Journal of Economic Theory and Econometrics, Vol. 37, No. 2, Jun. 2026, 43–88

The Fact-Checking Game: Fact-Checking Deters Misinformation, But When Does It Fail?

Sang-Hyun Kim* Hee Yeul Woo†

Abstract Misinformation remains a persistent global challenge in modern media systems. We investigate whether and under what conditions fact-checking effectively deters misinformation and promotes truth, focusing on the strategic interaction where a biased outlet decides whether to misreport and a fact-checker decides whether to scrutinize the report. Our findings suggest that the deterrent effect of fact-checking hinges critically on the degree of media market polarization. Provided that the market is not excessively polarized, fact-checking strictly reduces misreporting by biased outlets, thereby improving information welfare—a result that echoes the classical views of John Milton and John Stuart Mill that the competition between opposing claims facilitates the discovery of truth. We also show that the presence of an active fact-checker can make biased outlets better off by serving as an external commitment device, thereby resolving the dynamic inconsistency problem.

Keywords Contest over truth, deterrent effect, fact-checking, fake news, information welfare, news-market competition.

JEL Classification D82, D83, L82.

*School of Economics, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul, Republic of Korea 03722. E-mail: sang.kim@yonsei.ac.kr.

†School of Economics, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul, Republic of Korea 03722. E-mail: trhywoo@yonsei.ac.kr.

Received May 4, 2026, Revised June 20, 2026, Accepted June 22, 2026

1. INTRODUCTION

Misinformation is a persistent and global challenge in modern media systems, with far-reaching consequences for public health, democratic governance, and social cohesion. From the false claims about weapons of mass destruction preceding the 2003 Iraq War, to conspiracy theories surrounding COVID-19 vaccines, to baseless allegations of election fraud in the 2020 U.S. presidential election, widely disseminated falsehoods have distorted public discourse and eroded institutional trust. In response, fact-checking organizations have emerged as crucial actors in the media ecosystem, tasked with identifying and correcting misinformation. Two fundamental questions arise naturally: Does fact-checking actually deter misinformation and promote truth? And when does it succeed or fail in doing so?

While theoretical studies on fact-checking remain limited, a substantial body of research on competition in the market for news has examined whether media competition promotes accurate reporting (e.g., Gentzkow *et al.* (2015); Prat and Strömberg (2013)). A key limitation of this literature, however, is that it models competition primarily as a rivalry over market shares. Outlets compete for subscribers, and equilibrium audience composition is determined endogenously by readers choosing which outlet best serves their preferences. This approach implicitly assumes that readers know, prior to consuming any given report, which outlet is likely to provide the news they prefer, an assumption that sits uneasily with the nature of news as an experience good or credence good. Readers often cannot evaluate the accuracy of a report until after consuming it, and in many cases not even then. A media outlet that happens to report the truth may fail to attract a larger audience simply because readers cannot tell, at the time of consumption, whether the truth is being told or by whom. In practice, therefore, audience shares tend to be relatively stable in the short run, shaped by habits, accumulated trust, and prior beliefs rather than by continuous reoptimization over outlet choice. In such an environment, models of competition over market shares may be poorly suited to capturing the strategic forces that actually govern media conduct.

This paper complements the existing literature by taking a different approach. Rather than modeling competition over market shares, we take the distribution of subscribers across outlets as given and study the strategic interaction between a biased media outlet and a fact-checker over the content of a specific report. This conduct-level focus captures an important dimension of media competition that structural models abstract away from: the short-run contest over truth, in which a biased outlet decides whether to misreport a particular signal and a fact-checker decides whether to scrutinize it. By holding market structure fixed, we isolate the strategic incentives that govern this contest and examine how they are shaped within a framework that incorporates the institutional details of the fact-checking process—including the informational asymmetry between the outlet and the fact-checker, the heterogeneity of the audience, and the belief updating of subscribers.

In particular, our model features two firms. Firm A, a biased media outlet, receives a noisy private signal about the true state of the world and chooses whether to report it truthfully or fabricate news at a cost. Firm B, a fact-checker, observes firm A's report and chooses whether to verify it at a cost. Firm A's objective is to maximize the extent to which subscribers believe a particular (potentially false) state is true, while firm B aims to detect and expose false reporting. Subscribers are heterogeneous and fall into three groups: those who follow only firm A, those who follow only firm B, and those who consume content from both. Subscribers update their beliefs rationally based on the news they receive and on whether fact-checking occurs.

Our analysis yields two central findings. The first is that fact-checking deters misreporting, but its effectiveness depends critically on the structure of the audience. The deterrent effect operates unless the media environment is sufficiently polarized; for instance, when enough subscribers multihome. When the media environment is highly polarized, with most consumers following a single outlet exclusively, fact-checking loses its disciplining power. A biased outlet facing a predominantly captive audience has little to fear from exposure, since corrections reach few subscribers. Conversely, when the audience is well integrated, the threat of fact-checking is credible and effective, deterring misreporting and improving public beliefs. This result has direct implications for media policy: interventions that reduce audience polarization may do more to promote informational welfare than strengthening fact-checking institutions per se.

The model provides a formal foundation for the classical view, associated with Milton (1644) and Mill (1859), that competition between opposing viewpoints promotes the discovery of truth. Even as the fake news outlet's audience grows, overall informativeness can increase, provided the fact-checker retains enough influence to act as an effective counterweight. Yet our results also reveal the limits of this optimistic view: when the audience is sufficiently polarized, or when the fact-checker's reach is too narrow, the Millian mechanism breaks down. Competitive pressure alone, in other words, is insufficient to promote truth in polarized media environments. Even when audiences are willing to update their beliefs in response to accurate reporting, the credence good nature of news may prevent truth-telling outlets from being immediately rewarded with greater readership. Consequently, this may prevent a significant improvement in subscribers' information welfare in the absence of suitable policy intervention.

The second finding is more surprising. While fact-checking intuitively benefits subscribers, the biased outlet too may be strictly better off ex ante in the presence of an active fact-checker. The mechanism driving this result lies in the biased outlet's lack of commitment power. Firm A would earn higher payoffs if it could credibly commit to honest reporting, thereby avoiding the costs associated with misreporting. In the absence of external discipline, however, firm A faces a dynamic inconsistency problem: it always retains a short-run incentive to misreport whenever its signal is unfavorable. The fact-checker serves as an external commitment device. By imposing sufficient discipline to restrain the outlet's opportunistic behavior, the fact-checker effectively steers firm A toward a more profitable reporting strategy that it could not sustain on its own. This interdependence between two ostensibly adversarial firms—each made better off by the other's presence—illustrates the subtle and sometimes counterintuitive nature of epistemic competition in media markets.

This paper makes three contributions to the existing literature. First, as discussed above, a central question in the literature on competition in media markets is whether and when competition promotes truth. The prevailing finding is that it need not. When readers demand confirmatory news, competing outlets may slant their reporting toward their respective audiences rather than converging on accuracy (Mullainathan and Shleifer (2005); Gentzkow and Shapiro (2006)). Market structure, ownership concentration, and the threat of media capture further shape the incentives of news producers in ways that may undermine rather than enhance the quality of public information (Besley and Prat (2006); Anderson and McLaren (2012); Strömberg (2004)). For surveys of this literature, see Gentzkow *et al.* (2015) and Prat and Strömberg (2013). Our model differs from these studies in that we study, to take into account the credence good nature of news, the strategic interaction between a biased media outlet and a fact-checker, taking the composition of the audience as given rather than modeling it as an equilibrium outcome. We show that competitive pressure alone is insufficient to promote truth when media environments are sufficiently polarized.

Second, this paper contributes to the growing body of work on the economics of misinformation. A key empirical contribution by Allcott and Gentzkow (2017) documents the spread of fake news and highlights the role of consumer biases and selective exposure. Experimental studies have shown that fact-checking can improve belief accuracy and reduce misperceptions (Nyhan and Reifler (2010);

Chan *et al.* (2017); Walter *et al.* (2020)), though its effects are often modest and context-dependent (Guess *et al.* (2020)). Platform-level interventions such as content warnings have also been shown to reduce belief in false claims (Pennycook and Rand (2020); Allcott *et al.* (2020)). On the theoretical side, formal models of misinformation remain relatively limited. Kartal and Tyran (2022) examine deception in political competition, while Levy and Razin (2015) study biased information transmission under reputational constraints. Eliaz and Spiegler (2020) model competing narratives as frames that select which aspects of reality audiences attend to, with strategic firms choosing narratives to attract followers; their focus on audience composition and belief formation is closely related to ours. Although inspection games have been used to model regulatory enforcement (Tsebelis (1990); Graetz *et al.* (1986)), few models have applied this framework to media environments in which both misreporting and fact-checking are strategic, interdependent, and shaped by audience composition. Our model addresses this gap by explicitly analyzing the strategic interaction between a fake news publisher and a fact-checker, where incentives are determined by the distribution of subscribers across the two outlets.

Third, the model connects to the literature on strategic information transmission. Crawford and Sobel (1982) establish the foundational framework for cheap talk, in which a sender with misaligned incentives transmits information strategically to a receiver. Our model shares this basic architecture but differs in two important respects: misreporting is costly, and the audience is not a single decision-maker but a heterogeneous population whose belief updating shapes the sender's payoffs. The costly misreporting specification relates most directly to Kartik (2009), who studies cheap talk when the sender incurs a lying cost and shows that such costs can sustain more informative equilibria than standard cheap talk allows. The reputational dimension of our model connects to Ottaviani and Sørensen (2006), who examine how reputational concerns discipline information transmission. The commitment problem at the heart of our second finding—that firm A would benefit from committing to honest reporting but cannot do so unilaterally—also resonates with the broader literature on the value of commitment in sender-receiver games, as illustrated by Kamenica and Gentzkow (2011) in the context of Bayesian persuasion. Unlike that literature, however, our focus is not on optimal information design but on the strategic interaction between two competing firms, neither of whom can unilaterally control the information environment.

The remainder of the paper is organized as follows. Section 2 presents the model, and Section 3 characterizes the equilibria. We show that five types of equilibria arise depending on the composition of the audience. Section 4 examines the effects of fact-checking, with particular attention to the payoffs of firm A. Section 5 investigates the conditions under which fact-checking is effective, and Section 6 concludes.

2. MODEL

Primitives. Let $\Omega = \{\omega_1, \omega_2\}$ denote the set of possible states of the world. The prior belief that the true state is ω_2 is $\pi \in (0,1)$, and thus the prior belief of ω_1 is $1-\pi$.

There are two media firms, A and B. Firm A represents a biased media outlet, whereas firm B acts as a fact-checker. When gathering news materials, firm A receives a binary signal $s \in \{s_1, s_2\}$, whose distribution depends on the true state. In state ω_1, A receives signal s_1 with probability $p \in (\frac{1}{2}, 1]$ and signal s_2 with probability $1-p \in [0, \frac{1}{2})$. Conversely, in state ω_2, A receives s_1 with probability $1-p$ and s_2 with probability p. Since $p \in (\frac{1}{2}, 1]$, each signal s_i is a noisy but informative indicator of the corresponding state ω_i.

A proportion $a \in [0,1)$ of subscribers consume only firm A's news, and a proportion $b \in [0,1)$ consume only firm B's. A proportion $m \in (0,1]$ consists of multihoming subscribers who consume news from both firms. Since all subscribers follow at least one outlet, $a+b+m=1$. These proportions are common knowledge among the players. We exclude degenerate cases in which a single firm holds full monopoly power in news consumption (i.e., $a=1$ or $b=1$), since these eliminate strategic interaction between the firms. We also exclude the case in which no subscriber consumes news from both firms (i.e., $m=0$), as such environment is unrealistic.

Firm A's news reports. After receiving a signal $s \in \{s_1, s_2\}$, firm A reports news $n_A \in \{s_1, s_2\}$, which may or may not match the received signal. If firm A *reports truthfully* ($n_A = s$), it incurs no cost. If firm A *misreports* ($n_A \neq s$), it incurs a misreporting cost c_A. Firm A's pure strategy is a mapping from signals to news reports. Its behavioral strategy is given by $\sigma_A : \{s_1, s_2\} \to [0,1]$, where for each signal s, $\sigma_A(s)$ denotes the probability that firm A misreports upon receiving s.

Firm B's fact-check. After observing firm A's news $n_A \in \{s_1, s_2\}$, firm B decides whether to fact-check (f) or not. Firm B's pure strategy is a mapping from observed reports to fact-checking decisions. Its behavioral strategy is given by

$\sigma_B : \{s_1, s_2\} \to [0,1]$, where for each report n_A, $\sigma_B(n_A)$ denotes the probability that firm B fact-checks upon observing n_A.

If firm B chooses to fact-check, it accurately observes the signal received by firm A at a cost c_B, while the true state remains unknown. Crucially, the fact-checker does not have direct access to the true state. This reflects the realistic limitation that underlying states are typically unverifiable, whereas news reports themselves are subject to scrutiny.[1] Upon fact-checking, firm B truthfully reports the uncovered signal. As a result, subscribers regard firm B as credible. We implicitly assume that this credibility is valuable to firm B. Misreporting as a fact-checker undermines subscribers' trust and impose prohibitively high reputation costs, thereby disciplining firm B's reporting behavior.

If firm B does not fact-check, it provides no additional information about the state. In this case, it issues a generic and non-informative signal s_0. Let $n_B \in \{s_0, s_1, s_2\}$ denote the news report by firm B. When $n_B \in \{s_1, s_2\}$, this indicates that firm B has conducted a fact-check and truthfully reported the uncovered signal. In contrast, $n_B = s_0$ corresponds to the case in which no fact-check was conducted.

A useful way to interpret the three subscriber groups is by their channel of exposure. A-only subscribers follow the biased outlet directly. B-only subscribers do not consume firm A's report at all; their only source of information is the fact-checker's verdict—a correction when A is caught misreporting, and otherwise receives no news from anywhere. They are the analogue of readers who rely on a debunking service or watchdog outlet (e.g., PolitiFact or Snopes) rather than on the partisan source itself. Multihoming subscribers observe both.

Subscribers' belief updates. Subscribers update their beliefs about the true state based on the news they receive (n_A and/or n_B).[2] Let $\rho_A(n_A)$ denote the posterior belief of A-only subscribers that the true state is ω_1, formed after observing news n_A. Similarly, let $\rho_B(n_B)$ denote the posterior belief of B-only sub-

[1] In practice, fact-checking serves as a "lie detector" rather than a "state-revelation device". Verifying the true state is often prohibitively costly or infeasible: many underlying states are revealed only ex post (e.g., policy outcomes or economic forecasts), inherently unobservable (e.g., intentions or private information), or require substantial resources to verify. As a result, fact-checking typically evaluates the internal consistency and credibility of a news report—such as the reliability of sources, the appropriate use of statistics, the presence of selective or distorted interpretations—rather than directly verifying the true state.

[2] We implicitly assume that when firm B conducts a fact-check, its subscribers learn firm A's original report n_A. This additional access to n_A does not affect subscribers' updated beliefs, because the fact-checked report n_B fully reveals the signal underlying n_A and is therefore more informative about the true state.

scribers about state ω_1 after observing news n_B. For multihoming subscribers, let $\rho_M(n_A, n_B)$ denote the posterior belief that the state is ω_1 after observing both n_A and n_B.

