Communication with Strategic Fact-checking

Aleksandr Levkun

UC, San Diego

This version: May 25, 2022

Click here for the latest version

Abstract

We examine communication between an informed sender and an uninformed receiver with a presence of a strategic fact-checker. The sender makes a claim about an issue to persuade the receiver to approve the sender's proposal. The fact-checker has its own goal and chooses a stochastic fact-checking policy that checks sender's claims. Checking a claim is costly and, with some probability, can fail to verify whether the claim is true or false. Full fact-checking is optimal when the cost is below a threshold. Otherwise, no fact-checking is optimal. We characterize the cost threshold as a function of fact-checker's preferences. The receiver need not prefer a fact-checker with preferences aligned with the receiver to one with opposed preferences. Adding multiple fact-checkers does not necessarily improve communication even when all fact-checkers are willing to fully check by themselves. For intermediate cost of checking, having multiple fact-checkers can lead to 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 greatly appreciate the comments and suggestions from Florian Ederer, Shachar Kariv, Mark Machina, Denis Shishkin, Maria Titova, Joel Watson, and participants of the UCSD Workshop in Theory and Behavioral Economics. All errors are mine.

1 Introduction

Fact-checking of prominent public figures has become ubiquitous. Initially 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 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 hold politicians accountable for spreading deceitful claims (Graves, 2016). However, the fact-checkers' role of "arbiters of truth" has drawn criticism on the multiple counts including the fact-checkers' bias.[3] Ostermeier (2011) points out the lacking transparency in how the fact-checked claims get selected. The selection effect may lead to a biased perception of a politician's credibility: 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).

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 receiver and does adding fact-checkers help this receiver 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 receiver has to accept or reject a sender's proposal, but does not know whether it is good or bad for her. The sender is informed about a binary receiver's value of acceptance and can either convey this to the receiver using cheap-talk claims or stay silent. However, the sender would like the

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

[2]See Facebook (2021) and Reuters (2021).

[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).

receiver to always accept and, thus, makes a claim in an attempt to persuade the receiver. The fact-checker may verify truthfulness of a sender's claim by employing a fact-checking technology at a cost. This technology is subject to a potential failure to verify a claim.[4] The fact-checker commits to a stochastic fact-checking policy that initiates checks of sender's potential claims.[5,6] The fact-checker chooses a fact-checking policy to maximize its expected payoff net of the fact-checking cost. The fact-checker's payoff function is the central factor in our analysis. We consider three natural examples of this payoff function. First, the pro-receiver fact-checker maximizes the receiver's payoff. Second, the pro-sender fact-checker wishes for the sender's proposal to be accepted. Third, the anti-sender fact-checker wants the sender's proposal to be rejected.

Without fact-checking, the issue of pooling compromises communication: the bad sender's type pretends to be the good sender. The fact-checker is able to provide separation, thereby increasing the receiver's payoff relative to the sender-receiver cheap-talk game. However, the benefits of fact-checking for the sender depend on whether sender's information is persuasive, that is, whether the receiver accepts the sender's proposal under no communication. We show that when sender's information is not persuasive, fact-checking determines the extent to which the good sender can convince the receiver to accept. Consequently, more frequent fact-checking increases the sender's payoff. In any equilibrium, the good sender simply sends the most checked claim and the players' equilibrium payoffs are unique. When sender's information is persuasive, fact-checking determines the extent to which the bad sender can dupe the receiver into accepting. As

[4] In reality, the failure probability and the fact-checking cost can depend on the issue under consideration. Claims can be hard to verify because of the insufficient or lacking evidence on the issue (Graves, 2016). The paper focuses on one issue at a time, for which there is a given probability of failure.

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

[6] 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 based on the author's field experience with three major fact-checking organizations: PolitiFact, FactCheck.org, and Washington Post's Fact Checker. The first step identifies claims to check. Then the fact-checkers gather the evidence, assess the claim veracity, and publish the fact-check output in a transparent manner. We allow the fact-checker to be strategic only about the first step of this process.

a result, more frequent fact-checking can only harm the sender. The defining property of any equilibrium in this case is that the bad sender prefers as little fact-checking as possible but still needs to mimic the good sender's type. The good sender always gets his proposal accepted, with an opportunity to make any claim. Subsequently, various good sender's behavior corresponds to different players' payoffs.

Our first main result shows that the optimal fact-checking policy is a threshold policy in terms of the fact-checking cost. When the cost is above the threshold, the fact-checker never checks. When the cost is below the threshold, the fact-checker initiates checks with probability one. Even though varying a fact-checking policy affects sender's incentives, we get a bang-bang solution. The reason is that the fact-checker's objective can be written as a linear function of a single input, the maximal probability of checking across claims. Only when the fact-checking technology is perfect in a sense that it never fails, the fact-checker is able to deter the sender from producing false claims. Otherwise, the bad sender attempts to mimic the good type, and the separation is achieved only by successful fact-checking.

The cost threshold is given by the fact-checker's preferences. In particular, the pro-sender (anti-sender) fact-checker never checks when the sender's information is (not) persuasive, since uninformative communication makes the receiver choose the fact-checker's preferred action. Having an access to the pro-receiver fact-checker is not always the best option for the receiver. We can always find a fact-checker caring exclusively about the sender's payoff that fact-checks for a greater range of the fact-checking cost. The sufficient condition for this implication is that the sender gains more by persuading the receiver than does the receiver by learning the truth.

Our second main result addresses the question of whether having multiple fact-checkers improves communication. To study this, the paper considers a situation with two fact-checkers choosing fact-checking policies simultaneously. A fact-checker that would not check were it alone continues not to do so in this setting, since more frequent fact-checking can only decrease its payoff. Interesting equilibrium policies arise when both fact-checkers are willing to check by themselves. The free-riding motive arises: a fact-checker would like to delegate the need to check to another fact-checker enjoying the benefits of more informative communication at no cost. When the fact-checking cost is intermediate, this incentive shapes an equilibrium in which fact-checking is underprovided and the receiver's payoff decreases compared to the case of a single fact-checker. In this equilibrium, each fact-checker initiates checks with a nontrivial probability which depends on the other fact-checker's cost threshold. When the cost is low enough, the free-riding motive is weak and both fact-checkers check to the full extent in the unique equilibrium. The composite fact-checking policy checks more frequently and the failure of the fact-checking technology is mitigated.

