Communication with Strategic Fact-checking

Aleksandr Levkun

UC, San Diego

This version: December 2021

Click here for the latest version

Abstract

I examine communication between an informed sender and an uninformed receiver with a presence of a strategic fact-checker. The sender makes a claim about a certain issue to persuade the receiver to approve the sender's proposition. The fact-checker has its own interests and chooses a stochastic fact-checking policy that initiates checks of sender's claims. The fact-checking technology is subject to a potential failure to produce a fact-check and the usage of this technology is costly. We show that the optimal fact-checking policy is a threshold policy in terms of the fact-checking cost. Full fact-checking is optimal when the cost is below the threshold. Otherwise, no fact-checking is optimal. We characterize the cost threshold as a function of fact-checker's preferences. Interestingly, the receiver need not prefer a fact-checker with preferences aligned with the receiver to one with opposed preferences. The addition of the fact-checker does not necessarily improve communication even when both fact-checkers are willing to fully check by themselves. When the fact-checking cost is high enough, we find an equilibrium in which there is an underprovision of fact-checking due to free-riding.

JEL Classification Numbers: C72, D82, D83

Keywords: Fact-checking, Information, Lie detection, Mediation

I would like to thank Renee Bowen, Simone Galperti, and Joel Sobel for their support and guidance. I greatle appreciate the comments and suggestions from Florian Ederer, Shachar Kariv, Denis Shishkin, Maria Titova, and participants of the UCSD Workshop in Theory and Behavioral Economics. All errors are mine.

1 Introduction

The systematic fact-checking of prolific public figures has become ubiquitous over the course of last twenty years. At the beginning, the fact-checkers mostly devoted attention to the US elections. Now they constantly check political claims over the variety of challenging topics. Undoubtedly, fact-checking has become an integral and self-sufficient part of political discussion in the US (Graves, 2016).[1] The major social media companies such as Facebook and Twitter now flag suspicious and misleading content on their websites and accompany the conclusions by fact-checkers' reports.[2] The goal of fact-checking is to correct beliefs and hold politicians accountable for deceitful claims. However, the fact-checkers' role of "arbiters of truth" has drawn criticism on the multiple counts including the fact-checkers' bias.[3] Ostermeier (2010) points out the lacking transparency in how the fact-checked claims get selected. The selection effect may create false impressions about the overall honesty of politicians: actors who receive more negative fact-checking ratings deemed less truthful than those who are checked rarely and receive fewer negative ratings (Uscinski and Butler, 2013, Uscinski, 2015). For these reasons, our understanding of the effects of potentially biased fact-checking is important, especially in the age of fake news and alternative facts (Allcott and Gentzkow, 2017, Barrera et al., 2020). This paper takes the possibility of a strategically motivated fact-checker seriously. We ask following questions. Who benefits from fact-checking? How do these benefits depend on the fact-checker's preferences? Is fact-checking effective in preventing the speaker from spreading false claims? What kind of a fact-checker is preferred by a target of speaker's claims and does adding fact-checkers help this target to learn the truth more often?

To answer these questions, we incorporate a strategic fact-checker in a model of cheap-talk communication between a sender and a receiver. The sender observes the realization of a binary issue but his preferences do not depend on it. His objective is to sway the receiver toward accepting his proposition. To achieve this, the sender makes a claim about the issue real-

[1] Graves and Cherubini (2016) document the rise of fact-checking in Europe.

[2] See https://www.facebook.com/journalismproject/programs/third-party-fact-checking/how-it-works and https://www.reuters.com/technology/twitter-partners-with-ap-reuters-battle-misinformation-its-site-2021-08-02/.

[3] Examples of other critiques include an inability of fact-checkers to fight motivated reasoning (Walter et al., 2020) and the choice to examine claims that cannot be checked reliably (Uscinski and Butler, 2013).

ization or stays silent. The fact-checker has an access to a fact-checking technology that verifies the truthfulness of sender's claims. The fact-checking technology is subject to a potential failure to produce a fact-check output and the usage of this technology is costly.[4] The fact-checker chooses a stochastic fact-checking policy that initiates a fact-check of sender's potential claims. We assume that the fact-checker has commitment power and the fact-checking policy is announced at the beginning of the game.[5] The receiver wishes to match her decision with the issue realization. The receiver accumulates the sender's claim and the fact-check output and then makes a decision whether to accept or reject the sender's proposition. The payoffs then realize for all involved parties.

The fact-checker maximizes its expected payoff net of the fact-checking cost. The fact-checker's payoff function is the central object of our analysis. We allow it to depend on the decision and issue pairs. We also consider three natural examples of this payoff function. The pro-receiver fact-checker wishes to maximize the receiver's payoff. The pro-sender fact-checker wishes for the sender's proposition to be accepted, while the anti-sender fact-checker wants it rejected. We take a stance on the capacity of the fact-checker to make strategic decisions. Graves (2017) itemizes typical steps of a process of fact-checking. The first step identifies claims to check. Then the fact-checkers gather the evidence and assess the claim veracity. Finally, they publish the fact-check output in a transparent manner.[6] We allow the fact-checker to be strategic only about the first step of this process, the kind of claims that the fact-checker attempts to check. Once the fact-check is triggered, we assume that the fact-checking technology runs its course towards completion and the receiver observes the fact-check output.

Our first main result characterizes the optimal fact-checking policy for any fact-checker's preferences. We indicate that the characterization depends on the receiver's decision under the

[4]The failure probability and the usage cost of the fact-checking technology is necessarily a function of the issue under consideration. Claims can be hard to verify because of the insufficient or lacking evidence on the issue (Graves, 2016). We abstract away from specifics of the issue and consider an exogenous probability of the technology failure.

[5]Commitment can be made credible if the fact-checker strives for reputation in repeated interactions with speakers and audience.

[6]This systematization is based on the author's field experience with three major fact-checking organizations: PolitiFact, FactCheck.org, and Washington Post's Fact Checker.

prior. We say that the environment is sender-favorable (sender-unfavorable), if the receiver accepts (rejects) the sender's proposition without communication. The starting point of our characterization fixes a fact-checking policy and shows the equilibrium behavior and implied players' payoffs in the resulting subgame. Players' payoffs are unique in the sender-unfavorable environment. The sender who wants the issue to be known simply sends the most checked claim. On the other hand, there is a multiplicity in the equilibrium behavior and payoffs in the sender-favorable environment. The reason is the sender who wants to misrepresent the issue prefers as little fact-checking of his claim as possible but he is still confined to mimicking. Consequently, the equilibrium behavior we characterized tells us the direction of the change in payoffs for a more aggressive fact-checking policy that initiates fact-checking with greater frequency. A more aggressive fact-checking policy improves the ex ante sender's payoff in the sender-unfavorable environment, but can only hurt the sender in the sender-favorable environment. The receiver can only benefit from a more aggressive fact-checking policy, as the information transmission improves.

We show that the optimal fact-checking policy is a threshold policy in terms of the fact-checking cost. When the fact-checking cost is above the threshold, the fact-checker never checks. When the fact-checking cost is below the threshold, the fact-checker initiates fact-checks with probability one. The reason is that we can represent both the expected fact-checker's payoff and the minimal cost of fact-checking as linear functions of the maximal probability of checking across claims. We characterize the cost threshold in terms of fact-checker's preferences. In particular, the anti-sender (pro-sender) fact-checker never checks in the sender-unfavorable (sender-favorable) environment. Indeed, uninformative communication makes the receiver choose the fact-checker's preferred action. The cost threshold depends nontrivially on the prior beliefs. The underlying reason for this is the different structure of the cost-minimizing subgame equilibria depending on the environment. Notably, in the sender-unfavorable environment, the silent message is utilized by the sender wishing to misrepresent the claim realization. Furthermore, only when the fact-checking technology is perfect in a sense that it never fails, the fact-checker is able to make the sender completely refrain from producing false claims. The implication is a higher cost threshold for the perfect fact-checking technology.

One may expect that having an access to the fact-checker that maximizes the receiver's payoff

is always the best option for the receiver. An interesting corollary of our characterization is that this is not always the case. We can always find a fact-checker caring exclusively about the sender's payoff that will fact-check more aggressively than the pro-receiver fact-checker for a high enough fact-checking cost. Whether it is a pro-sender or an anti-sender fact-checker depends on the environment. The sufficient condition for this implication is that the sender gains more by persuading the receiver than the receiver by learning the truth.

Our second main result delivers the conditions under which two fact-checkers check more than one fact-checker. The choice of fact-checking policies is simultaneous and fact-checkers best respond to each other. A fact-checker that does not check at all when it is the only fact-checker available continues to do so in this setting, since a more aggressive fact-checking policy can only decrease its payoff. Interesting equilibrium policies arise then both fact-checkers are willing to fact-check by themselves. We show that when the fact-checking cost is low enough, both fact-checkers check to the full extent in the unique equilibrium. The composite fact-checking policy becomes more aggressive and the failure of the fact-checking technology is mitigated. Alternatively, there are equilibria in which only one fact-checker checks, while another fact-checker enjoys the benefit of more information communication at no cost. The strong free-riding motive imposed by the fact-checking cost then results in no improvements relative to the case of only one fact-checker available. Moreover, when the fact-checking technology is not perfect and the fact-checking cost is high enough, there is an equilibrium in which fact-checking is underprovided and the receiver's payoff decreases compared to the case of one fact-checker. In this equilibrium, each fact-checker initiates a fact-check with a nontrivial probability which depends on the other fact-checker's cost threshold.

We hope that our analysis can speak to the discussion of the effects of fact-checking in the real world. Several authors suggested that "partisan" fact-checkers can be harmful for more informative political discourse (Ostermeier, 2010, Graves, 2016).[7] We show that this is not necessarily the case.[8] The partisan fact-checker may be willing to fact-check the claim to help or hurt the sender, while the fact-checking cost may prevent the non-partisan fact-checker

[7] See also https://www.scientificamerican.com/article/the-psychology-of-fact-checking1/.

[8] One clear example of a partisan fact-checker is StopFake, Ukrainian fact-checking organization devoted to refutation of Russian propaganda.

from selecting this claim. As for the fact-checking cost, the automated fact-checking will necessarily drive down the fact-checking cost. While most fact-checking efforts are currently made by journalists and experts, there is hope for systematic computer-assisted fact-checking (Hassan et al., 2017, Graves, 2018). Our results suggest that the decrease in the fact-checking cost can only sustain more informative communication. However, currently researchers and practitioners agree that the real promise of the automated fact-checking lies in methods to assist human fact-checkers in selecting the claims for verification (Graves, 2018). This may "debias" the fact-checker, which in our setting can have adverse effects for information transmission.

Related Literature. Our paper contributes to the growing literature on communication with detectable deception. Three recent papers explore the implications of lie detection in a cheap-talk setting. Balbuzanov (2019) studies the parametrized version of Crawford and Sobel (1982) model. The message space is equal to the state space, a unit interval. If the sender's message does not correspond to the true state, the receiver observes a private signal pointing out a sender's lie with an exogenous probability. Due to sender's type-dependent preferences, fully revealing equilibria exist, even for small probabilities of lie detection. The main driver of this result is that the receiver is able to condition punishing actions based on the message. Dzuida and Salas (2018) analyze the implication of having the same lie detection technology in a communication game with no common interests between the sender and the receiver, as in our setup. In informative equilibria, low sender's types lie and a positive measure of high types reveal the truth. An increased probability of lie detection necessarily increases information transmission. Holm (2010) investigates the role of the truth and lie detection in binary bluffing games, where the sender's goal is to deceive the receiver. Truth (lie) detection corresponds to the receiver observing a perfect signal with a fixed probability if the sender's statement is true (false). In the considered bluffing game, truth or lie detection shrinks the set of equilibria. The equilibrium is unique if the probability of detection is sufficiently high. These papers differ from ours in two ways. First, our fact-checking technology allows for catching lies and pointing out truths simultaneously. Second, these papers study communication with exogenously provided lie detection. However, our fact-checking policy is not exogenously given but it is chosen by a strategic agent incurring the fact-checking cost. Our focus is the implications of fact-checker's incentives on the equilibrium outcomes and players' welfare. Besides

the cheap-talk setting, Ederer and Min (2021) study the consequences of the lie detection presence in a binary Bayesian persuasion model of Kamenica and Gentzkow (2011). Ederer and Min (2021) show that the sender lies more often and the sender's payoff weakly decreases with the improvement of the lie detection technology. Interestingly, for their environment we show that if the fact-checker checks more aggressively, then the sender's payoff increases, as it helps the high sender's type to separate more often.

Our work is related to the literature on optimal auditing. This strand of literature pioneered by Townsend (1979) studies the effects of auditing on the sender's incentives to misrepresent private information. The auditor commits to an auditing scheme specifying auditing probabilities for sender's claims and additional transfers when the sender's claim is checked. A fact-checking policy chosen by the fact-checker in our setting can be seen as an auditing scheme. As in our paper, Border and Sobel (1987) and Mookherjee and Png (1989) allow for stochastic auditing schemes. Also Baron and Besanko (1984) and Laffont and Tirole (1986) present models in which auditing cannot guarantee learning of sender's private information because of an exogenous noise, which corresponds to our imperfect fact-checking technology. The auditor relies on transfers to induce truth-telling by the sender. However, our fact-checker does not have an access to transfers. Instead, the fact-checker has to respect the constraints of the resulting sender-receiver game altered by a fact-checking policy. In this sense, our model is purely informational as our strategic intermediary can only use informational tools to affect the outcomes of the game. In this light, we view our paper as a bridge between literatures on communication with detectable deceit and optimal auditing.

Our paper can also be linked to the literature on the strategic mediation. Ivanov (2010), Ambrus, Azevedo, and Kamada (2013), and Salamanca (2021) allow for the possibility of the biased mediator in a cheap-talk model. The closest paper to ours is Ivanov (2010) who introduces the strategic mediator into an otherwise standard uniform-quadratic setting of the Crawford and Sobel (1982). This paper shows that there exists a strategic mediator that delivers the highest possible receiver's payoff. Importantly, the optimal mediator for the receiver is not pro-receiver, with the bias opposed to the sender's bias. Relative to this paper, our fact-checker

acts as the strategic mediator who has commitment power.[9] Moreover, the fact-checker is unable to send arbitrary messages and restricted to the usage of the fact-checking technology. The strategic mediator in Ivanov (2010) may increase the noise in communication, whereas the fact-checking technology can only decrease the noise.