Firm A' payoff. Firm A strategically selects a news report to maximize weighted average belief among subscribers that the true state is ω_1, while taking into account the misreporting cost. For each $i \in \{1,2\}$, firm A's expected payoff from reporting news s_i, given firm B's strategy σ_B and received signal s, is denoted by $v_A(n_A = s_i \mid \sigma_B, s)$ and is defined as follows.

$$\begin{aligned} v_A(n_A = s_i \mid \sigma_B, s) = \sigma_B(s_i)\,&[a\rho_A(s_i) + b\rho_B(s) + m\rho_M(s_i, s)] \\ &+ \{1 - \sigma_B(s_i)\}\,[a\rho_A(s_i) + b\rho_B(s_0) + m\rho_M(s_i, s_0)] \\ &- c_A. \end{aligned}$$

This expected payoff depends on the received signal s through the beliefs that subscribers form upon consuming firm B's news conditional on fact-checking. The misreporting cost c_A is incurred whenever firm A misreports. It captures the internal cost of fabricating news—such as the editorial effort required to construct a false narrative or the legal exposure from publishing inaccurate information.[3]

Payoffs of firm B. Firm B receives a payoff of one if it successfully detects false news, and zero otherwise, that is, if the fact-checked news is truthful or if it does not fact-check. For any strategy σ_A of firm A and observed news n_A, firm B's payoff from fact-checking is given by

$$v_B(f \mid \sigma_A, n_A) = Q(n_A) - c_B,$$

where $Q(n_A)$ denotes firm B's belief, formed after observing report n_A, that firm A has issued false news. We assume that the fact-checking cost satisfies $c_B < \pi p + (1-\pi)(1-p)$, so that firm B prefers to fact-check whenever firm A's misreporting probability is sufficiently high in equilibrium.[4]

Subscribers do not make strategic decisions. Since the equilibrium analysis does not depend on their payoffs, we do not specify them explicitly.

[3]In Appendix C, a reputation cost is considered, which is incurred only if the misreport is detected. This cost is relevant only for firm B's subscribers, since A-only subscribers remain unaware of firm A's misreporting. The two cost specifications should not be interpreted as mutually exclusive descriptions of reality; instead, each offers a complementary perspective on how different misreporting costs shape equilibrium outcomes and information welfare.

[4]If $c_B \geq \pi p + (1-\pi)(1-p)$, it is always optimal for firm B not to fact-check; see equation (1).

Equilibrium. We adopt Perfect Bayesian equilibrium as the solution concept and focus on a class of equilibria in which firm A misreports only upon receiving signal s_2, information unfavorable to firm A. In such equilibria, firm A conceals relevant information from subscribers, potentially reducing their informational welfare. By construction, firm A's news s_2 is truthful. As a result, firm B has no incentive to fact-check this truthful report at a positive cost.

For simplicity, we use "misreporting" to refer to the case in which firm A misreports upon receiving signal s_2, and "fact-checking" to refer to the case in which firm B verifies firm A's news s_1 and thereby reveals the uncovered signal. For notational brevity, we represent equilibrium strategies by the pair (α^*, β^*), where α^* denotes firm A's misreporting probability and β^* denotes firm B's fact-checking probability.

We restrict attention to the *concealment* class of equilibria, in which firm A misreports only upon receiving the unfavorable signal s_2. Two remarks justify the reduction to a pair (α, β). First, the restriction on firm B is without loss: since A's report s_2 is always truthful, firm B never fact-checks this truthful report at a positive cost, and hence B's strategy is summarized by the single probability of fact-checking upon receiving news s_1. Second, the restriction on firm A is a deliberate selection. Equilibria in which A misreports after s_1 may exist for large b, but in this case, misreporting is intended to induce verification and thereby *raises* information welfare; since our focus is on the failure of fact-checking to deter concealment, we set those equilibria aside.

Definition 1 (Equilibrium)**.** A strategy profile (α^*, β^*) of the firms, together with firm B's belief function Q and subscribers' belief functions (ρ_A, ρ_B, ρ_M) constitutes an equilibrium if the following requirements are satisfied.

(i) Firm A's strategy satisfies $\alpha^* \in \arg\max_{\alpha \in [0,1]} \alpha v_A(s_1 \mid \alpha^*, \beta^*, s_2) + (1 - \alpha) v_A(s_2 \mid \alpha^*, , \beta^*, s_2)$, where the expected payoffs $v_A(s_1 \mid \alpha^*, \beta^*, s_2)$ and $v_A(s_2 \mid \alpha^*, \beta^*, s_2)$ reflect subscribers' beliefs satisfying part (iii).

(ii) Firm B's belief $Q(n_A)$ is consistent with firm A's strategy α^* on the equilibrium path. In equilibrium with $\alpha^* = 1$, if firm A's news s_2 is observed unexpectedly, probability one is assigned to the event that firm A received signal s_2; that is, $Q(s_2) = 0$. Firm B's strategy satisfies $\beta^* \in \arg\max_{\beta \in [0,1]} \beta v_B(f \mid \alpha^*, s_1)$.

(iii) Subscribers update their beliefs, $\rho_A(n_A)$, $\rho_B(n_B)$, and $\rho_M(n_A, n_B)$, according to Bayes rule whenever possible. Under the equilibrium $(\alpha^*, \beta^*) = (1, 1)$, two events occur off the equilibrium path: (1) firm A's deviation

$n_A = s_2$, and (2) firm B's deviation yielding $(n_A, n_B) = (s_1, s_0)$. In case (1), all subscribers assign probability one to the event that firm A received signal s_2. In case (2), A-only subscribers and multihomers form beliefs $\rho_A(s_1)$ and $\rho_M(s_1, s_0)$ according to Bayes rule; B-only subscribers assign probability one to the event that firm A received signal s_2.

Part (i) implies that, given subscribers' beliefs satisfying part (iii), firm A selects an optimal reporting strategy. Here, firm A takes the subscribers' updates induced by α^* as given when evaluating its payoff; it does not commit to a reporting strategy with subscribers' beliefs varying accordingly. Similarly, part (ii) states that firm B forms a consistent belief about whether firm A's news is false and makes an optimal fact-checking decision given the report provided by firm A. Part (iii) requires that subscribers' posterior beliefs be consistent with the two firms' equilibrium strategies along the equilibrium path. For a deviation with $n_A = s_2$ under equilibrium $(\alpha^*, \beta^*) = (1, 1)$, we adopt off-path beliefs that impose the maximal "punishment" on firm A; that is, subscribers assign probability one to the event that firm A received signal s_2.[5] Moreover, these beliefs satisfy the $D1$ criterion: any off-equilibrium belief under which deviation yields a strictly higher payoff after receiving signal s_1 also yields a strictly higher payoff after receiving signal s_2. Consequently, the set of beliefs under which deviation is optimal following s_1 is contained in that for s_2, and the deviation is therefore attributed to signal s_2. Appendix A provides the detailed formulas and a full characterization of the beliefs of both subscribers and firm B.

3. EQUILIBRIUM ANALYSIS

3.1. EQUILIBRIUM CHARACTERIZATION

We start by describing the key properties of subscribers' posterior beliefs. When firm B conducts a fact-check, all of its subscribers hold the same beliefs: $\rho_B(n_B) = \rho_M(n_A, n_B)$ for any n_A and $n_B \in \{s_1, s_2\}$, since belief updates are based solely on the verified information. Conversely, when firm B does not fact-check, multihoming subscribers align their beliefs with those of A-only subscribers: $\rho_A(n_A) = \rho_M(n_A, s_0)$ for any n_A, since for these multihomers, firm B's report s_0 is uninformative.

In particular, when $n_A = s_1$, an expected proportion $a + m(1 - \beta)$ of subscribers relies exclusively on firm A's news in forming their posterior beliefs.

[5]Firm B's payoffs are independent of subscribers' beliefs and hence any off-equilibrium belief is equally plausible from firm B's perspective.

This proportion takes values from a (when $\beta = 1$) to $a+m$ (when $\beta = 0$). We refer to this proportion as firm A's effective share, capturing its "monopolistic influence" on subscribers' belief formation. As firm A's effective share increases, a larger fraction of subscribers update their beliefs solely based on firm A's news $n_A = s_1$.

Now, we determine firm B's best-responses upon observing firm A's report s_1. Firm B optimally conducts a fact-check if

$$c_B \leq Q(s_1) \iff \alpha \geq \frac{c_B}{1-c_B}\frac{(1-\pi)p+\pi(1-p)}{(1-\pi)(1-p)+\pi p} \equiv \bar{\alpha}. \tag{1}$$

In essence, firm B fact-checks report s_1 if firm A misreports with sufficiently high probability.

Let $\bar{\alpha}$ denote the threshold probability satisfying $c_B = Q(s_1)$. Under the assumption that $0 < c_B < (1-\pi)(1-p)+\pi p$, this threshold lies in $(0,1)$ and is given by $\bar{\alpha} = \frac{c_B}{1-c_B}\frac{(1-\pi)p+\pi(1-p)}{(1-\pi)(1-p)+\pi p}$. It is increasing in c_B and decreasing in the probability $(1-\pi)(1-p)+\pi p$ that firm A receives signal s_2. Intuitively, a higher cost requires a greater likelihood of misreporting to justify fact-checking. Moreover, since firm A misreports only upon receiving s_2, a higher probability of s_2 raises the expected benefit from verification, so a lower misreporting probability suffices to induce fact-checking.[6]

The best response function $BR_B(\alpha)$ of firm B upon observing A's news s_1 specifies the set of optimal fact-checking probabilities as a function of misreporting probability α:

$$BR_B(\alpha) \in \begin{cases} \{1\} & \text{if } \alpha > \bar{\alpha}, \\ [0,1] & \text{if } \alpha = \bar{\alpha}, \\ \{0\} & \text{if } \alpha < \bar{\alpha}. \end{cases}$$

Based on firm B's best responses, the equilibrium strategy pairs can be classified into five types, summarized in Table 1. Each ordered pair describes the firms' strategies: the first component denotes firm A's misreporting probability, and the second denotes firm B's fact-checking probability. Types 1 and 5 are pure-strategy equilibria; types 2 and 4 are mixed-strategy equilibria in which only firm A randomizes; and type 3 is a mixed-strategy equilibrium in which both firms randomize. We will show that, given the parameter values (π, p, c_A, c_B), the equilibrium type is determined by the subscriber share (a, b, m). In stan-

[6]When $\pi > \frac{1}{2}$, the threshold probability $\bar{\alpha}$ decreases as the signal accuracy p increases; the opposite holds when $\pi < \frac{1}{2}$.

Type 1	Type 2	Type 3	Type 4	Type 5
$(1,1)$	$(\alpha,1);\ \alpha\in[\bar{\alpha},1)$	$(\bar{\alpha},\beta);\ \beta\in(0,1)$	$(\alpha,0);\ \alpha\in(0,\bar{\alpha}]$	$(0,0)$

Table 1: FIVE POSSIBLE EQUILIBRIUM TYPES. The table summarizes the five possible equilibrium strategy profiles, with the first component denoting firm A's misreporting probability and the second component denoting firm B's fact-checking probability.

dard games, a player's payoff depends on others' beliefs only indirectly, through their strategy choices. In our model, however, firm A's expected payoff is directly influenced by subscribers' beliefs about the state, and therefore by how they interpret the behavior of the two firms. Specifically, the same news pair $(n_A,n_B)=(s_1,s_0)$ yields different payoffs for firm A, depending on subscribers' beliefs about firm A's misreporting probability. Thus, although part (iii) of the equilibrium definition requires consistency between firm A's reporting strategy and subscribers' beliefs, firm A takes those beliefs as given when optimizing its reporting strategy.

To characterize firm A's incentive to misreport, consider firm B's strategy β and subscribers' belief $\hat{\alpha}$ about firm A's misreporting. Given the environment, firm A prefers to misreport upon receiving signal s_2 if and only if

$$\underbrace{\{a+m(1-\beta)\}\underbrace{[\rho_A(s_1;\hat{\alpha})-\rho_A(s_2)]}_{\equiv\gamma(\hat{\alpha})}}_{\equiv G(\hat{\alpha},\beta)} \ \geq\ \underbrace{b\beta\underbrace{[\rho_B(s_0;\hat{\alpha})-\rho_B(s_2)]}_{\equiv\lambda(\hat{\alpha},\beta)}+c_A}\,. \qquad (2)$$

Here $\rho_A(s_1;\hat{\alpha})$ and $\rho_B(s_0;\hat{\alpha})$ denote the updated beliefs about state ω_1 obtained by Bayes rule under their perceived misreporting probability $\hat{\alpha}$. Firm A takes $\hat{\alpha}$ as exogenous, and it enters directly into the objective function through subscribers' belief updating. In essence, firm A misreports if the expected gain from misreporting exceeds the corresponding loss.

The left-hand side of expression (2) represents the expected gain from providing false news s_1. Recall that subscribers regard A's news s_2 as truthful. For the effective share $a+m(1-\beta)$ of firm A, false news s_1 increases their belief in ω_1 by $\big(\rho_A(s_1;\hat{\alpha})-\rho_A(s_2)\big)$. In contrast, the right-hand side represents the expected loss from misreporting. When false news s_1 is fact-checked, B-only subscribers learn that firm A received signal s_2 and accordingly update their belief about ω_1 downward.[7]

[7]When false news s_1 is fact-checked, multihoming subscribers form belief $\rho_M(s_1,s_2)$, which coincides with $\rho_M(s_2,s_0)$, the belief formed when firm A reports truthfully.

We denote the left-hand side as the gain function $G(\hat{\alpha},\beta)$ and the right-hand side as the loss function $L(\hat{\alpha},\beta)$. Specifically, $G(\hat{\alpha},\beta)$ depends on $\hat{\alpha}$ through $\rho_A(s_1;\hat{\alpha})$, while $L(\hat{\alpha},\beta)$ depends on both $\hat{\alpha}$ and β through $\rho_B(s_0;\hat{\alpha})$. For ease of exposition, let $\gamma(\hat{\alpha}) = \rho_A(s_1;\hat{\alpha}) - \rho_A(s_2)$ and $\lambda(\hat{\alpha},\beta) = \rho_B(s_0;\hat{\alpha}) - \rho_B(s_2)$.

In the following example, we illustrate how firm A's best responses, and hence the resulting equilibria, vary with subscribers' belief $\hat{\alpha}$ in a setting.

Example 1. We examine the following parameter values: $\pi = \frac{1}{2}$, $p = \frac{3}{4}$, $c_A = \frac{1}{5}$, $c_B = \frac{1}{5}$, and $(a,b,m) = (\frac{11}{20}, \frac{2}{5}, \frac{1}{20})$. Under these values, the threshold misreporting probability is $\bar{\alpha} = \frac{1}{4}$.

First, suppose that subscribers believe that firm A misreports with probability $\frac{1}{4}$. Figure 1a plots $G(\bar{\alpha},\beta)$ as an orange dashed line and $L(\bar{\alpha},\beta)$ as a blue solid line, and the two curves intersect at $\beta = \frac{2}{5}$ and $\beta = \frac{8}{9}$. When firm B fact-checks with either a sufficiently low or a sufficiently high probability, that is, when $\beta < \frac{2}{5}$ or $\beta > \frac{8}{9}$, the gain from misreporting strictly exceeds the corresponding loss, so firm A optimally misreports. By contrast, for intermediate values of β, firm A prefers not to misreport. These yield firm A's best responses, represented by the blue curve in Figure 1b, while firm B's best-response curve is shown in red. The two best-response curves intersect three times. Among these candidates, $(1,1)$ cannot constitute an equilibrium because it violates the consistency requirement: although $\alpha^* = 1$, subscribers hold the belief that firm A misreports with probability $\frac{1}{4}$.