Several authors suggested that "partisan" fact-checkers can be harmful for more informative political discourse (Ostermeier, 2011; 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 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

This 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 a version of Crawford and Sobel (1982) model. 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. Fully revealing equilibria exist, even

[7]See also Scientific American (2020).

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

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. Dziuda and Salas (2018) analyze the implication of having the same lie detection technology as in Balbuzanov (2019) 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 (2022) study the consequences of the lie detection presence in a binary Bayesian persuasion model of Kamenica and Gentzkow (2011). Ederer and Min (2022) 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 good sender's type to separate more often.

This paper 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). Ivanov (2010) shows that there exists a strategic mediator that delivers the highest possible receiver's payoff, as if communication happened through an optimal non-strategic mediator. 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, recently surveyed by Nieminen and Rapeli (2019). 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 online experiment 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

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

the candidate.[10] Nyhan and Reifler (2015) demonstrate that the fact-checking efforts may discourage politicians from spreading false claims. 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] We show that the free-riding motive may induce fact-checkers to "divide the market" among themselves, as the benefits of double-checking are swamped by the fact-checking cost.

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. The receiver can choose between accepting or rejecting a sender's proposal. The receiver's payoff depends on a state of the world, whereas the sender has state-independent preference for approval. The receiver's decision relies on information contained in a sender's claim and a fact-check output. A fact-checking policy assigns to each sender's claim the probability that the claim is checked. Successful fact-checking reveals whether the sender's claim is truthful or not, while unsuccessful fact-checking generates the empty output. We seek to solve the problem of the fact-checker who can commit to a fact-checking policy to maximize its payoff.

[10] Ideology and political affiliation with a speaker may decrease the effectiveness of fact-checking in adjusting beliefs (Nyhan and Reifler, 2010; Jarman, 2016). 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; Nyhan, Porter, 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 (2015) reports variations of the fact-check outputs for the claims on topics of climate change, racism, and consequences of the national debt.

Players and information

There is an issue $\theta \in \{0,1\}$ that is relevant for a receiver's decision between accepting, $a = A$, or rejecting, $a = R$, the sender's proposal. Nature picks θ from the prior distribution with probability $\mu(\theta)$, where $\mu(1) = \mu \in (0,1)$, with a slight abuse of notation. The privately informed sender learns θ and makes a claim about the issue in a form of a costless message $m \in \mathcal{M} = \{0,1,m_s\}$. Message $m = m_s$ is a *silent message*. Non-silent message $m \in \{0,1\}$ corresponds to a sender's claim that $\theta = m$. The fact-checker decides whether to check sender's message m for veracity by means of a fact-checking technology described below. Successful fact-checking generates the fact-check output $\mathcal{O} = 1$ if m is truthful and $\mathcal{O} = 0$ if m is deceitful. Unsuccessful fact-checking generates an empty output, $\mathcal{O} = \varnothing$. The receiver observes message m and fact-check output $\mathcal{O}$ and then acts, $a \in \{A,R\}$.

Fact-checking technology

The fact-checker has an access to a technology that verifies truthfulness of sender's claims. The usage of this technology has a cost of $c \geq 0$. If the fact-checker initiates a check of non-silent message m, then the technology produces fact-check output $\mathcal{O} \in \{0,1,\varnothing\}$ in the following way. With probability p, verification fails and $\mathcal{O} = \varnothing$. With probability $1 - p$, the generated fact-check output is $\mathcal{O} = 1$ when $\theta = m$ and $\mathcal{O} = 0$ when $\theta \neq m$. If m is silent or the fact-checker does not initiate a check of m, then the output is empty, $\mathcal{O} = \varnothing$. In what follows, we consider the imperfect fact-checking technology, that is, $p \in (0,1)$. Section 6 discusses the perfect fact-checking technology ($p = 0$).

Strategies

We will refer to a sender with knowledge θ as θ-sender. A sender's strategy is a probability distribution $\sigma(\cdot|\theta)$ over messages $m \in \mathcal{M}$ sent by θ-sender. The fact-checker selects $\chi : \mathcal{M} \to [0,1]$, where $\chi(m)$ specifies the probability of initiating a check of sender's claim m. Without loss of generality, we can set $\chi(m_s) = 0$. Message m is successfully checked with probability $\chi_p(m) := (1-p)\chi(m)$.[12] A fact-checker's strategy is a choice of a *fact-checking policy* $\chi_p(m) \in [0,1-p]$ for $m \in \{0,1\}$. Finally, a receiver's acceptance strategy $\alpha(m,\mathcal{O})$ specifies the probability of choosing $a = A$ after observing message m and fact-check output $\mathcal{O}$. The receiver's posterior belief that $\theta = 1$ is denoted as $\pi(m,\mathcal{O})$.

Payoffs

The sender's goal is to convince the receiver to accept, that is, the sender's payoff is $u_S(a) = 1\{a = A\}$. 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 proposal.[13] The parameter $\omega \in (0,1)$ tracks the minimal belief that $\theta = 1$ for the receiver to be willing to accept the sender's proposal. The fact-checker has preferences over action-issue pairs, $u_F(a,\theta)$, net of the fact-checking cost. We will consider three natural variations of fact-checker's preferences: the fact-checker is *pro-receiver* if $u_F(a,\theta) = u_R(a,\theta)$, *pro-sender* if $u_F(a,\theta) = u_S(a)$, and *anti-sender* if $u_F(a,\theta) = -u_S(a)$. Fact-checker's preferences are fixed, parameters ω, μ, and p are common knowledge, and all players are expected utility maximizers.

