due to differences in the number of false URLs originating from news domains frequented by those groups. This is consistent with Part (a) of Hypothesis W. Moreover, copartisan political content receives more shares without clicks, with partisans engaging in it more than politically neutral users. Guess et al. (2019) and Guess et al. (2023) come to a similar conclusion, also using data from Facebook. This is consistent with Part (b) of Hypothesis W.

In surveys run on MTurk, Pennycook et al. (2021) find further evidence for this hypothesis, asking individuals if they intend to share a certain piece of information: “*Whether the headline was politically concordant or discordant had a significantly larger effect on sharing intentions [...] than whether the headline was true.*” Moreover, “*participants were more than twice as likely to consider sharing false but politically concordant headlines as they*

²⁵Greater network centrality has been found to correlate positively with educational and academic achievement; see, for example, Park et al. (2020), Williams et al. (2019), and Vignery and Laurier (2020).

were to rate such headlines as accurate.” It appears that Pennycook et al. (2021) did not clearly specify on which social network participants were to share these headlines. Therefore, it seems likely that a majority of the subjects had a closed network in mind, as the by far largest network in the U.S. at the time was Facebook, with more than three times as many users as Twitter.²⁶ This suggests worldview signaling was more salient at the time of the survey, and their results are consistent with Hypothesis W. However, they also show that by explicitly increasing the salience of the ability motive through accuracy nudges, the willingness to share misinformation decreased significantly among survey respondents. An accompanying Twitter intervention had the same result.

The effectiveness of such accuracy nudges can be explained through different mechanisms. Stein et al. (2024) show that accuracy beliefs are resistant to accuracy incentives. Instead, we argue that the salience of ability signaling increases through accuracy nudges. In a follow-up study to Pennycook et al. (2021), Lin et al. (2023) show exactly that: these behavioral changes are achieved by increasing “*the weight participants put on accuracy while deliberating.*” Guriev et al. (2026) study a similar accuracy nudge intervention (among others) and show through structural estimation that it indeed works by shifting weight toward reputational (ability) concerns in the sharing decision.

In line with these findings, our model suggests that in situations where fake news is shared disproportionately under a worldview motive, nudging towards accuracy—and thus an ability motive—may increase the quality of shared information. This will be the case if the bias of the fake news aligns with the receiver’s beliefs; in such a situation, under a worldview motive fake news sharing can be significant (Hypothesis W), whereas it tends to be more limited under an ability motive (Hypothesis A). However, in situations where fake news tends to be surprising, worldview signaling may lead to higher overall quality of shared information than ability signaling. However, note that this is driven to a large extent by fake *decision-irrelevant* signals—that is, signals that do not change the receiver’s action irrespective of their veracity—which are shared under a worldview motive but not under an ability motive. The quality of information *conditional on relevance* should generally increase

²⁶See <https://www.pewresearch.org/short-reads/2019/04/10/share-of-u-s-adults-using-socialmedia-including-facebook-is-mostly-unchanged-since-2018/>.

under such a nudge. Therefore, even though accuracy nudges may lead to greater spread of fake information, the receiver’s decision quality should improve.

Another implication of our analysis is that one should be cautious to use institutional fact-checking, because such efforts can have nuanced effects. As long as the ability motive is salient, incentives will lead senders to fact-check. Hence, institutional fact-checking can improve the quality of information by increasing the number of “high-ability” senders. But if institutional fact-checking is available to everyone, it may shut down ability signaling altogether. This way, institutional fact-checking may increase the salience of worldview signaling, and thus potentially decrease the quality of shared information.

Finally, our analysis also highlights the consequences of different social image concerns for the type of information individuals receive on social media. Under an ability motive, social media users predominantly receive decision-relevant information—that is, information that challenges their prior beliefs. This differs under a worldview motive, where audiences receive information likely aligned with their worldview, and information that challenges it is less prevalent. While under both motives senders tailor their strategy to their audience, this tendency is especially pronounced under the worldview motive, where individuals with extreme worldviews also receive the most unbalanced news diets through their networks. So far, the empirical literature has typically focused on sender beliefs and behavior. However, studying the beliefs and polarization levels of social media audiences under different information-sharing motives also seems a fruitful avenue for future research. Apart from the information contained in the news shared, particularly under the worldview motive, individuals also receive information about others’ worldviews and beliefs. The consequences of this for belief polarization and network structure have so far received relatively little attention in the literature.

A Mathematical Appendix

A.1 Proof of Proposition 1

Throughout this proof, let $\tilde{q}^S \equiv \tilde{q}^S(0, p_S, U)$ denote the belief of a low-ability sender that a surprising signal is fake. We first prove an important lemma:

Lemma 1. Δ is strictly decreasing in $\kappa(0, L, \tilde{q})$.

Proof. Both π_{0P} and π_{0U} are strictly decreasing in $\kappa(0, L, \tilde{q})$, so $u_{0U} = \lambda_R(1 - \tilde{q}^S)\pi_{0P} + (1 - \lambda_R)\pi_{0U} - c_S$ is strictly decreasing in it as well. Silence pools more low ability senders as $\kappa(0, L, \tilde{q})$ rises, so $u_\emptyset = \pi_\emptyset$ is strictly increasing. Hence $\Delta = u_{0U} - u_\emptyset$ is strictly decreasing. $\square$

Now define the probability estimate of player $i \in \{S, R\}$ that the signal generation process leads to a surprising signal, z_0^i , and a proper surprising signal, z_{0P}^i , by

$$\begin{aligned} z_0^i &= q(1 - \beta) + (1 - q)(p_i(1 - \eta) + (1 - p_i)\eta), \\ z_{0P}^i &= (1 - q)(p_i(1 - \eta) + (1 - p_i)\eta). \end{aligned}$$

The probability of receiving a fake and surprising signal follows from these two probabilities and is $z_{0F}^i = z_0^i - z_{0P}^i = q(1 - \beta)$. Given our above-defined notation, and assuming the claimed equilibrium probabilities $\kappa(1, H, \tilde{q}^S) = \kappa(0, H, 1) = \kappa(1, L, \tilde{q}^S) = 0$ and $\kappa(0, H, 0) = 1$, we can now define the relevant beliefs held by receivers about sender's ability θ_S . If a surprising signal is relayed, there are three different beliefs to keep track of. A high ability receiver will check the signal's veracity and thus holds belief $\pi_{0F} = 0$ if the signal is fake. If the signal is proper, her belief is

$$\pi_{0P} = \frac{\lambda_S z_{0P}^R}{\lambda_S z_{0P}^R + (1 - \lambda_S) z_0^R \kappa(0, L, \tilde{q})} = \frac{\lambda_S}{\lambda_S + (1 - \lambda_S) \kappa(0, L, \tilde{q})}.$$

A low ability receiver does not know the signal's veracity and hence holds belief

$$\pi_{0U} = \frac{\lambda_S z_0^R}{\lambda_S z_0^R + (1 - \lambda_S) z_0^R \kappa(0, L, \tilde{q})}.$$