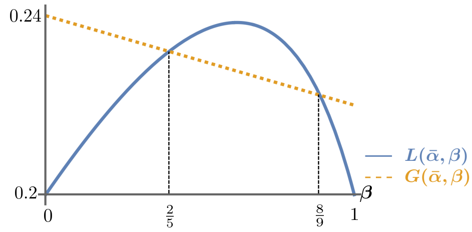
Now suppose that subscribers believe that firm A misreports with probability $\frac{3}{8}$. As in the previous case, Figure 1c plots the gain and loss functions for $\hat{\alpha} = \frac{3}{8}$, and Figure 1d displays the best-response functions of firms A and B, shown in blue and red, respectively. Although the two curves intersect infinitely many times at $(\alpha,1)$ for $\alpha \geq \frac{1}{4}$, only one of these intersections satisfies the consistency requirement. In equilibrium, firm A misreports with probability $\frac{3}{8}$ and firm B fact-checks with certainty. □

The following lemma summarizes key properties of the gain and loss functions, which would be useful for characterizing equilibrium conditions.

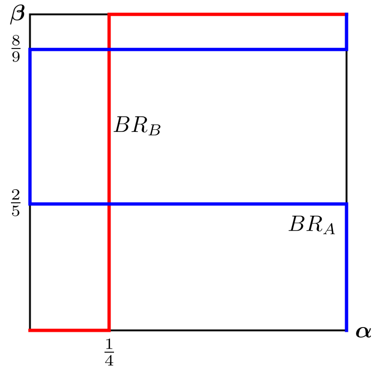
Lemma 1. *The gain and loss functions have the following properties.*

(i) We have $\gamma(\hat{\alpha}) > 0$ for all $\hat{\alpha} \in [0,1]$, and $\gamma(\hat{\alpha})$ is strictly decreasing in $\hat{\alpha}$. $G(\hat{\alpha},\beta)$ is strictly decreasing and linear in β.

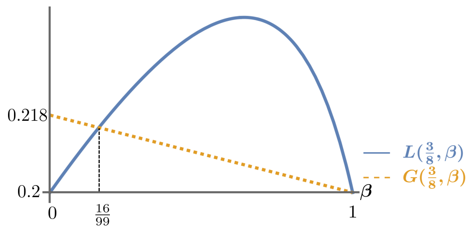
(ii) When $b = 0$ or $\beta \in \{0,1\}$, we have $L(\hat{\alpha},\beta) = c_A(s_1 \mid s_2)$ for all $\hat{\alpha}$. When $b \neq 0$ and $\beta \in (0,1)$, $L(\hat{\alpha},\beta)$ is strictly increasing in $\hat{\alpha}$. Moreover, it is strictly concave in β for any $\hat{\alpha} \neq 1$ and weakly concave in β for $\hat{\alpha} = 1$.



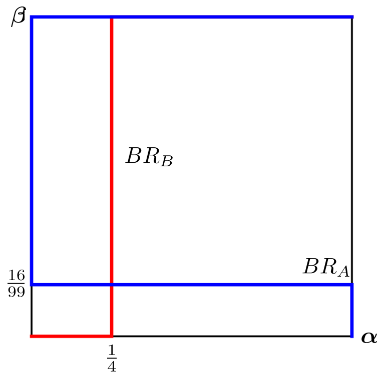
(a) $G(\bar{\alpha},\beta)$ AND $L(\bar{\alpha},\beta)$. This panel plots the gain and loss functions when subscribers believe that firm A misreports with probability $\bar{\alpha}$.



(b) BR CURVES WITH BELIEFS $\hat{\alpha} = \frac{1}{4}$. This panel shows the best-response curves of firms A and B under the belief $\hat{\alpha} = \frac{1}{4}$.



(c) $G(\frac{3}{8},\beta)$ AND $L(\frac{3}{8},\beta)$. This panel plots the gain and loss functions when subscribers believe that firm A misreports with probability $\frac{3}{8}$.



(d) BR CURVES WITH BELIEFS $\hat{\alpha} = \frac{3}{8}$. This panel shows the best-response curves of firms A and B under the belief $\hat{\alpha} = \frac{3}{8}$.

Figure 1: BEST-RESPONSE CURVES FOR FIRM A DEPENDING ON SUBSCRIBERS' BELIEFS $\hat{\alpha}$. The panels compare the gain and loss functions and the associated best-response curves under alternative beliefs about firm A's misreporting probability.

Table 2 provides a set of conditions under which each equilibrium type arises. The first column lists the condition labels, and the second column presents the corresponding algebraic conditions in terms of subscriber shares.[8]

[8]We exclude $m = 0$ mainly on interpretive grounds, as multihoming is present in virtually any

C_1	$a \geq \frac{c_A}{\gamma(1)}$
C_2	$\frac{c_A}{\gamma(\bar{\alpha})} \leq a < \frac{c_A}{\gamma(1)}$
C_3	$\exists \beta \in (0,1); G(\bar{\alpha},\beta) = L(\bar{\alpha},\beta)$
C_4	$\frac{c_A}{\gamma(0)} < a+m \leq \frac{c_A}{\gamma(\bar{\alpha})}$
C_5	$a+m \leq \frac{c_A}{\gamma(0)}$

Table 2: CONDITIONS ON SUBSCRIBER SHARES (a,b,m) FOR THE EXISTENCE OF EACH EQUILIBRIUM TYPE. The table lists the algebraic conditions on subscriber shares corresponding to the five equilibrium types described in Table 1.

Proposition 1. *An equilibrium of type i exists if and only if condition C_i holds for $i \in \{1,2,3,4,5\}$.*

Condition C_1 implies that, even when firm B fact-checks with certainty, the gain from misreporting is at least as large as the corresponding loss. This condition holds when a sufficiently large fraction of subscribers consumes firm A's news exclusively. By contrast, condition C_5 implies the opposite extreme: even when firm B never fact-checks, the gain from misreporting does not exceed the associated loss. This condition holds when firm A's effective share $a+m$ is sufficiently small, or equivalently, when a sufficiently large fraction of subscribers consumes firm B's news exclusively.

Condition C_2 implies that, when firm B fact-checks with certainty, firm A's effective share a is moderately large, rendering firm A indifferent between its two reporting options for some belief $\hat{\alpha} \in [\bar{\alpha},1)$. By contrast, condition C_4 implies that, when firm B never fact-checks, firm A's effective share $a+m$ is moderately small, again rendering firm A indifferent between its two reporting options for some belief $\hat{\alpha} \in (0,\bar{\alpha}]$.

While conditions C_3 does not admit a simple closed-form characterization, it requires the existence of a fact-checking probability $\beta^* \in (0,1)$ that equates the expected gain and loss from misreporting, given that subscribers believe that firm A misreports with probability $\bar{\alpha} \in (0,1)$.

Figure 2 illustrates the regions of subscriber shares satisfying conditions C_1 through C_5. Near the lower-left corner, a large fraction of subscribers consumes

media market. The only peculiarity is that, with $m = 0$, conditions C_2 and C_4 are simultaneously satisfied at a non-generic, single point ($a = c_A/\gamma(\bar{\alpha})$), which carries no economic significance but would require separate treatment.

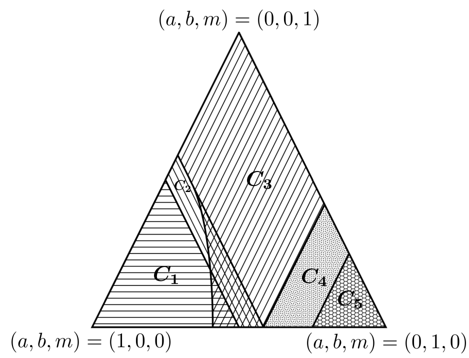


Figure 2: SUBSCRIBER SHARES SATISFYING $C_1 - C_5$: $\pi = \frac{1}{2}$, $p = \frac{3}{4}$, $c_A = \frac{1}{8}$, $c_B = \frac{2}{5}$. The figure illustrates how the subscriber-share simplex is partitioned into regions corresponding to the five equilibrium conditions.

firm A's news exclusively, and an equilibrium in which firm A misreports with certainty arises. As this fraction decreases, firm A's misreporting probability declines—from one under C_1 to an intermediate probability under C_2, and to $\bar{\alpha}$ under C_3. As a larger fraction of subscribers consumes firm B's news exclusively, firm A's misreporting probability falls below $\bar{\alpha}$ and eventually to zero near the lower-right corner.

Figure 2 further shows that the equilibrium conditions collectively partition the entire space of subscriber shares, ensuring that for any (a, b, m), at least one equilibrium type in Table 1 arises. This result holds generally and is not specific to the parameter values used in Figure 2, and is formally stated in Proposition 2. Importantly, although all conditions other than C_3 are mutually exclusive, C_3 may hold simultaneously with C_1 or C_2, giving rise to multiple equilibria.

Proposition 2. *For every subscriber share (a, b, m), an equilibrium of one of the types in Table 1 exists; equivalently, the conditions from C_1 to C_5 cover the simplex of subscriber shares.*

In the proof, we show that for any subscriber share satisfying none of the conditions C_1, C_2, C_4, and C_5 satisfies C_3. Hence, from Table 2, the region in which C_3 holds uniquely is characterized by $a < c_A/\gamma(\bar{\alpha})$ and $b < 1 - c_A/\gamma(\bar{\alpha})$, which excludes the two corners of the simplex where $a \approx 1$ or $b \approx 1$.

3.2. MULTIPLE EQUILIBRIA

When multiple equilibria exist, we first exclude unstable equilibria. Among the remaining candidates, we select the equilibrium that yields the highest interim payoff to firm A upon receiving signal s_2. We regard this equilibrium as plausible because, as the first mover, firm A can influence the expectations of firm B and subscribers through its choice of misreporting probability.

Multiple equilibria arise precisely when $G(\bar{\alpha},1) \geq L(\bar{\alpha},1)$ and either (i) the graphs of $G(\bar{\alpha},\beta)$ and $L(\bar{\alpha},\beta)$ are tangent to each other, or (ii) they intersect at two distinct points. The former case occurs only for specific subscriber shares and is therefore non-generic.[9] In the latter case, we select the type-3 equilibrium $(\bar{\alpha},\beta^*)$ such that $G(\bar{\alpha},\beta)$ intersects $L(\bar{\alpha},\beta)$ from above at β^*. A more detailed discussion of equilibrium selection is provided in Appendix B.

Recall that in Example 1, there are three equilibria $(\alpha^*,\beta^*) \in \{(\frac{1}{4},\frac{2}{5}),(\frac{1}{4},\frac{8}{9}),(\frac{3}{8},1)\}$, one type-2 and two type-3 equilibria. Among these, we exclude the unstable equilibrium $(\frac{1}{4},\frac{8}{9})$, and then select $(\frac{1}{4},\frac{2}{5})$, since it yields a higher interim payoff conditional on receiving s_2 than $(\frac{3}{8},1)$.

We refine conditions C_1 and C_2 by restricting attention to cases in which C_3 does not hold. That is, we interpret condition C_i for $i \in \{1,2\}$ as applying only when C_3 fails. By construction, each condition then corresponds to a unique equilibrium type.

All results below are reported for the selected equilibrium. The selection matters only within the overlap of C_3 with C_1 or C_2: an alternative rule selecting the type-1 or type-2 stable equilibrium would relabel the overlap accordingly, and every result in the paper continues to hold unchanged under the relabeled partition (See Appendix B). The selection rule is therefore a matter of labeling, not a choice among qualitatively distinct predictions. However, pointwise predictions inside the overlap may differ across selections; for instance, the same change in subscriber shares may leave misreporting unchanged under our selection and increase it under the alternative.

4. THE EFFECTS OF FACT-CHECKING

In this section, we study the effects of fact-checking on A's misreporting behavior and payoff. We begin by comparing A's equilibrium misreporting probability in the presence and absence of B, establishing that fact-checking deters

[9] Only when $G(\bar{\alpha},1) \geq L(\bar{\alpha},1)$, either condition C_1^z or C_2^z holds. If instead $G(\bar{\alpha},1) < L(\bar{\alpha},1)$, the equilibrium is unique. In the tangency case, we exclude the unstable type-3 equilibrium and select the remaining equilibrium (of type 1 or type 2).

misreporting. We then turn to the more surprising implication: that A may itself be strictly better off when an active fact-checker is present.

The key to understanding this result is a dynamic inconsistency problem. Firm A would benefit from committing to a lower misreporting probability, since doing so avoids the costs associated with misreporting. However, such a commitment is not self-enforceable: whenever A receives an unfavorable signal, it faces an incentive to misreport. The fact-checker resolves this problem partially by acting as an external commitment device. The threat of detection raises the expected cost of misreporting, disciplining A in a way it cannot discipline itself. This restraint, by preserving credibility, can translate into a higher ex-ante payoff for A—giving rise to what we term adversarial symbiosis, wherein the fact-checker and the biased outlet each benefit from the other's presence.

4.1. THE DETERRENT EFFECT

We begin by considering a benchmark case in which firm B does not engage in fact-checking. In this setting, B-only subscribers do not regard firm A's news as an information source and therefore do not consume A's news as in the baseline specification. As a result, they retain their prior belief and do not update. Thus, the prior belief may be viewed as the probability that subscribers assign to states based on all media sources other than firm A. This benchmark captures environments in which news content is government-controlled or in which the fact-checking cost is prohibitively high.[10] We continue to focus on equilibria in which firm A misreports only upon receiving signal s_2.

Proposition 3. *Suppose that firm B is inactive. Firm A's misreporting probability is weakly higher than when firm B is active, and strictly higher under conditions C_2 and C_3.*

Figure 3 illustrates the equilibrium misreporting probabilities under the same parameter values as in Example 1. The blue solid line corresponds to the case in which firm B is active, whereas the orange dashed line corresponds to the case in which firm B is inactive. Along the horizontal axis, a increases, holding m fixed. For each subscriber share (a,b,m), we indicate the equilibrium condition it satisfies. Fact-checking deters misreporting in the sense that firm A misreports (weakly) less frequently when firm B is active. This deterrent effect, however, is limited. First, fact-checking fully eliminates misreporting only under condition

[10]If the fact-checking cost is prohibitively high, firm$B$ would never engage in fact-checking and therefore is effectively inactive. This provides an alternative justification for assuming a low fact-checking cost in the main analysis.

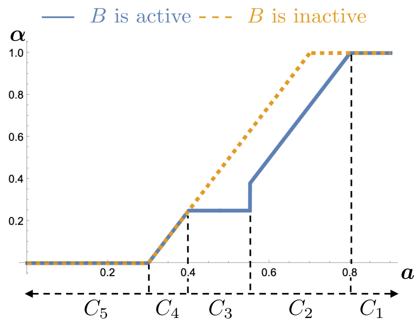


Figure 3: MISREPORTING BEHAVIOR WHEN B IS ACTIVE AND NOT: $\pi = \frac{1}{2}$, $p = \frac{3}{4}$, $c_A = c_B = \frac{1}{5}$, AND $m = \frac{1}{10}$. The figure compares firm A's equilibrium misreporting probability when firm B is active with the benchmark case in which firm B is inactive.

C_5—a case in which firm A never misreports regardless of whether firn B is present. Second, firm A continues to misreport with positive probability under all conditions other than C_5.