Solution concept and equilibrium

We assume that the fact-checker has commitment power. Accordingly, the fact-checker chooses the fact-checking policy χ_p at the outset of the game.[14] Each fact-checker's choice of fact-checking policy χ_p initiates a subgame between the sender and the receiver for which we require standard perfect Bayesian equilibrium conditions and an additional requirement of *consistency with fact-checking technology*:

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

[13] 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) among others. This specification effectively makes $a = A$ a "risky" action with a state-dependent payoff for the receiver, while $a = R$ is a "safe" action.

[14] 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_p. 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 search for a fact-check will take χ_p into account.

1. If at least one of $\sigma(m|0)$ or $\sigma(m|1)$ is non-zero, then $\pi(m,\varnothing) = \frac{\mu\sigma(m|1)}{\mu\sigma(m|1)+(1-\mu)\sigma(m|0)}$.

2. For $m \in \{0,1\}$ and $\mathcal{O} \in \{0,1\}$, $\pi(m,\mathcal{O}) = 1\{m = \mathcal{O}\}$.

3. If $\pi(m,\mathcal{O}) > \omega$, then $\alpha(m,\mathcal{O}) = 1$. If $\pi(m,\mathcal{O}) < \omega$, then $\alpha(m,\mathcal{O}) = 0$.

4. $\sigma(\cdot|\theta)$ is supported on $\underset{m\in\mathcal{M}}{\operatorname{argmax}} \{\chi_p(m)\alpha(m,1\{\theta = m\}) + (1-\chi_p(m))\alpha(m,\varnothing)\}$.[15]

The first requirement is a standard Bayesian updating of receiver's beliefs after observing on-path messages. Consistency with fact-checking technology requires receiver's understanding of a nonempty fact-check output for both on-path and off-path messages.[16] The third requirement states that the receiver's decision is optimal given her beliefs. The final requirement 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_p, we refer to a triple (σ,α,π) that satisfies conditions above as a χ_p-equilibrium. Let $\mathcal{E}(\chi_p)$ denote the set of χ_p-equilibria, with a typical element ε. Each χ_p-equilibrium ε is associated with the joint distribution of decisions and issues $\lambda(a,\theta|\varepsilon,\chi_p)$.[17] The fact-checker's problem is to choose fact-checking policy χ_p and χ_p-equilibrium jointly to maximize its expected payoff net of the fact-checking cost. Specifically, the fact-checker solves

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

A solution to this problem, χ_p^* and $\varepsilon^* \in \mathcal{E}(\chi_p^*)$, 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.[18]

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

[16] Our definitions of on-path and off-path messages are standard. Fixing equilibrium σ, the on-path messages satisfy $\sigma(m|1) > 0$ or $\sigma(m|0) > 0$. The off-path messages are messages that are not on-path.

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

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

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

[18] This is a standard assumption in the information design literature for an agent with commitment power

Fixing a fact-checking policy χ_p and a χ_p-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_p and a χ_p-equilibrium that generate those payoffs.

We refer to a pair (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 proposal, we refer to (μ, ω) as a *sender-unfavorable environment* (SUE). When the receiver chooses to accept under the prior, that is, $\mu > \omega$, we refer to (μ, ω) as a *sender-favorable environment* (SFE).

3 Feasible Payoffs and Subgame Equilibria

In this section, we describe properties of the feasible payoffs across all possible fact-checking policies. We also characterize χ_p-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*. Considering extreme policies helps us to identify lower and upper bounds on the feasible payoffs. We focus on the sender's incentives first and characterize feasible payoffs of 0- and 1-senders, while delegating the discussion of receiver's feasible payoffs to the end of this section.

The no fact-checking policy corresponds to $\chi_p(0) = \chi_p(1) = 0$. 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: $U_S(1) = U_S(0) = 1$.[19]

(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.

[19] Irrespective of the environment, some equilibria can still be informative, with some messages revealing

Consider now the full fact-checking policy, that is, $\chi_p(0) = \chi_p(1) = 1 - p$. Then after observing a non-silent message, the receiver learns the issue with probability $1 - p$. 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 of at least $1 - p$. 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 $(m, \mathcal{O}) = (1,1)$ and chooses the sender-preferred action. However, in a χ_p-equilibrium, it is impossible to have $\alpha(m_s, \varnothing) \geq 1 - p$, 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 learns the issue in SUE conditional on the successful fact-check. Corresponding sender's payoffs are $U_S(1) = 1 - p$ and $U_S(0) = 0$. The situation is different in SFE. Here pooling on the silent message survives as a χ_p-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 when the fact-check is successful. In another, the receiver does not fully learn after observing the silent message but still accepts the sender's proposition. The sender's payoffs are $U_S(1) = 1$ and $U_S(0) \in \{p, 1\}$ in SFE, depending on the equilibrium pattern.

We now proceed to characterizing feasible payoffs spanned by all fact-checking policies. We show that two insights from extreme 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.*

the issue. However, the receiver's payoff is fixed across all χ_p-equilibria at $U_R = \max\{0, \mu - \omega\}$. Indeed, additional information does not increase the receiver's payoff, since her optimal action remains unchanged conditional on receiving or not receiving this information.

All proofs are in the appendix. This result has several implications. 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_p and a corresponding χ_p-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_p, with $\chi_p(1) \geq \chi_p(0)$. 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 $\chi_p(1)$ and makes an optimal choice. With probability $1 - \chi_p(1)$, 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_p-equilibrium. The sender's payoffs are $U_S(1) = \chi_p(1)$ and $U_S(0) = 0$ in SUE, whereas $U_S(1) = 1$ and $U_S(0) = 1 - \chi_p(1)$ 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. 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 $a = A$ and $a = R$ 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.[20] As a result, fact-checking may enable commitment in a setting with a continuous state space. 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 of θ-sender's payoffs. First, 0-sender is not able to 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 by asking 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_p, 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 $\overline{\chi}_p = \max\{\chi_p(0), \chi_p(1)\}$. Note that $\overline{\chi}_p$ is bounded above by $1-p$. Similarly, we define $\underline{m}$ as a non-silent message that is checked with the probability $\underline{\chi}_p = \min\{\chi_p(0), \chi_p(1)\}$.[21] We say that a fact-checking policy χ_p is *more aggressive* than χ'_p if $\overline{\chi}_p > \overline{\chi}'_p$.[22] 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,*

[20] 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.