Finally, our paper relates to the empirical literature on of fact-checking. Nieminen and Rapeli (2019) provide a survey of this literature and outline certain empirical regularities. The evidence on the effects of fact-checking is mixed. Weeks and Garrett (2014) and Weeks (2015) show that the corrections to false information improve the belief accuracy of the receivers of information. By the means of a randomized during the 2017 French presidential election campaign, Barrera et al. (2020) find that the fact-checking of "alternative facts" by Marine Le Pen shifted voters' posteriors on facts towards the truth but did not affect policy conclusions or support for the candidate. Ideology and political affiliation with a speaker may decrease the effectiveness of fact-checking in adjusting beliefs (Nyhan and Reifler, 2010, Jarman, 2016).[10] Recent experimental evidence suggests that retractions of earlier information are hard to process and subjects do not fully "unlearn" upon receiving the retraction (Gonçalves, Libgober, and Willis, 2021). To focus on the fact-checkers' incentives, we assume that the receivers of information are able to fully process the fact-check outputs. Nyhan and Reifler (2015) demonstrate that the fact-checking efforts may discourage politicians from spreading false claims. We show that the sender's incentive to spread false claims may be shut down under the perfect fact-checking technology. Concerning the influence of the fact-checker's identity on the effects of fact-checking, Wintersieck, Fridkin, and Kenney (2021) find that the source of the fact-check only modestly impacts assessments of the fact-check output. Lim (2018) suggests that different fact-checkers rarely check the same claims: only one in 10 statements was found to be fact-checked by both the Washington Post Fact Checker and Politifact.[11] Our paper provides con-

[9] The mediator in Salamanca (2021) maximizes the sender's payoff and also has commitment power.

[10] Nyhan and Reifler (2010) demonstrate a "backfire" effect: corrections may increase the belief in false claims among some ideological groups. The importance of the backfire effect is disputed as many following studies found no evidence for the backfire effect (Weeks and Garrett, 2014, Wood and Porter, 2018, Nyhan et al., 2020).

[11] Amazeen (2015) and Amazeen (2016) provide an evidence of the consistency of the fact-check output for the same claim for different fact-checkers. At the same time, Marietta, Barker, and Bowser (2016) reports variations of the fact-check outputs for the claims on topics of climate change, racism, and consequences of the national debt.

ditions on the fact-checker's preferences such that the claim will not be checked. Additionally, even if both fact-checkers would like to check the claim by themselves, the coordination problem fostered by free-riding may result in the underprovision of fact-checking.

2 Model

There are three players, a sender (he), a fact-checker (it), and a receiver (she), who participate in a one-round communication game.

There is an issue $\theta \in \{0,1\}$ that is relevant for a receiver's decision between accepting, $a = 1$, or rejecting, $a = 0$, the sender's proposition. The prior probabilities are summarized by $\mu(\theta)$, with $\mu(1) = \mu \in (0,1)$, with a slight abuse of notation. The privately informed sender learns the truth about the issue θ. I will refer to a sender with knowledge θ as θ-sender. The sender makes a claim about the issue in a form of a costless message $m \in \mathcal{M} = \{0,1,m_s\}$. The message $m \in \{0,1\}$ is referred to as non-silent and corresponds to a sender's claim that $\theta = m$. The message $m = m_s$ is referred to as a silent message. The sender's goal is to convince the receiver to accept, that is, the sender's payoff is $u_S(a) = a$. The receiver's payoff $u_R(a,\theta)$ is $\theta - \omega$ if the receiver chooses to accept and 0 if the receiver decides to reject the sender's proposition.[12] The parameter $\omega \in (0,1)$ tracks the minimal receiver's belief that $\theta = 1$ for the receiver to be willing to accept the sender's proposition.

A key novelty is an introduction of the strategic fact-checker. The fact-checker has an access to a technology that checks the veracity of sender's claims. The usage of this technology has a cost of $c \geq 0$. The technology operates in the following way. If claim $m \in \{0,1\}$ is checked, then the fact-check output $\mathcal{O} \in \{0,1\}$ is generated. The fact-check output is $\mathcal{O} = 1$ when $\theta = m$ and $\mathcal{O} = 0$ when $\theta \neq m$. If claim $m \in \mathcal{M}$ is not checked, then the fact-check output is empty, $\mathcal{O} = \varnothing$. The fact-checker selects a *fact-checking policy* $\chi : \mathcal{M} \to [0,1]$, where $\chi(m)$ specifies the probability of the fact-checker initiating a check of sender's claim m. Without loss of generality, we can set $\chi(m_s) = 0$, since a silent message does not have an informational

[12] For θ being an element of the unit interval, the same payoff structure for the receiver is adopted in Kolotilin et al. (2017), Shishkin (2021), Mylovanov and Zapechelnyuk (2021), among others. This specification effectively makes $a = 1$ a "risky" action with a state-dependent payoff for the receiver, while $a = 0$ is a "safe" action.

content for a check. Importantly, we allow for a failure of the fact-checking technology to produce a fact-check with probability $p \in [0,1)$. Specifically, when the fact-checker initiates a fact-check of m with probability $\chi(m)$, claim m is checked with probability $\hat{\chi}(m) := (1-p)\chi(m)$.[13] The total probability of an empty fact-check output $\mathcal{O} = \varnothing$ following claim m is $1-\hat{\chi}(m)$. When $p = 0$, we call the fact-checking technology *perfect*. Otherwise, we call the fact-checking technology *imperfect*.

We assume that the fact-checker has commitment power. Accordingly, the fact-checker chooses the fact-checking policy χ at the outset of the game. Before making the decision a, the receiver observes sender's claim m and fact-check output $\mathcal{O}$. Note that in this setting, the fact-checking policy is only relevant for the sender's strategy, whereas the receiver may potentially not even observe χ.[14]

The fact-checker has preferences over decision-issue pairs, $u_F(a,\theta)$, net of the fact-checking cost. Sometimes we will focus on the three natural variations of fact-checker's preferences. The fact-checker is *pro-receiver* if $u_F(a,\theta) = u_R(a,\theta)$. The fact-checker is *pro-sender* if $u_F(a,\theta) = u_S(a)$. Finally, the fact-checker is *anti-sender* if $u_F(a,\theta) = -u_S(a)$. Fact-checker's preferences are fixed throughout the game, parameters ω, μ, and p are common knowledge, and all players are expected utility maximizers.

We will refer to the parameters (μ,ω) as an environment. It will be useful to distinguish whether the environment is predisposed toward the sender or not. Specifically, when $\mu < \omega$, that is, under no information the receiver chooses to reject the sender's proposition, we refer to the parameters of the game as a *sender-unfavorable environment* (SUE). When the receiver chooses to accept under the prior, that is, $\mu > \omega$, we refer to the parameters of the game as a *sender-favorable environment* (SFE).

We now define an equilibrium of this game. The fact-checking policy χ selected by the fact-checker specifies the probabilities of initiating a fact-check of claims $m = 0$ and $m = 1$. The actual probability of claim m checked is then $\hat{\chi}(m) = (1-p)\chi(m)$. A sender's strategy is a

[13] We can allow the failure probability of the fact-checking technology to vary across m, but that would not change our results qualitatively.

[14] The situation will change if the receiver has an option to search for a fact-check at some non-zero search cost. Then the decision whether to seach for a fact-check will take χ into account.

probability distribution $\sigma(\cdot|\theta)$ over messages $m \in \mathcal{M}$ sent by θ-sender. A receiver's acceptance strategy $\alpha(m, \mathcal{O})$ specifies the receiver's probability of accepting the sender's proposition based on message m and fact-check output $\mathcal{O}$ she gets. A receiver's posterior belief that $\theta = 1$ after observing pair $(m, \mathcal{O})$ is $\pi(m, \mathcal{O})$. Each fact-checker's choice of fact-checking policy χ initiates a subgame between the sender and the receiver for which I require standard perfect Bayesian equilibrium conditions and an additional requirement of *consistency with fact-checking technology*:

1. (belief evolution) Whenever at least one of $\sigma(m|0)$ or $\sigma(m|1)$ is non-zero,
$$\pi(m, \varnothing) = \frac{\mu\sigma(m|1)}{\mu\sigma(m|1) + (1-\mu)\sigma(m|0)};$$
2. (consistency with fact-checking technology) For $m \in \{0,1\}$ and $\mathcal{O} \in \{0,1\}$,
$$\pi(m, \mathcal{O}) = \begin{cases} 1, & \text{if } m = \mathcal{O}; \\ 0, & \text{if } m \neq \mathcal{O}; \end{cases}$$
3. (the receiver behaves optimally) If $\pi(m, \mathcal{O}) > \omega$, then $\alpha(m, \mathcal{O}) = 1$. If $\pi(m, \mathcal{O}) < \omega$, then $\alpha(m, \mathcal{O}) = 0$;[15]
4. (the sender behaves optimally) $\sigma(\cdot|\theta)$ is supported on
$$\arg\max_{m \in \mathcal{M}} \left\{\hat{\chi}(m)\alpha(m, 1\{m = \theta\}) + (1 - \hat{\chi}(m))\alpha(m, \varnothing)\right\}.$$[16,17]

The first requirement is a standard Bayesian updating of receiver's beliefs after observing on-path messages. Consistency with fact-checking technology requires the receiver's understanding of the nonempty fact-check output both for on-path and off-path messages.[18] The third requirement states that the receiver's decision is optimal given her beliefs. The final requirement

[15] When $\pi(m, \mathcal{O}) = \omega$, this requirement does not restrict the receiver's acceptance probability.

[16] Given that $\chi(m_s) = 0$, the value assigned to $1\{m_s = \theta\}$ is irrelevant.

[17] Even though the sender has state-independent preferences, he has to consider the implications of a potential fact-check output which is a function of the true θ.

[18] Our definitions of on-path and off- path messages are standard. Fixing equilibrium sender's strategy σ, the on-path messages are messages that satisfy at least one of the inequalities $\sigma(m|1) > 0$ and $\sigma(m|0) > 0$. The off-path messages are messages that are not on-path.

prescribes that the sender sends only messages that lead to the highest probability of acceptance, with an understanding that these messages can be fact-checked.

Given χ, we refer to a triple (σ, α, π) that satisfies conditions above as a χ-equilibrium. Let $\mathcal{E}(\chi)$ denote the set of χ-equilibria, with a typical element ε. Each χ-equilibrium $\varepsilon = (\sigma, \alpha, \pi)$ is associated with the joint distribution of decisions and issues $\lambda(a, \theta|\varepsilon, \chi)$.[19] The fact-checker's problem is to choose fact-checking policy χ and χ-equilibrium jointly to maximize its expected payoff net of the fact-checking cost. Specifically, the fact-checker solves

$$\max_{\chi} \max_{\varepsilon \in \mathcal{E}(\chi)} \left\{ \sum_{a,\theta} u_F(a, \theta) \lambda(a, \theta|\varepsilon, \chi) - c \sum_{\theta, m \in \{0,1\}} \chi(m) \sigma(m|\theta) \mu(\theta) \right\}.$$

A solution to this problem is an equilibrium.

In our definition of the equilibrium, we view the fact-checker as a principal who is able to select among its favorite equilibria.[20] Nonetheless, in our analysis we discuss the properties of all χ-equilibria available to the fact-checker upon choosing fact-checking policy χ.

We characterize χ-equilibria of this game using corresponding sender's and receiver's payoffs. Fixing a fact-checking policy χ and a χ-equilibrium, $U_S(\theta)$ stands for the payoff of θ-sender, $U_S = \mu U_S(1) + (1 - \mu) U_S(0)$ is the sender's ex ante payoff, and U_R is the receiver's ex ante payoff. We say that equilibrium payoffs $U_S(\theta)$ and U_R are *feasible* if there is a fact-checking policy χ and a χ-equilibrium that generate those payoffs.

[19] Formally, $\varepsilon = (\sigma, \alpha, \pi)$ generates a joint decision-issue distribution as follows:

$$\lambda(a = 1, \theta|\varepsilon, \chi) = \mu(\theta) \sum_{m \in \mathcal{M}} \sigma(m|\theta) \left[\hat{\chi}(m) \alpha(m, 1\{m = \theta\}) + (1 - \hat{\chi}(m)) \alpha(m, \varnothing) \right],$$

$$\lambda(a = 0, \theta|\varepsilon, \chi) = \mu(\theta) \sum_{m \in \mathcal{M}} \sigma(m|\theta) \left[\hat{\chi}(m)(1 - \alpha(m, 1\{m = \theta\})) + (1 - \hat{\chi}(m))(1 - \alpha(m, \varnothing)) \right].$$

[20] This is a standard assumption in the information design literature for an agent with commitment power (e.g., Kamenica and Gentzkow, 2011). Mathevet, Perego, and Taneva (2020) analyze the information design framework under various selection rules, including the worst-equilibrium selection.

3 Feasible Payoffs and χ-equilibria

In this section, we deliver properties of the feasible payoffs across all possible fact-checking policies. We also characterize χ-equilibria depending on the environment (μ, ω) and the failure probability of the fact-checking technology p. We start our analysis by considering two extreme cases of the fact-checking policies: *no fact-checking* and *full fact-checking*.