4.2. THE EFFECT ON FIRM A'S PAYOFF

Figure 4 presents the ex-ante payoffs of firm A under the same parameter values as in Example 1 and follows the same format as Figure 3.

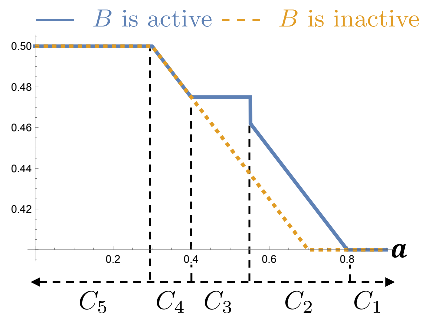


Figure 4: FIRM A'S EX-ANTE PAYOFFS: $\pi = \frac{1}{2}$, $p = \frac{3}{4}$, $c_A = c_B = \frac{1}{5}$, AND $m = \frac{1}{10}$. The figure shows how firm A's ex-ante payoff varies with subscriber shares.

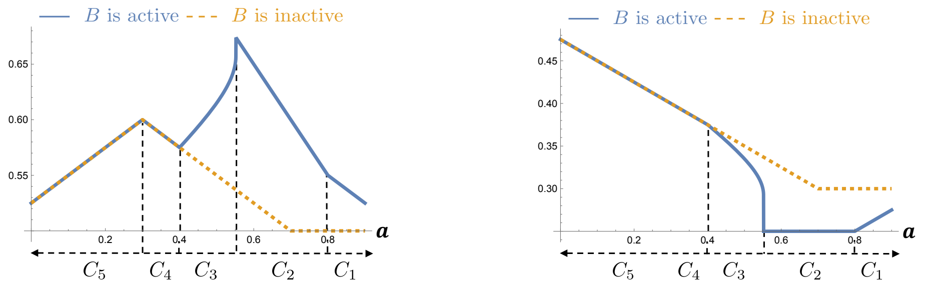
Two observations are noteworthy. First, firm A is weakly better off when firm B is active, and is strictly better off under conditions C_2 and C_3. Second, Figure 4 is essentially the mirror image of Figure 3, reflected vertically. This is because firm A's ex-ante payoff is given by $(1-\pi)-\phi_2\alpha^* c_A$, where ϕ_2 denotes the probability that firm A receives signal s_2 and α^* is the equilibrium misreporting probability. Therefore, the improvement in firm A's ex-ante payoff directly reflects the reduction in misreporting induced by fact-checking.

By construction, when firm B is inactive, its ex-ante payoff is identically zero. By contrast, when it fact-checks with a positive probability, its payoff must be non-negative, and thus weakly improve. Combined with the improvement in firm A's welfare, this indicates that the interaction between the two firms exhibits adversarial symbiosis. The above observations are not driven by the specific parameter values used for Figure 4 and are formalized in the corollary below.

Corollary 1. *Firm A is weakly better off—and strictly so under conditions C_2 and C_3—when firm B is active. The interaction between the two firms exhibits an adversarial symbiosis.*

We next decompose the effect of fact-checking on firm A's payoff by comparing its interim payoffs when firm B is active and inactive. Figure 5 plots these interim payoffs under the same parameter values as in Example 1 and follows the same format as Figure 4. When firm B is active, firm A's payoff is weakly higher upon receiving signal s_1 and weakly lower upon receiving signal s_2 relative to the case in which firm B is inactive. The underlying intuition is as follows. After receiving signal s_1, firm A benefits from the increased credibility resulting from a lower misreporting probability, which is further reinforced by a higher likelihood of fact-checking. By contrast, after receiving signal s_2, although firm A likewise benefits when misreporting goes undetected, this effect is outweighed by the higher likelihood of having to report the truthful news s_2, together with the greater fact-checking intensity. As a result, firm A is worse off under s_2 when B is active.

One might conjecture that fact-checking enhances the credibility of firm A's news and thereby creates an environment conducive to misreporting. This reasoning, however, reverses the direction of causality. In our model, improved credibility is a consequence of reduced misreporting rather than a driver of misreporting. Moreover, the payoff structure itself contradicts this conjecture. As shown in Figure 5, when firm A misreports after receiving s_2, its payoff is strictly lower when firm B is active. Hence, fact-checking does not create a misreporting-friendly environment; rather, it reduces the profitability of misreporting.



(a) FIRM A'S PAYOFF UPON RECEIVING s_1. This panel compares firm A's interim payoff after the favorable signal.

(b) FIRM A'S PAYOFF UPON RECEIVING s_2. This panel compares firm A's interim payoff after the unfavorable signal.

Figure 5: INTERIM PAYOFFS OF FIRM A: $\pi = \frac{1}{2}$, $p = \frac{3}{4}$, $c_A = c_B = \frac{1}{5}$, AND FIXED $m = \frac{1}{10}$. The two panels report firm A's interim payoffs after receiving each signal, comparing the case with an active fact-checker to the benchmark without fact-checking.

5. THE LIMITS OF FACT-CHECKING

The deterrent effect established in Section 4 is not unconditional; the success of fact-checking in disciplining firm A depends crucially on the degree of market polarization. We say the market is highly polarized when most subscribers follow a single outlet exclusively, i.e. the share lies near a corner ($a \approx 1$ or $b \approx 1$) of the simplex.; this is distinct from fragmentation ($a+b \approx 1$, i.e. $m \approx 0$). When the media market is highly polarized, fact-checking loses its disciplining power. Otherwise, for instance, in a market where multihoming is sufficiently prevalent, firm B 's verification effectively restrains misreporting compared to the scenario where firm B is absent.

The intuition is straightforward. When the market is highly polarized, two distinct forces neutralize the impact of fact-checking. First, firm A's subscriber base is too limited to justify the costs of misreporting. Second, since most subscribers consume exclusively firm A's news, the threat of detection remains weak, and firm A misreports with probability one even when B is active. Conversely, when the market is not sufficiently polarized, a moderate fraction consumes A's news, leaving room for strategic manipulation. At the same time, a moderate fraction receives B's corrections, making detection costly and therefore, keeping the misreporting probability below one.

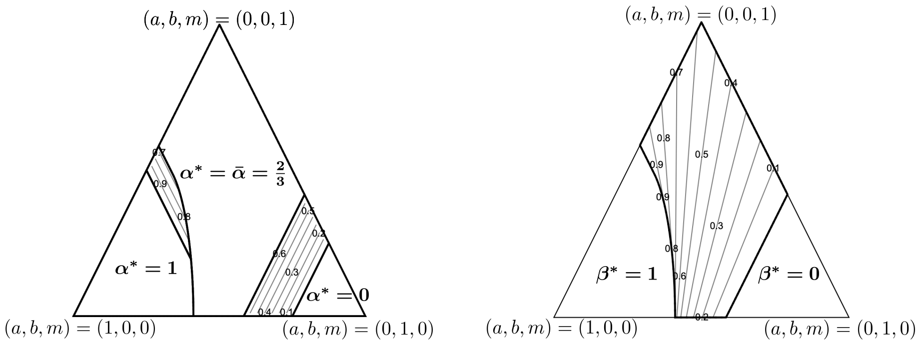
How changes in audience composition affect equilibrium behavior depends on where the subscriber shares lie in the parameter space. Under condition C_3, where both firms randomize, B responds to stronger misreporting incentives by increasing its fact-checking probability, keeping A's misreporting probability fixed at $\bar{\alpha}$. Informational welfare can then improve as the intensity of the contest between misreporting and verification increases, directly reflecting Mill's argument. Under conditions C_2 and C_4, by contrast, B cannot adjust its fact-checking intensity, and the same shift results in higher misreporting without any compensating increase in verification. Section 5.1 characterizes these equilibrium responses, and Section 5.2 traces the implications for subscribers' informational welfare.

5.1. EQUILIBRIUM RESPONSES TO AUDIENCE COMPOSITION

This section analyzes how changes in subscriber shares affect the equilibrium behavior of the two firms. Since changes in subscriber shares within the regions characterized by conditions C_1 or C_5 do not alter the firms' equilibrium behavior, we focus on subscriber shares satisfying C_2, C_3, or C_4. Before analyzing each case in detail, Figure 6 displays equilibrium misreporting and fact-checking probabilities over the subscriber-share simplex. Inside the simplex, the thick lines partition the space into regions corresponding to conditions C_1 to C_5. In particular, in Figure 6b, these lines separate the subscriber shares associated with equilibria in which $\beta^* = 0$, $\beta^* \in (0,1)$, and $\beta^* = 1$. This figure would serve as a visual guide to how equilibrium behavior varies with (a,b,m), and the subsequent discussion explains the forces underlying these patterns.

We begin by examining firm A, focusing on changes in subscriber shares within the regions characterized by conditions conditions C_2 and C_4, respectively. Under condition C_2, an equilibrium is characterized by $\beta^* = 1$, and B-only and multihoming subscribers correctly infer firm A's signal. As a result, firm A's misreporting behavior depends only on the proportion of A-only subscribers, that is, on firm A's effective share. As a increases, the gain from misreporting rises, leading firm A to misreport more frequently. Under condition C_4, equilibrium is characterized by $\beta^* = 0$, and firm A's effective share is given by $a+m$. Hence, as $a+m$ increases (equivalently, as b decreases), the gain from misreporting rises, again leading firm A to misreport more frequently.

We now turn to firm B and focus on changes in subscriber shares within the region characterized by condition C_3. In this region, equilibrium is characterized by $\bar{\alpha} \in (0,1)$. The key question is how such changes affect firm B's fact-checking probability while maintaining firm A's indifference between its two reporting



(a) FIRM A'S MISREPORTING PROBABILITIES. This panel reports firm A's equilibrium misreporting probability over the subscriber-share simplex.

(b) FIRM B'S FACT-CHECKING PROBABILITIES. This panel reports firm B's equilibrium fact-checking probability over the subscriber-share simplex.

Figure 6: EQUILIBRIUM BEHAVIORS: $\pi = \frac{1}{2}$, $p = \frac{3}{4}$, $_A = \frac{1}{8}$, AND $c_B = \frac{2}{5}$. The panels display equilibrium misreporting and fact-checking probabilities over the subscriber-share simplex, with the thick lines indicating the boundaries between equilibrium regions.

options under subscribers' belief $\bar{\alpha}$. Proposition 4 shows that when changes in subscriber shares favor firm A—for example, when a increases and b decreases—firm B responds by fact-checking with higher probability.

Proposition 4. *The equilibrium responses of the two firms to changes in subscriber shares can be summarized as follows.*

(i) Firm A misreports with higher probability when a increases under condition C_2, and when b decreases under C_4.

(ii) Let $(\bar{\alpha}, \beta^*)$ be an equilibrium strategy profile under condition C_3. A change $(\Delta a, \Delta b, \Delta m)$ in subscriber shares induces firm B to increase its fact-checking probability if

$$\Delta a \beta^* \cdot \gamma(\bar{\alpha}) > \Delta b \left[\beta^* \lambda(\bar{\alpha}, \beta^*) + (1-\beta^*)\gamma(\bar{\alpha})\right].$$

Proposition 4 further highlights the deterrent effect of fact-checking. Suppose that $\Delta a > 0$ and $\Delta b < 0$ within the region characterized by condition C_3. Holding β^* fixed, such a change increases $G(\bar{\alpha}, \beta^*)$ and decreases $L(\bar{\alpha}, \beta^*)$, so that $G(\bar{\alpha}, \beta^*) > L(\bar{\alpha}, \beta^*)$. Firm A would therefore prefer to raise its misreporting probability. However, in the new equilibrium, firm A's misreporting probability remains at $\bar{\alpha}$. This outcome reflects the deterrent effect: firm B responds to the strengthened misreporting incentives of firm A by raising its fact-checking probability above β^*, thereby preventing firm A from misreporting more frequently.

By contrast, under conditions C_2 and C_4, the same change in subscriber shares, $\Delta a > 0$ and $\Delta b < 0$, results in a higher misreporting probability. The difference arises because firm B is unable to adjust its fact-checking behavior in response. Under condition C_2, firm B's deterrence capacity is already fully utilized, leaving no scope to increase its fact-checking intensity. Under condition C_4, fact-checking is not triggered because firm A's misreporting probability remains low. As a result, firm B continues to refrain from fact-checking.

We say that a subscriber share becomes *more fragmented* if the proportion of multihoming subscribers decreases while the ratio of A-only to B-only subscribers remains constant. Proposition 4 implies that under condition C_3, the effect of fragmentation on fact-checking depends on the relative sizes of the A-only and B-only subscriber bases. Fragmentation therefore does not have a uniform effect on fact-checking; what matters is not fragmentation per se, but the relative size of the subscriber bases.

Figure 7 plots the equilibrium fact-checking probabilities for two cases, $\frac{a}{b} = \frac{3}{7}$ and $\frac{a}{b} = \frac{7}{3}$, holding π, p, c_A, and c_B fixed as in Example 1. The horizontal axis measures the degree of fragmentation, with larger values of m corresponding to a less fragmented subscriber share. As fragmentation increases (i.e., as m falls), firm B reduces its fact-checking probability when $\frac{a}{b} = \frac{3}{7}$, but increases it when $\frac{a}{b} = \frac{7}{3}$.[11]

Corollary 2. *When subscriber shares become more fragmented under condition C_3, firm B raises its fact-checking probability if and only if*

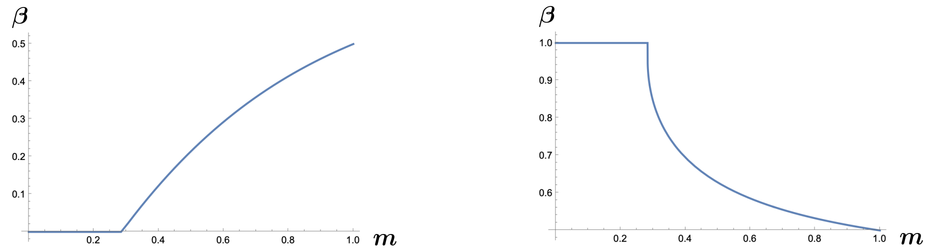
$$\frac{a}{b} > \frac{\beta^*\lambda(\bar{\alpha},\beta^*) + (1-\beta^*)\gamma(\bar{\alpha})}{\beta^*\gamma(\bar{\alpha})},$$

If subscriber shares become less fragmented, firm B raises its fact-checking probability if and only if the inequality is reversed.

5.2. INFORMATION WELFARE

This section analyzes how subscriber shares shape their informedness. To this end, we introduce a measure of informedness, which we interpret as sub-

[11] When $\frac{a}{b} = \frac{3}{7}$, the subscriber shares pass through the regions characterized by condition C_3 and C_4 and then by C_5, as fragmentation increases. When $\frac{a}{b} = \frac{7}{3}$, they pass through the regions characterized by C_3 and C_2.



(a) WHEN $\frac{a}{b} = \frac{3}{7}$. This panel traces fact-checking behavior as fragmentation changes while the subscriber-share ratio is fixed at $\frac{3}{7}$.

(b) WHEN $\frac{a}{b} = \frac{7}{3}$. This panel traces fact-checking behavior as fragmentation changes while the subscriber-share ratio is fixed at $\frac{7}{3}$.