Note that $\pi_{0P} > \pi_{0U}$ because $z_0^R > z_{0P}^R$. Moreover, $\pi_{0P} > \lambda_S$, because $\kappa(0, L, \tilde{q}) < 1$. If no

signal is shared, the belief is

$$\pi_\emptyset = \frac{\lambda_S(1 - z_{0\mathcal{P}}^R)}{\lambda_S(1 - z_{0\mathcal{P}}^R) + (1 - \lambda_S)[(1 - z_0^R + z_0^R(1 - \kappa(0, L, \tilde{q}))]}.$$

Finally, the belief a low ability sender holds regarding the veracity of a surprising signal she holds is

$$\tilde{q}^S = \frac{z_{0\mathcal{F}}^S}{z_{0\mathcal{F}}^S + z_{0\mathcal{P}}^S} = \frac{z_{0\mathcal{F}}^S}{z_0^S}. \quad (7)$$

Thus, $\partial \tilde{q}^S / \partial q > 0$, because

$$\frac{\partial \tilde{q}^S}{\partial q} = \frac{(1 - \beta)((1 - 2\eta)p_S + \eta)}{(p_S(2\eta - 1)(1 - q) + q(\beta + \eta - 1) - \eta)^2} > 0 \Leftrightarrow (1 - 2\eta)p_S + \eta > 0.$$

The last inequality follows from the LHS being decreasing in p_S , and when $p_S = 1$ we have $1 - 2\eta + \eta = 1 - \eta > 0$. We need these different beliefs to define S 's expected utility from sharing a given signal. If the sender knows a surprising signal to be proper, she receives expected utility from sharing it equal to $u_{0\mathcal{P}}$ while sharing a signal that is known to be fake yields $u_{0\mathcal{F}}$ with $u_{0\mathcal{P}} = \lambda_R\pi_{0\mathcal{P}} + (1 - \lambda_R)\pi_{0\mathcal{U}} - c_S$ and $u_{0\mathcal{F}} = (1 - \lambda_R)\pi_{0\mathcal{U}} - c_S$. If the signal's veracity is not known, sharing a surprising signal yields an expected utility of

$$\begin{aligned} u_{0\mathcal{U}} &= \lambda_R(\tilde{q}^S \cdot 0 + (1 - \tilde{q}^S)\pi_{0\mathcal{P}}) + (1 - \lambda_R)\pi_{0\mathcal{U}} - c_S \\ &= \lambda_R(1 - \tilde{q}^S)\pi_{0\mathcal{P}} + (1 - \lambda_R)\pi_{0\mathcal{U}} - c_S. \end{aligned}$$

Not sharing a signal implies no sharing cost c_S , and hence yields expected utility $u_\emptyset = \pi_\emptyset$. Finally, off-equilibrium sharing of a non-surprising signal yields $u_1^D = \pi_1 - c_S$.

Define

$$\Delta \equiv u_{0\mathcal{U}} - u_\emptyset. \quad (8)$$

In equilibrium, it must hold that the low ability sender is indifferent between sharing and keeping her signal, $\Delta = 0$. Moreover, she must weakly prefer not to share a non-surprising signal to deviating and sharing it, $u_\emptyset \geq u_1^D$. Hence, we need $u_{0\mathcal{U}} = u_\emptyset \geq u_1^D$.

First assume $\kappa(0, L, \tilde{q}) = 1$.

$$\begin{aligned} \Delta|_{\kappa(0,L,\tilde{q})=1} &= \lambda_R(1 - \tilde{q}^S)\lambda_S + (1 - \lambda_R)\frac{\lambda_S z_{0P}^R}{\lambda_S z_{0P}^R + (1 - \lambda_S)z_0^R} - c_S \\ &\quad - \frac{\lambda_S(1 - z_{0P}^R)}{\lambda_S(1 - z_{0P}^R) + (1 - \lambda_S)(1 - z_0^R)} \end{aligned}$$

This expression is decreasing in both c_S and $\tilde{q}^S$. We have

$$\begin{aligned} \Delta|_{\kappa(0,L,\tilde{q})=1 \wedge c_S=\tilde{q}^S=0} &= \lambda_R\lambda_S + (1 - \lambda_R)\frac{\lambda_S z_{0P}^R}{\lambda_S z_{0P}^R + (1 - \lambda_S)z_0^R} \\ &\quad - \frac{\lambda_S(1 - z_{0P}^R)}{\lambda_S(1 - z_{0P}^R) + (1 - \lambda_S)(1 - z_0^R)} < 0 \\ \Leftrightarrow \lambda_R + (1 - \lambda_R)\frac{z_{0P}^R}{\lambda_S z_{0P}^R + (1 - \lambda_S)z_0^R} &< \frac{(1 - z_{0P}^R)}{\lambda_S(1 - z_{0P}^R) + (1 - \lambda_S)(1 - z_0^R)} \end{aligned}$$

The LHS increases in z_{0P}^R , while the RHS decreases in it. The greatest possible value it can take is z_0^R . Then the LHS becomes 1, and so does the RHS. But this implies that for any $z_{0P}^R < z_0^R$, $\Delta|_{\kappa(0,L,\tilde{q})=1 \wedge c_S=\tilde{q}^S=0} < 0$ holds, and thus also $\Delta|_{\kappa(0,L,\tilde{q})=1} < 0$ holds true. Hence, even if sharing is costless, the low ability sender will keep some signal to herself.

Next assume $\kappa(0, L, \tilde{q}) = 0$. Then

$$\Delta|_{\kappa(0,L,\tilde{q})=0} = \lambda_R(1 - \tilde{q}^S) + (1 - \lambda_R) - c_S - \frac{\lambda_S(1 - z_{0P}^R)}{\lambda_S(1 - z_{0P}^R) + (1 - \lambda_S)} \quad (9)$$

Our goal is to show that this is positive when c_S is sufficiently small, hence let $c_S = 0$. Then

$$\begin{aligned} \Delta|_{\kappa(0,L,\tilde{q})=0 \wedge c_S=0} &= \lambda_R(1 - \tilde{q}^S) + (1 - \lambda_R) - \frac{\lambda_S(1 - z_{0P}^R)}{\lambda_S(1 - z_{0P}^R) + (1 - \lambda_S)} > 0 \\ \Leftrightarrow 1 - \frac{\lambda_S(1 - z_{0P}^R)}{\lambda_S(1 - z_{0P}^R) + (1 - \lambda_S)} &> \lambda_R \tilde{q}^S \Leftrightarrow \frac{1}{\lambda_R} \frac{1 - \lambda_S}{\lambda_S(1 - z_{0P}^R) + (1 - \lambda_S)} > \tilde{q}^S \end{aligned}$$