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

[22] 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 $\chi_p(0) \geq \chi'_p(0)$ and $\chi_p(1) \geq \chi'_p(1)$, with at least one strict inequality.

- *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_p-equilibria as a correspondence with a single input $\overline{\chi}_p$. In SUE, the payoffs U_S and U_R are unique for all fact-checking policies with the same $\overline{\chi}_p$. In SFE, this is no longer the case. Still we can characterize the bounds of the payoff range with $\overline{\chi}_p$ 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 thereby increasing the ex ante sender's payoff. In SFE, 0-sender's claims can be checked more frequently. However, SFE allows for a χ_p-equilibrium, in which 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_p-equilibria for any fact-checking policy χ_p. Here to eliminate the consideration of multiple cases, suppose that $\overline{\chi}_p > \underline{\chi}_p > 0$ for the sake of clarity. 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 probability of at least $\sigma(\overline{m}|0) \geq \frac{\mu}{1-\mu} \cdot \frac{1-\omega}{\omega}$, 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 $\underline{m}$. These messages reveal 0-sender. However, this additional information does not affect the receiver's payoff, since her optimal action stays unchanged.

Table 2 presents potential supports of sender's equilibrium strategies in SFE. There are three equilibrium patterns 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

Table 1: The support of sender's equilibrium strategy $\sigma(m|\theta)$ in the sender-unfavorable environment.

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

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.

Table 2: Potential supports of sender's equilibrium strategy $\sigma(m|\theta)$ in the sender-favorable environment.

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

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

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

The equilibrium pattern is unique in SUE in the sense that the strategy of one of θ-senders is fixed across χ_p-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. The inclusion of messages m_s and $\underline{m}$ in the strategy of 1-sender generates additional equilibrium patterns producing multiplicity.

The characterization of χ_p-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 $U_S = \mu\overline{\chi}_p$ and $U_R = \mu(1-\omega)\overline{\chi}_p$. 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 $\overline{\chi}_p$ 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 $(1-\mu)\chi_p(m)$, making the receiver change her optimal action to $a = 0$. Hence, the sender's payoff is $U_S = 1 - (1-\mu)\chi_p(m)$. The receiver's payoff is $U_R = \mu - \omega + (1-\mu)\omega\chi_p(m)$, 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 $\overline{\chi}_p$ only:

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

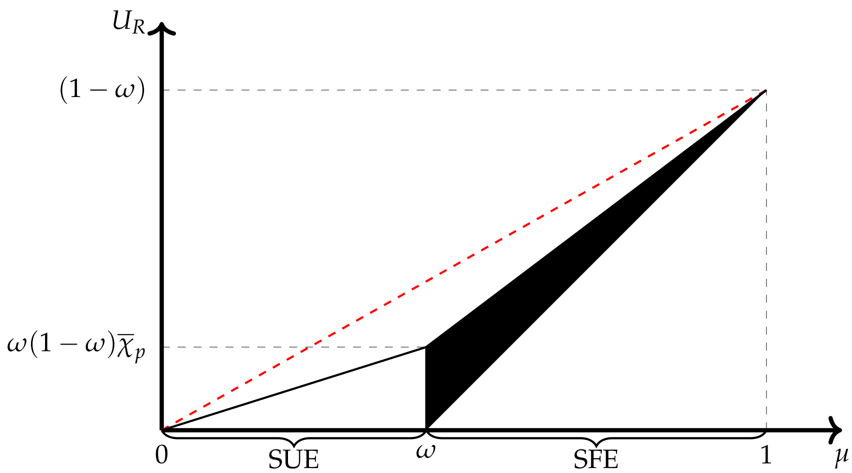

Figure 1: Feasible U_R depending on prior μ for fact-checking policies with fixed $\overline{\chi}_p$. The dashed red line corresponds to the receiver's payoff under complete information.

Figure 1 provides an illustration of the part of Proposition 2 on the receiver's payoff. The receiver is better off when the fact-checking policy is more aggressive as it sustains more informative communication. The receiver's payoff under complete information is attainable only when $\overline{\chi}_p$ approaches one, which can be achieved by the full fact-checking policy and only when p approaches zero, that is, the fact-checking technology is perfect.

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_p-equilibrium for the fact-checker.

The optimal fact-checking policy is characterized by a cost threshold. 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 $\bar{c}(u_F) > 0$, such that $\overline{\chi}_p = 0$ is optimal for $c > \bar{c}(u_F)$ and $\overline{\chi}_p = 1 - p$ is optimal for $c < \bar{c}(u_F)$. Furthermore,*

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

Intuitively, when the fact-checking cost is too high, the no fact-checking policy is optimal. The fact-checker is also more likely to do no fact-checking, when the initiated fact-checks are less likely to produce a check, that is, p increases. When p approaches one, the cost threshold goes to zero, since the fact-checking technology that always fails is not worth to use for any fact-checker.

Proposition 3 tells us that if the fact-checking cost becomes sufficiently low, then the full fact-checking policy becomes optimal.[23] Following our characterization of χ_p-equilibria, the joint distribution of decisions and issues $\lambda(a,\theta|\varepsilon,\chi_p)$ can be summarized by the maximal probability of fact-checking $\overline{\chi}_p$ for any χ_p-equilibrium ε. We can then find the minimal cost of fact-checking that supports distribution $\lambda(a,\theta|\varepsilon,\chi_p)$ as a function of $\overline{\chi}_p$. We show that

[23] Note that picking $\underline{\chi}_p = 1 - p$ is not necessary for optimality. However, when there are multiple equilibrium patterns, the fact-checker is able to steer players toward the preferred χ_p-equilibrium in which $\underline{m}$ is never played.

the fact-checker's benefit $\sum_{a,\theta} u_F(a,\theta)\lambda(a,\theta|\varepsilon,\chi_p)$ and the minimal cost of fact-checking are linear functions of $\overline{\chi}_p$ in the interior. 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 $\lambda(a,\theta'|\varepsilon,\chi_p)$ is fixed for issue θ' over all fact-checking policies χ_p and χ_p-equilibria. Indeed, $U_S(\theta')$ can be written as $\lambda(a = A,\theta'|\varepsilon,\chi_p)$ in χ_p-equilibrium ε. Therefore, different fact-checking policies can only affect the fact-checker's payoff in issue θ.