Figure 7: FRAGMENTATION EFFECT ON FACT-CHECKING: $\pi = \frac{1}{2}$, $p = \frac{3}{4}$, $c_A = \frac{1}{8}$, AND $c_B = \frac{2}{5}$. The figure compares how firm B's equilibrium fact-checking probability changes with fragmentation under two different ratios of A-only to B-only subscribers.

scribers' utility. For each subscriber type $k \in \{A,B,M\}$, utility u_k is defined as one minus the expected error in beliefs about the state. Formally, for an equilibrium (α^*,β^*), the utility of type-k subscribers is:

$$\begin{aligned}
u_k(\alpha^*,\beta^*) = 1 - \Bigg\{ &(1-\pi) \sum_{n_k \in N_k} pr(n_k \mid \omega_1)\,(1-\rho_k(n_k)) \\
&+ \pi \sum_{n_k \in N_k} pr(n_k \mid \omega_2)\rho_k(n_k) \Bigg\} \\
= 1 - 2 \sum_{n_k \in N_k} &\{(1-\pi)pr(n_k \mid \omega_1) + \pi pr(n_k \mid \omega_2)\} \\
&\times \rho_k(n_k)\,(1-\rho_k(n_k)),
\end{aligned}$$

where N_k denotes the set of news items (or pairs of items) observable to type-k subscribers, and $pr(n_k \mid \omega)$ denotes the probability that a type-k subscriber observes news $n_k \in N_k$ in state $\omega \in \{\omega_1, \omega_2\}$, under the firms' equilibrium strategies. The second line follows from the fact that $(1-\pi)pr(n_k \mid \omega_1)\,(1-\rho_k(n_k))$ $= \pi pr(n_k \mid \omega_2)\rho_k(n_k) = [(1-\pi)pr(n_k \mid \omega_1) + \pi pr(n_k \mid \omega_2)]\,\rho_k(n_k)(1-\rho_k(n_k))$. For subscriber share (a,b,m), the aggregate information welfare is given by

$$U(a,b,m) = au_A(\alpha^*,\beta^*) + bu_B(\alpha^*,\beta^*) + mu_M(\alpha^*,\beta^*),$$

where (α^*,β^*) is the equilibrium induced by the subscriber distribution (a,b,m).

Although subscribers take no action, the measure u_k has a decision-theoretic foundation. Suppose a type-k subscriber later chooses an action $x \in [0,1]$ and incurs loss $(\mathbf{1}\{\omega = \omega_1\} - x)^2$, where $\mathbf{1}$ is an indicator function. The optimal action is $x = \rho_k$ and the expected loss is $\rho_k(1-\rho_k)$; integrating over the realized news and taking an affine transformation yields exactly u_k as defined above. Belief accuracy is thus a relevant welfare criterion because it directly measures the expected payoff of a decision-maker acting on those beliefs.

Two effects are at work. First, because subscribers receive different news depending on their subscription patterns, a change in shares directly affects information welfare; we call this as the composition effect. Second, because the same change alters the firms' equilibrium behavior, it indirectly affects information welfare; we call this as the equilibrium effect. For certain changes in subscriber shares, the impact on information welfare is straightforward for one of two reasons: either the equilibrium effect is absent, as under conditions C_1 and C_5, or two effects operate in the same direction, as under C_2.

Under condition C_1, firm A always reports news s_1, and firm B always fact-checks. As a result, firm B's subscribers are fully informed about firm A's signal, whereas A-only subscribers gain no information beyond their prior belief. Therefore, an increase in a reduces information welfare.

Under C_5, firm B does not engage in fact-checking, so B-only subscribers suffer from informational vacuum. In contrast, firm A's subscribers learn its signal. Consequently, information welfare improves as fewer subscribers rely exclusively on firm B's news.

Now, we turn to condition C_2, under which firm B's subscribers are fully informed about firm A's signal. An increase in a reduces information welfare for two reasons. First, the proportion of less-informed A-only subscribers rises. Second, firm A increases its misreporting probability in new equilibrium, making A-only subscribers even less informed.

Proposition 5. *Information welfare is strictly decreasing in a under C_1 and C_2, strictly increasing in b under C_4, strictly decreasing in b under C_5. Under C_3, information welfare increases in a, with b held fixed, strictly if $b \neq 0$ and weakly if $b = 0$, and strictly decreases in b, with a held fixed.*

Proposition 5 implies that information welfare is maximized at subscriber shares satisfying (i) $b = 0$ under C_3, (ii) the smallest a under C_2, or (iii) on the boundary of the regions characterized by C_4 and C_5, where a type-5 equilibrium arises.

Figure 8 depicts information welfare across the space of subscriber shares using a blue-green-yellow color scale, where blue represents lower welfare and

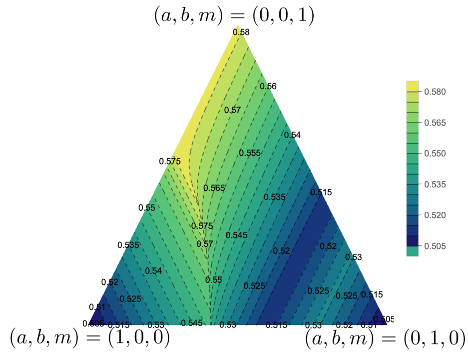


Figure 8: INFORMATION WELFARE: $\pi=\frac{1}{2}$, $p=\frac{3}{4}$, $c_R=\frac{1}{8}$, $c_B=\frac{2}{5}$. The figure depicts aggregate information welfare across the subscriber-share simplex, with lighter regions representing higher levels of informedness.

yellow represents higher welfare. Subscribers are least informed when shares are highly "polarized", that is, when most subscribers consume a single firm's news exclusively. In such cases, the firm that most subscribers rely on provides little information. Near the lower-left corner, the majority of subscribers consume only firm A's news. In type-1 equilibrium, they receive only news s_1 regardless of the realized signal and therefore cannot update beyond their prior belief. By contrast, near the lower-right corner, most subscribers are B-only subscribers. In a type-5 equilibrium with fact-checking probability equal to zero, they receive virtually no news from either firm.

Figure 8 further demonstrates that subscribers are better informed when subscriber shares are neither highly fragmented nor highly polarized (for example, when $m=1$). Intuitively, one subscriber share satisfying the requirement (i) is $m=1$. In this case, information welfare is maximized, as all subscribers access all available information sources. The subscriber share satisfying the requirement (ii) leads to a type-2 equilibrium with $(\bar{\alpha},1)$. When firm A misreports with probability $\bar{\alpha}$, information welfare is maximized if firm B fact-checks with maximal probability. Indeed, by the continuity of welfare measure in (a,b,m), the shares satisfying (i) and (ii) yield the same level of information welfare.

Information welfare can be maximized at subscriber shares satisfying requirement (iii), particularly when the fact-checking cost is high. In this case, $\bar{\alpha}$ is close to one, and at the shares satisfying (ii), the majority of subscribers consume firm A's news exclusively. Hence, in a type-2 equilibrium $(\bar{\alpha},1)$, only a small fraction of firm B's subscribers learn firm A's signal, while the remaining subscribers, who constitute the majority, receive almost no information, resulting in a low level of information welfare.

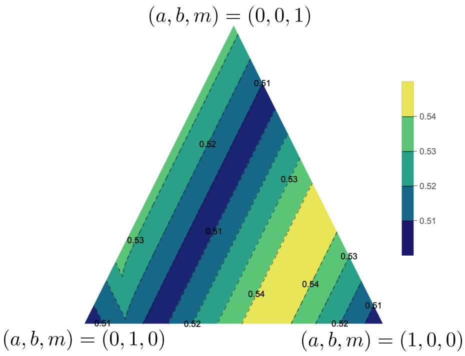


Figure 9: INFORMATION WELFARE: $\pi = \frac{1}{2}$, $p = \frac{3}{4}$, $c_A = \frac{1}{5}$, $c_B = \frac{19}{40}$. The figure illustrates information welfare when the fact-checking cost is relatively high.

Figure 9 depicts information welfare with a relatively high fact-checking cost $c_B = \frac{19}{40}$ and follows the same format as Figure 8. Given the parameter values, the threshold probability is $\bar{\alpha} = \frac{19}{21}$, which is close to one. A type-2 equilibrium $(\bar{\alpha}, 1)$ arises at some subscriber shares with $a = \frac{4}{5}$. In this case, only a fraction $\frac{1}{5}$ of subscribers learn firm A's signal, while all remaining subscribers receive almost no information due to the high value of $\bar{\alpha}$. A type-5 equilibrium $(0, 0)$ arises at subscriber shares with $a + m = \frac{2}{5}$ along the boundary between C_4 and C_5. In this case, a fraction $\frac{2}{5}$ of subscribers learn firm A's signal, while the remaining B-only subscribers receive almost no information. Therefore, information welfare is maximized at subscriber shares satisfying requirement (iii).

5.3. MILL'S ARGUMENT AND POLICY IMPLICATIONS

Proposition 5 echoes John Stuart Mill's argument in *On liberty*—that truth becomes clearer through its collision with falsehood. In our setting, a change in subscriber shares—either $\Delta a > 0$ with $\Delta b = 0$ or $\Delta b < 0$ with $\Delta a = 0$—increases firm A's effective share given the fact-checking probability, thereby strengthening its incentive to misreport. The resulting welfare implications, however, differ depending critically on the subscriber shares, and thus on the equilibrium types.

Under C_3, such changes prompt firm B to raise its fact-checking probability. The more intense collision between falsehood and verification enhances subscribers' information welfare, directly reflecting Mill's argument. By contrast, under conditions C_2 and C_4, the same changes instead increase firm A's misreporting probability without inducing a stronger fact-checking response. In these cases, the collision between falsehood and verification is weaker, and subscribers are worse informed. This contrast highlights Mill's insight: the effectiveness of truth-revealing mechanisms depends on the intensity of the interplay between accurate reporting and misreporting.

From this perspective, one may ask whether information welfare necessarily improves when a society is highly attentive to potential misinformation. To address this question, we impose $\beta = 1$, so that fact-checking occurs across all reports. At first glance, such comprehensive fact-checking is expected to reduce misreporting and thus enhance information welfare.[12] However, as illustrated in Example 2, this intuition does not necessarily hold.

Example 2. We consider the parameter values: $\pi = \frac{1}{2}$, $p = \frac{3}{4}$, $c_R = \frac{1}{5}$, and $c_B = \frac{9}{20}$, with subscriber share $(a,b,m) = (\frac{4}{5}, \frac{4}{25}, \frac{1}{25})$. Without imposing $\beta = 1$, the equilibrium is characterized by $(\alpha^*, \beta^*) = (\frac{9}{11}, 0.7083)$, yielding information welfare of 0.5268. By contrast, under full fact-checking, the equilibrium is characterized by $(\alpha^*, \beta^*) = (1,1)$, and information welfare falls by 0.0018. Although the decrease in information welfare is small, it underscores an important policy lesson: promoting fact-checking can sometimes backfire, as additional verification efforts may unintentionally reduce overall information welfare. □

In Example 2, imposing full fact-checking induces a new equilibrium in which firm A raises its misreporting frequency. The misreporting probability increases by $\frac{2}{11}$, and the fact checking probability increases by 0.2917. The adverse effect of the increased misreporting is further reinforced by the relatively large share of A-only subscribers, so that the negative impact of misreporting outweighs the benefit of additional verification.

6. CONCLUSION

This paper develops a game-theoretic model of the strategic interaction between a biased media outlet and a fact-checker, taking the composition of the audience as given. By focusing on conduct rather than market structure, the model captures the short-run contest over truth that structural models of media competition abstract away from—a dimension that becomes especially important once we recognize the credence good nature of news, which prevents audience shares from adjusting fluidly in response to reporting quality.

[12]Extensive fact-checking may entail social costs, as resources are expanded to verifying even truthful reports. We abstract from these verification costs and focus solely on information welfare.

Our analysis yields two central findings. First, fact-checking deters misreporting, but only when the two firms' market influences are more or less balanced. When most subscribers follow a biased outlet exclusively, the fact-checker's reach is too limited to impose meaningful discipline: a biased outlet facing a captive audience has little to fear from corrections that most of its readers will never encounter. Effective fact-checking therefore requires not only a credible verification mechanism, but an audience structure in which that mechanism can actually reach the outlet's subscribers. This result formalizes a version of Mill's argument that competition between opposing viewpoints promotes truth, while also identifying its limits: when the audience is sufficiently polarized, or the fact-checker's influence too narrow, the Millian mechanism breaks down and stronger misinformation incentives go unchecked.

Second, and more surprisingly, fact-checking can benefit the biased outlet itself. This counterintuitive result reflects a dynamic inconsistency problem. Without external discipline, the biased outlet cannot credibly commit to honest reporting—it always has a short-run incentive to misreport unfavorable signals—even though such commitment would raise its own payoffs. The fact-checker acts as an external commitment device, restraining the outlet's opportunistic behavior and steering it toward a more profitable reporting strategy that it could not sustain on its own. This finding implies that opposition to fact-checking from biased media outlets may be strategically shortsighted: the disciplining effect of an active fact-checker can, under the right conditions, raise the outlet's own payoffs rather than reduce them.

Together, these findings point to a set of policy implications that are more nuanced than the conventional case for fact-checking suggests. Strengthening fact-checking institutions is unlikely to improve information welfare if the audience is highly polarized, since the deterrent effect is limited in this case. Policies that reduce audience polarization—for instance, by promoting cross-outlet exposure or reducing algorithmic filter bubbles—may therefore be more effective complements to fact-checking than increases in verification capacity alone. At the same time, our analysis cautions against mandating comprehensive fact-checking across all reports. As Example 2 illustrates, imposing full verification can induce a strategic response from the biased outlet—raising its misreporting frequency—that more than offsets the informational benefits of additional scrutiny.

Several limitations of the current model point toward directions for future research. First, we treat the audience composition as fixed, abstracting from the long-run dynamics through which subscriber shares evolve in response to reporting quality and fact-checking activity. Integrating the short-run conduct analysis developed here with a model of endogenous audience formation would allow a richer examination of how fact-checking shapes the media market over time. Second, our model features a single biased outlet and a single fact-checker, whereas real media environments involve multiple outlets with varying degrees of bias and multiple competing fact-checking organizations. Extending the model to allow for richer competitive structures on both sides would likely generate additional insights into the conditions under which fact-checking ecosystems are self-sustaining. Third, we assume that subscribers update their beliefs rationally. Incorporating behavioral features such as confirmation bias or motivated reasoning could generate departures from the rational benchmark and enrich the welfare analysis. We leave these extensions for future work.

REFERENCES

Allcott, H. and Gentzkow, M. (2017). “Social media and fake news in the 2016 election,” *Journal of Economic Perspectives* 31, 211–236.

Allcott, H., Braghieri, L., Eichmeyer, S. and Gentzkow, M. (2020). “The welfare effects of social media,” *American Economic Review* 110, 629–676.

Anderson, S. P. and McLaren, J. (2012). “Media mergers and media bias with rational consumers,” *Economic Journal* 122, 271–296.

Besley, T. and Prat, A. (2006). “Handcuffs for the grabbing hand? Media capture and government accountability,” *American Economic Review* 96, 720–736.

Chan, M. S., Jones, C. R., Jamieson, K. H. and Albarrac'in, D. (2017). “Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation,” *Psychological Science* 28, 1531–1546.

Crawford, V. P. and Sobel, J. (1982). “Strategic information transmission,” *Econometrica* 50, 1431–1451.

Eliaz, K. and Spiegler, R. (2020). “A model of competing narratives,” *American Economic Review* 110, 3786–3816.

Gentzkow, M. and Shapiro, J. M. (2006). “Media bias and reputation,” *Journal of Political Economy* 114, 280–316.