The LHS increases in z_{0P}^R , while the RHS is independent of it. Hence, let $z_{0P}^R = 0$. Then it must hold that $\frac{1}{\lambda_R}(1 - \lambda_S) > \tilde{q}^S$. Note that because $\lambda_S + \lambda_R > 0$ we have $\frac{1}{\lambda_R}(1 - \lambda_S) > 1$. Moreover, because $\tilde{q}^S < 1$, the inequality is satisfied. This implies that as long as c_S is small enough, $\Delta|_{\kappa(0,L,\tilde{q})=0} > 0$. However, when c_S becomes large, this changes. Hence, there exists $\bar{c}_S$ such that if $c_S < \bar{c}_S$, then $\Delta|_{\kappa(0,L,\tilde{q})=0} > 0$. Because for all c_S it holds that

$\Delta|_{\kappa(0,L,\tilde{q})=1} < 0$, and because Δ is a continuous function of $\kappa(0,L,\tilde{q})$, it follows from the intermediate value theorem that there exists $\kappa^*(0,L,\tilde{q}) \in (0,1)$ such that $\Delta(\kappa^*(0,L,\tilde{q})) = 0$. Now assume $c_S > 0$. To prove existence of the equilibrium discussed in the proposition, we need to show that (9) is positive. Note that q enters (9) through $\tilde{q}^S$, and $\tilde{q}^S$ increases in q . It also enters through z_{0P}^R which decreases in q , while (9) increases in z_{0P}^R . Therefore, when $\Delta|_{\kappa(0,L,\tilde{q})=0} > 0$ for some q' , then the same is true for all $q < q'$. But does such q' always exist? When $q \rightarrow 0$, (9) becomes $\Delta|_{\kappa(0,L,\tilde{q})=0 \wedge q=0} = 1 - c_S - \frac{\lambda_S(1-z_{0P}^R)}{\lambda_S(1-z_{0P}^R) + (1-\lambda_S)} > 0$. As long as $c_S < c_L \equiv \frac{1-\lambda_S}{1-\lambda_S z_{0P}^R} \in (0,1)$, this is true. This implies that for all $c_S \in [0, c_L)$ there exists $\bar{q}(c_S)$ such that if $q \leq \bar{q}(c_S)$, then the equilibrium exists. Moreover, our above analysis reveals that $\bar{q}(0) = 1$ and $\bar{q}(c_L) = 0$. That $\bar{q}(c_S)$ decreases in c_S follows from

$$\frac{\partial \bar{q}(c_S)}{\partial c_S} = -\frac{\frac{\partial \Delta|_{\kappa(0,L,\tilde{q})=0}}{\partial c_S}}{\frac{\partial \Delta|_{\kappa(0,L,\tilde{q})=0}}{\partial q}} = -\frac{-1}{-\lambda_S \frac{\partial \tilde{q}^S}{\partial q}} = -\frac{1}{\lambda_S \frac{\partial \tilde{q}^S}{\partial q}} < 0,$$

because $\frac{\partial \tilde{q}^S}{\partial q} > 0$ (see (7)). When the low ability sender is indifferent between sharing and not sharing the signal, the high ability sender has a strict incentive to share a proper and surprising signal, because $u_{0P} > u_{0U}$. Moreover, because $u_{0F} < u_{0U} = u_{\emptyset}$, the high ability sender also keeps a fake but surprising signal to herself. Finally, no sender type has an incentive to share a non-surprising signal if $\pi_1 - c_S \leq u_{\emptyset}$, or equivalently if $\pi_1 \leq c_S + u_{\emptyset}$. Hence, there exists a threshold $\bar{\pi}_1 \in (0,1)$ such that the equilibrium exists for any off-equilibrium belief $\pi_1 \leq \bar{\pi}_1$.

So far we have characterized the equilibrium for $c_S < c_L$. We next consider $c_S \in [c_L, c_H]$. By Lemma 1, Δ is strictly decreasing in $\kappa(0,L,\tilde{q})$, so $\Delta|_{\kappa=0}$ is its largest value. Since $\Delta|_{\kappa=0} > 0$ if and only if $c_S < c_L$, for $c_S \geq c_L$ we have $\Delta \leq 0$ throughout and the low ability sender remains silent, $\kappa^*(0,L,\tilde{q}) = 0$. A shared surprising signal then originates only from the high ability type sharing proper signals, so $\pi_{0P} = \pi_{0U} = 1$ and she shares if and only if $1 - c_S \geq \pi_{\emptyset}|_{\kappa=0}$. This defines $c_H \equiv 1 - \pi_{\emptyset}|_{\kappa=0} \in (c_L, 1)$. Hence the high ability sender shares every proper surprising signal for $c_S \in [c_L, c_H]$ and remains silent for $c_S > c_H$. $\square$

A.2 Proof of Fact 1

Without loss of generality consider $\sigma = 0$. The sender's belief that $\sigma = 0$ is fake is $\tilde{q}^S = z_{0F}^S/z_0^S$. Since z_0^S is decreasing in p_S for all $\eta > \frac{1}{2}$ while z_{0F}^S is independent of p_S , we have $\partial \tilde{q}^S / \partial p_S > 0$, so $\tilde{q}^S$ increases in $|p_S - \sigma| = p_S$. The case $\sigma = 1$ is analogous.

Since $\partial \Delta / \partial c_S < 0$ and $\partial \Delta / \partial p_S < 0$, while $\partial \Delta / \partial \kappa < 0$ by Lemma 1, the indifference condition $\Delta = 0$ yields $\partial \kappa^*(0, L, \tilde{q}) / \partial c_S < 0$ and $\partial \kappa^*(0, L, \tilde{q}) / \partial p_S < 0$. $\square$

A.3 Proof of Proposition 2

Recall that the quality of information is defined as $\gamma = \sigma^F / (\sigma^F + \sigma^P)$. The probability that proper news is shared is $\sigma^P = (1 - q)((1 - p_T)\eta + p_T(1 - \eta))[\lambda_S + (1 - \lambda_S)\kappa_0]$, whereas the probability that fake news is shared is $\sigma^F = q(1 - \lambda_S)(1 - \beta)\kappa_0$. Moreover, from the proof of Fact 1 we know that κ_0^* decreases in c_S and p_S , proving the proposition. $\square$

A.4 Proof of Proposition 3

Recall that $\gamma = \frac{\sigma^F}{\sigma^F + \sigma^P}$, with $\sigma^P = (1 - q)[p_T(1 - \eta) + (1 - p_T)\eta][(1 - \lambda_S)\kappa_0^* + \lambda_S]$ and $\sigma^F = q(1 - \beta)(1 - \lambda_S)\kappa_0^*$. We know that γ increases in $\kappa^*(0, L, \tilde{q})$. Hence, let $\kappa^*(0, L, \tilde{q}) = 1$. Then we have $\sigma^P|_{\kappa^*(0, L, \tilde{q})=1} = (1 - q)[p_T(1 - \eta) + (1 - p_T)\eta]$ and $\sigma^F|_{\kappa^*(0, L, \tilde{q})=1} = q(1 - \beta)(1 - \lambda_S)$, and hence $\gamma|_{\kappa^*(0, L, \tilde{q})=1} = \frac{q(1 - \beta)(1 - \lambda_S)}{(1 - q)[p_T(1 - \eta) + (1 - p_T)\eta] + q(1 - \beta)(1 - \lambda_S)}$. This is monotone decreasing in β as $\frac{\partial \gamma|_{\kappa^*(0, L, \tilde{q})=1}}{\partial \beta} = \frac{(1 - \lambda)(1 - q)q((2\eta - 1)p_T - \eta)}{(\eta + (2\eta - 1)p_T(q - 1) - q(\beta(-\lambda) + \beta + \eta + \lambda - 1))^2} < 0 \Leftrightarrow (2\eta - 1)p_T - \eta < 0$, which is always true. To see this note that the LHS is maximized if $\eta = 1$ when $p_T > \frac{1}{2}$ and when $\eta = \frac{1}{2}$ when $p_T < \frac{1}{2}$. In the first case, the LHS is $p_T - \eta < 0$, in the latter $-p_T < 0$. Finally, if $p_T = \frac{1}{2}$, the LHS equals $-\frac{1}{2} < 0$.