When $\bar{c}(u_F) \leq 0$, the no fact-checking policy is always optimal for the fact-checker with preferences u_F. The fact-checker that prefers $a = R$ when the issue $\theta = 1$ never fact-checks in SUE. Similarly, the fact-checker that prefers $a = A$ 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. Moreover, $\bar{c}(u_F)$ goes to zero when μ approaches one. This follows from the set of χ_p-equilibria available to the fact-checker depending on the environment. In SUE, distribution $\lambda(a,\theta|\varepsilon,\chi_p)$ is uniquely pinned down by $\overline{\chi}_p$. The question is what χ_p-equilibrium for a fact-checking policy with $\overline{\chi}_p$ is associated with the minimal cost of fact-checking. The answer to this question is χ_p-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 $\bar{c}(u_F)$ is independent of the prior. In SFE, the minimal cost of implementing any equilibrium pattern from Table 2 is achieved by implementing χ_p-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_p-equilibrium results in more fact-checking without changing the distribution of decision and issues. The fact-checker that desires to implement a more aggressive fact-checking policy has to pay a cost in the size of $\frac{c\overline{\chi}_p}{1-p}$, 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.

Proposition 3 allows us to describe receiver's preferences over settings with different fact-checker's payoffs 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: $u_F(a,\theta) = \beta_S u_S(a) + \beta_R u_R(a,\theta) = \beta_S a + \beta_R 1\{a = A\}(\theta - \omega)$. 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 following corollary 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 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 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, then 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.

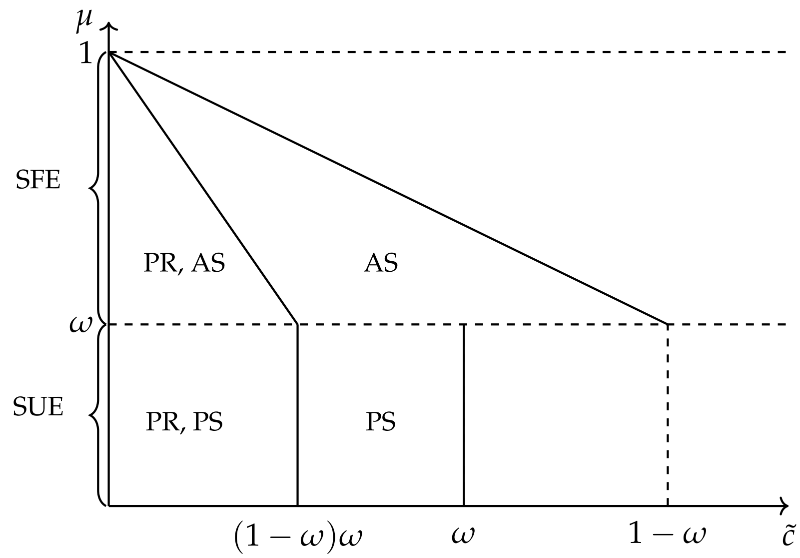

Figure 2: This figure shows the regions in the $(\tilde{c}, \mu)$ space, where $\tilde{c} = \frac{c}{1-p}$, for fixed $\omega < \frac{1}{2}$, where pro-receiver (PR), pro-sender (PS), and anti-sender (AS) fact-checkers choose the full fact-checking policy.

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 more often.[24] 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 steer the sender and the receiver toward its favorite χ_p-equilibrium. Suppose instead that the

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

worst χ_p-equilibrium for the fact-checker is played out by the sender and the receiver after it chooses a fact-checking policy χ_p. In SUE, we know that the distribution of decisions and issues $\lambda(a,\theta|\chi_p,\varepsilon)$ is uniquely pinned down by the maximal probability of fact-checking $\overline{\chi}_p$ in fact-checking policy χ_p. Thus, the fact-checker's expected benefit does not depend on the selection of a specific χ_p-equilibrium ε. The worst-equilibrium selection can only drive up the minimal cost of fact-checking by selecting χ_p-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 $\mu(1-p)[u_F(A,1)-u_F(R,1)]$. 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, since SFE permits silence by both 0-sender and 1-sender as an equilibrium, the fact-checker cannot do better than the no fact-checking policy.

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 $u_{F,2}$.[25] At the beginning of the game, each fact-checker decides on the fact-checking policies, $\chi_{p,1},\chi_{p,2}\in[0,1-p]^2$. Note that the probability of message m checked is

[25]When there are more than two fact-checkers, the equilibrium structure remains qualitatively the same.

$\chi_p(m) := 1 - (1 - \chi_{p,1}(m))(1 - \chi_{p,2}(m))$. Then the game unfolds as in our baseline model. The sender observes the issue $\theta \in \{0,1\}$ and sends message m. The receiver sees sender's message $m \in \{0,1,m_s\}$ and realized fact-check outputs $\mathcal{O}_1, \mathcal{O}_2 \in \{0,1,\varnothing\}$.[26] Based on the observed message and fact-check outputs, the receiver makes the decision a.