Gentzkow, M., Shapiro, J. M. and Stone, D. F. (2015). "Media bias in the news industry," in *Handbook of Media Economics*, Vol. 1, eds., S. P. Anderson, J. Waldfogel, and D. Strömberg, North-Holland, 623–738.

Graetz, M. J., Reinganum, J. F. and Wilde, L. L. (1986). "The tax compliance game: Toward an interactive theory of law enforcement," *Journal of Law, Economics, and Organization* 2, 1–32.

Guess, A. M., Lyons, B. A., Montgomery, J. M., Nyhan, B. and Reifler, J. (2020). "Fake news, Facebook ads, and misperceptions: Assessing information quality in the age of social media," *American Journal of Political Science* 64, 880–895.

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

Kartik, N. (2009). "Strategic communication with lying costs," *Review of Economic Studies* 76, 1359–1395.

Kartal, M. and Tyran, J.-R. (2022). "Deception and democracy," Working Paper.

Levy, G. and Razin, R. (2015). "Correlation neglect, voting behavior, and information aggregation," *American Economic Review* 105, 1634–1645.

Milton, J. (1644). *Areopagitica*, Yale University Press. Reprinted in *Complete Prose Works of John Milton*, Yale University Press, 1959.

Mill, J. S. (1859). *On Liberty*, John W. Parker and Son, London.

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

Nyhan, B. and Reifler, J. (2010). "When corrections fail: The persistence of political misperceptions," *Political Behavior* 32, 303–330.

Ottaviani, M. and Sørensen, P. N. (2006). "Reputational cheap talk," *RAND Journal of Economics* 37, 155–175.

Pennycook, G. and Rand, D. G. (2020). "Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy nudge intervention," *Psychological Science* 31, 770–780.

Prat, A. and Strömberg, D. (2013). “The political economy of mass media,” in *Advances in Economics and Econometrics: Theory and Applications, Tenth World Congress*, Vol. 2, eds., D. Acemoglu, M. Arellano, and E. Dekel, Cambridge University Press, Cambridge, 135–187.

Strömberg, D. (2004). “Mass media competition, political competition, and public policy,” *Review of Economic Studies* 71, 265–284.

Tsebelis, G. (1990). “Are sanctions effective? A game-theoretic analysis,” *Journal of Conflict Resolution* 34, 3–28.

Walter, N., Cohen, J., Holbert, R. L. and Morag, Y. (2020). “Fact-checking: A meta-analysis of what works and for whom,” *Political Communication* 37, 350–375.

A. BELIEF UPDATES

Suppose that a strategy profile (α^*, β^*) is played in an equilibrium. We first characterize firm B's belief about whether firm A's news is false. We then describe how subscribers update their beliefs about the true state according to Bayes rule.

Firm B's belief. In any equilibrium with $\alpha^* \in [0,1)$, firm A reports $n_A = s_2$ only after receiving signal s_2. Therefore, firm B assigns probability one to the event that the news is truthful when $n_A = s_2$, implying $Q(s_2) = 0$. In an equilibrium with $\alpha^* = 1$, we assume that firm B maintains the belief $Q(s_2) = 0$ off the equilibrium path. When $n_A = s_1$, firm B's updated belief is given by

$$Q(s_1) = \frac{\alpha^* [(1-\pi)(1-p) + \pi p]}{[(1-\pi)p + \pi(1-p)] + \alpha^* [(1-\pi)(1-p) + \pi p]},$$

for any $\alpha^* \in [0,1]$. This belief is strictly increasing in α^*, implying that a higher misreporting probability makes firm B more suspicious of the report.

Subscribers' beliefs. After observing news n_A, A-only subscribers form the belief $\rho_A(n_A)$ that the true state is ω_1. If they observe s_2 (including in an equilibrium with $\alpha^* = 1$), they assign probability one to the event that firm A received signal s_2. Accordingly, their updated belief in ω_1 is given by $\frac{(1-\pi)(1-p)}{(1-\pi)(1-p)+\pi p}$. For any $\alpha^* \in [0,1]$, A-only subscribers ' belief is given by

$$\rho_A(n_A) = \begin{cases} \frac{(1-\pi)[p+(1-p)\alpha^*]}{(1-\pi)[p+(1-p)\alpha^*]+\pi[1-p+p\alpha^*]} & \text{if } n_A = s_1, \\ \frac{(1-\pi)(1-p)}{(1-\pi)(1-p)+\pi p} & \text{if } n_A = s_2. \end{cases}$$

When firm B provides fact-checked news $n_B \in \{s_1, s_2\}$, B-only and multihoming subscribers update their beliefs based solely on the verified information. This updating rule applies even when fact-checking occurs off the equilibrium path ($\beta^* = 0$) since subscribers regard fact-checked news as fully credible. Accordingly, their posterior beliefs are given by

$$\rho_B(n_B) = \rho_M(n_A, n_B) = \begin{cases} \frac{(1-\pi)p}{(1-\pi)p+\pi(1-p)} & \text{if } n_B = s_1, \\ \frac{(1-\pi)(1-p)}{(1-\pi)(1-p)+\pi p} & \text{if } n_B = s_2. \end{cases}$$

After observing firm B's news s_0, B-only subscribers form the belief $\rho_B(s_0)$ that the true state is ω_1. If they unexpectedly observe news s_0 in an equilibrium with $(\alpha^*, \beta^*) = (1,1)$, they assign probability one to the joint event that firm A received signal s_2, reported truthfully, and firm B subsequently chose not to conduct a fact-check. Accordingly, their posterior belief is given by

$$\rho_B(s_0) = \begin{cases} \frac{(1-\pi)[p(1-\beta^*)+(1-p)(1-\alpha^*\beta^*)]}{(1-\pi)[p(1-\beta^*)+(1-p)(1-\alpha^*\beta^*)]+\pi[(1-p)(1-\beta^*)+p(1-\alpha^*\beta^*)]} & \text{if } \alpha^* \in [0,1) \text{ or } \beta^* \in [0,1), \\ \frac{(1-\pi)(1-p)}{(1-\pi)(1-p)+\pi p} & \text{if } \alpha^* = 1 \text{ and } \beta^* = 1. \end{cases}$$

When firm B does not fact-check, multihoming subscribers align their beliefs with those of A-only subscribers: $\rho_M(n_A, s_0) = \rho_A(n_A)$ for any n_A. If they unexpectedly observe $(n_A, n_B) = (s_1, s_0)$ in an equilibrium with $\beta^* = 1$, they update their beliefs according to Bayes rule based solely on firm A's report s_1. If they unexpectedly observe $(n_A, n_B) = (s_2, s_0)$ in an equilibrium $(\alpha^*, \beta^*) = (1,1)$, they assign probability one to the event that firm A received signal s_2.

Lemma A.1 shows that subscribers' updated beliefs are monotone in the misreporting probability. In particular, $\rho_A(s_1)$ decreases in α, because as firm A misreports more frequently, A-only subscribers increasingly believe that news s_1 was generated after signal s_2. However, $\rho_B(s_0)$ increases in α. Conditional on signal s_2 being realized, more frequent misreporting induces firm B to fact-check more often. As a result, after observing unverified news s_0, B-only subscribers place less weight on the event that firm A received signal s_2.

Lemma A.1. *Subscribers' updated beliefs have the following properties.*

(i) *$\rho_A(s_1)$ and $\rho_M(s_1,s_0)$ are strictly decreasing in $\alpha \in [0,1]$. They vary from $1-\pi$ to $\frac{(1-\pi)p}{(1-\pi)p+\pi(1-p)}$.*

(ii) *For $\beta \in (0,1)$, $\rho_B(s_0)$ is strictly increasing in $\alpha \in [0,1]$. Regardless of $\alpha \in [0,1]$, $\rho_B(s_0) = 1-\pi$ when $\beta = 0$, and $\rho_B(s_0) = \frac{(1-\pi)(1-p)}{(1-\pi)(1-p)+\pi p}$ when $\beta = 1$.*

(iii) *When $\alpha \neq 1$, $\rho_B(s_0)$ is strictly decreasing and strictly concave in β. When $\alpha = 1$, $\rho_B(s_0)$ is equal to $1-\pi$ if $\beta \neq 1$ and to $\frac{(1-\pi)(1-p)}{(1-\pi)(1-p)+\pi p}$ if $\beta = 1$. $\rho_B(s_0)$ varies from $\frac{(1-\pi)(1-p)}{(1-\pi)(1-p)+\pi p}$ to $1-\pi$.*

Proof. Let $\phi_3 = (1-\pi)\{p+(1-p)\alpha\} + \pi\{1-p+p\alpha\}$ and $\phi_4 = (1-\pi)\{p(1-\beta)+(1-p)(1-\alpha\beta)\} + \pi\{(1-p)(1-\beta)+p(1-\alpha\beta)\}$. The former is the probability that A-only subscribers receives news s_1. The latter is the probability B-only subscribers receives news s_0, which is positive when $(\alpha,\beta) \neq (1,1)$. (i): Recall that $\rho_A(s_1) = \rho_M(s_1,s_0)$. By taking derivative with respect to α, we obtain $\frac{\partial \rho_A(s_1)}{\partial \alpha} = -\left(\frac{1}{\phi_3}\right)^2 \{\pi(1-\pi)(2p-1)\}$. This expression is negative for any $\alpha \in [0,1]$, since $p \in (\frac{1}{2},1)$.
(ii): If $\beta \in (0,1)$, then $\phi_4 > 0$. By taking derivative with respect to α, we obtain $\frac{\partial \rho_B(s_0)}{\partial \alpha} = \left(\frac{1}{\phi_4}\right)^2 \{\pi(1-\pi)\beta(1-\beta)(2p-1)\}$. This expression is positive for any $\beta \in (0,1)$.
(iii): Suppose that $\alpha \in [0,1)$. Then, $\phi_4 > 0$. By taking derivative with respect to β, we obtain $\frac{\partial \rho_B(s_0)}{\partial \beta} = -\left(\frac{1}{\phi_4}\right)^2 \{\pi(1-\pi)(2p-1)(1-\alpha)\}$. This expression is negative. Since ϕ_4 is strictly decreasing in β, $\rho_B(s_0)$ is strictly concave in β. □

B. MULTIPLE EQUILIBRIA

First, we claim that when multiple equilibria arise, some equilibria are of type 3, while the remaining equilibria are of type 1 or 2. Since all conditions other than C_3 are mutually exclusive by Table 2, it is sufficient to show that conditions C_4 and C_5 cannot hold simultaneously with C_3. If either C_4 or C_5 holds, then $(a+m)\gamma(\bar{\alpha}) \leq c_L$, which is equivalent to $G(\bar{\alpha},0) \leq c_L$. By Lemma 1, $G(\bar{\alpha},\beta)$ strictly decreases in β, implying that $G(\bar{\alpha},\beta) < c_L < L(\bar{\alpha},\beta)$ for all $\beta \in (0,1)$, so C_3 cannot be satisfied.

Suppose now that condition C_3 holds. Recall from Lemma 1 that $G(\bar{\alpha},\beta)$ is linear in β and $L(\bar{\alpha},\beta)$ is strictly concave in β . Then, one of the following must hold: (1) $G(\bar{\alpha},\beta)$ intersects $L(\bar{\alpha},\beta)$ at two distinct points, (2) $G(\bar{\alpha},\beta)$ is tangent to $L(\bar{\alpha},\beta)$ at some $\hat{\beta}\in(0,1)$, or (3) $G(\bar{\alpha},\beta)$ intersects $L(\bar{\alpha},\beta)$ once from above at some $\hat{\beta}\in(0,1)$.[13] If $G(\bar{\alpha},\beta)$ intersects $L(\bar{\alpha},\beta)$ once from above at some $\hat{\beta}$, we must have $G(\bar{\alpha},1)<L(\bar{\alpha},1)$. In this case, neither C_1 nor C_2 can be satisfied because for all $\alpha\in[\bar{\alpha},1]$, we have $G(\alpha,1)\leq G(\bar{\alpha},1)<L(\bar{\alpha},1)=L(\alpha,1)$ from Lemma 1. Therefore, when multiple equilibria arise, $G(\bar{\alpha},\beta)$ must intersect $L(\bar{\alpha},\beta)$ at two distinct points or be tangent to $L(\bar{\alpha},\beta)$.

Case (1): If $G(\bar{\alpha},\beta)$ intersects $L(\bar{\alpha},\beta)$ at two distinct points, three equilibria exist: $(\bar{\alpha},\beta_1)$, $(\bar{\alpha},\beta_2)$, $(\alpha^*,1)$, where $\beta_1<\beta_2$ and $\alpha^*\in[\bar{\alpha},1]$. Since $\beta_1<\beta_2$, $G(\bar{\alpha},\beta)$ crosses $L(\bar{\alpha},\beta)$ from above at β_1 and from below at β_2. Among these equilibria, we exclude the unstable equilibrium $(\bar{\alpha},\beta_2)$. If firm B slightly increases its fact-checking probability from β_2, the gain from misreporting exceeds the corresponding loss, inducing firm A to raise its misreporting probability. This, in turn, leads firm B to further increase its fact-checking probability, thereby reinforcing the deviation from equilibrium.

Between the two remaining equilibria $(\bar{\alpha},\beta_1)$ and $(\alpha^*,1)$, we argue that firm A obtains a higher interim payoff in the former, conditional on receiving signal s_2. We first consider the subcase in which conditions C_3 and C_2 hold simultaneously, so that $\alpha^*\in[\bar{\alpha},1)$. We then turn to the remaining subcase in which conditions C_3 and C_1 hold simultaneously, so that $\alpha^*=1$.

Suppose that conditions C_3 and C_2 hold. In equilibrium $(\alpha^*,1)$ with $\alpha^*\in(\bar{\alpha},1)$, firm A is indifferent between its two reporting options. When it provides truthful news s_2 upon receiving signal s_2, all subscribers infer that firm A received the signal s_2, and thus assign probability $\frac{(1-\pi)(1-p_A)}{(1-\pi)(1-p_A)+\pi p_A}$ to state ω_1. In the alternative equilibrium $(\bar{\alpha},\beta_1)$, when firm A likewise reports truthful news s_2, subscribers consuming firm A's news again infer that firm A received the signal s_2 and form the same posterior belief as in the first equilibrium. However, because firm B fact-checks news s_1 with a probability less than one, B-only subscribers may fail to learn that firm A received signal s_2. As a result, they assign a higher probability to state ω_1, that is, $\rho_B(s_0)>\rho_B(s_2)$.

Suppose now that conditions C_3 and C_1 hold. In the equilibrium $(1,1)$, when firm A misreports, its false news s_1 conveys no information to A-only subscribers, in the sense that $\rho_A(s_1)=1-\pi$. However, subscribers consuming firm B's news learn that firm A received signal s_2. In the alternative equilibrium $(\bar{\alpha},\beta_1)$, when firm A misreports, A-only subscribers assign a strictly higher probability to state ω_1, since $\rho_A(s_1)$ strictly decreases in α. Moreover, if firm B does not fact-check

[13] We say that given α, $G(\alpha,\beta)$ intersects $L(\alpha,\beta)$ from above at $\hat{\beta}$ if $G(\alpha,\hat{\beta}+\varepsilon)-L(\alpha,\hat{\beta}+\varepsilon)<0<G(\alpha,\hat{\beta}-\varepsilon)-L(\alpha,\hat{\beta}-\varepsilon)$ for sufficiently small $\varepsilon>0$. Moreover, we say that $G(\alpha,\beta)$ intersects $L(\alpha,\beta)$ from below at $\hat{\beta}$ if $G(\alpha,\hat{\beta}-\varepsilon)-L(\alpha,\hat{\beta}-\varepsilon)<0<G(\alpha,\hat{\beta}+\varepsilon)-L(\alpha,\hat{\beta}+\varepsilon)$.

news s_1, its subscribers assign a higher probability to state ω_1 than in the first equilibrium; that is $\rho_B(s_0) > \rho_B(s_2)$ and $\rho_M(s_1, s_0) > \rho_M(s_1, s_2)$. Finally, firm A's expected misreporting cost in the type-1 equilibrium is higher than in a type-3 equilibrium, regardless of the cost specifications.