To prove the proposition we hence only need to set $\gamma|_{\kappa^*(0, L, \tilde{q})=1} = q$ and solve for β , which yields $\tilde{\beta} = 1 - \frac{(1 - p_T)\eta + p_T(1 - \eta)}{1 - \lambda_S} < 1$. If $\beta > \tilde{\beta}$, then $\gamma < q$, independent of $\kappa^*(0, L, \tilde{q})$. This proves the proposition. $\square$

A.5 Proof of Proposition 4

In equilibrium it has to hold that $C_l = C_h = 0$ as defined in Conditions (2) and (3). We next prove that there always exists a constant $\xi \in (0, 1)$ such that there exists $(p_{Sl}, p_{Sh}) \in$

$(0, \xi) \times (\xi, 1)$ that simultaneously solves (2) and (3) if $c_S < \bar{c}_S \equiv \min \left\{ 1 - \xi, \frac{\int_0^\xi z f(z) dz}{\int_0^\xi f(z) dz} \right\}$. If instead $c_S \geq \bar{c}_S$, we may have $p_{Sl} = 0$ or $p_{Sh} = 1$.

First consider C_l and let $p_{Sl} = 0$. Then:

$$C_l|_{p_{Sl}=0} = -c_S + \int_0^1 \left| \frac{\hat{p}_R \int_0^{p_{Sh}} z f_S(z) dz + (1 - \hat{p}_R) \int_0^1 z f_S(z) dz}{\hat{p}_R F_S(p_{Sh}) + (1 - \hat{p}_R)(1 - F_S(0))} \right| dF_R(p_R).$$

This expression is strictly positive if $c_S = 0$. We will now construct an upper bound on c_S to ensure this expression is positive. Note that $p_{Sh} \leq 1$, and therefore the expression in the integral weakly decreases in $\hat{p}_R$. Thus, let $\hat{p}_R = 1$, yielding $C_l|_{p_{Sl}=0 \wedge \hat{p}_R=1} = -c_S + \frac{\int_0^{p_{Sh}} z f_S(z) dz}{F_S(p_{Sh})}$. This expression increases in p_{Sh} , and hence if $p_{Sh} = \xi$, it is least likely to be positive. Letting $p_{Sh} = \xi$ gives us one of the conditions on c_S from the proposition.

Next let $p_{Sl} = \xi$. Then:

$$\begin{aligned} C_l|_{p_{Sl}=\xi} &= - \left(\xi - \frac{\int_0^\xi z f_S(z) dz}{\int_0^\xi f_S(z) dz} \right) - c_S \\ &+ \int_0^1 \left| \xi - \frac{\hat{p}_R \int_0^{p_{Sh}} z f_S(z) dz + (1 - \hat{p}_R) \int_\xi^1 z f_S(z) dz}{\hat{p}_R F_S(p_{Sh}) + (1 - \hat{p}_R)(1 - F_S(\xi))} \right| dF_R(p_R). \end{aligned}$$

If the integral becomes very large, then this is not negative. How does $\hat{p}_S(\emptyset)$ change with $\hat{p}_R$?

$$\frac{\partial \hat{p}_S(\emptyset)}{\partial \hat{p}_R} = \frac{\left(\int_0^{p_{Sh}} z f_S(z) dz \right) (1 - F(p_{Sl})) - \left(\int_{p_{Sl}}^1 z f_S(z) dz \right) F(p_{Sh})}{(\hat{p}_R F_S(p_{Sh}) + (1 - \hat{p}_R)(1 - F_S(p_{Sl})))^2}$$

Thus, $\text{Sign} \left[\frac{\partial \hat{p}_S(\emptyset)}{\partial \hat{p}_R} \right] = \text{Sign} \left[\frac{\int_0^{p_{Sh}} z f_S(z) dz}{F_S(p_{Sh})} - \frac{\int_{p_{Sl}}^1 z f_S(z) dz}{(1 - F_S(p_{Sl}))} \right]$. Since $p_{Sh} \geq p_{Sl}$, $\hat{p}_S(\emptyset)$ is decreasing in $\hat{p}_R$, we can find an upper bound for the integral in $C_l|_{p_{Sl}=\eta}$ either when $\hat{p}_R = 1$ or when $\hat{p}_R = 0$ for all p_R . If $\hat{p}_R = 1$, then

$$\begin{aligned} C_l|_{p_{Sl}=\xi \wedge \hat{p}_R=1} &= - \left(\xi - \frac{\int_0^\xi z f_S(z) dz}{\int_0^\xi f_S(z) dz} \right) - c_S + \left| \xi - \frac{\int_0^{p_{Sh}} z f_S(z) dz}{F_S(p_{Sh})} \right| \\ &= - \left(\xi - \frac{\int_0^\xi z f_S(z) dz}{\int_0^\xi f_S(z) dz} \right) - c_S + \left(\xi - \frac{\int_0^{p_{Sh}} z f_S(z) dz}{F_S(p_{Sh})} \right) \\ &= \frac{\int_0^\xi z f_S(z) dz}{\int_0^\xi f_S(z) dz} - \frac{\int_0^{p_{Sh}} z f_S(z) dz}{F_S(p_{Sh})} - c_S \leq 0. \end{aligned}$$

The inequality follows from $\frac{\int_0^{p_{Sh}} z f_S(z) dz}{\int_0^{p_{Sh}} f_S(z) dz} \geq \frac{\int_0^\xi z f_S(z) dz}{\int_0^\xi f_S(z) dz}$ because $p_{Sh} \geq \xi$. If instead $\hat{p}_R = 0$, then

$$\begin{aligned} C_l|_{p_{Sl}=\xi \wedge \hat{p}_R=0} &= -\left(\xi - \frac{\int_0^\xi z f_S(z) dz}{\int_0^\xi f_S(z) dz}\right) + \left|\xi - \frac{\int_\xi^1 z f_S(z) dz}{(1 - F_S(\xi))}\right| - c_S \\ &= -\left(\xi - \frac{\int_0^\xi z f_S(z) dz}{\int_0^\xi f_S(z) dz}\right) + \left(\frac{\int_\xi^1 z f_S(z) dz}{(1 - F_S(\xi))} - \xi\right) - c_S. \end{aligned}$$

Hence, if

$$-2\xi + \frac{\int_0^\xi z f_S(z) dz}{F_S(\xi)} + \frac{\int_\xi^1 z f_S(z) dz}{(1 - F_S(\xi))} - c_S \leq 0, \quad (10)$$

then there exists ξ and p_{Sl} such that $p_{Sl} \in (0, \xi)$ solves $C_l = 0$ for all $p_{Sh} \in (\xi, 1)$.