The no fact-checking policy corresponds to $\chi(0) = \chi(1) = 0$, with no messages ever checked. Without fact-checking, messages do not have an intrinsic meaning. Our game collapses to the cheap-talk game with a binary state of the world and state-independent sender's preferences. In SUE, the equilibrium sender's strategy is such that any message leads to the receiver rejecting the sender's proposition. Consequently, $U_S(1) = U_S(0) = 0$. On the other hand, in SFE, the receiver accepts the sender's proposition after observing any on-path message: $U_S(1) = U_S(0) = 1$. Irrespective of the environment, some equilibria can still be informative, with some messages revealing the issue. However, the receiver's payoff is fixed across all χ-equilibria at *no-communication payoff* $U_R = \max\{0, \mu - \omega\}$.[21]

Consider now the full fact-checking policy, that is, $\chi(0) = \chi(1) = 1$, and suppose that the fact-checking technology is perfect, $p = 0$. Then both non-silent messages $m = 0$ and $m = 1$ have an intrinsic meaning: they are always checked, and after observing a non-silent message, the receiver is guaranteed to learn the issue. In SUE, such fact-checking policy prevents 0-sender and 1-sender from pooling on the silent message. In fact, 1-sender never sends the silent message. Indeed, for 1-sender to be willing to send m_s, the receiver needs to accept after this message with probability one. This is because 1-sender can always send only a true message $m = 1$: by consistency with fact-checking technology and under given fact-checking policy, the receiver understands the implications of observing $(m, \emptyset) = (1, 1)$ and chooses the sender-preferred action. However, in the equilibrium, it is impossible to have $\alpha(m_s, \varnothing) = 1$, since the condition of the sender-unfavorable environment would require 0-sender to place some weight on fully checked non-silent messages creating profitable deviations for him. Thus, the receiver always learns the issue in SUE, generating equilibrium payoffs of $U_S(1) = 1$, $U_S(0) = 0$,

[21] Additional information does not increase the receiver's payoff, since her optimal action remains unchanged conditional on receiving or not receiving this information.

and $U_R = \mu(1-\omega)$. The situation is different in SFE. Here pooling on the silent message is an imperishable equilibrium. Due to 1-sender's indifference between revealing himself and being pooled with 0-sender, two equilibrium patterns persist. In one, as in SUE, 1-sender never sends m_s and the receiver learns the issue. In another, the receiver does not fully learn after observing the silent message but still accepts the sender's proposition. The equilibrium payoffs are $U_S(1) = 1$, $U_S(0) \in \{0,1\}$, and $U_R \in \{\mu-\omega, \mu(1-\omega)\}$ in SFE, depending on the equilibrium pattern. When the fact-checking technology is imperfect, $p > 0$, the probability of a non-silent message checked is bounded above by $1-p$. As a result, the receiver cannot be assured to fully learn, since 0-sender's mimicking of 1-sender stays undetected with non-zero probability. Then the payoff of 1-sender is not greater than $1-p$ in SUE. Similarly, 0-sender is guaranteed to get a payoff of p in SFE.

We now proceed to characterizing feasible payoffs spanned by all fact-checking policies. We show that two insights from aforementioned fact-checking policies generalize to any fact-checking policy χ. First, 0-sender's proposition is always rejected by the receiver in SUE. Second, 1-sender always gets his proposition accepted in SFE.

Proposition 1. *The feasible sender's payoffs are*

- $U_S(1) \in [0, 1-p]$ *and* $U_S(0) = 0$ *in the sender-unfavorable environment,*
- $U_S(1) = 1$ *and* $U_S(0) \in [p, 1]$ *in the sender-favorable environment.*

All proofs are in Appendix. This result has several constituencies. First, no fact-checking and full fact-checking policies deliver the extremes of the range of sender's feasible payoffs. Second, we can always construct a fact-checking policy χ and a corresponding χ-equilibrium that generate an interior 1-sender's payoff in SUE and 0-sender's payoff in SFE. One such construction is as follows. Suppose the fact-checker chooses a fact-checking policy χ, with $\chi(1) \geq \chi(0)$. Both 0-sender and 1-sender completely pool on $m = 1$, that is, $\sigma(1|1) = \sigma(1|0) = 1$. The receiver learns the issue with probability $\hat{\chi}(1) = (1-p)\chi(1)$ and makes an optimal choice. With probability $1-\hat{\chi}(1)$, message $m = 1$ is not checked. In such an event, the receiver chooses to reject in SUE and accept in SFE. With appropriately chosen receiver's posterior beliefs after off-path messages, we show that this is indeed a χ-equilibrium.

The sender's payoffs are $U_S(1) = \hat{\chi}(1)$ and $U_S(0) = 0$ in SUE, whereas $U_S(1) = 1$ and $U_S(0) = 1 - \hat{\chi}(1)$ in SFE. Finally, this result shows that no other sender's payoffs are feasible. Intuitively, with probability of at least p, the fact-checking technology fails to produce a fact-check, and the game unfolds as if the no fact-checking policy is in place. Moreover, in SUE, fact-checking can only help 1-sender to separate himself from 0-sender. On the other hand, fact-checking only detects 0-sender's mimicking in SFE.

Note that Proposition 1 implies that the receiver always plays a pure strategy after on-path messages in both SUE and SFE. Indeed, if the receiver was mixing on the equilibrium path, the payoffs of both 0-sender and 1-sender would be strictly between 0 and 1, which contradicts Proposition 1.

We now relate the result to the best possible communication outcome for the sender. In a setting without the fact-checker but with the sender's commitment power as in Kamenica and Gentzkow (2011), the sender can obtain the ex ante payoff of $\frac{\mu}{\omega}$ in SUE. To achieve this, 1-sender always sends a "winning" message $m_w \in \mathcal{M}$ and 0-sender sends m_w with probability $\frac{\mu}{1-\mu} \cdot \frac{1-\omega}{\omega}$ to make the receiver exactly indifferent between taking actions $a = 1$ and $a = 0$ upon observing m_w. The tie is broken in the sender's favor. In our setting, even when the fact-checking technology never fails, the maximum ex ante payoff is $U_S = \mu$ achieved by the full fact-checking policy. The sender's commitment payoff is not achievable, since it requires an undetectable randomization on the side of 0-sender. Our sender lacks commitment power. If θ-sender sends multiple messages, then he is indifferent between sending any one of them. Fact-checking cannot make 0-sender randomize without revealing him. We note that for large state space $\theta \in [0,1]$, this is no longer true. The reason is that the best communication outcome for the sender no longer requires randomization on his side.[22] We discuss this in more detail in Section 6.

Proposition 1 tells us that fact-checking affects ex ante sender's payoff by varying only one

[22] Titova (2021) shows that in a sender-receiver game with a large state space, the sender can achieve the commitment outcome with verifiable information only. Also related is Guo and Shmaya (2021) who study a cheap-talk game in which the sender incurs "miscalibration cost" for undermining the meaning of a certain claim. They show that high miscalibration cost acts as a substitute for commitment and the sender can achieve the commitment outcome.

of θ-sender's payoffs. First, 0-sender is not able escape the zero payoff in SUE regardless of whether his messages get checked or not. Additional fact-checking can only help 1-sender to get his messages verified. Second, 1-sender is always capable to get his proposition accepted irrespective of a 0-sender's strategy and a fact-checking policy. Additional fact-checking can only reveal 0-sender more frequently. We now formalize this logic.

We ask a natural question: when the fact-checker checks more aggressively, how are the sender's and the receiver's payoffs affected? For a fixed fact-checking policy χ, let us denote a non-silent message that is checked with the highest probability as $\overline{m} \in \{0,1\}$ and the corresponding probability as $\overline{\chi} = \max\{\hat{\chi}(0), \hat{\chi}(1)\}$. Note that $\overline{\chi}$ is bounded above by $1-p$. Similarly, we define $\underline{m}$ as a non-silent message that is checked with the probability $\underline{\chi} = \min\{\hat{\chi}(0), \hat{\chi}(1)\}$.[23] We say that a fact-checking policy χ is *more aggressive* than χ' if $\overline{\chi} > \overline{\chi}'$.[24] The following proposition shows how the ex ante payoffs of the sender and the receiver alter for a more aggressive fact-checking policy.

Proposition 2. *When the fact-checking policy is more aggressive:*

- *both the sender and the receiver benefit in the sender-unfavorable environment,*
- *the lower bound on the sender's payoff decreases and the upper bound on the receiver's payoff increases in the sender-favorable environment.*

The key insight behind Proposition 2 is that we can characterize the range of sender's and receiver's payoffs in all χ-equilibria as a correspondence with a single input $\overline{\chi}$. In SUE, the payoffs U_S and U_R are unique for all fact-checking policies with the same $\overline{\chi}$. In SFE, this is no longer the case. Still we can characterize the bounds of the payoff range with $\overline{\chi}$ and we show that the set of sender's and receiver's payoffs is greater in the strong set order for a more aggressive fact-checking policy.

Proposition 2 delivers a comparative statics on U_S and U_R for different fact-checking policies. In SUE, 1-sender gets verified more often with a more aggressive fact-checking policy

[23] If $\chi(0) = \chi(1)$, messages $m = 0$ and $m = 1$ can be assigned to $\overline{m}$ and $\underline{m}$ arbitrarily.

[24] This order is chosen primarily for expository purposes. Our results could be presented for an alternative definition of a more aggressive fact-checking policy that would require $\chi(0) \geq \chi'(0)$ and $\chi(1) \geq \chi'(1)$, with at least one strict inequality.

thereby increasing the ex ante sender's payoff. In SFE, 0-sender's claims can be checked more frequently. However, there is an imperishable χ-equilibrium, in which both 0-sender and 1-sender pool on the silent message. Thus, we need to make use of the comparative statics on sets for SFE. The part of Proposition 2 that concerns the receiver is intuitive. A more aggressive fact-checking policy leads to more informative communication, with the same caveat for SFE.

As a by-product, the proof of Proposition 2 characterizes χ-equilibria for any fact-checking policy χ. Here to eliminate the consideration of multiple cases, suppose that $\overline{\chi} > \underline{\chi} > 0$ for the sake of clarity. First, consider the case of the imperfect fact-checking technology, $p > 0$. Table 1 presents the support of sender's equilibrium strategies in SUE. We can see that 1-sender only sends the message that is checked the most. In turn, 0-sender sends $\overline{m}$ with the high enough probability. Specifically, it has to be the case that $\sigma(\overline{m}|0) \geq \frac{\mu}{1-\mu} \cdot \frac{1-\omega}{\omega}$, so that the receiver decides to reject the sender's proposition upon seeing message $\overline{m}$ and an empty fact-check output $\mathcal{O} = \varnothing$. Otherwise, 0-sender would get the positive payoff which contradicts Proposition 1. The remaining weight of $\sigma(\cdot|0)$ can be placed arbitrarily on m_s and $\overline{m}$. These messages reveal 0-sender. However, this additional information does not affect the receiver's payoff, since her optimal action stays unchanged.

$\sigma(m\|\theta)$	$\theta = 0$	$\theta = 1$
$m = m_s$	$\cdot$	0
$m = \underline{m}$	$\cdot$	0
$m = \overline{m}$	$\cdot$	1

Table 1: The support of sender's equilibrium strategy $\sigma(m|\theta)$ in SUE, when $p > 0$.

Table 2 presents potential supports of sender's equilibrium strategies in SFE. There are three types of equilibria depending on which message m is sent by 0-sender. For this message, it has to be the case that $\sigma(m|1) \geq \frac{1-\mu}{\mu} \cdot \frac{\omega}{1-\omega}$, so that the receiver decides to accept the sender's proposition upon seeing message m and an empty fact-check output $\mathcal{O} = \varnothing$. Otherwise, either 1-sender does not get a payoff of one which contradicts Proposition 1, or 0-sender has a profitable deviation. The remaining weight of $\sigma(\cdot|1)$ an be placed arbitrarily on the messages that are checked more frequently than m. These messages reveal 1-sender.

$\sigma(m\|\theta)$	$\theta=0$	$\theta=1$
$m=m_s$	1	$\cdot$
$m=\underline{m}$	0	$\cdot$
$m=\underline{m}$	0	$\cdot$

$\sigma(m\|\theta)$	$\theta=0$	$\theta=1$
$m=m_s$	0	0
$m=\underline{m}$	1	$\cdot$
$m=\underline{m}$	0	$\cdot$

$\sigma(m\|\theta)$	$\theta=0$	$\theta=1$
$m=m_s$	0	0
$m=\underline{m}$	0	0
$m=\underline{m}$	1	1

Table 2: Potential supports of sender's equilibrium strategy $\sigma(m|\theta)$ in SFE, when $p > 0$.

The equilibrium pattern is unique in SUE in the sense that the strategy of one of θ-senders is fixed across χ-equilibria. In SFE, we have multiple equilibrium patterns. This difference stems from the sender's incentives depending on the environment. Indeed, 1-sender simply sends the most checked message in SUE, since he can get a positive payoff only when fact-checked. In SFE, 0-sender only sends the message that is checked the least out of the messages played by 1-sender. In other words, 0-sender wants as little fact-checking as possible but he still needs to mimick 1-sender. This additional requirement generates multiplicity.

When the fact-checking technology is perfect, $p = 0$, it is possible to have a message checked with probability one, that is, $\overline{\chi} = 1$. If $\overline{\chi} = 1$, then there is an additional equilibrium pattern in both SUE and SFE, where 1-sender only sends $\overline{m}$ and 0-sender can play any strategy. In words, 1-sender sending only fully checked messages leaves no option for 0-sender to extract a positive payoff. Then any 0-sender's strategy is an equilibrium strategy. We highlight one of these equilibria, where 0-sender plays the silent message m_s with probability one, and we call this χ-equilibrium *completely separating*. The completely separating equilibrium reveals the issue, while only the claim made by 1-sender gets fact-checked. This will be useful to us, when we characterize the optimal fact-checking policy. We note that the completely separating equilibrium is available only when the fact-checking technology is perfect.

The characterization of χ-equilibria presented above allows us to calculate the ex ante payoffs U_S and U_R for both environments. In SUE, the equilibrium payoffs are unique and equal to $U_S = \mu\overline{\chi}$ and $U_R = \mu(1-\omega)\overline{\chi}$. Intuitively, both the sender and the receiver get the positive payoff only when the message $\overline{m}$ gets fact-checked and the receiver accepts the sender's proposition.