The fact-checking policies chosen by fact-checkers generate probabilities $\chi_p(m)$ of each non-silent message m checked. Then the game continues with a one of χ_p-equilibria, which we already conveniently characterized in Section 3 with $\overline{\chi}_p = \max\{\chi_p(0), \chi_p(1)\}$. We can also define $\overline{\chi}_{p,i} = \max\{\chi_{p,i}(0), \chi_{p,i}(1)\}$ as before. To make predictions about the fact-checkers' choice of $\chi_{p,1}$ and $\chi_{p,2}$, we need to make a stance on the selection of χ_p-equilibria. We make two assumptions. First, we assume that if there are two available χ_p-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 be played.[27] Second, in SFE, we assume that 1-sender sends only the most checked message. In other words, we assume that the most informative χ_p-equilibrium is played.[28] These assumptions guarantee that after fact-checkers choose their fact-checking policies, they know that the game will continue in accordance with a specific χ_p-equilibrium. If fact-checkers select their fact-checking policies $\chi_{p,1}$ and $\chi_{p,2}$ by best responding to each other, then we call $\chi_{p,1}$ and $\chi_{p,2}$ equilibrium fact-checking policies. Equilibrium fact-checking policies and succeeding χ_p-equilibrium constitute an equilibrium. In what follows, we characterize equilibrium fact-checking policies.

Suppose $\overline{c}(\cdot)$ as given by Proposition 3 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 optimal policy involves no fact-checking by fact-checker i. Indeed, if fact-checker i does not want to

[26]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.

[27]That is, we assume that the chosen χ_p-equilibrium has to be Pareto-undominated for fact-checkers. We view this requirement as a logical extension of the best-equilibrium selection in the case of one fact-checker.

[28]If we allow for a small fine for the sender that is caught in a lie, this extension would select the most informative equilibrium pattern. Interestingly, 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.

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 $\bar{c}(u_{F,i}) < 0$ or $c \geq \bar{c}(u_{F,i})$ 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* $c < \bar{c}(u_{F,i})$ *for both fact-checkers. In the equilibrium:*

- *if* $c < p\bar{c}(u_{F,i})$ *for both fact-checkers, then* $\overline{\chi}_{p,1} = \overline{\chi}_{p,2} = 1 - p$*;*
- *if* $c < p\bar{c}(u_{F,i})$ *and* $c > p\bar{c}(u_{F,j})$, $j \neq i$, *then* $\overline{\chi}_{p,i} = 1 - p$ *and* $\overline{\chi}_{p,j} = 0$*;*
- *if* $c > p\bar{c}(u_{F,i})$ *for both fact-checkers, then there are three equilibria: (1)* $\overline{\chi}_{p,1} = 1 - p$, $\overline{\chi}_{p,2} = 0$, *(2)* $\overline{\chi}_{p,1} = 0$, $\overline{\chi}_{p,2} = 1 - p$, *and (3)* $\overline{\chi}_{p,i} = 1 - \frac{c}{\bar{c}(u_{F,j})}$, $j \neq i$.

Importantly, when $\overline{\chi}_{p,i} > 0$ for both fact-checkers in the equilibrium, they check the same non-silent message with their own maximal probability to save on fact-checking cost. Proposition 4 holds for both SUE and SFE, with cost threshold $\bar{c}(\cdot)$ given by Proposition 3. If the fact-checking cost is low enough, then both fact-checkers select the full fact-checking policy, thereby increasing the maximal probability of fact-checking $\overline{\chi}_p$ 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 benefiting the receiver. 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 $\overline{\chi}_p$ at $1 - p$, as if there is only one fact-checker present.[29] In this case an

[29] In a different setting, Carletti, Cerasi, and Daltung (2007) examine a bank's choice between lending to

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 cost is intermediate, 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 $\overline{\chi}_p = 1 - \frac{c}{\overline{c}(u_{F,1})} \cdot \frac{c}{\overline{c}(u_{F,2})}$. 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 intermediate, $c > \sqrt{p}\sqrt{\overline{c}(u_{F,1})\overline{c}(u_{F,2})}$, both fact-checkers want to implement the full fact-checking policy by themselves, but the composite fact-checking policy is less aggressive, $\overline{\chi}_p < 1 - p$. 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.

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.[30] Moreover, our setting does not allow for repeated checks in case of the technology failure. We view both of these restrictions as reflecting time-pressure conditions of real-world competition between fact-checking organizations. As pointed out by Graves (2016), "editors at FactCheck.org have remarked several times on the sharper deadline pressure the group faced once its national rivals appeared". FactCheck.org responded to new market conditions by introducing the "FactCheck Wire" in 2009 with a purpose to deliver shorter fact-checks in a timely manner.

6 Discussion

This section considers two variations of our baseline model that allows us to discuss facts that can be checked perfectly and are not binary in nature.

---

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.

[30]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.

Perfect fact-checking technology

When the fact-checking technology is perfect, $p = 0$, it is possible to have a message checked with probability one, that is, $\overline{\chi}_0 = 1$. If $p = 0$ and $\overline{\chi}_0 = 1$, 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_0-equilibrium *completely separating*. The completely separating equilibrium reveals the issue, while only the claim made by 1-sender gets fact-checked. The following proposition shows that the optimal policy is still a threshold policy that utilizes the availability of separation at a lower minimal fact-checking cost.

Proposition 5. *Suppose that $p = 0$. For the fact-checker with preferences $u_F(a,\theta)$, there exists $\overline{c}(u_F) > 0$, such that $\overline{\chi}_0 = 0$ is optimal for $c > \overline{c}(u_F)$ and $\overline{\chi}_0 = 1$ is optimal for $c < \overline{c}(u_F)$. Furthermore,*

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

It is evident that the cost threshold $\overline{c}(u_F)$ is discontinuous at $p = 0$. The reason behind this result is a discontinuity of the minimal cost of fact-checking: it is more costly to detect pooling than sustain separation by making use of the silent message. The completely separating equilibrium is preferred by the fact-checker that wishes to fact-check fully. 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.

In the setting with multiple fact-checkers, 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 linear in $\overline{\chi}_0$ for $\overline{\chi}_0 \in [0,1)$ and subject to a downward jump at $\overline{\chi}_0 = 1$, since the completely separating equilibrium becomes available.

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.[31] 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.[32] For such state space, Titova (2021) shows that the sender can achieve the commitment outcome in SUE with verifiable information only.[33] 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 $a = A$ and $a = R$ 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 $a = A$ and $a = R$ upon observing m_l and the tie is broken against the sender.