Now consider C_h and let $p_{Sh} = 1$:

$$C_h|_{p_{Sh}=1} = -c_S + \int_0^1 \left| 1 - \frac{\hat{p}_R \int_0^{p_{Sh}} z f_S(z) dz + (1 - \hat{p}_R) \int_{p_{Sl}}^1 z f_S(z) dz}{\hat{p}_R F_S(p_{Sh}) + (1 - \hat{p}_R)(1 - F_S(p_{Sl}))} \right| dF_R(p_R).$$

If c_S is small, this is positive. As before, the integral is minimized if either $\hat{p}_R = 0$ or $\hat{p}_R = 1$. If $\hat{p}_R = 0$, then $C_h|_{p_{Sh}=1 \wedge \hat{p}_R=0} = -c_S + 1 - \frac{\int_{p_{Sl}}^1 z f_S(z) dz}{(1 - F_S(p_{Sl}))}$. This is smallest if $p_{Sl} = 0$, in which case we get $C_h|_{p_{Sh}=1 \wedge \hat{p}_R=0 \wedge p_{Sl}=0} = -c_S + 1 - \xi$. This is positive if $c_S < 1 - \xi$, which is the second condition stated in the proposition.

Finally, let $p_{Sh} = \xi$:

$$C_h|_{p_{Sh}=\xi} = \xi - \frac{\int_\xi^1 z f_S(z) dz}{\int_\xi^1 f_S(z) dz} - c_S + \int_0^1 \left| \xi - \frac{\hat{p}_R \int_0^\xi z f_S(z) dz + (1 - \hat{p}_R) \int_{p_{Sl}}^1 z f_S(z) dz}{\hat{p}_R F_S(\xi) + (1 - \hat{p}_R)(1 - F_S(p_{Sl}))} \right| dF_R(p_R)$$

The expression is maximized if either $\hat{p}_R = 0$ or $\hat{p}_R = 1$. Letting $\hat{p}_R = 0$,

$$\begin{aligned} C_h|_{p_{Sh}=\xi \wedge \hat{p}_R=0} &= -\left(\frac{\int_\xi^1 z f_S(z) dz}{\int_\xi^1 f_S(z) dz} - \xi\right) - c_S + \left|\xi - \frac{\int_{p_{Sl}}^1 z f_S(z) dz}{(1 - F_S(p_{Sl}))}\right| \\ &= -\left(\frac{\int_\xi^1 z f_S(z) dz}{\int_\xi^1 f_S(z) dz} - \xi\right) - c_S + \left(\frac{\int_{p_{Sl}}^1 z f_S(z) dz}{(1 - F_S(p_{Sl}))} - \xi\right) \\ &= \frac{\int_{p_{Sl}}^1 z f_S(z) dz}{(1 - F_S(p_{Sl}))} - \frac{\int_\xi^1 z f_S(z) dz}{(1 - F_S(\xi))} - c_S < 0 \end{aligned}$$

If instead $\hat{p}_R = 1$, then

$$\begin{aligned} C_h|_{p_{Sh}=\xi \wedge \hat{p}_R=1} &= -\left(\frac{\int_{\xi}^1 z f_S(z) dz}{\int_{\xi}^1 f_S(z) dz} - \xi\right) - c_S + \left|\xi - \frac{\int_0^{\xi} z f_S(z) dz}{F_S(\xi)}\right| \\ &= -\left(\frac{\int_{\xi}^1 z f_S(z) dz}{\int_{\xi}^1 f_S(z) dz} - \xi\right) - c_S + \left(\xi - \frac{\int_0^{\xi} z f_S(z) dz}{F_S(\xi)}\right). \end{aligned}$$

Hence, if

$$2\xi - \frac{\int_{\xi}^1 z f_S(z) dz}{\int_{\xi}^1 f_S(z) dz} - \frac{\int_0^{\xi} z f_S(z) dz}{F_S(\xi)} - c_S \leq 0 \quad (11)$$

and (10), then there exist ξ and p_{Sh} such that $p_{Sh} \in (\xi, 1)$ solves $C_h = 0$ for any $p_{Sl} \in (0, \xi)$. Therefore, what is left to prove is that there exists $\xi \in (0, 1)$ guaranteeing that both (10) and (11) are satisfied simultaneously. This is indeed the case for all $c_S \geq 0$ if $2\xi - \frac{\int_{\xi}^1 z f_S(z) dz}{\int_{\xi}^1 f_S(z) dz} - \frac{\int_0^{\xi} z f_S(z) dz}{F_S(\xi)} = 0$. If $\xi = 0$, then the LHS is $-\mathbb{E}[p_S] < 0$, while if $\xi = 1$ the LHS is $2 - \mathbb{E}[p_S] > 0$. It follows from continuity of the LHS that there exists $\xi \in (0, 1)$ such that the equation is satisfied. Because $[0, \xi] \times [\xi, 1]$ is compact, and because both C_l and C_h are continuous, it follows that there must exist $(p_{Sl}, p_{Sh}) \in [0, \xi] \times [\xi, 1]$ such that $C_l = C_h = 0$. $\square$

A.6 Proof of Proposition 5

First, we prove a lemma that is important to simplify the analysis:

Lemma 2. *In any interior equilibrium with known receiver type, it holds that*

$$\left| \frac{\hat{p}_R \int_0^{p_{Sh}^*} z f_S(z) dz + (1 - \hat{p}_R) \int_{p_{Sl}^*}^1 z f_S(z) dz}{\hat{p}_R F_S(p_{Sh}^*) + (1 - \hat{p}_R)(1 - F_S(p_{Sl}^*))} - p_{Sl}^* \right| = \frac{\hat{p}_R \int_0^{p_{Sh}^*} z f_S(z) dz + (1 - \hat{p}_R) \int_{p_{Sl}^*}^1 z f_S(z) dz}{\hat{p}_R F_S(p_{Sh}^*) + (1 - \hat{p}_R)(1 - F_S(p_{Sl}^*))} - p_{Sl}^* \quad (12)$$

and

$$\left| p_{Sh}^* - \frac{\hat{p}_R \int_0^{p_{Sh}^*} z f_S(z) dz + (1 - \hat{p}_R) \int_{p_{Sl}^*}^1 z f_S(z) dz}{\hat{p}_R F_S(p_{Sh}^*) + (1 - \hat{p}_R)(1 - F_S(p_{Sl}^*))} \right| = p_{Sh}^* - \frac{\hat{p}_R \int_0^{p_{Sh}^*} z f_S(z) dz + (1 - \hat{p}_R) \int_{p_{Sl}^*}^1 z f_S(z) dz}{\hat{p}_R F_S(p_{Sh}^*) + (1 - \hat{p}_R)(1 - F_S(p_{Sl}^*))}. \quad (13)$$