In SFE, the equilibrium payoffs are not unique for fixed $\overline{\chi}$ anymore and they depend on the

equilibrium pattern as presented in Table 2. We can summarize these patterns by message m that 0-sender plays with probability one. If m is the silent message m_s, then the sender always gets his proposition accepted, $U_S = 1$, and the receiver's payoff is equal to the no-communication payoff $U_R = \mu - \omega$. If m is a non-silent message, then 0-sender is revealed with probability $(1-\mu)\chi(m)$, making the receiver change her optimal action to $a = 0$. Hence, the sender's payoff is $U_S = 1 - (1-\mu)\chi(m)$. The receiver's payoff is $U_R = \mu - \omega + (1-\mu)\omega\chi(m)$, the no-communication payoff plus an additional benefit of not making a wrong decision with payoff $-\omega$ when 0-sender gets revealed by fact-checking. We can describe the range of equilibrium payoffs in SFE with $\overline{\chi}$ only:

$$U_S \in [1-(1-\mu)\overline{\chi}, 1] \text{ and } U_R \in [\mu - \omega, \mu - \omega + (1-\mu)\omega\overline{\chi}]\,.$$

Figure 1 provides an illustration of the part of Proposition 2 on the receiver's payoff. The Receiver's payoff under complete information is attainable only with $\overline{\chi} = 1$, which can be achieved by the full fact-checking policy and only when the fact-checking technology is perfect.

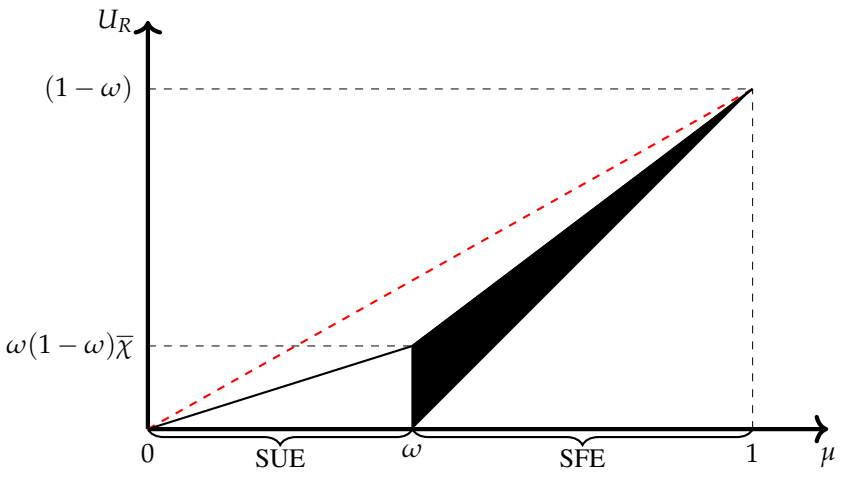

Figure 1: Feasible U_R depending on prior μ for fact-checking policies with fixed $\overline{\chi}$. The dashed red line corresponds to the Receiver's payoff under complete information.

4 Optimal Fact-checking

In this section, we characterize the optimal fact-checking policy for the fact-checker with arbitrary preferences over receiver's decisions and issues. This allows us to generate receiver's preferences over different fact-checkers. We also discuss how our predictions change under the selection of the worst χ-equilibrium for the fact-checker.

The optimal fact-checking policy is characterized by a cost threshold $\overline{c}$. For the fact-checking cost higher than the threshold, no fact-checking is one of the optimal policies. For fact-checking cost lower than the threshold, full fact-checking is one of the optimal policies. We are able to represent the cost threshold in terms of the fact-checker's preferences, as the following proposition shows.

Proposition 3. *For the fact-checker with preferences $u_F(a,\theta)$, there exists $\overline{c}(u_F) > 0$, such that $\overline{\chi} = 0$ is optimal for $c > \overline{c}(u_F)$ and $\overline{\chi} = 1 - p$ is optimal for $c < \overline{c}(u_F)$. Furthermore, when the fact-checking technology is imperfect,*

- $\overline{c}(u_F) = \omega\left[u_F(1,1) - u_F(0,1)\right]$ *in the sender-unfavorable environment,*
- $\overline{c}(u_F) = (1-\mu)\left[u_F(0,0) - u_F(1,0)\right]$ *in the sender-favorable environment.*

When the fact-checking technology is perfect,

- $\overline{c}(u_F) = u_F(1,1) - u_F(0,1)$ *in the sender-unfavorable environment,*
- $\overline{c}(u_F) = \frac{1-\mu}{\mu} \cdot \left[u_F(0,0) - u_F(1,0)\right]$ *in the sender-favorable environment.*

Intuitively, when the fact-checking cost is too high, the no fact-checking policy is optimal. Proposition 3 tells us that if the fact-checking cost becomes sufficiently low, then the full fact-checking policy may become optimal. Following our characterization of χ-equilibria, the joint distribution of decisions and issues $\lambda(a,\theta|\varepsilon,\chi)$ can be summarized by the maximal probability of fact-checking $\overline{\chi}$ for any χ-equilibrium ε. We can then find the minimal cost of fact-checking that supports distribution $\lambda(a,\theta|\varepsilon,\chi)$ as a function of $\overline{\chi}$. We show that the fact-checker's benefit $\sum_{a,\theta} u_F(a,\theta)\lambda(a,\theta|\varepsilon,\chi)$ and the minimal cost of fact-checking are linear functions of $\overline{\chi}$ in the

interior.[25] This linearity generates the threshold policy, making either no fact-checking or full fact-checking optimal depending on the fact-checking cost.

The cost threshold depends only on the fact-checker's preferences $u_F(\cdot,\theta)$ in issue θ, for which $U_S(\theta)$ is varying across different fact-checking policies. By Proposition 1, it is $\theta = 1$ in SUE and $\theta = 0$ in SFE. The reason is $U_S(\theta')$ is fixed for $\theta' \neq \theta$ and thus the distribution of decisions and issues $\lambda(a,\theta'|\varepsilon,\chi)$ is fixed for issue θ' over all fact-checking policies χ and χ-equilibria. Indeed, $U_S(\theta')$ can be written as $\lambda(a=1,\theta'|\varepsilon,\chi)$ in χ-equilibrium ε. Therefore, different fact-checking policies can only affect the fact-checker's payoff in issue θ.

When $\overline{c}(u_F) \leq 0$, the no fact-checking policy is always optimal for the fact-checker with preferences u_F. The fact-checker that prefers $a = 0$ when the issue $\theta = 1$ never fact-checks in SUE. Similarly, the fact-checker that prefers $a = 1$ when the issue $\theta = 0$ plays the no fact-checking policy in SFE. This is intuitive, since the no fact-checking policy effectively shuts down informative communication. Without communication, the receiver already makes a decision preferred by the fact-checker.

The cost threshold depends on the prior only in SFE but stays constant in SUE. Moreover, $\overline{c}(u_F) \to 0$ when $\mu \to 1$. This follows from the set of χ-equilibria available to the fact-checker depending on the environment. Consider the case of the imperfect fact-checking technology. In SUE, distribution $\lambda(a,\theta|\varepsilon,\chi)$ is uniquely pinned down by $\overline{\chi}$. The question is what χ-equilibrium for a fact-checking policy with $\overline{\chi}$ is associated with the minimal cost of fact-checking. The answer to this question is χ-equilibrium in which the maximal weight of 0-sender's strategy is put on an unchecked message, $\sigma(\overline{m}|0) = \frac{\mu}{1-\mu} \cdot \frac{1-\omega}{\omega}$ and $\sigma(m_s|0) = 1 - \sigma(\overline{m}|0)$, such that the receiver's incentive constraints are intact. As a consequence, the fact-checker's benefit and the minimal cost of fact-checking are linear in $\mu\overline{\chi}$, and $\overline{c}(u_F)$ is independent of the prior. In SFE, the minimal cost of implementing any equilibrium pattern from Table 2 is achieved by implementing χ-equilibrium in which 0-sender and 1-sender pool on the same message m, that is, $\sigma(m|0) = \sigma(m|1) = 1$. Any other χ-equilibrium results in more fact-checking without changing the distribution of decision and issues. The fact-checker that desires

[25] There is a caveat concerning multiplicity of χ-equilibria in SFE. Still, as the proof of Proposition 2 indicates, we can represent the fact-checker's objective as a linear function of the parameter that tracks an equilibrium pattern presented in Table 2.

to implement a more aggressive fact-checking policy has to pay a cost in the size of $c\overline{\chi}$, while the fact-checker's benefit is linear in $1-\mu$. As an implication, $\overline{c}(u_F)$ is linear in $1-\mu$ as well.

The perfect fact-checking technology makes a completely separating equilibrium available. In this equilibrium achievable by the full fact-checking policy, 1-sender sends only $\overline{m}$ and 0-sender sends m_s exclusively. In this equilibrium, incentives of 0-sender do not have to be respected. Consequently, the minimal cost of fact-checking has a discontinuity at $\overline{\chi} = 1$. Hence, the fact-checker that wants to implement a full fact-checking policy can do so for a larger range of the fact-checking cost.

Proposition 3 allows us to describe receiver's preferences over settings with different fact-checker's payoffs u_F. Alternatively, if multiple fact-checkers with different payoffs are available in our setting, but the receiver can only listen to one of them[26], we characterize the fact-checker that will be chosen by the receiver in terms of u_F. To fix ideas, suppose that the fact-checker's payoff u_F is a weighted sum of the sender's and the receiver's payoffs: $u_F(a,\theta) = \beta_S u_S(a) + \beta_R u_R(a,\theta) = \beta_S a + \beta_R 1\{a=1\}(\theta-\omega)$. This allows us to deduce the receiver's preferences over different kinds of fact-checkers in terms of weights β_S and β_R, as the corollary below shows.

Corollary 1. *Suppose $u_F(a,\theta) = \beta_S u_S(a) + \beta_R u_R(a,\theta)$. Then the receiver weakly benefits when*

- *β_S increases and β_R increases in the sender-unfavorable environment,*
- *β_S decreases and β_R increases in the sender-favorable environment.*

By Proposition 2, the receiver prefers a more aggressive fact-checking policy. By Proposition 3, the fact-checker is guaranteed to implement either the no fact-checking policy or the full fact-checking policy for almost every fact-checking cost c. Thus, the comparative statics provided in Corollary 1 speaks to the range of the fact-checking cost for which the full fact-checking

[26]The receiver can be unable to observe multiple fact-checkers' outputs due to rational inattention. For example, the receiver can choose how much of the costly attention to pay to each fact-checker as in Myatt and Wallace (2012) and Galperti and Trevino (2020). Under suitable assumptions on the cost of acquiring information, the receiver will pay attention to only one fact-checker.

policy is implemented. This range can only expand when the fact-checker puts more weight on the receiver's payoff. The fact-checker that cares less about the sender, that is, β_S decreases, is more likely to implement the no fact-checking policy in SUE, as under no information the receiver decides to the reject the sender's proposal. Similar logic tells us that if β_S increases in SFE, the fact-checker chooses the no fact-checking policy for a greater range of fact-checking cost.

We can specialize even more and consider the receiver's preferences over pro-receiver, pro-sender, and anti-sender fact-checkers. The pro-receiver fact-checker puts a weight of $\beta_S = 0$ on the sender's payoff and a weight of $\beta_R = 1$. The pro-sender's (anti-sender's) weights are $\beta_S = 1$ ($\beta_S = -1$) and $\beta_R = 0$. By Corollary 1, we can immediately conclude that the receiver prefers the pro-receiver fact-checker over the anti-sender (pro-sender) fact-checker in SUE (SFE). Figure 2 presents the range of the fact-checking cost for which pro-receiver, pro-sender, and anti-sender fact-checkers implement the full fact-checking policy for different prior probabilities μ on $\theta = 1$ and under the imperfect fact-checking technology. When the fact-checking technology is perfect, Figure 2 remains qualitatively similar, with the expanded ranges of fact-checking cost for which the fact-checkers implement the full fact-checking policy.

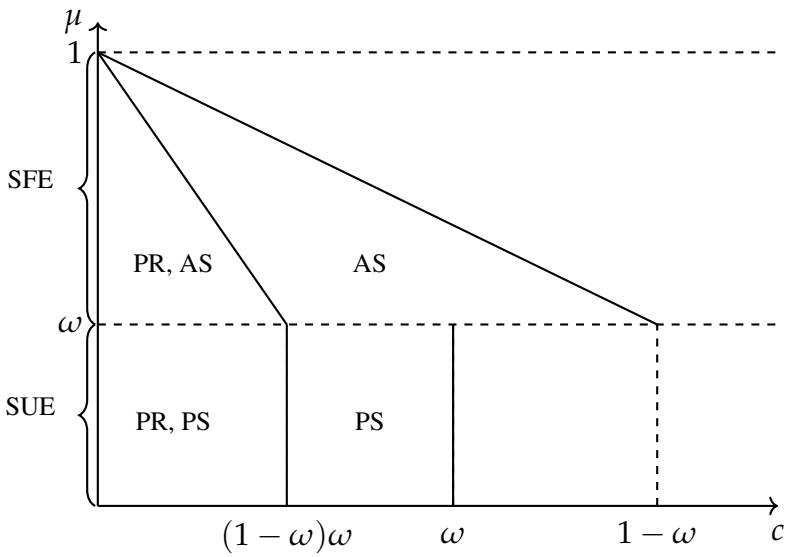

Figure 2: This figure shows the regions in the (c, μ) space for fixed $\omega < \frac{1}{2}$ and the imperfect fact-checking technology, where pro-receiver (PR), pro-sender (PS), and anti-sender (AS) fact-checkers choose the full fact-checking policy.

Figure 2 shows that the anti-sender fact-checker never checks in SUE and the pro-sender fact-checker implements the no fact-checking policy in SFE. Indeed, uninformative communication makes the receiver choose the fact-checker's preferred action. Interestingly, there is a range of the fact-checking cost, for which the receiver's best fact-checker is not pro-receiver. We note that this result is not robust to the linear transformation of u_F: we could rescale u_F for the pro-receiver fact-checker, so that it implements the full fact-checking policy for a larger range of the fact-checking cost.[27] However, our main point is we can always find a fact-checker caring exclusively about the sender's payoff that will be more likely to implement the full fact-checking policy than the fact-checker maximizing the receiver's payoff. The receiver prefers the pro-sender (anti-sender) fact-checker in SUE (SFE) if the following cardinal condition holds for payoff functions u_S and u_R: the sender gains more by persuading the receiver than the receiver by learning the truth.