[31]For example, Donald Trump famously spread information about US unemployment rates that received negative fact-checking ratings (National Public Radio, 2017).

[32]The variation of this setting is analyzed in Balbuzanov (2019).

[33]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, HUNT and MATTHEW GENTZKOW (2017), “Social Media and Fake News in the 2016 Election”, *Journal of Economic Perspectives*, 31, 2, pp. 211-36.

AMAZEEN, MICHELLE A. (2015), “Revisiting the Epistemology of Fact-Checking”, *Critical Review*, 27, 1, pp. 1-22.

— (2016), “Checking the Fact-Checkers in 2008: Predicting Political Ad Scrutiny and Assessing Consistency”, *Journal of Political Marketing*, 15, 4, pp. 433-464.

AMBRUS, ATTILA, EDUARDO M. AZEVEDO, and YUICHIRO KAMADA (2013), “Hierarchical Cheap Talk”, *Theoretical Economics*, 8, 1, pp. 233-261.

BALBUZANOV, IVAN (2019), “Lies and Consequences”, *International Journal of Game Theory*, 48, 4, pp. 1203-1240.

BARON, DAVID P. and DAVID BESANKO (1984), “Regulation, Asymmetric Information, and Auditing”, *The RAND Journal of Economics*, pp. 447-470.

BARRERA, OSCAR, SERGEI GURIEV, EMERIC HENRY, and EKATERINA ZHURAVSKAYA (2020), “Facts, Alternative Facts, and Fact Checking in Times of Post-Truth Politics”, *Journal of Public Economics*, 182, p. 104123.

BORDER, KIM C. and JOEL SOBEL (1987), “Samurai Accountant: A Theory of Auditing and Plunder”, *The Review of Economic Studies*, 54, 4, pp. 525-540.

CARLETTI, ELENA, VITTORIA CERASI, and SONJA DALTUNG (2007), “Multiple-Bank Lending: Diversification and Free-Riding in Monitoring”, *Journal of Financial Intermediation*, 16, 3, pp. 425-451.

CRAWFORD, VINCENT P. and JOEL SOBEL (1982), “Strategic Information Transmission”, *Econometrica*, 50, 6, p. 1431.

DZIUDA, WIOLETTA and CHRISTIAN SALAS (2018), “Communication with Detectable Deceit”, *Working Paper*.

EDERER, FLORIAN and WEICHENG MIN (2022), “Bayesian Persuasion with Lie Detection”, *Working Paper*.

FACEBOOK (2021), *How Facebook’s Third-Party Fact-Checking Program Works*, `https://www.facebook.com/journalismproject/programs/third-party-fact-checking/how-it-works`, Last accessed on 2021-12-30.

GRAVES, LUCAS (2016), *Deciding What's True: The Rise of Political Fact-Checking in American Journalism*, Columbia University Press, p. 336.

— (2017), "Anatomy of a Fact Check: Objective Practice and the Contested Epistemology of Fact Checking", *Communication, Culture & Critique*, 10, 3, pp. 518-537.

— (2018), "Understanding the Promise and Limits of Automated Fact-Checking", *Reuters Institute for the Study of Journalism*.

GRAVES, LUCAS and FEDERICA CHERUBINI (2016), "The Rise of Fact-Checking Sites in Europe", *Reuters Institute for the Study of Journalism*.

GUO, YINGNI and ERAN SHMAYA (2021), "Costly Miscalibration", *Theoretical Economics*, 16, 2, pp. 477-506.

HASSAN, NAEEMUL, FATMA ARSLAN, CHENGKAI LI, and MARK TREMAYNE (2017), "Toward Automated Fact-Checking: Detecting Check-Worthy Factual Claims by Claimbuster", *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1803-1812.

HOLM, HÅKAN J. (2010), "Truth and Lie Detection in Bluffing", *Journal of Economic Behavior & Organization*, 76, 2, pp. 318-324.

IVANOV, MAXIM (2010), "Communication via a Strategic Mediator", *Journal of Economic Theory*, 145, 2, pp. 869-884.

JARMAN, JEFFREY W. (2016), "Influence of Political Affiliation and Criticism on the Effectiveness of Political Fact-Checking", *Communication Research Reports*, 33, 1, pp. 9-15.

KAMENICA, EMIR and MATTHEW GENTZKOW (2011), "Bayesian Persuasion", *American Economic Review*, 101, 6 (Oct. 2011), pp. 2590-2615.

KOLOTILIN, ANTON, TYMOFIY MYLOVANOV, ANDRIY ZAPECHELNYUK, and MING LI (2017), "Persuasion of a Privately Informed Receiver", *Econometrica*, 85, 6, pp. 1949-1964.

LAFFONT, JEAN-JACQUES and JEAN TIROLE (1986), "Using Cost Observation to Regulate Firms", *Journal of political Economy*, 94, 3, Part 1, pp. 614-641.

LIM, CHLOE (2018), "Checking how Fact-Checkers Check", *Research & Politics*, 5, 3.

MARIETTA, MORGAN, DAVID C. BARKER, and TODD BOWSER (2015), "Fact-checking Polarized Politics: Does the Fact-check Industry Provide Consistent Guidance on Disputed Realities?", *The Forum*, 4, De Gruyter, vol. 13, pp. 577-596.

MATHEVET, LAURENT, JACOPO PEREGO, and INA TANEVA (2020), "On Information Design in Games", *Journal of Political Economy*, 128, 4, pp. 1370-1404.

MOOKHERJEE, DILIP and IVAN PNG (1989), "Optimal Auditing, Insurance, and Redistribution", *The Quarterly Journal of Economics*, 104, 2, pp. 399-415.

NATIONAL PUBLIC RADIO (2017), *Ahead Of Trump's First Jobs Report, A Look At His Remarks On The Numbers*, https://www.npr.org/2017/01/29/511493685/ahead-of-trumps-first-jobs-report-a-look-at-his-remarks-on-the-numbers, Last accessed on 2021-12-30.