Proof. To prove (12), recall that $\hat{p}_S(\emptyset)$ is monotone in $\hat{p}_R$. If $\hat{p}_R = 0$, the LHS of (12) becomes $\left| \frac{\int_{p_{Sl}^*}^1 z f_S(z) dz}{1 - F_S(p_{Sl}^*)} - p_{Sl}^* \right| = \frac{\int_{p_{Sl}^*}^1 z f_S(z) dz}{1 - F_S(p_{Sl}^*)} - p_{Sl}^* > 0$. If $\hat{p}_R = 1$, it becomes $\left| \frac{\int_0^{p_{Sh}^*} z f_S(z) dz}{F_S(p_{Sh}^*)} - p_{Sl}^* \right|$.

If the interior of the absolute value is also positive, we proved (12). Thus, assume it was negative. Then C_l becomes $\frac{\int_0^{p_{Sl}^*} z f_S(z) dz}{F_S(p_{Sl}^*)} - \frac{\int_0^{p_{Sh}^*} z f_S(z) dz}{F_S(p_{Sh}^*)} - c_S < 0$. This is negative, and hence the equilibrium condition can never be satisfied with known receiver worldview. Thus, for known p_R , the interior of the absolute value must be positive in equilibrium. The proof of (13) follows the same line of reasoning. $\square$

From the lemma it follows that in a Responsive Equilibrium with one receiver with known type, the equilibrium conditions in (2) and (3) simplify to

$$\begin{aligned} C_l &= \hat{p}_S(0) + \hat{p}_S(\emptyset) - 2p_{Sl} - c_S, \\ C_h &= 2p_{Sh} - \hat{p}_S(1) - \hat{p}_S(\emptyset) - c_S. \end{aligned} \tag{14}$$

We know from the proof of Proposition 4 that C_l is positive for $p_{Sl} = 0$ and negative for $p_{Sl} = \xi$. From Assumption 1 it follows that $\frac{\partial C_l}{\partial p_{Sl}} = \frac{\partial \hat{p}_S(0)}{\partial p_{Sl}} + \frac{\partial \hat{p}_S(\emptyset)}{\partial p_{Sl}} - 2 < 0$ for all p_{Sl} , see Assumption 1. Similarly, we know from the proof of Proposition 4 that C_h is negative for $p_{Sh} = \xi$ and positive for $p_{Sh} = 1$. Thus, to guarantee uniqueness of equilibrium, we assume $\frac{\partial C_h}{\partial p_{Sh}} = 2 - \frac{\partial \hat{p}_S(1)}{\partial p_{Sh}} + \frac{\partial \hat{p}_S(\emptyset)}{\partial p_{Sh}} > 0$ for all p_{Sh} (Assumption 1 as well). Thus, we assume that beliefs do not change too abruptly with p_{Sl} and p_{Sh} .

Under Assumption 1 there is a unique $(p_{Sl}, p_{Sh}) \in [0, \xi] \times [\xi, 1]$ solving the equilibrium conditions (Gale and Nikaido, 1965). To derive comparative statics, define

$$\begin{aligned} J &= \begin{pmatrix} \frac{\partial C_l}{\partial p_{Sl}} & \frac{\partial C_l}{\partial p_{Sh}} \\ \frac{\partial C_h}{\partial p_{Sl}} & \frac{\partial C_h}{\partial p_{Sh}} \end{pmatrix} & J_L^{p_R} &= \begin{pmatrix} -\frac{\partial C_l}{\partial p_R} & \frac{\partial C_l}{\partial p_{Sh}} \\ -\frac{\partial C_h}{\partial p_R} & \frac{\partial C_h}{\partial p_{Sh}} \end{pmatrix} & J_H^{p_R} &= \begin{pmatrix} \frac{\partial C_l}{\partial p_{Sl}} & -\frac{\partial C_l}{\partial p_R} \\ \frac{\partial C_h}{\partial p_{Sl}} & -\frac{\partial C_h}{\partial p_R} \end{pmatrix} \\ J_L^c &= \begin{pmatrix} -\frac{\partial C_l}{\partial c_S} & \frac{\partial C_l}{\partial p_{Sh}} \\ -\frac{\partial C_h}{\partial c_S} & \frac{\partial C_h}{\partial p_{Sh}} \end{pmatrix} & J_H^c &= \begin{pmatrix} \frac{\partial C_l}{\partial p_{Sl}} & -\frac{\partial C_l}{\partial c_S} \\ \frac{\partial C_h}{\partial p_{Sl}} & -\frac{\partial C_h}{\partial c_S} \end{pmatrix}. \end{aligned}$$

Then

$$\begin{aligned}
\frac{\partial p_{Sl}}{\partial p_R} &= \frac{|J_L^{p_R}|}{|J|} = \frac{-\frac{\partial C_l}{\partial p_R} \frac{\partial C_h}{\partial p_{Sh}} + \frac{\partial C_l}{\partial p_{Sh}} \frac{\partial C_h}{\partial p_R}}{|J|} = \frac{-\frac{\partial \hat{p}_S(\emptyset)}{\partial p_R} \left(\frac{\partial C_h}{\partial p_{Sh}} + \frac{\partial C_l}{\partial p_{Sh}} \right)}{|J|} \\
\frac{\partial p_{Sh}}{\partial p_R} &= \frac{|J_H^{p_R}|}{|J|} = \frac{-\frac{\partial C_h}{\partial p_R} \frac{\partial C_l}{\partial p_{Sl}} + \frac{\partial C_h}{\partial p_{Sl}} \frac{\partial C_l}{\partial p_R}}{|J|} = \frac{\frac{\partial \hat{p}_S(\emptyset)}{\partial p_R} \left(\frac{\partial C_l}{\partial p_{Sl}} + \frac{\partial C_h}{\partial p_{Sl}} \right)}{|J|} \\
\frac{\partial p_{Sl}}{\partial c_S} &= \frac{|J_L^c|}{|J|} = \frac{-\frac{\partial C_l}{\partial c_S} \frac{\partial C_h}{\partial p_{Sh}} + \frac{\partial C_l}{\partial p_{Sh}} \frac{\partial C_h}{\partial c_S}}{|J|} = \frac{\frac{\partial C_h}{\partial p_{Sh}} - \frac{\partial C_l}{\partial p_{Sh}}}{|J|} \\
\frac{\partial p_{Sh}}{\partial c_S} &= \frac{|J_H^c|}{|J|} = \frac{-\frac{\partial C_h}{\partial c_S} \frac{\partial C_l}{\partial p_{Sl}} + \frac{\partial C_h}{\partial p_{Sl}} \frac{\partial C_l}{\partial c_S}}{|J|} = \frac{\frac{\partial C_l}{\partial p_{Sl}} - \frac{\partial C_h}{\partial p_{Sl}}}{|J|},
\end{aligned}$$