We now comment on the equilibrium selection. We assume that the fact-checker can stir the sender and the receiver toward its favorite χ-equilibrium. Suppose instead that the worst χ-equilibrium for the fact-checker is played out by the sender and the receiver after it chooses a fact-checking policy χ. In SUE, we know that the distribution of decisions and issues $\lambda(a, \theta|\chi, \varepsilon)$ is uniquely pinned down by the maximal probability of fact-checking $\overline{\chi}$ in fact-checking policy χ. Thus, the fact-checker's benefit does not depend on the selection of a specific χ-equilibrium ε. The worst-equilibrium selection can only drive up the minimal cost of fact-checking by selecting χ-equilibrium, in which both 1-sender and 0-sender only send $\overline{m}$. This would lead to a decrease in the fact-checking threshold $\overline{c}(u_F)$ to the level of $\mu[u_F(1,1) - u_F(0,1)]$. In SFE, the fact-checker that desires to implement a more aggressive fact-checking policy will not be able to sustain informative communication under the worst-equilibrium selection. Indeed, there is an equilibrium in which 1-sender and 0-sender only send the silent message m_s. We conclude that in this case the fact-checker cannot do better than the no fact-checking policy.

[27] If we rescale both u_F and c, then clearly the cost threshold is unaffected.

5 Many Fact-Checkers

This section is devoted to the extension of our baseline model which allows the possibility of multiple fact-checkers available to the receiver. We characterize equilibrium fact-checking policies and their implications for the provision of fact-checking and players' payoffs. We showcase an equilibrium which results in the underprovision of fact-checking relative to the case of only one fact-checker present. We provide conditions for the existence of this equilibrium.

Up until now, we assumed that there is a single fact-checker. In reality, there are many fact-checking institutions available to the receiver, each with a potentially different agenda. What happens to the provision of fact-checking and players' payoffs in our setting if there are several fact-checkers each choosing its own fact-checking policy? To answer this question, we modify our model as follows. Suppose there are two fact-checkers with payoffs $u_{F,1}$ and $u_{F,2}$.[28] At the beginning of the game, each fact-checker decides on the fact-checking policies, χ_1 and χ_2. Note that the probability of message m checked is $\hat{\chi}(m) := 1 - (1 - \hat{\chi}_1(m))(1 - \hat{\chi}_2(m))$. Then the game unfolds as in our baseline model. The sender observes the issue θ and sends message m. The receiver sees sender's message m and realized fact-check outputs $\mathcal{O}_1, \mathcal{O}_2 \in \{0, 1, \varnothing\}$.[29] Then the receiver makes the decision a.

The fact-checking policies chosen by fact-checkers generate probabilities $\hat{\chi}(m)$ of each non-silent message m checked. Then the game continues with a one of χ-equilibria, which we already conveniently characterized in Section 3 with $\overline{\chi} = \max\{\hat{\chi}(0), \hat{\chi}(1)\}$. We can also define $\overline{\chi}_1$ and $\overline{\chi}_2$ as before. To make predictions about the fact-checkers' choice of χ_1 and χ_2, we need to make a stance on the selection of χ-equilibria. We make two assumptions. First, we assume that if there are two available χ-equilibria ε_1 and ε_2 that generate the same joint distribution of decisions and issues but ε_2 is associated with a weakly greater fact-checking cost for both fact-checkers than ε_1 and strictly greater for at least one of them, then ε_2 cannot

[28] When there are more than two fact-checkers, the analysis remains qualitatively the same.

[29] The informational content of two nonempty fact-check outputs is the same. Therefore, the receiver makes the same decision irrespective of whether she observed one or two nonempty fact-check outputs.

be played.[30] Second, in SFE, we assume that 1-sender sends only the most checked message.[31] These assumptions guarantee that after fact-checkers chose their fact-checking policies, they know that the game will continue in accordance with a specific χ-equilibrium. If fact-checkers select their fact-checking policies χ_1 and χ_2 by best responding to each other, then we call χ_1 and χ_2 equilibrium fact-checking policies. Equilibrium fact-checking policies and succeeding χ-equilibrium constitute an equilibrium. In what follows, we characterize equilibrium fact-checking policies.

Suppose $\overline{c}(\cdot)$ is fixed, that is, we fix parameters μ, ω, and p. First, note that either of conditions $\overline{c}(u_{F,i}) < 0$ or $c \geq \overline{c}(u_{F,i})$ imply that the no fact-checking policy is optimal for fact-checker i. Indeed, if fact-checker i does not want to provide information to the receiver when it is alone, a more aggressive fact-checking policy can only negatively affect its payoff. Then the equilibrium fact-checking policy for fact-checker $j \neq i$ is given by Proposition 3. For a more interesting case, suppose that conditions $\overline{c}(u_{F,i}) < 0$ or $c \geq \overline{c}(u_{F,i})$ do not hold for both fact-checkers. In words, both fact-checkers would select the full fact-checking policy if they were an only fact-checker available. Then the following proposition characterizes all equilibrium fact-checking policies.

Proposition 4. *Fix the environment (μ, ω) and the failure probability of the fact-checking technology p. Suppose that $c < \overline{c}(u_{F,i})$ for both fact-checkers. In the equilibrium:*

- *if $c < p\overline{c}(u_{F,i})$ for both fact-checkers, then $\overline{\chi}_1 = \overline{\chi}_2 = 1 - p$ uniquely;*
- *if $c < p\overline{c}(u_{F,i})$ and $c > p\overline{c}(u_{F,j})$, $j \neq i$, then $\overline{\chi}_i = 1 - p$ and $\overline{\chi}_j = 0$ uniquely;*
- *if $c > p\overline{c}(u_{F,i})$ for both fact-checkers, then $\overline{\chi}_1 = 1 - p$, $\overline{\chi}_2 = 0$ or $\overline{\chi}_1 = 0$, $\overline{\chi}_2 = 1 - p$.*

[30] That is, we assume that the chosen χ-equilibrium has to be Pareto-undominated for fact-checkers. While this cooperative selection may seem quaint, we view it as a logical extension of the best-equilibrium selection we adopted for the case of one fact-checker.

[31] In other words, we assume that the third equilibrium pattern from Table 2 or the completely separating equilibrium is played when the fact-checking technology is imperfect (when the first pattern is played, then there is no job for fact-checkers to do). In Section 6, we discuss a possibility of adding a small fine for the sender that is caught in a lie. This extension would select this equilibrium pattern.

Additionally, when the fact-checking technology is imperfect and $c > p\overline{c}(u_{F,i})$ *for both fact-checkers, there is an equilibrium in which* $\overline{\chi}_i = 1 - \frac{c}{\overline{c}(u_{F,j})}$, $j \neq i$.

Importantly, this proposition holds for both SUE and SFE, with $\overline{c}(\cdot)$ given by Proposition 3. If the fact-checking cost is low enough and the fact-checking technology is imperfect, then both fact-checkers select the full fact-checking policy, thereby increasing the maximal probability of fact-checking $\overline{\chi}$ to $1 - p^2$. Thus, the composite fact-checking policy created by two fact-checkers becomes more aggressive than in the case of only one fact-checker. The presence of multiple fact-checkers helps to alleviate the failure of fact-checking technology in this case and increases the provision of fact-checking. The receiver can only benefit from this by Proposition 2. The sender benefits from the added fact-checker only in SUE, as it makes more likely for 1-sender to get his proposition accepted when he is verified by fact-checking.

Alternatively, there are equilibria in which only one fact-checker carries out the full fact-checking policy. In an anticipation of this, another fact-checker prefers to not fact-check at all enjoying the benefit of more informative communication at no cost. This free-riding motive keeps $\overline{\chi}$ at $1 - p$, as if there is only one fact-checker present.[32] When the fact-checking technology is perfect, the maximal probability of fact-checking is one, achieved by these equilibria. However, when the fact-checking technology is imperfect, an additional fact-checker does not assist in overcoming a failure of fact-checking technology. The payoffs of the sender and the receiver remain unaffected.

Moreover, when the fact-checking technology is imperfect and the fact-checking cost is not low enough, there is an equilibrium which may promote the underprovision of fact-checking relative to the case of one fact-checker. In this equilibrium, both fact-checkers do not check to the full extent and the maximal probability of fact-checking is $\overline{\chi} = 1 - \frac{c}{\overline{c}(u_{F,1})} \cdot \frac{c}{\overline{c}(u_{F,2})}$. When $c < \sqrt{p}\sqrt{\overline{c}(u_{F,1})\overline{c}(u_{F,2})}$, the composite fact-checking policy is more aggressive than there is only one fact-checker present. However, when the fact-checking cost is sufficiently high, $c > \sqrt{p}\sqrt{\overline{c}(u_{F,1})\overline{c}(u_{F,2})}$, both fact-checkers want to implement the full fact-checking policy

[32] In a different setting, Carletti, Cerasi, and Daltung (2007) examine a bank's choice between lending to firms individually or in cooperation with other banks. Their setting features a similar free-riding problem due to the need to monitor bank-firm relationships at a cost.

by themselves, but the composite fact-checking policy is less aggressive, $\overline{\chi} < 1 - p$. The coordination problem stimulated by a strong free-riding motive results into the underprovision of fact-checking. In this case, less informative communication hurts the receiver. The sender can enjoy less informative communication only in SFE, as 0-sender gets his proposition accepted more often. We point out that the underprovision of fact-checking can only occur under the imperfect fact-checking technology. When the fact-checking technology is perfect, both fact-checkers never choose the fact-checking technology other than no fact-checking or full fact-checking. This is because the minimal cost of fact-checking is not continuous when $\overline{\chi} = 1$, since the competely separating equilibrium becomes available.

Finally, we point out that the existence of the equilibrium with the underprovision of fact-checking relies on our assumption of the simultaneous fact-checkers' moves. If the fact-checkers moved sequentially, then the first fact-checker would have a first-mover advantage adopting a no fact-checking policy, passing the need to fact-check to the second fact-checker.

6 Discussion

Silent Message. The importance of the right to silence has been recognized by Seidmann (2005) and Ioannidis, Offerman, and Sloof (2020) in the investigator-suspect games. The main insight is that innocent suspects can gain from criminals' exercise of the right. In our model, this insight holds when the fact-checker is able to implement the completely separating equilibrium. This equilibrium is associated with the minimal cost for the fact-checker.

It is important to point out that the presence of the silent message does not affect most of our results. Indeed, the fact-checker can always make message $\underline{m}$ to be essentially silent by letting $\underline{\chi} = 0$. We keep the silent message in our model to illustrate that pooling survives in SFE even if the full fact-checking policy is adopted. Moreover, we view it as our result that the equilibrium payoffs can be presented in terms of $\overline{\chi}$ and do not rely on the presence of silent message making m_s and $\underline{m}$ interchangeable when $\underline{\chi} = 0$.

Resistance to Retractions. Gonçalves, Libgober, and Willis (2021) experimentally show that subjects do not fully unlearn from retractions. Specifically, after getting information and faced

with a retraction, subjects update less than from equivalent new information. This phenomenon may undermine the ability of fact-checking to correct receiver's beliefs. Our parameter p can be seen as a probabilistic way to model this resistance to retractions if we look at it form a different perspective. Suppose that the fact-checking technology is perfect, but the informational content of the fact-check is ignored by the receiver with probability p, as if no fact-check output is observed. Our model is clearly unaffected. Our results suggest that the resistance to retractions puts the receiver at a disadvantage. More interestingly, the sender may benefit from resistance to retractions in SFE as the fact-checking policy becomes effectively less aggressive.

Exogeneous Punishment. In our setting, the distinction between sender's "truths" and "lies" in their conventional meaning does not matter. Indeed, for 1-sender there is no difference whether his message $m = 0$ or $m = 1$ gets fact-checked as the receiver learns the same information about the issue. We can fix this by adding a small exogenous punishment for being caught in a lie.[33] Suppose that if θ-sender sends a non-silent message $m \neq \theta$ and gets fact-checked, then a small fine $f > 0$ is imposed on him. If $\overline{\chi} = \hat{\chi}(1) > 0$, then this addition would eliminate multiplicity in SFE: indeed, 1-sender simply sends the most checked message $m = 1$ and avoids a potential fine. Moreover, 0-sender trades-off the benefits of mimicking 1-sender and the expected fine for being caught in a lie. As a result, $\sigma(1|0) < 1$. The implication for the optimal fact-checking is that the fact-checker no longer needs to check to the full extent to fully reveal the issue.

Larger State Space. Our model considers only claims about the binary issues. In practice, fact-checkers check variety of statements, some of them quantitative in nature.[34] One way to allow for such statements is to enlarge the state space, so that $\theta \in [0,1]$. For simplicity, suppose that the prior is uniform on $[0,1]$ and the message space may contain only the closed intervals that are subsets of the unit interval. For such state space, Titova (2021) shows that the sender

[33] For example, Nyhan and Reifler (2015) provide results for a field experiment suggesting that the speaker is less likely to receive negative fact-checking rating when fact-checking poses a salient threat in a form of reputational risks.

[34] For example, Donald Trump famously spread information about US unemployment rates that received negative fact-checking ratings. See https://www.npr.org/2017/01/29/511493685/ahead-of-trumps-first-jobs-report-a-look-at-his-remarks-on-the-numbers.

can achieve the commitment outcome in SUE with verifiable information only.[35] The solution involves a winning message $m_w = [\theta^*, 1]$ and a losing message $m_l = [0, \theta^*]$, where the cutoff value θ^* is chosen to make the receiver exactly indifferent between taking actions $a = 1$ and $a = 0$ upon observing m_w. The tie is broken in the sender's favor. In our setting, messages are cheap but the fact-checker can provide their verification. Thus, the pro-sender fact-checker is able to deliver the sender's commitment payoff in SUE, if the fact-checking cost is low enough. In particular, the fact-checker only checks m_w with probability one. The outcome does not rely on the selection and does not involve randomization on the sender's side. In our binary setting, the commitment payoff is not achievable, since it requires undetectable randomization by 0-sender which the fact-checker cannot sustain without revealing him. Note that the similar construction to Titova (2021) can show that the anti-sender fact-checker uses the same structure to implement the sender-worst outcome in SFE, with a difference that the cutoff value θ^* for messages m_w and m_l is chosen to make the receiver exactly indifferent between taking actions $a = 1$ and $a = 0$ upon observing m_l and the tie is broken against the sender.

[35] The definitions of SUE and SFE remain the same. In SUE (SFE), the receiver rejects (accepts) the sender's proposition under the prior.