Case (2): If $G(\bar{\alpha}, \beta)$ is tangent to $L(\bar{\alpha}, \beta)$ at some $\hat{\beta} \in (0,1)$, two equilibria exist: $(\bar{\alpha}, \hat{\beta})$ and $(\alpha^*, 1)$, where $\alpha^* \in (\bar{\alpha}, 1]$. Among the two equilibrium, $(\bar{\alpha}, \hat{\beta}_1)$ is unstable, as in case (1). We claim that the remaining equilibrium is stable. If firm A slightly deviates to a misreporting probability α sufficiently close to α^*, firm B continues to fact-check with certainty. Consequently, firm A's best response is to return to α^*, reinforcing the stability of the equilibrium.

Robustness of the selection. All qualitative results in the paper hold under both stable equilibria in the overlap region. To see this, consider an alternative selection rule that picks type-1 or type-2 equilibrium in place of the type-3 euqilibrium, and let $\{X_i^z\}$ denote the resulting partition of the simplex, in analogy to $\{C_i^z\}$. The two partitions differ only in how the overlap region is labeled: under our rule it belongs to C_3^z, under the alternative to X_1^z or X_2^z. Since every result is stated region by region, each statement carries over without modification with $\{X_i^z\}$ in place of $\{C_i^z\}$. The selection rule is therefore a matter of labeling, not a choice among qualitatively distinct predictions.

C. THE REPUTATION-COST SCENARIO

This appendix analyzes the reputation-cost scenario, in which firm A's misreporting cost is incurred only upon detection. We present the model difference, characterize the equilibria, and restate the main results, highlighting where they depart from the baseline model. The propositions and corollary in this appendix are numbered to match their counterparts in the main text.

Setup. The only modification to the model is the misreporting cost. When firm A misreport given the received signal s, the cost of misreporting is given by $\sigma_B(s_i)(b+m)c_R$: a reputation cost c_R is incurred only if the misreport is detected, and it is relevant only for firm B's subscribers, since A-only subscribers remain unaware of firm A's misreporting. All other elements of the model are unchanged, and firm A's best responses depend on subscribers' belief $\hat{\alpha}$ in the same manner as in the main text, although the cost—and hence the associated loss function—differs.

Equilibrium characterization. A key difference is immediate: equilibrium types 4 and 5 never arise. If firm B never fact-checks, the expected loss from misreporting is zero, i.e., $L(\hat{\alpha},0)=0$ for all $\hat{\alpha}$, while the expected gain $G(\hat{\alpha},0)=(a+m)\gamma(\hat{\alpha})$ is strictly positive. Firm A therefore optimally misreports, and in any equilibrium, firm B must fact-check with positive probability. Table 3 provides the conditions on subscriber shares under which each of the remaining equilibrium types arises.

C_1^R	C_2^R	C_3^R
$a \geq \frac{c_R}{c_R+\gamma(1)}$	$\frac{c_R}{c_R+\gamma(\bar{\alpha})} \leq a < \frac{c_R}{c_R+\gamma(1)}$	$\exists \beta \in (0,1); G(\bar{\alpha},\beta)=L(\bar{\alpha},\beta)$

Table 3: CONDITIONS ON SUBSCRIBER SHARES (a,b,m) UNDER A REPUTATION COST. The table reports the conditions under which the remaining equilibrium types arise when the misreporting cost is incurred only after detection.

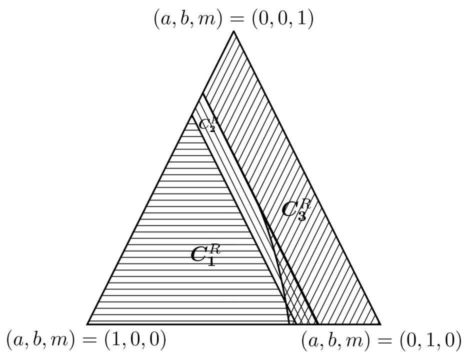


Figure 10: CONDITIONS $C_1^R - C_3^R$ UNDER THE REPUTATION COST: $\pi=\frac{1}{2}$, $p=\frac{3}{4}$, $c_R=\frac{1}{10}$, $c_B=\frac{1}{4}$. The figure shows the subscriber-share regions associated with the three equilibrium conditions in the reputation-cost scenario.

Proposition C.1. *An equilibrium of type i exists if and only if condition C_i^R holds, for $i \in \{1,2,3\}$.*

The conditions parallel those in the main text. Conditions C_1^R and C_2^R carry the same interpretation as C_1 and C_2: when firm B fact-checks with certainty, firm A misreports with probability one if its effective share a is sufficiently large, and randomizes if a is moderately large. Condition C_3^R likewise requires a fact-checking probability $\beta^* \in (0,1)$ equating the expected gain and loss. The contrast with the main text appears when b is large: whereas a type-5 equilibrium with no misreporting arises under the baseline model, here fact-checking occurs with low probability and—since the expected loss from detection is small—firm A misreports with the relatively high probability $\bar{\alpha}$.

The deterrent effect and firm A's payoff. Consider the benchmark in which firm B is inactive. In contrast to the baseline model, only a babbling equilibrium arises: since fact-checking never occurs and misreporting entails no cost, firm A reports only news s_1, and its news is uninformative.

Proposition C.3. *Suppose that firm B is inactive. Firm A's misreporting probability is weakly higher than when firm B is active, and strictly higher under conditions C_2^R and C_3^R.*

The deterrent effect therefore operates as in the main text. However, the effect on firm A's payoff differs sharply: firm A's payoff does *not* improve when firm B is active. Although firm A misreports less frequently, this does not translate into a lower expected misreporting cost, because when B is inactive, no fact-checking occurs and thus no misreporting cost is incurred. The adversarial symbiosis identified in Corollary 1 is thus specific to the baseline model.

Equilibrium responses and information welfare. The comparative statics of the main text carry over with the obvious modifications.

Proposition C.4. *(i) Firm A misreports with higher probability when a increases under condition C_2^R. (ii) Let $(\bar{\alpha}, \beta^*)$ be an equilibrium under condition C_3^R. A change $(\Delta a, \Delta b, \Delta m)$ in subscriber shares induces firm B to increase its fact-checking probability if*

$$\Delta a \beta^* [\gamma(\bar{\alpha}) + c_R] > \Delta b [\beta^* \lambda(\bar{\alpha}, \beta^*) + (1 - \beta^*)\gamma(\bar{\alpha})].$$

Corollary C.2. *When subscriber shares become more fragmented under condition C_3^R, firm B raises its fact-checking probability if and only if*

$$\frac{a}{b} > \frac{\beta^* \lambda(\bar{\alpha}, \beta^*) + (1 - \beta^*)\gamma(\bar{\alpha})}{\beta^* [\gamma(\bar{\alpha}) + c_R]}.$$

If subscriber shares become less fragmented, firm B raises its fact-checking probability if and only if the inequality is reversed.

Proposition C.5. *Information welfare is strictly decreasing in a under C_1^R and C_2^R. Under C_3^R, information welfare increases in a, with b held fixed, strictly if $b \neq 0$ and weakly if $b = 0$, and strictly decreases in b, with a held fixed.*

Since types 4 and 5 do not arise, information welfare is maximized at subscriber shares satisfying (i) $b = 0$ under C_3^R or (ii) the smallest a under C_2^R; the third possibility in the main text—the boundary of C_4 and C_5—has no counterpart here.

D. PROOFS

The proofs below cover both cost specifications. We refer to the baseline model as the lying-cost scenario and to the alternative introduced in Appendix C as the reputation-cost scenario. Table 4 reproduce conditions on subscriber shares for each cost specification: Table 4a for the lying-cost scenario and Table 4b for the reputation-cost scenario.

C_1^L	$a \geq \frac{c_L}{\gamma(1)}$
C_2^L	$\frac{c_L}{\gamma(\bar{\alpha})} \leq a < \frac{c_L}{\gamma(1)}$
C_3^L	$\exists \beta \in (0,1);\, G(\bar{\alpha},\beta) = L(\bar{\alpha},\beta)$
C_4^L	$\frac{c_L}{\gamma(0)} < a+m \leq \frac{c_L}{\gamma(\bar{\alpha})}$
C_5^L	$a+m \leq \frac{c_L}{\gamma(0)}$

(a) UNDER A LYING COST. The subtable lists the equilibrium conditions for the baseline specification in which firm A incurs a direct lying cost.

C_1^R	$a \geq \frac{c_R}{c_R+\gamma(1)}$
C_2^R	$\frac{c_R}{c_R+\gamma(\bar{\alpha})} \leq a < \frac{c_R}{c_R+\gamma(1)}$
C_3^R	$\exists \beta \in (0,1);\, G(\bar{\alpha},\beta) = L(\bar{\alpha},\beta)$

(b) UNDER A REPUTATION COST. The subtable lists the equilibrium conditions for the alternative specification in which firm A incurs a cost only when misreporting is detected.

Table 4: CONDITIONS ON SUBSCRIBER SHARES (a,b,m) FOR THE EXISTENCE OF EACH EQUILIBRIUM TYPE. The table reproduces the equilibrium condition of subscriber-shares for each cost specification.

D.1. LEMMA 1

We omit the proof of Lemma 1, as it follows directly from Lemma A.1.

D.2. PROPOSITION 1

For each cost specification $z \in \{L, R\}$, we claim that an equilibrium of type $i \in \{1,2,3\}$ exists if and only if condition C_i^z holds. We omit the proofs for conditions C_4^L and C_5^L in the lying-cost scenario, as they are analogous to those of conditions C_2^z and C_1^z, respectively.

Condition C_1^z: Condition C_1^z is equivalent to $G(1,1) \geq L(1,1)$ since this inequality reduces to $a \geq \frac{c_L}{\gamma(1)}$ in the lying-cost scenario and $a \geq \frac{c_R}{c_R+\gamma(1)}$ in the reputation-cost scenario. If a type-1 equilibrium exists, then $G(1,1) \geq L(1,1)$, so condition C_1^z must hold. Conversely, if C_1^z holds, firm A optimally misreports even when firm B fact-checks with certainty. Since this misreporting probability exceeds the threshold $\bar{\alpha}$, firm B optimally fact-checks. Hence $(1,1)$ constitutes a type-1 equilibrium under C_1^z.

Condition C_2^z: Evaluated at $\beta = 1$, condition C_2^z is equivalent to $G(1,1) < c_A(s_1|s_2) \leq G(\bar{\alpha},1)$, regardless of the cost specification. If a type-2 equilibrium exists, $G(\alpha^*,1) = L(\alpha^*,1)$ for some $\alpha^* \in [\bar{\alpha},1)$. This equality reduces to $a = \frac{c_L}{\gamma(\alpha^*)}$ in the lying-cost scenario, and $a = \frac{c_R}{c_R+\gamma(\alpha^*)}$ in the reputation-cost scenario. Since γ is strictly decreasing by Lemma 1, condition C_2^z holds. Conversely, if C_2^z holds, there exists $\alpha^* \in [\bar{\alpha},1)$ such that $G(\alpha^*,1) = L(\alpha^*,1)$. Firm A optimally misreports with probability α^* when firm B fact-checks with certainty. Since this misreporting probability is no smaller than the threshold $\bar{\alpha}$, firm B optimally fact-checks. Hence $(\alpha^*,1)$ constitutes a type-2 equilibrium under C_2^z.

Condition C_3^z: If a type-3 equilibrium $(\bar{\alpha},\beta^*)$ with $\bar{\alpha} \in (0,1)$ and $\beta^* \in (0,1)$ exists, then, $G(\bar{\alpha},\beta^*) = L(\bar{\alpha},\beta^*)$, implying that condition C_3^z holds. Conversely, if $G(\bar{\alpha},\beta^*) = L(\bar{\alpha},\beta^*)$ for some $\beta^* \in (0,1)$, firm A is indifferent and therefore willing to misreport with probability $\bar{\alpha}$ when firm B fact-checks with probability β^*. Given that firm A misreports with probability $\bar{\alpha}$, firm B is indifferent and optimally fact-check with probability β^*. Therefore, $(\bar{\alpha},\beta^*)$ constitutes a type-3 equilibrium under C_3^z.

D.3. PROPOSITION 2

Since $\gamma(\hat{\alpha})$ is strictly decreasing in $\hat{\alpha}$ from Lemma 1, Table 2 implies that conditions from C_1, C_2, C_4, and C_5 are mutually exclusive. Therefore, it is sufficient to show that when none of the four conditions holds, condition C_3 is satisfied.

Suppose that none of C_1, C_2, C_4, and C_5 holds. Then, we have $a + m > \frac{c_A}{\gamma(\bar{\alpha})}$ and $a < \frac{c_A}{\gamma(\bar{\alpha})}$, which are equivalent to $G(\bar{\alpha},0) > L(\bar{\alpha},0) = c_A$ and $G(\bar{\alpha},1) < L(\bar{\alpha},1) = c_A$. Then, the continuity of $G(\bar{\alpha},\beta)$ and $L(\bar{\alpha},\beta)$ in β ensures condition C_3.

D.4. PROPOSITION 3

Suppose that firm B is inactive. Then, equilibrium can be characterized by firm A's misreporting probability only. In an equilibrium with misreporting probability $\tilde{\alpha}$, the gain from misreporting is $G(\tilde{\alpha},0)$, and the loss from misreporting is given by $L(\tilde{\alpha},0) = c_L$ in the lying-cost scenario and $L(\tilde{\alpha},0) = 0$ in the reputation-cost scenario.

Since misreporting entails no loss in the reputation-cost scenario and $G(\tilde{\alpha},0) > 0$ for all $\tilde{\alpha}$, firm A misreports with probability one under all conditions C_1^R, C_2^R, and C_3^R. Therefore, firm A's misreporting probability is strictly higher under conditions C_2^R and C_3^R than in the basic model with active fact-checking, and remains unchanged under C_1^R.

In the lying-cost scenario, the equilibrium misreporting probability remains unchanged under conditions C_4^L and C_5^L, because no fact-checking occurs even in type-4 and type-5 equilibria of the baseline model.

Under condition C_2^L, an equilibrium of the baseline model is characterized by $(\alpha^*,1)$ such that $G(\alpha^*,1) = L(\alpha^*,1) = c_L$ and $\alpha^* \in [\bar{\alpha},1)$. For the same α^*, we have $L(\alpha^*,0) = c_L = G(\alpha^*,1) < G(\alpha^*,0)$. Therefore, when firm B is inactive, the gain from misreporting exceeds the corresponding loss at α^*, implying that firm A misreports with strictly higher probability when B is inactive than in the basic model. An analogous argument applies under C_1^L. However, firm A already misreports with probability one in the basic model, no further increase in the misreporting probability is possible.