where we used $\frac{\partial C_l}{\partial p_R} = -\frac{\partial C_h}{\partial p_R} = \frac{\partial \hat{p}_S(\emptyset)}{\partial p_R}$ and $\frac{\partial C_l}{\partial c_S} = \frac{\partial C_h}{\partial c_S} = -1$. In the proof of Proposition 4 we have shown that $\frac{\partial \hat{p}_S(\emptyset)}{\partial p_R} < 0$ and Assumption 1 states that $|J| < 0$. This implies that $\text{Sign} \left[\frac{\partial p_{Sl}}{\partial p_R} \right] = \text{Sign} \left[- \left(\frac{\partial C_h}{\partial p_{Sh}} + \frac{\partial C_l}{\partial p_{Sh}} \right) \right]$, $\text{Sign} \left[\frac{\partial p_{Sh}}{\partial p_R} \right] = \text{Sign} \left[\left(\frac{\partial C_l}{\partial p_{Sl}} + \frac{\partial C_h}{\partial p_{Sl}} \right) \right]$, $\text{Sign} \left[\frac{\partial p_{Sl}}{\partial c_S} \right] = \text{Sign} \left[\frac{\partial C_l}{\partial p_{Sh}} - \frac{\partial C_h}{\partial p_{Sh}} \right]$, and $\text{Sign} \left[\frac{\partial p_{Sh}}{\partial c_S} \right] = \text{Sign} \left[\frac{\partial C_h}{\partial p_{Sl}} - \frac{\partial C_l}{\partial p_{Sl}} \right]$. Next note that $\frac{\partial C_l}{\partial p_{Sh}} = \frac{\partial \hat{p}_S(\emptyset)}{\partial p_{Sh}} > 0$ and $\frac{\partial C_h}{\partial p_{Sl}} = -\frac{\partial \hat{p}_S(\emptyset)}{\partial p_{Sl}} < 0$. Together with our assumptions of $\frac{\partial C_l}{\partial p_{Sl}} < 0$ and $\frac{\partial C_h}{\partial p_{Sh}} > 0$, we can sign $\frac{\partial p_{Sl}}{\partial p_R} < 0$ and $\frac{\partial p_{Sh}}{\partial p_R} < 0$. Thus, both equilibrium thresholds decrease in p_R . Moreover, $|J| < 0$, $\frac{\partial C_l}{\partial p_{Sl}} < 0$ and $\frac{\partial C_h}{\partial p_{Sh}} > 0$ imply that $\frac{\partial p_{Sh}}{\partial c_S} > 0$ and $\frac{\partial p_{Sl}}{\partial c_S} < 0$. Thus, higher costs of sharing decrease the amount of signals shared from both types of signals.

A.7 Proof of Proposition 6

Let $\beta = \hat{\beta} \equiv p_T \eta + (1 - p_T)(1 - \eta) \in (0, 1)$. Then (6) reduces to $\gamma = q$. Next, notice that the only expressions in (6) that depend on p_R are $F_S(p_{Sl})$ and $F_S(p_{Sh})$, implicitly through the thresholds p_{Sl} and p_{Sh} , respectively. Hence, the derivative of (6) is

$$\begin{aligned}
\frac{\partial \gamma}{\partial p_R} &= - \frac{(1 - q)q(\beta - \hat{\beta})}{(F_S(p_{Sl})\mathbb{P}(\sigma = 0) + (1 - F_S(p_{Sh}))\mathbb{P}(\sigma = 1))^2} \\
&\quad \times \left[F_S(p_{Sl})f_S(p_{Sh})\frac{\partial p_{Sh}}{\partial p_R} + f_S(p_{Sl})(1 - F_S(p_{Sh}))\frac{\partial p_{Sl}}{\partial p_R} \right].
\end{aligned}$$

Clearly, if $\beta = \hat{\beta}$, this derivative equals zero. If $\beta > \hat{\beta}$, the derivative is positive iff the expression in squared parentheses is negative. We know from Proposition 5 that both p_{Sl} and p_{Sh} are decreasing in p_R , and hence the derivative is indeed positive when $\beta > \hat{\beta}$. Similarly, if $\beta < \hat{\beta}$, the derivative is negative. This proves items (i) and (ii) of the proposition. $\square$

A.8 Proof of Proposition 7

That $q = \gamma$ when $\beta = \hat{\beta}$ follows from Proposition 6. For the other cases, recall that

$$\gamma - q = q \left(\frac{(1 - \beta)F_S(p_{Sl}) + \beta(1 - F_S(p_{Sh}))}{\mathbb{P}(\sigma = 0)F_S(p_{Sl}) + \mathbb{P}(\sigma = 1)(1 - F_S(p_{Sh}))} - 1 \right),$$

where $\mathbb{P}(\sigma = 1) = q\beta + (1 - q)[p_T\eta + (1 - p_T)(1 - \eta)]$ and $\mathbb{P}(\sigma = 0) = 1 - \mathbb{P}(\sigma = 1)$. Thus, $\gamma > q$ iff

$$\begin{aligned} (1 - \beta)F_S(p_{Sl}) + \beta(1 - F_S(p_{Sh})) - \mathbb{P}(\sigma = 0)F_S(p_{Sl}) - \mathbb{P}(\sigma = 1)(1 - F_S(p_{Sh})) &> 0 \\ \Leftrightarrow (\hat{\beta} - \beta)(F_S(p_{Sl}) - (1 - F_S(p_{Sh}))) &> 0. \end{aligned}$$

This proves the proposition. □

References

- Acemoglu, D., Ozdaglar, A., and Gheibi, A. P. (2010). Spread of (mis)information in social networks. *Games and Economic Behavior*, 70(2):194 – 227.
- Acemoglu, D., Ozdaglar, A., and Siderius, J. (2023). A model of online misinformation. *The Review of Economic Studies*, 91:3117–3150.
- Allcott, H. and Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2):211–36.
- Altman, I. and Taylor, D. A. (1973). *Social penetration: The development of interpersonal relationships*. Holt, Rinehart & Winston.
- Angelucci, C. and Prat, A. (2024). Is journalistic truth dead? Measuring how informed voters are about political news. *American Economic Review*, 114(4):887–925.
- Banerjee, A., Chandrasekhar, A. G., Duflo, E., and Jackson, M. O. (2013). The diffusion of microfinance. *Science*, 341(6144):1236498.