References

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2), 211–236.

Amazeen, M. A. (2015). Revisiting the epistemology of fact-checking. *Critical Review*, 27(1), 1–22.

Amazeen, M. A. (2016). Checking the fact-checkers in 2008: Predicting political ad scrutiny and assessing consistency. *Journal of Political Marketing*, 15(4), 433–464.

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

Balbuzanov, I. (2019). Lies and consequences. *International Journal of Game Theory*, 48(4), 1203–1240.

Baron, D. P., & Besanko, D. (1984). Regulation, asymmetric information, and auditing. *The RAND Journal of Economics*, 15, 447–470.

Barrera, O., Guriev, S., Henry, E., & Zhuravskaya, E. (2020). Facts, alternative facts, and fact checking in times of post-truth politics. *Journal of Public Economics*, 182, 104–123.

Border, K. C., & Sobel, J. (1987). Samurai accountant: A theory of auditing and plunder. *The Review of Economic Studies*, 54(4), 525–540.

Carletti, E., Cerasi, V., & Daltung, S. (2007). Multiple-bank lending: Diversification and free-riding in monitoring. *Journal of Financial Intermediation*, 16(3), 425–451.

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

Dziuda, W., & Salas, C. (2018). Communication with detectable deceit. Available at SSRN 3234695.

Ederer, F., & Min, W. (2020). Bayesian persuasion with lie detection. Available at SSRN 3732910.

Galperti, S., & Trevino, I. (2020). Coordination motives and competition for attention in information markets. *Journal of Economic Theory*, 188, 105039.

Gonçalves, D., Libgober, J., & Willis, J. (2021). Learning versus unlearning: An experiment on retractions. arXiv preprint arXiv:2106.11433.

Graves, L (2016) Deciding what's true: The rise of political fact-checking in American journalism. Columbia University Press.

Graves, L. (2017). Anatomy of a fact check: Objective practice and the contested epistemology of fact checking. *Communication, Culture & Critique*, 10(3), 518–537.

Graves, D. (2018). Understanding the promise and limits of automated fact-checking. Technical report.

Graves, L., & Cherubini, F. (2016). The rise of fact-checking sites in Europe. Reuters Institute for the Study of Journalism, University of Oxford.

Guo, Y., & Shmaya, E. (2021). Costly miscalibration. *Theoretical Economics*, 16(2), 477–506.

Hassan, N., Arslan, F., Li, C., & Tremayne, M. (2017). Toward automated fact-checking: Detecting check-worthy factual claims by claimbuster. *In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1803–1812.

Ioannidis, K., Offerman, T., & Sloof, R. (2020). Lie detection: A strategic analysis of the Verifiability Approach. Working paper.

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

Jarman, J. W. (2016). Influence of political affiliation and criticism on the effectiveness of political fact-checking. *Communication Research Reports*, 33(1), 9–15.

Holm, H. J. (2010). Truth and lie detection in bluffing. *Journal of Economic Behavior & Organization*, 76(2), 318–324.

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

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

Laffont, J. J., & Tirole, J. (1986). Using cost observation to regulate firms. *Journal of Political Economy*, 94(3, Part 1), 614–641.

Lim, C. (2018). Checking how fact-checkers check. *Research & Politics*, 5(3).

Marietta, M., Barker, D. C., & Bowser, T. (2015). Fact-checking polarized politics: Does the fact-check industry provide consistent guidance on disputed realities?. *The Forum*, 13(4), 577–596.

Mathevet, L., Perego, J., & Taneva, I. (2020). On information design in games. *Journal of Political Economy*, 128(4), 1370–1404.

Mookherjee, D., & Png, I. (1989). Optimal auditing, insurance, and redistribution. *The Quarterly Journal of Economics*, 104(2), 399–415.

Myatt, D. P., & Wallace, C. (2012). Endogenous information acquisition in coordination games. *The*

Review of Economic Studies, 79(1), 340–374.

Mylovanov, T., & Zapechelnyuk, A. (2021). A model of debates: Moderation vs free speech. Working paper.

Ostermeier, E. (2011). Selection bias? PolitiFact rates Republican statements as false at three times the rate of Democrats. *Smart Politics*, 10.

Nieminen, S., & Rapeli, L. (2019). Fighting misperceptions and doubting journalists' objectivity: A review of fact-checking literature. *Political Studies Review*, 17(3), 296–309.

Nyhan, B., Porter, E., Reifler, J., & Wood, T. J. (2020). Taking fact-checks literally but not seriously? The effects of journalistic fact-checking on factual beliefs and candidate favorability. *Political Behavior*, 42(3), 939–960.

Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, 32(2), 303–330.

Nyhan, B., & Reifler, J. (2015). The effect of fact-checking on elites: A field experiment on US state legislators. *American Journal of Political Science*, 59(3), 628–640.

Salamanca, A. (2021). The value of mediated communication. *Journal of Economic Theory*, 192, 105191.

Seidmann, D. J. (2005). The effects of a right to silence. *The Review of Economic Studies*, 72(2), 593–614.

Shishkin, D. (2021). Evidence acquisition and voluntary disclosure. Working paper.

Titova, M. (2021). Persuasion with verifiable information. Working Paper.

Townsend, R. M. (1979). Optimal contracts and competitive markets with costly state verification. *Journal of Economic Theory*, 21(2), 265–293.

Uscinski, J. E., & Butler, R. W. (2013). The epistemology of fact checking. *Critical Review*, 25(2), 162–180.

Uscinski, J. E. (2015). The epistemology of fact checking (is still naìve): Rejoinder to Amazeen. *Critical Review*,, 27(2), 243–252.

Walter, N., Cohen, J., Holbert, R. L., & Morag, Y. (2020). Fact-checking: A meta-analysis of what works and for whom. *Political Communication*, 37(3), 350–375.

Weeks, B. E. (2015). Emotions, partisanship, and misperceptions: How anger and anxiety moderate the

effect of partisan bias on susceptibility to political misinformation. *Journal of Communication*, 65(4), 699–719.

Weeks, B. E., & Garrett, R. K. (2014). Electoral consequences of political rumors: Motivated reasoning, candidate rumors, and vote choice during the 2008 US presidential election. *International Journal of Public Opinion Research*, 26(4), 401–422.

Wintersieck, A., Fridkin, K., & Kenney, P. (2021). The message matters: The influence of fact-checking on evaluations of political messages. *Journal of Political Marketing*, 20(2), 93–120.

Appendix

Proof of Proposition 1. We start by showing that $U_S(0) = 0$ for every χ and χ-equilibrium in SUE. Fix the environment (μ, ω), such that $\mu < \omega$. Fix a fact-checking policy χ and χ-equilibrium (σ, α, π). Suppose towards the contrary that $U_S(0) > 0$. This means that there exists an on-path message $m \in \mathcal{M}$, such that $\sigma(m|0) > 0$, $\hat{\chi}(m) < 1$, and $\alpha(m, \varnothing) > 0$. The latter inequality implies that the receiver's posterior belief for message m and empty fact-check output $\mathcal{O} = \varnothing$ satisfies $\pi(m, \varnothing) \geq \omega$. We can represent this condition in terms of the sender's strategy:

$$\sigma(m|1) \geq \frac{1-\mu}{\mu} \cdot \frac{\omega}{1-\omega} \cdot \sigma(m|0) > \sigma(m|0),$$

where the second inequality follows from $\mu < \omega$ and $\sigma(m|0) > 0$. Then $\sigma(m|0) < 1$, and there exists $m' \neq m$, such that $\sigma(m'|0) > 0$. Then for 0-sender to behave optimally, it has to be the case that $\hat{\chi}(m') < 1$ and $\alpha(m', \varnothing) > 0$. Following the same reasoning as for m, we need to have $\sigma(m'|1) > \sigma(m'|0)$. We arrive at a contradiction, since exhausting the probability constraint for 0-sender, $\sum_{m \in \mathcal{M}} \sigma(m|0) = 1$, will violate the probability constraint for 1-sender, $\sum_{m \in \mathcal{M}} \sigma(m|1) = 1$.

For any χ and χ-equilibrium, we now show that $U_S(1) \leq 1 - p$ in SUE. If $p = 0$, then there is nothing to prove, which is why we suppose that $p > 0$. Suppose that there exists a fact-checking policy χ and χ-equilibrium (σ, α, π), such that $U_S(1) > 1 - p$. Since the maximal probability of any message m checked $\hat{\chi}(m)$ is bounded above by $1 - p$, this implies that there exists an on-path message $m \in \mathcal{M}$, such that $\sigma(m|1) > 0$ and $\alpha(m, \varnothing) > 0$. However, this would imply that 0-sender can guarantee himself a non-zero payoff by playing $\sigma(m|0) = 1$. We arrive at a contradiction, since $U_S(0) = 0$.

Finally, we construct a fact-checking policy χ and a χ-equilibrum (σ, α, π) that delivers a payoff in the $[0, 1 - p]$ interval to 1-sender in SUE. Select a fact-checking policy χ, with $\chi(1) \geq \chi(0)$. Consider the sender's strategy that satisfies $\sigma(1|1) = \sigma(1|0) = 1$. Then $\pi(1, \varnothing) < \omega$. Let the posterior belief after off-path messages $m \in \{0, m_s\}$ satisfy $\pi(m, \varnothing) < \omega$. This is an equilibrium. Indeed, 0-sender is indifferent between playing any $m \in \mathcal{M}$. 1-sender does not have a profitable deviation, since he only gets a positive payoff in the event of his non-silent message checked, and $m = 1$ is associated with the maximal probability of

checking $\hat{\chi}(1) = (1-p)\chi(1)$. The payoff of 1-sender in the constructed equilibrium is then $(1-p)\chi(1)$. Therefore, by controlling $\chi(1)$ and respecting the inequality $\chi(1) \geq \chi(0)$, we can produce any $U_S(1) \in [0, 1-p]$.

We now switch to SFE. We start by showing that $U_S(1) = 1$ for every χ and χ-equilibrium in SFE. Fix the environment (μ, ω), such that $\mu > \omega$. Fix a fact-checking policy χ and χ-equilibrium (σ, α, π). Suppose towards the contrary that $U_S(1) < 1$. This means that there exists an on-path message $m \in \mathcal{M}$, such that $\sigma(m|1) > 0$, $\hat{\chi}(m) < 1$, and $\alpha(m, \varnothing) < 1$. The latter inequality implies that the receiver's posterior belief for message m and empty fact-check output $\mathcal{O} = \varnothing$ satisfies $\pi(m, \varnothing) \leq \omega$. We can represent this condition in terms of the sender's strategy:

$$\sigma(m|0) \geq \frac{\mu}{1-\mu} \cdot \frac{1-\omega}{\omega} \cdot \sigma(m|1) > \sigma(m|1),$$

where the second inequality follows from $\mu > \omega$ and $\sigma(m|1) > 0$. However, this implies that there exists an on-path message $m' \neq m$ that satisfies

$$\sigma(m'|0) < \frac{\mu}{1-\mu} \cdot \frac{1-\omega}{\omega} \cdot \sigma(m'|1).$$

This inequality results in $\pi(m', \varnothing) > \omega$. Then 1-sender fails to optimize and we arrive at a contradiction.

For any χ and χ-equilibrium, we now show that $U_S(0) \geq p$ in SFE. If $p = 0$, then there is nothing to prove, which is why we suppose that $p > 0$. Fix a fact-checking policy χ and χ-equilibrium (σ, α, π). We know that $U_S(1) = 1$. Thus, there exists an on-path message $m \in \mathcal{M}$ that satisfies $\sigma(m|1) > 0$, $\hat{\chi}(m) < 1$, and $\alpha(m, \varnothing) = 1$. Then 0-sender can always guarantee himself at least a payoff of $1 - \hat{\chi}(m)$ by playing $\sigma(m|0) = 1$. Since $\hat{\chi}(m) \leq 1-p$, we have $U_S(0) \geq p$.

Finally, we construct a fact-checking policy χ and a χ-equilibrum (σ, α, π) that delivers a payoff in the $[p, 1]$ interval to 0-sender in SFE. Fix a fact-checking policy χ and consider the sender's strategy that satisfies $\sigma(1|1) = \sigma(1|0) = 1$. Then $\pi(1, \varnothing) > \omega$. Let the posterior belief after off-path messages $m \in \{0, m_s\}$ satisfy $\pi(m, \varnothing) < \omega$. This is an equilibrium. Indeed, 1-sender achieves the maximum attainable payoff of 1. 0-sender does not have a profitable deviation, since only sending $m = 1$ brings him a non-zero payoff. The payoff of 0-sender in the

constructed equilibrium is $1-(1-p)\chi(1) \in [p,1]$. Therefore, by controlling $\chi(1)$, we can produce any $U_S(0) \in [p,1]$.

Proof of Proposition 2. Fix a fact-checking policy χ. Let $\overline{\chi} = \max\{\hat{\chi}(1), \hat{\chi}(0)\}$ and $\underline{\chi} = \min\{\hat{\chi}(1), \hat{\chi}(0)\}$. Let $\overline{m}$ ($\underline{m}$) denote a non-silent message that is checked with probability $\overline{\chi}$ ($\underline{\chi}$). If $\chi(1) = \chi(0)$, messages $m = 0$ and $m = 1$ can be assigned to $\overline{m}$ and $\underline{m}$ in an arbitrary way.

We start by characterizing χ-equilibria in SUE. By Proposition 1, we know that $U_S(0) = 0$. This implies that for any on-path message m, we have $\hat{\chi}(m) = 1$ or $\alpha(m, \varnothing) = 0$. If $\overline{\chi} > 0$ and $\overline{\chi} > \underline{\chi}$, then the optimal behavior for 1-sender prescribes $\sigma(\overline{m}|1) = 1$. If $\overline{\chi} = 1$, then any $\sigma(\cdot|0)$ is an equilibrium strategy of 0-sender, with the restriction $\pi(m, \varnothing) < \omega$ on the receiver's posterior belief after an off-path message m. If $\overline{\chi} < 1$, then it has to be the case that $\alpha(\overline{m}, \varnothing) = 0$, or in terms of the 0-sender's strategy, $\sigma(\overline{m}|0) \geq \frac{\mu}{1-\mu} \cdot \frac{1-\omega}{\omega}$. The remaining weight of $\sigma(\cdot|0)$ can be placed arbitrarily on the messages other than $\overline{m}$. The restriction $\pi(m, \varnothing) < \omega$ on the receiver's posterior belief after an off-path message m ensures that we have an equilibrium.