Under condition C_3^L, an equilibrium of the basic model is characterized by $(\bar{\alpha},\beta^*)$ such that $G(\bar{\alpha},\beta^*) = L(\bar{\alpha},\beta^*)$ and $\beta^* \in (0,1)$. Then, we have $L(\bar{\alpha},0) = c_L \leq L(\bar{\alpha},\beta^*) = G(\bar{\alpha},\beta^*) < G(\bar{\alpha},0)$, implying that firm A misreports with strictly higher probability when B is inactive than in the basic model.

D.5. PROPOSITION 4

Suppose that an equilibrium (α^*,β^*) arises at subscriber share (a,b,m), and consider a small change $(\Delta a,\Delta b,\Delta m)$. By taking the total derivative of expression (2), we obtain

$$\{\Delta a + \Delta m(1-\beta^*)\}\gamma - \beta^*\lambda\Delta b = \left\{m\gamma + b\lambda + b\beta^*\frac{\partial\lambda}{\partial\beta}\right\}\Delta\beta^* + \left\{b\beta^*\frac{\partial\lambda}{\partial\alpha} - \{a+m(1-\beta^*)\}\frac{\partial\gamma}{\partial\alpha}\right\}\Delta\alpha^* \quad (3)$$

in the lying-cost scenario, and in the reputation-cost scenario, we obtain

$$\{\Delta a+\Delta m(1-\beta^*)\}\gamma-\beta^*\lambda\Delta b+\beta^* c_R\Delta a = \left\{\begin{array}{l} m\gamma+b\lambda+b\beta^*\dfrac{\partial\lambda}{\partial\beta} \\ +(1-a)c_R \end{array}\right\}\Delta\beta^* + \left\{\begin{array}{l} b\beta^*\dfrac{\partial\lambda}{\partial\alpha} \\ -\{a+m(1-\beta^*)\}\dfrac{\partial\gamma}{\partial\alpha} \end{array}\right\}\Delta\alpha^* \tag{4}$$

, where $\Delta\beta^* = \frac{\partial\beta^*}{\partial a}\Delta a+\frac{\partial\beta^*}{\partial b}\Delta b+\frac{\partial\beta^*}{\partial m}\Delta m$ and $\Delta\alpha^* = \frac{\partial\alpha^*}{\partial a}\Delta a+\frac{\partial\alpha^*}{\partial b}\Delta b+\frac{\partial\alpha^*}{\partial m}\Delta m$.

First, we examine the region characterized by condition C_2^z for $z\in\{L,R\}$. In a type-2 equilibrium, $\beta^*=1$, implying that $\Delta\beta^*=0$. Since $\lambda(\alpha,1)=0$ for any α, $\frac{\partial\lambda}{\partial\alpha}=0$. Using these observations, expressions (3) and (4) simplify to $\Delta a\gamma=-a\frac{\partial\gamma}{\partial\alpha}\Delta\alpha^*$ and $\Delta a(\gamma+c_R)=-a\frac{\partial\gamma}{\partial\alpha}\Delta\alpha^*$, respectively. Since $a>0$ under C_2^z, $\gamma>0$ and $\frac{\partial\gamma}{\partial\alpha}<0$ by Lemma 1, $\Delta a>0$ implies that $\Delta\alpha^*>0$.

Now, we examine the region characterized by condition C_4^L. In a type-4 equilibrium, $\beta^*=0$, implying that $\Delta\beta^*=0$. Using these observations, expression (3) simplifies to $(\Delta a+\Delta m)\gamma=-(a+m)\frac{\partial\gamma}{\partial\alpha}\Delta\alpha^*$. Therefore, $\Delta a+\Delta m>0$ (equivalently $\Delta b<0$) implies that $\Delta\alpha^*>0$.

Finally, we examine the region characterized by condition C_3^z for $z\in\{L,R\}$. In a type-3 equilibrium, $\alpha^*=\bar{\alpha}$, implying that $\Delta\alpha^*=0$. According to the equilibrium selection rule, at $(\bar{\alpha},\beta^*)$, $G(\bar{\alpha},\beta)$ crosses $L(\bar{\alpha},\beta)$ from above at β^*. Using these observations, expressions (3) and (4) simplify to

$$\Delta a\beta^*\gamma-\Delta b(\beta^*\lambda+(1-\beta^*)\gamma)=\left\{m\gamma+b\lambda+b\beta^*\frac{\partial\lambda}{\partial\beta}\right\}\Delta\beta^*$$

and

$$\Delta a\beta^*(\gamma+c_R)-\Delta b(\beta^*\lambda+(1-\beta^*)\gamma)=\left\{m\gamma+b\lambda+b\beta^*\frac{\partial\lambda}{\partial\beta}+(1-a)c_R\right\}\Delta\beta^*$$

respectively. In the lying-cost scenario, $\Delta a\beta^*\gamma-\Delta b(\beta^*\lambda+(1-\beta^*)\gamma)>0$ implies that $\Delta\beta^*>0$, because $m\gamma+b\lambda+b\beta^*\frac{\partial\lambda}{\partial\beta}>0$ under the equilibrium selection rule. Analogously, in the reputation-cost scenario, $\Delta a\beta^*(\gamma+c_R)-\Delta b(\beta^*\lambda+(1-\beta^*)\gamma)>0$ implies that $\Delta\beta^*>0$.

D.6. PROPOSITION 5

Lemma D.1. *The following results hold for any* $\alpha \in (0,1)$.

(i) $u_M \geq u_A$ *with equality only when* $\beta = 0$ *and* $u_M \geq u_B$ *with equality only when* $\beta = 1$.

(ii) For any β, $u'_{A,\alpha} < 0$ *and* $0 < u'_{M,\beta} < u'_{B,\beta}$. *Moreover,* $u_M - u_A = \beta u'_{M,\beta}$.

(iii) $u_A - u_B - u'_{A,\alpha}\frac{\gamma}{\gamma'} < 0$.

Proof. All claims are proved through straightforward calculations. Let $\phi_1 = (1-\pi)p + \pi(1-p)$, $\phi_2 = (1-\pi)(1-p) + \pi p$, $\phi_3 = (1-\pi)(p+(1-p)\alpha) + \pi(1-p+p\alpha)$, and $\phi_4 = (1-\pi)(p(1-\beta)+(1-p)(1-\alpha\beta)) + \pi((1-p)(1-\beta)+p(1-\alpha\beta)$. Here, ϕ_1 and ϕ_2 are the probabilities of signals s_1 and s_2, respectively. ϕ_3 is the probability that firm A's subscribers receive news $n_A = s_1$, and ϕ_4 is the probability that firm B's subscribers receive news $n_B = s_0$. When $\alpha \in (0,1)$, ϕ_1, ϕ_2, ϕ_3, and ϕ_4 are positive.
(i): $u_M - u_A = \frac{2\alpha\beta\pi^2(1-\pi)^2(2p-1)^2}{\phi_1\phi_2\phi_3} \geq 0$ and $u_M - u_B = \frac{2(1-\alpha)(1-\beta)\pi^2(1-\pi)^2(2p-1)^2}{\phi_2\phi_3\phi_4} \geq 0$. The first holds with equality only when $\beta = 0$ and the second holds with equality only when $\beta = 1$.

(ii): $u'_{A,\alpha} = -\frac{2\pi^2(1-\pi)^2(2p-1)^2}{\phi_2(\phi_3)^2} < 0$. $u'_{M,\beta} = \frac{2\alpha\pi^2(1-\pi)^2(2p-1)^2}{\phi_1\phi_2\phi_3} > 0$ and therefore $u_M - u_A = \beta u'_{M,\beta}$. $u'_{B,\beta} - u'_{M,\beta} = \frac{2(1-\alpha)^2\pi^2(1-\pi)^2(2p-1)^2}{\phi_3(\phi_4)^2} > 0$.
(iii): $u_A - u_B - u'_{A,\alpha}\frac{\gamma}{\gamma'} = -\frac{2\pi^2(1-\pi)^2(2p-1)^2\{\phi_1(1-\beta)+\phi_2\beta(1-\alpha\beta)\}}{\phi_1(\phi_2)^2\phi_4} < 0$. In particular, this reduces to $-\frac{2\pi^2(1-\pi)^2(2p-1)^2}{(\phi_2)^2}$ if $\beta = 0$ and $-\frac{2\pi^2(1-\pi)^2(2p-1)^2}{\phi_1(\phi_2)^2}$ if $\beta = 1$. □

Suppose that an equilibrium (α^*, β^*) arises at subscriber distribution (a,b,m) and consider a small change $(\Delta a, \Delta b, \Delta m)$. The total derivative of $U(a,b,m)$ yields

$$\Delta U = (u_A\Delta a + u_B\Delta b + u_M\Delta m) + \left(au'_{A,\alpha} + bu'_{B,\alpha} + mu'_{M,\alpha}\right)\Delta\alpha^* + \left(bu'_{B,\beta} + mu'_{M,\beta}\right)\Delta\beta^* \tag{5}$$

where

$$\Delta\alpha^* = \frac{\partial\alpha^*}{\partial a}\Delta a + \frac{\partial\alpha^*}{\partial b}\Delta b + \frac{\partial\alpha^*}{\partial m}\Delta m,$$
$$\Delta\beta^* = \frac{\partial\beta^*}{\partial a}\Delta a + \frac{\partial\beta^*}{\partial b}\Delta b + \frac{\partial\beta^*}{\partial m}\Delta m.$$

First, we examine the region characterized by condition C_2^z for $z \in \{L, R\}$, where an equilibrium $(\alpha^*, 1)$ with $\alpha^* \in [\bar{\alpha}, 1)$ arises. Here, $\Delta\beta^* = 0$, $u_B = u_M$, and $u'_{B,\alpha} = u'_{M,\alpha} = 0$. From the proof of Proposition 4, we have $a\Delta\alpha^* = -\frac{\Delta a\gamma}{\gamma'}$ in the lying-cost scenario and $a\Delta\alpha^* = -\frac{\Delta a(\gamma + c_R)}{\gamma'}$ in the reputation-cost scenario. Substituting into expression (5) yields $\Delta U = \left(u_A - u_B - u'_{A,\alpha}\frac{\gamma}{\gamma'}\right)\Delta a$ in the lying-cost scenario and $\Delta U = \left(u_A - u_B - u'_{A,\alpha}\frac{\gamma + c_R}{\gamma'}\right)\Delta a$ in the reputation-cost scenario. Since $u'_{A,\alpha} < 0$ and $\frac{\partial\gamma}{\partial\alpha} < 0$, we have $-u'_{A,\alpha}\frac{c_R}{\gamma'} < 0$. Since $u_A - u_B - u'_{A,\alpha}\frac{\gamma}{\gamma'} < 0$, $\Delta a > 0$ implies $\Delta U < 0$ in both scenarios.

Now, examine the region characterized by condition C_4^L, where $\beta^* = 0$. Thus, $\Delta\beta^* = 0$, $u_A = u_M$, and $u_B = 1 - 2\pi(1-\pi)$, implying $u'_{A,\alpha} = u'_{M,\alpha}$ and $u'_{B,\alpha} = 0$. From the proof of Proposition 4, we obtain $(a+m)\Delta\alpha^* = \frac{\Delta b\gamma}{\gamma'}$. Substituting into expression (5) yields $\Delta U = \left(u_B - u_A + u'_{A,\alpha}\frac{\gamma}{\gamma'}\right)\Delta b$. Since $u_A - u_B - u'_{A,\alpha}\frac{\gamma}{\gamma'} < 0$, $\Delta b > 0$ implies $\Delta U > 0$.

Lastly, we examine the region characterized by condition C_3^z for $z \in \{L, R\}$, where $\alpha^* = \bar{\alpha}_2$ and $\Delta\alpha^* = 0$. From the proof of Proposition 4, we obtain

$$\Delta a\beta^*\gamma - \Delta b\left(\beta^*\lambda + (1-\beta^*)\gamma\right) = \left\{m\gamma + b\lambda + b\beta^*\frac{\partial\lambda}{\partial\beta}\right\}\Delta\beta^*$$

in the lying-cost scenario and

$$\Delta a\beta^*\left(\gamma + c_R\right) - \Delta b\left(\beta^*\lambda + (1-\beta^*)\gamma\right) = \left\{m\gamma + b\lambda + b\beta^*\frac{\partial\lambda}{\partial\beta} + (1-a)c_R\right\}\Delta\beta^*$$

in the reputation-cost scenario.

If $\Delta a = 0$, expression (5) is simplified into

$$\Delta U = \left\{ u_B - u_M - \left(bu'_{B,\beta} + mu'_{M,\beta}\right) \frac{\beta^*\lambda + (1-\beta^*)\gamma}{m\gamma + b\lambda + b\beta^*\frac{\partial\lambda}{\partial\beta}} \right\}\Delta b$$

in the lying-cost scenario and

$$\Delta U = \left\{ u_B - u_M - \left(bu'_{B,\beta} + mu'_{M,\beta}\right) \frac{\beta^*\lambda + (1-\beta^*)\gamma}{m\gamma + b\lambda + b\beta^*\frac{\partial\lambda}{\partial\beta} + (1-a)c_R} \right\}\Delta b$$

in the reputation-cost scenario. The terms in braces are negative since $u_B < u_M$, $u'_{B,\beta} > 0$, and $u'_{M,\beta} > 0$ from Lemma D.1. Therefore, $\Delta b < 0$ implies $\Delta U > 0$ in both scenarios.

If $\Delta b = 0$, expression (5) is written as

$$\Delta U = \left\{ \begin{matrix} u_A - u_M + \left(bu'_{B,\beta} + mu'_{M,\beta} \right) \\ \dfrac{\beta^* \gamma}{m\gamma + b\lambda + b\beta^* \frac{\partial \lambda}{\partial \beta}} \end{matrix} \right\} \Delta a$$

in the lying-cost scenario, which is simplified into

$$\Delta U = \frac{b\beta^*}{m\gamma + b\lambda + b\beta^* \frac{\partial \lambda}{\partial \beta}} \left\{ \begin{matrix} \gamma u'_{B,\beta} - \lambda u'_{M,\beta} \\ - \beta^* \dfrac{\partial \lambda}{\partial \beta} u'_{M,\beta} \end{matrix} \right\} \Delta a,$$

using the fact that $u_M - u_A = \beta u'_{M,\beta}$. In the reputation-cost scenario, we analogously obtain

$$\Delta U = \frac{b\beta^*}{m\gamma + b\lambda + b\beta^* \frac{\partial \lambda}{\partial \beta}} \left\{ \begin{matrix} (\gamma + c_R) u'_{B,\beta} - (\lambda + c_R) u'_{M,\beta} \\ - \beta^* \dfrac{\partial \lambda}{\partial \beta} u'_{M,\beta} \end{matrix} \right\} \Delta a.$$

Therefore, $b = 0$ implies that $\Delta U = 0$ in both scenarios. Now suppose that $b \neq 0$. It is sufficient to show that the first two terms in braces together are positive (i.e., $\gamma u'_{B,\beta} - \lambda u'_{M,\beta} > 0$ in the lying-cost scenario and $(\gamma + c_R) u'_{B,\beta} - (\lambda + c_R) u'_{M,\beta} > 0$ in the reputation-cost scenario). This is because the last term is positive from Lemma A.1 (i.e., $-\beta^* \frac{\partial \lambda}{\partial \beta} u'_{M,\beta} > 0$). The result follows since $0 < \lambda < \gamma$ from Lemma A.1 and $0 < u'_{M,\beta} < u'_{B,\beta}$ from Lemma D.1.