- Banerjee, A., Chandrasekhar, A. G., Duflo, E., and Jackson, M. O. (2019). Using gossips to spread information: Theory and evidence from two randomized controlled trials. *The Review of Economic Studies*, 86(6):2453–2490.
- Bazarova, N. N. and Choi, Y. H. (2014). Self-disclosure in social media: Extending the functional approach to disclosure motivations and characteristics on social network sites. *Journal of Communication*, 64(4):635–657.
- Bowen, T. R., Dmitriev, D., and Galperti, S. (2023). Learning from shared news: When abundant information leads to belief polarization. *The Quarterly Journal of Economics*, 138(2):955–1000.
- Buisseret, P. and Van Weelden, R. (2022). Polarization, valence, and policy competition. *American Economic Review: Insights*, 4(3):341–352.
- Cantarella, M., Fraccaroli, N., and Volpe, R. (2023). Does fake news affect voting behaviour? *Research Policy*, 52(1):104628.
- Che, Y. and Kartik, N. (2009). Opinions as incentives. *Journal of Political Economy*, 117(5):815–860.
- Crawford, V. P. and Sobel, J. (1982). Strategic information transmission. *Econometrica*, 50(6):1431–1451.
- Denter, P., Dumay, M., and Ginzburg, B. (2021). Social connectivity, media bias, and correlation neglect. *The Economic Journal*, 131:2033–2057.
- Denter, P. and Ginzburg, B. (2026). Troll farms. Technical report, arXiv:2411.03241.
- Downs, A. (1957). An economic theory of political action in a democracy. *Journal of political economy*, 65(2):135–150.
- Enkhbayar, A., Haustein, S., Barata, G., and Alperin, J. P. (2020). How much research shared on Facebook happens outside of public pages and groups? A comparison of public and private online activity around PLOS ONE papers. *Quantitative Science Studies*, 1(2):749–770.

- Erdt, M., Nagarajan, A., Sin, S.-C. J., and Theng, Y.-L. (2016). Altmetrics: an analysis of the state-of-the-art in measuring research impact on social media. *Scientometrics*, 109(2):1117–1166.
- Gale, D. and Nikaido, H. (1965). The jacobian matrix and global univalence of mappings. *Mathematische Annalen*, 159:81–93.
- Germano, F., Gómez, V., and Sobbrío, F. (2026). Ranking for engagement: How social media algorithms fuel misinformation and polarization. *Journal of Public Economics*, 255:105589.
- Greene, K., Derlega, V. J., and Mathews, A. (2006). Self-disclosure in personal relationships. *The Cambridge handbook of personal relationships*, 409:427.
- Grossman, G. M. and Helpman, E. (2023). Electoral competition with fake news. *European Journal of Political Economy*, 77:102315.
- Guess, A., Nagler, J., and Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances*, 5(1).
- Guess, A. M., Malhotra, N., Pan, J., Barberá, P., Allcott, H., Brown, T., Crespo-Tenorio, A., Dimmery, D., Freelon, D., Gentzkow, M., González-Bailón, S., Kennedy, E., Kim, Y. M., Lazer, D., Moehler, D., Nyhan, B., Rivera, C. V., Settle, J., Thomas, D. R., Thorson, E., Tromble, R., Wilkins, A., Wojcieszak, M., Xiong, B., de Jonge, C. K., Franco, A., Mason, W., Stroud, N. J., and Tucker, J. A. (2023). Reshares on social media amplify political news but do not detectably affect beliefs or opinions. *Science*, 381(6656):404–408.
- Guriev, S., Henry, E., Marquis, T., and Zhuravskaya, E. (2026). Curtailing false news, amplifying truth. *Econometrica*, 94(5):1563–1601.
- Kranton, R. and McAdams, D. (2024). Social connectedness and information markets. *American Economic Journal: Microeconomics*, 16:33–62.
- Kümpel, A. S., Karnowski, V., and Keyling, T. (2015). News sharing in social media: A review of current research on news sharing users, content, and networks. *Social Media + Society*, 1(2):1–14.

- Lee, C. S. and Ma, L. (2012). News sharing in social media: The effect of gratifications and prior experience. *Computers in Human Behavior*, 28(2):331–339.
- Lee, C. S., Ma, L., and Goh, D. H.-L. (2011). Why do people share news in social media? In Zhong, N., Callaghan, V., Ghorbani, A. A., and Hu, B., editors, *Active Media Technology*, pages 129–140, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Levy, G. and Razin, R. (2018). Information diffusion in networks with the bayesian peer influence heuristic. *Games and Economic Behavior*, 109:262–270.
- Lin, H., Pennycook, G., and Rand, D. G. (2023). Thinking more or thinking differently? using drift-diffusion modeling to illuminate why accuracy prompts decrease misinformation sharing. *Cognition*, 230:105312.
- Marwick, A. E. and boyd, d. (2011). I tweet honestly, i tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, 13(1):114–133.
- Montagni, I., Ouazzani-Touhami, K., Mebarki, A., Texier, N., Schück, S., Tzourio, C., and the CONFINS group (2021). Acceptance of a Covid-19 vaccine is associated with ability to detect fake news and health literacy. *Journal of Public Health*, 43(4):695–702.
- Osmundsen, M., Bor, A., Vahlstrup, P. B., Bechmann, A., and Petersen, M. B. (2021). Partisan polarization is the primary psychological motivation behind political fake news sharing on twitter. *American Political Science Review*, 115(3):999–1015.
- Papanastasiou, Y. (2020). Fake news propagation and detection: A sequential model. *Management Science*, 66(5):1826–1846.
- Park, K., Kwak, H., Song, H., and Cha, M. (2020). “trust me, i have a ph. d.”: A propensity score analysis on the halo effect of disclosing one’s offline social status in online communities. In *Proceedings of the international AAAI conference on web and social media*, volume 14, pages 534–544.
- Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., and Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature*, 592:590 – 595.

- Pogorelskiy, K. and Shum, M. (2019). News we like to share: How news sharing on social networks influences voting outcomes. Technical report.
- Stein, J., Keuschnigg, M., and van de Rijdt, A. (2024). Partisan belief in new misinformation is resistant to accuracy incentives. *PNAS nexus*, 3(11):pgae506.
- Sundar, S. S., Snyder, E. C., Liao, M., Yin, J., Wang, J., and Chi, G. (2025). Sharing without clicking on news in social media. *Nature Human Behavior*, 9:156–168.
- Taber, C. S. and Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, 50(3):755–769.
- van der Does, T., Galesic, M., Dunivin, Z. O., and Smaldino, P. E. (2022). Strategic identity signaling in heterogeneous networks. *Proceedings of the National Academy of Sciences*, 119(10):e2117898119.
- Vignery, K. and Laurier, W. (2020). Achievement in student peer networks: A study of the selection process, peer effects and student centrality. *International Journal of Educational Research*, 99:101499.
- Vosoughi, S., Roy, D., and Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380):1146–1151.
- Williams, E. A., Zwolak, J. P., Dou, R., and Brewe, E. (2019). Linking engagement and performance: The social network analysis perspective. *Physical review physics education research*, 15(2):020150.
- Zimmermann, F. and Kohring, M. (2020). Mistrust, disinforming news, and vote choice: A panel survey on the origins and consequences of believing disinformation in the 2017 german parliamentary election. *Political communication*, 37(2):215–237.