If $\overline{\chi} = \underline{\chi} > 0$, then the optimality for 1-sender prescribes $\sigma(1|1) + \sigma(0|1) = 1$, that is, m_s is never sent by 1-sender. If $\overline{\chi} = 1$, then any $\sigma(\cdot|0)$ is an equilibrium strategy of 0-sender, with the restriction $\pi(m, \varnothing) < \omega$ on the receiver's posterior belief after an off-path message m. If $\overline{\chi} < 1$, then for any m, such that $\sigma(m|1) > 0$, we need to have $\sigma(m|0) \geq \frac{\mu}{1-\mu} \cdot \frac{1-\omega}{\omega} \cdot \sigma(m|1)$. The restriction $\pi(m, \varnothing) < \omega$ on the receiver's posterior belief after an off-path message m ensures that we have an equilibrium.

If $\overline{\chi} = 0$, then for any on-path message m, we have $\alpha(m, \varnothing) = 0$. Thus, any σ that satisfies $\sigma(m|0) \geq \frac{\mu}{1-\mu} \cdot \frac{1-\omega}{\omega} \cdot \sigma(m|1)$ for every on-path message m can be an equilibrium sender's strategy. The restriction $\pi(m, \varnothing) < \omega$ on the receiver's posterior belief after an off-path message m ensures that we have an equilibrium.

Now we characterize χ-equilibria in SFE. By Proposition 1, we know that $U_S(1) = 1$. This implies that for any message m that satisfies $\sigma(m|1) > 0$, we have $\hat{\chi}(m) = 1$ or $\alpha(m, \varnothing) = 1$. In terms of the sender's strategy, $\alpha(m, \varnothing) = 1$ corresponds to the condition $\sigma(m|1) \geq \frac{1-\mu}{\mu} \cdot \frac{\omega}{1-\omega} \cdot \sigma(m|0)$. The optimality for 0-sender prescribes that $\sigma(m|0) > 0$ only if $\sigma(m|1) > 0$

and $m \in \arg\min \hat{\chi}(\cdot)$.

Suppose $\sigma(m_s|1) > 0$. First, consider $\underline{\chi} > 0$. Then $\sigma(m_s|0) = 1$ and $\sigma(m_s|1) \geq \frac{1-\mu}{\mu} \cdot \frac{\omega}{1-\omega}$. The remaining weight of $\sigma(\cdot|1)$ can be placed arbitrarily on non-silent messages. Now consider $\overline{\chi} > \underline{\chi} = 0$. Then in an equilibrium it has to be the case that $\sigma(m_s|0) + \sigma(\underline{m}|0) = 1$. For $m \in \{m_s, \underline{m}\}$, such that $\sigma(m|0) > 0$, we need to have $\sigma(m|1) \geq \frac{1-\mu}{\mu} \cdot \frac{\omega}{1-\omega} \cdot \sigma(m|0)$. Finally, consider $\overline{\chi} = 0$. For $m \in \mathcal{M}$, such that $\sigma(m|0) > 0$, we need to have $\sigma(m|1) \geq \frac{1-\mu}{\mu} \cdot \frac{\omega}{1-\omega} \cdot \sigma(m|0)$. The restriction $\pi(m, \varnothing) < \omega$ is set for off-path messages m in all cases.

Now suppose that $\sigma(m_s|1) = 0$ and $\sigma(\underline{m}|1) > 0$. Suppose $\underline{\chi} = 1$. Then any $\sigma(\cdot|0)$ is an equilibrium strategy of 0-sender, since any strategy brings him the payoff of zero. Now suppose that $\underline{\chi} \in [0, 1)$ and $\overline{\chi} > \underline{\chi}$. Then $\sigma(\underline{m}|0) = 1$ and $\sigma(\underline{m}|1) \geq \frac{1-\mu}{\mu} \cdot \frac{\omega}{1-\omega}$. The remaining weight of $\sigma(\cdot|1)$ can be placed on $\overline{m}$. If $\overline{\chi} = \underline{\chi} \in [0, 1)$, then $\sigma(\underline{m}|0) + \sigma(\overline{m}|0) = 1$ and for $m \in \{\underline{m}, \overline{m}\}$, such that $\sigma(m|0) > 0$, we need to have $\sigma(m|1) \geq \frac{1-\mu}{\mu} \cdot \frac{\omega}{1-\omega} \cdot \sigma(m|0)$. The restriction $\pi(m, \varnothing) < \omega$ is set for off-path messages m in all cases.

Now suppose that $\sigma(\overline{m}|1) = 1$. If $\overline{\chi} = 1$, then any $\sigma(\cdot|0)$ is an equilibrium strategy of 0-sender, since any strategy brings him the payoff of zero. If $\overline{\chi} < 1$, then the optimality for 0-sender prescribes that $\sigma(\overline{m}|0) = 1$. The restriction $\pi(m, \varnothing) < \omega$ is set for off-path messages m in all cases. This completes the characterization of χ-equilibria, since we exhausted all possibilities.

We can calculate the sender's and receiver's payoffs in χ-equilibria we characterized in terms of $\overline{\chi}$ and $\underline{\chi}$.

In SUE, $U_S(1) = \overline{\chi}$, since 1-sender plays messages that are checked the most and he gets a payoff of 1 only when fact-checked. Thus, the sender's ex ante payoff is $U_S = \mu\overline{\chi}$. The receiver's payoff is $U_R = \mu(1-\omega)\overline{\chi}$. Indeed, the receiver plays $a = 1$ only when 1-sender's message gets fact-checked.

In SFE, the sender's and receiver's payoffs depend on the support of equilibrium 1-sender's strategy $\sigma(\cdot|1)$. Suppose the support of $\sigma(\cdot|1)$ includes a message that is checked with probability zero (m_s is one such message irrespective of a fact-checking policy). Then 0-sender only sends such messages. The payoff of 0-sender is $U_S(0) = 1$ and the sender's ex ante payoff is then $U_S = 1$. The receiver's payoff is the payoff without communication $U_R = \mu - \omega$. In-

stead, suppose that the support of $\sigma(\cdot|1)$ does not include m_s but includes $\underline{m}$ that is checked with probability $\underline{\chi} \in [0, \overline{\chi}]$. Then the support of $\sigma(\cdot|0)$ only includes messages that are checked with probability $\underline{\chi}$. The payoff of 0-sender is $U_S(0) = 1 - \underline{\chi}$ and the sender's ex ante payoff is $U_S = 1 - (1-\mu)\underline{\chi}$. The receiver's payoff is $U_R = \mu(1-\omega) + (1-\mu)(1-\underline{\chi})(-\omega) = \mu - \omega + (1-\mu)\omega\underline{\chi}$. Finally, suppose that $\sigma(\overline{m}|1) = 1$. Then either $\overline{\chi} = 1$ or 0-sender pools on $\overline{m}$, $\sigma(\overline{m}|0) = 1$. In either case, the payoff of 0-sender can be summarized by $U_S(0) = 1 - \overline{\chi}$. The sender's ex ante payoff is $U_S = 1 - (1-\mu)\overline{\chi}$. A similar calculation as above demonstrates that $U_R = \mu - \omega + (1-\mu)\omega\overline{\chi}$.

We conclude that the range of payoffs U_S and U_R in all χ-equilibria for a fixed fact-checking policy χ can be summarized by a single parameter $\overline{\chi}$. In SUE, these payoffs are unique, $U_S = \mu\overline{\chi}$ and $U_R = \mu(1-\omega)\overline{\chi}$, both increasing in $\overline{\chi}$. In SFE, $U_S \in [1-(1-\mu)\overline{\chi}, 1]$ and $U_R \in [\mu - \omega, \mu - \omega + (1-\mu)\omega\overline{\chi}]$. The lower bound on the sender's payoff decreases in $\overline{\chi}$ and the upper bound on the receiver's payoff increases in $\overline{\chi}$.

Proof of Proposition 3. Fix a fact-checking policy χ. The characterization of χ-equilibria provided in the proof of Proposition 2 allows us to generate available distributions $\lambda(a, \theta|\varepsilon, \chi)$ for any χ-equilibrium ε.

We start from SUE. The joint distribution of actions and issues in SUE for a fixed fact-checking policy χ for any χ-equilibrium ε is given by

$\lambda(a,\theta\|\varepsilon,\chi)$	$\theta = 0$	$\theta = 1$
$a = 0$	$1-\mu$	$\mu(1-\overline{\chi})$
$a = 1$	0	$\mu\overline{\chi}$

For a fact-checking policy with fixed $\overline{\chi}$, the cheapest equilibrium to implement for the fact-checker depends on whether $\overline{\chi} = 1$ or $\overline{\chi} = 0$. If $\overline{\chi} = 1$, then an equilibrium that is associated with the minimal cost of fact-checking has $\sigma(\overline{m}|1) = 1$ and $\sigma(m_s|0) = 1$. Indeed, condition $\sigma(\overline{m}|1) = 1$ has to hold. Then if 0-sender is never checked, then the fact-checker minimizes cost of fact-checking. If $\overline{\chi} < 1$, then $\sigma(m_s|0) = 1$ is not available anymore. Indeed, in any χ-equilibrium, we need to have $\alpha(m, \varnothing) = 0$ for any m that is checked with probability $\overline{\chi}$. Then an equilibrium that is associated with the minimal cost of fact-checking has $\sigma(\overline{m}|1) = 1$,

$\sigma(\overline{m}|0) = \frac{\mu}{1-\mu} \cdot \frac{1-\omega}{\omega}$, and $\sigma(m_s|0) = \frac{\omega-\mu}{(1-\mu)\omega}$. The implied minimal cost of implementing a fact-checking policy with $\overline{\chi}$ is

$$C_{\text{SUE}}(\overline{\chi}) := \begin{cases} \mu c, & \text{if } \overline{\chi} = 1, \\ \frac{\mu\overline{\chi}}{\omega} \cdot c, & \text{if } \overline{\chi} < 1. \end{cases}$$

The problem of the fact-checker with preferences $u_F(a,\theta)$ is then given by

$$\max_{\overline{\chi} \in [0,1-p]} \left\{ \mu\overline{\chi} u_F(1,1) + \mu(1-\overline{\chi}) u_F(0,1) + (1-\mu) u_F(0,0) - C_{\text{SUE}}(\overline{\chi}) \right\}.$$

If $p > 0$, then the objective is a linear function of $\overline{\chi}$ with the following solution:

$$\overline{\chi} \begin{cases} = 0, & \text{if } c > \omega(u_F(1,1) - u_F(0,1)), \\ \in [0, 1-p], & \text{if } c = \omega(u_F(1,1) - u_F(0,1)), \\ = 1-p, & \text{if } c > \omega(u_F(1,1) - u_F(0,1)). \end{cases}$$

If $p = 0$, then the objective is a linear function of $\overline{\chi}$ with a discontinuity at $\overline{\chi} = 1$. The solution is then always a corner solution:

$$\overline{\chi} \begin{cases} = 0, & \text{if } c > u_F(1,1) - u_F(0,1), \\ \in \{0,1\}, & \text{if } c = u_F(1,1) - u_F(0,1), \\ = 1, & \text{if } c < u_F(1,1) - u_F(0,1). \end{cases}$$

Now consider SFE. Let $g(\overline{\chi}) \in [0, \overline{\chi}]$ be a function that tracks what type of χ-equilibrium is played. Specifically, when $g(\overline{\chi}) = 0$, 1-sender's strategy has a message checked with zero probability in its support. When $g(\overline{\chi}) = \underline{\chi} \in (0, \overline{\chi})$, 1-sender's strategy has a message checked with probability $\underline{\chi}$ in its support and $\sigma(m_s|1) = 0$. Finally, when $g(\overline{\chi}) = \overline{\chi}$, 1-sender's strategy only has messages checked with probability $\overline{\chi}$ in its support. The joint distribution of actions and issues in SUE for a fixed fact-checking policy χ for any χ-equilibrium ε that generates function $g(\cdot)$ as described above is given by

$\lambda(a,\theta\|\varepsilon,\chi)$	$\theta = 0$	$\theta = 1$
$a = 0$	$(1-\mu)g(\overline{\chi})$	0
$a = 1$	$(1-\mu)(1-g(\overline{\chi}))$	μ

We now fix g and discuss an equilibrium that minimizes the cost of fact-checking for fixed fact-checking policy with $\overline{\chi}$. When $\overline{\chi} = 1$ and $g(1) = 1$, an equilibrium that is associated with the minimal cost of fact-checking has $\sigma(\overline{m}|1) = 1$ and $\sigma(m_s|0) = 1$, similarly to SUE. When $g(\overline{\chi}) = 0$, an equilibrium that is associated with the minimal cost of fact-checking has $\sigma(m_s|1) = \sigma(m_s|0) = 1$, since all other equilibria of this type include checking non-silent messages of 1-sender. When $g(\overline{\chi}) = \underline{\chi} \in (0, \overline{\chi})$, an equilibrium that is associated with the minimal cost of fact-checking has $\sigma(\underline{m}|1) = \sigma(\underline{m}|0) = 1$, since all other equilibria of this type include checking message $\overline{m}$ of 1-sender which bears additional costs. Finally, when $g(\overline{\chi}) = \overline{\chi}$ and $\overline{\chi} < 1$, both 0-sender and 1-sender send only messages that are checked with probability $\overline{\chi}$. We conclude that the minimal cost of implementing a fact-checking policy with $\overline{\chi}$ is

$$C_{\text{SFE}}(\overline{\chi}, g(\cdot)) := \begin{cases} \mu c, & \text{if } \overline{\chi} = 1 \text{ and } g(1) = 1, \\ g(\overline{\chi})c, & \text{otherwise.} \end{cases}$$

The problem of the fact-checker with preferences $u_F(a, \theta)$ is then given by

$$\max_{\overline{\chi} \in [0, 1-p], g(\cdot)} \{\mu u_F(1,1) + (1-\mu)(1 - g(\overline{\chi}))u_F(1,0) + (1-\mu)g(\overline{\chi})u_F(0,0) - C_{\text{SFE}}(\overline{\chi}, g(\cdot))\},$$

subject to $g(\overline{\chi}) \in [0, \overline{\chi}]$.

If $p > 0$, then the objective is a linear function of $g(\overline{\chi})$ with the following solution:

$$g(\overline{\chi}) \begin{cases} = 0, & \text{if } c > (1-\mu)(u_F(0,0) - u_F(1,0)), \\ \in [0, 1-p], & \text{if } c = (1-\mu)(u_F(0,0) - u_F(1,0)), \\ = 1 - p, & \text{if } c < (1-\mu)(u_F(0,0) - u_F(1,0)). \end{cases}$$

This solution can be achieved by choosing $g(\chi) = \chi$ for all χ. Note that $g(\overline{\chi}) = 1 - p$ is only attainable by this choice of $g(\cdot)$.

If $p = 0$, then the objective is a linear function of $g(\overline{\chi})$ with a discontinuity at $g(\overline{\chi}) = 1$:

$$g(\overline{\chi}) \begin{cases} = 0, & \text{if } c > \frac{1-\mu}{\mu} \cdot (u_F(0,0) - u_F(1,0)), \\ \in \{0, 1\}, & \text{if } c = \frac{1-\mu}{\mu} \cdot (u_F(0,0) - u_F(1,0)), \\ = 1, & \text{if } c < \frac{1-\mu}{\mu} \cdot (u_F(0,0) - u_F(1,0)). \end{cases}$$

This solution can be achieved by choosing $g(\chi) = \chi$ for all χ. Note that $g(\overline{\chi}) = 1$ is only attainable by this choice of $g(\cdot)$.

This completes the proof, as the cost thresholds are inferred from the optimality considerations above.

Proof of Proposition 4. Our assumption of Pareto-undominated χ-equilibrium guarantees that for any χ, a subgame equilibrium for the sender and the receiver is chosen such that the fact-checking cost is minimized for both fact-checkers.

Consider SUE. Suppose that $p > 0$. In this case, the cost threshold in the case of one fact-checker is given by $\overline{c}(u_F) = \omega(u_F(1,1) - u_F(0,1))$ by Proposition 3. Fix the strategy of the second fact-checker χ_2. Note that $\overline{\chi}$ is bounded below by $\overline{\chi}_2 := \max\{\hat{\chi}_2(1), \hat{\chi}_2(0)\}$. Then the cheapest way to generate $\overline{\chi} \in [\overline{\chi}_2, 1-p^2]$ is to check the message $m \in \arg\max \overline{\chi}_2(\cdot)$ with probability $\overline{\chi}_1 = \frac{\overline{\chi}-\overline{\chi}_2}{1-\overline{\chi}_2}$. The problem of the first fact-checker is

$$\max_{\overline{\chi}_1 \in [0,1-p]} \left\{ \mu \overline{\chi} (u_{F,1}(1,1) - u_{F,1}(0,1)) - C_{\text{SUE}}(\overline{\chi}_1) \right\},$$

subject to $\overline{\chi} = 1 - (1-\overline{\chi}_1)(1-\overline{\chi}_2)$. If $\overline{c}(u_{F,1}) \leq 0$ or $c \geq \overline{c}(u_{F,1})$, then the no fact-checking policy is always optimal for the first fact-checker. Otherwise, the best response of the first fact-checker is

$$\overline{\chi}_1(\overline{\chi}_2) \begin{cases} = 0, & \text{if } \overline{\chi}_2 > 1 - \frac{c}{\overline{c}(u_{F,1})}, \\ \in [0, 1-p], & \text{if } \overline{\chi}_2 = 1 - \frac{c}{\overline{c}(u_{F,1})}, \\ = 1-p, & \text{if } \overline{\chi}_2 < 1 - \frac{c}{\overline{c}(u_{F,1})}. \end{cases}$$

Note that if $c < p\overline{c}(u_{F,1})$, then $\overline{\chi}_1(\cdot) = 1-p$ is always a best response.

Similar calculation delivers the best response of the second fact-checker. If $\overline{c}(u_{F,2}) \leq 0$ or $c \geq \overline{c}(u_{F,2})$, then $\overline{\chi}_2(\cdot) = 0$. Otherwise,

$$\overline{\chi}_2(\overline{\chi}_1) \begin{cases} = 0, & \text{if } \overline{\chi}_1 > 1 - \frac{c}{\overline{c}(u_{F,2})}, \\ \in [0, 1-p], & \text{if } \overline{\chi}_1 = 1 - \frac{c}{\overline{c}(u_{F,2})}, \\ = 1-p, & \text{if } \overline{\chi}_1 < 1 - \frac{c}{\overline{c}(u_{F,2})}. \end{cases}$$

If $\overline{c}(u_{F,i}) \leq 0$ or $c \geq \overline{c}(u_{F,i})$ is true for both $i \in \{1,2\}$, then $\overline{\chi}_1 = \overline{\chi}_2 = 0$ in the equilibrium. If $\overline{c}(u_{F,i}) \leq 0$ or $c \geq \overline{c}(u_{F,i})$ is true for one $i \in \{1,2\}$, but not for $j \neq i$, then

$\overline{\chi}_i = 0$ and $\overline{\chi}_j = 1 - p$. Now consider the case where $\overline{c}(u_{F,i}) \leq 0$ or $c \geq \overline{c}(u_{F,i})$ is false for both $i \in \{1,2\}$. If $c < p\overline{c}(u_{F,i})$ is true for both $i \in \{1,2\}$, then $\overline{\chi}_1 = \overline{\chi}_2 = 1 - p$. If $c < p\overline{c}(u_{F,i})$ is true for one $i \in \{1,2\}$, but not for $j \neq i$, then $\overline{\chi}_i = 1 - p$ and $\overline{\chi}_j = 0$ (when $c = p\overline{c}(u_{F,j})$, $\overline{\chi}_j \in [0, 1-p]$). Finally, suppose that $c < p\overline{c}(u_{F,i})$ is false for both $i \in \{1,2\}$. Then there are three equilibria: (1) $\overline{\chi}_1 = 0$, $\overline{\chi}_2 = 1 - p$; (2) $\overline{\chi}_1 = 1 - p$, $\overline{\chi}_2 = 0$; (3) $\overline{\chi}_1 = 1 - \frac{c}{\overline{c}(u_{F,2})}$, $\overline{\chi}_2 = 1 - \frac{c}{\overline{c}(u_{F,1})}$.

Suppose now that $p = 0$. In this case, the cost threshold in the case of one fact-checker is given by $\overline{c}(u_F) = u_F(1,1) - u_F(0,1)$ by Proposition 3. When $\overline{\chi}_2 = 1$, the best response for the first fact-checker is $\overline{\chi}_1 = 0$. As before, the first fact-checker can generate $\overline{\chi} \in [\overline{\chi}_2, 1)$ by checking message $m \in \arg\max \overline{\chi}_2(\cdot)$ with probability $\overline{\chi}_1 = \frac{\overline{\chi} - \overline{\chi}_2}{1 - \overline{\chi}_2} \in [0,1)$. The cost of doing so is $C_{\text{SUE}}(\overline{\chi}_1) = \frac{\mu\overline{\chi}_1}{\omega} \cdot c$. Alternatively, the fact-checker can generate $\overline{\chi} = 1$ by selecting $\overline{\chi}_1 = 1$ at a cost of μc. Note that if $\overline{\chi}_1 > \omega$, then the latter option is cheaper. The problem of the first fact-checker is

$$\max \left\{ \sup_{\overline{\chi}_1 \in [0,1)} \left\{ \mu\overline{\chi}(u_{F,1}(1,1) - u_{F,1}(0,1)) - C_{\text{SUE}}(\overline{\chi}_1) \right\}, \mu(u_{F,1}(1,1) - u_{F,1}(0,1)) - \mu c \right\},$$

subject to $\overline{\chi} = 1 - (1 - \overline{\chi}_1)(1 - \overline{\chi}_2)$. There cannot be an interior solution. Indeed, the objective in the inner problem is linear in $\overline{\chi}_1$. Thus, the supremum is achieved on either $\overline{\chi}_1 = 0$ or $\overline{\chi}_1 = 1$. If the supremum is achieved on $\overline{\chi}_1 = 1$, then $\mu(u_{F,1}(1,1) - u_{F,1}(0,1)) - \mu c$ is greater than this supremum due to the lower cost of fact-checking.

If $\overline{c}(u_{F,1}) \leq 0$ or $c \geq \overline{c}(u_{F,1})$, then the no fact-checking policy is always optimal for the first fact-checker. Otherwise, the best response of the first fact-checker is

$$\overline{\chi}_1(\overline{\chi}_2) \begin{cases} = 0, & \text{if } \overline{\chi}_2 > 1 - \frac{c}{\overline{c}(u_{F,1})}, \\ \in \{0,1\}, & \text{if } \overline{\chi}_2 = 1 - \frac{c}{\overline{c}(u_{F,1})}, \\ = 1, & \text{if } \overline{\chi}_2 < 1 - \frac{c}{\overline{c}(u_{F,1})}. \end{cases}$$

Similar calculation delivers the best response of the second fact-checker. If $\overline{c}(u_{F,2}) \leq 0$ or

$c \geq \overline{c}(u_{F,2})$, then $\overline{\chi}_2 = 0$. Otherwise, the best response of the second fact-checker is

$$\overline{\chi}_2(\overline{\chi}_1) \begin{cases} = 0, & \text{if } \overline{\chi}_1 > 1 - \frac{c}{\overline{c}(u_{F,2})}, \\ \in \{0,1\}, & \text{if } \overline{\chi}_1 = 1 - \frac{c}{\overline{c}(u_{F,2})}, \\ = 1, & \text{if } \overline{\chi}_1 < 1 - \frac{c}{\overline{c}(u_{F,2})}. \end{cases}$$

If $\overline{c}(u_{F,i}) \leq 0$ or $c \geq \overline{c}(u_{F,i})$ is true for both $i \in \{1,2\}$, then $\overline{\chi}_1 = \overline{\chi}_2 = 0$ in the equilibrium. If $\overline{c}(u_{F,i}) \leq 0$ or $c \geq \overline{c}(u_{F,i})$ is true for one $i \in \{1,2\}$, but not for $j \neq i$, then $\overline{\chi}_i = 0$ and $\overline{\chi}_j = 1$. Now consider the case where $\overline{c}(u_{F,i}) \leq 0$ or $c \geq \overline{c}(u_{F,i})$ is false for both $i \in \{1,2\}$. Then there are two equilibria: (1) $\overline{\chi}_1 = 0, \overline{\chi}_2 = 1$; (2) $\overline{\chi}_1 = 1, \overline{\chi}_2 = 0$.

Consider SFE. In this case, the cost threshold in the case of one fact-checker is given by $\overline{c}(u_F) = (1-\mu)(u_F(0,0) - u_F(1,0))$ by Proposition 3. When $p > 0$, in any χ-equilibrium, $\sigma(\overline{m}|1) = \sigma(\overline{m}|0) = 1$. When $p = 0$ and $\overline{\chi} = 1$, there are additional χ-equilibria, in which $\sigma(\overline{m}|1) = 1$ and $\sigma(\cdot|0)$ is arbitrary. Fix the strategy of the second fact-checker. Note that $\overline{\chi}$ is bounded below by $\overline{\chi}_2 := \max\{\hat{\chi}_2(1), \hat{\chi}_2(0)\}$. To generate $\overline{\chi} \in [\overline{\chi}_2, 1-p^2]$, the first fact-checker checks the message $m \in \arg\max \overline{\chi}_2(\cdot)$ with probability $\overline{\chi}_1 = \frac{\overline{\chi} - \overline{\chi}_2}{1-\overline{\chi}_2}$. The problem of the first fact-checker is

$$\max_{\overline{\chi}_1 \in [0,1-p]} \left\{ (1-\mu) g(\overline{\chi}) (u_{F,1}(0,0) - u_{F,1}(1,0)) - C_{\text{SFE}}(\overline{\chi}_1, g(\cdot)) \right\},$$

subject to $\overline{\chi} = 1 - (1-\overline{\chi}_1)(1-\overline{\chi}_2)$, where $g(\cdot)$ is defined on the page 39. Under our selection, $g(\overline{\chi}) = \overline{\chi}$, and

$$C_{\text{SFE}}(\overline{\chi}, \cdot) := \begin{cases} \mu c, & \text{if } \overline{\chi} = 1, \\ \overline{\chi} c, & \text{if } \overline{\chi} < 1. \end{cases}$$

When $p > 0$, the fact-checker's problem can be reduced to:

$$\max_{\overline{\chi}_1 \in [0,1-p]} \left\{ \overline{\chi}_1 \left((1-\overline{\chi}_2) \overline{c}(u_{F,1}) - c \right) \right\}.$$

Then the best responses are the same as in SUE, subject to pinned $\overline{c}(\cdot)$.

When $p = 0$, $\overline{c}(u_F) = \frac{1-\mu}{\mu} \cdot (u_F(0,0) - u_F(1,0))$ by Proposition 3. The problem of the first fact-checker can be written as

$$\max \left\{ \sup_{\overline{\chi}_1 \in [0,1)} \left\{ \mu \overline{c}(u_{F,1}) \overline{\chi} - \overline{\chi}_1 c \right\}, \mu \overline{c}(u_{F,1}) - \mu c \right\},$$

subject to $\overline{\chi} = 1 - (1 - \overline{\chi}_1)(1 - \overline{\chi}_2)$. There cannot be an interior solution for the same reason as in the problem in SUE under the perfect fact-checking technology. Then $\overline{\chi}_1 = 1$ is optimal when $\mu\overline{c}(u_{F,1}) - \mu c \geq \mu\overline{c}(u_{F,1})\overline{\chi}_2$, or $c \leq (1 - \overline{\chi}_2)\overline{c}(u_{F,1})$. When $c \geq (1 - \overline{\chi}_2)\overline{c}(u_{F,1})$, $\overline{\chi}_1 = 0$ is optimal. Then the best responses are the same as in SUE, subject to pinned $\overline{c}(\cdot)$. This completes the proof.