NIEMINEN, SAKARI and LAURI RAPELI (2019), "Fighting Misperceptions and Doubting Journalists' Objectivity: A Review of Fact-Checking Literature", *Political Studies Review*, 17, 3, pp. 296-309.

NYHAN, BRENDAN, ETHAN PORTER, JASON REIFLER, and THOMAS WOOD (2020), "Taking Fact-checks Literally but not Seriously? The Effects of Journalistic Fact-Checking on Factual Beliefs and Candidate Favorability", *Political Behavior*, 42, 3, pp. 939-960.

NYHAN, BRENDAN and JASON REIFLER (2010), "When Corrections Fail: The Persistence of Political Misperceptions", *Political Behavior*, 32, 2, pp. 303-330.

— (2015), "The Effect of Fact-checking on Elites: A Field Experiment on US State Legislators", *American Journal of Political Science*, 59, 3, pp. 628-640.

OSTERMEIER, ERIC (2011), "Selection Bias? PolitiFact Rates Republican Statements as False at Three Times the Rate of Democrats", *Smart Politics*, 10.

REUTERS (2021), *Twitter Partners with AP, Reuters to Battle Misinformation on Its Site*, https://www.reuters.com/technology/twitter-partners-with-ap-reuters-battle-misinformation-its-site-2021-08-02/, Last accessed on 2021-12-30.

SALAMANCA, ANDRÉS (2021), "The Value of Mediated Communication", *Journal of Economic Theory*, 192, p. 105191.

SCIENTIFIC AMERICAN (2020), *The Psychology of Fact-Checking*, https://www.scientificamerican.com/article/the-psychology-of-fact-checking1/, Last accessed on 2021-12-30.

SHISHKIN, DENIS (2021), "Evidence Acquisition and Voluntary Disclosure", *Proceedings of the 22nd ACM Conference on Economics and Computation*, pp. 817-818.

TITOVA, MARIA (2021), "Persuasion with Verifiable Information", *Working Paper*.

TOWNSEND, ROBERT M. (1979), "Optimal Contracts and Competitive Markets with Costly State Verification", *Journal of Economic theory*, 21, 2, pp. 265-293.

USCINSKI, JOSEPH (2015), "The Epistemology of Fact Checking (is Still Naive): Rejoinder to Amazeen", *Critical Review*, 27, 2, pp. 243-252.

USCINSKI, JOSEPH and RYDEN BUTLER (2013), "The Epistemology of Fact Checking", *Critical Review*, 25, 2, pp. 162-180.

WALTER, NATHAN, JONATHAN COHEN, R. LANCE HOLBERT, and YASMIN MORAG (2020), "Fact-Checking: A Meta-Analysis of What Works and for Whom", *Political Communication*, 37, 3, pp. 350-375.

WEEKS, BRIAN 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, pp. 699-719.

WEEKS, BRIAN E. and R. KELLY GARRETT (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, pp. 401-422.

WINTERSIECK, AMANDA, KIM FRIDKIN, and PATRICK KENNEY (2021), "The Message Matters: The Influence of Fact-Checking on Evaluations of Political Messages", *Journal of Political Marketing*, 20, 2, pp. 93-120.

Appendix

All the proofs provided cover both the cases of the imperfect, $p > 0$, and perfect, $p = 0$, fact-checking technologies.

Proof of Proposition 1

We start by showing that $U_S(0) = 0$ for every χ_p and χ_p-equilibrium in SUE. Fix the environment (μ, ω), such that $\mu < \omega$. Fix a fact-checking policy χ_p and χ_p-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 $\sigma(m|0) > 0$, $\chi_p(m) < 1$, 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 $\chi_p(m') < 1$ 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_p and χ_p-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_p and χ_p-equilibrium (σ, α, π), such that $U_S(1) > 1 - p$. Since the probability of any message m checked $\chi_p(m)$ 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_p and a χ_p-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_p, with $\chi_p(1) \geq \chi_p(0)$. 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 $\chi_p(1)$. The payoff of 1-sender in the constructed equilibrium is then $\chi_p(1)$. Therefore, by controlling $\chi_p(1) \in [0, 1-p]$ and respecting the inequality $\chi_p(1) \geq \chi_p(0)$, 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_p and χ_p-equilibrium in SFE. Fix the environment (μ, ω), such that $\mu > \omega$. Fix a fact-checking policy χ_p and χ_p-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 $\sigma(m|1) > 0$, $\chi_p(m) < 1$, 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_p and χ_p-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_p and χ_p-equilibrium (σ, α, π). We know that $U_S(1) = 1$. Thus, there exists an on-path message $m \in \mathcal{M}$ that satisfies $\sigma(m|1) > 0$, $\chi_p(m) < 1$, and $\alpha(m, \varnothing) = 1$. Then 0-sender can always guarantee himself at least a payoff of $1 - \chi_p(m)$ by playing $\sigma(m|0) = 1$. Since $\chi_p(m) \leq 1 - p$, we have $U_S(0) \geq p$.

Finally, we construct a fact-checking policy χ_p and a χ_p-equilibrum (σ, α, π) that delivers a payoff in the $[p,1]$ interval to 0-sender in SFE. Fix a fact-checking policy χ_p 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 $1 - \chi_p(1) \in [p,1]$. Therefore, by controlling $\chi_p(1)$, we can produce any $U_S(0) \in [p,1]$.

Proof of Proposition 2

Fix χ_p. Let $\overline{\chi}_p = \max\{\chi_p(1), \chi_p(0)\}$ and $\underline{\chi}_p = \min\{\chi_p(1), \chi_p(0)\}$. Let $\overline{m}$ ($\underline{m}$) denote a non-silent message that is checked with probability $\overline{\chi}_p$ ($\underline{\chi}_p$). If $\chi_p(1) = \chi_p(0)$, messages $m = 0$ and $m = 1$ can be assigned to $\overline{m}$ and $\underline{m}$ in an arbitrary way.

We start by characterizing χ_p-equilibria in SUE. By Proposition 1, we know that $U_S(0) = 0$. This implies that for any on-path message m, we have $\chi_p(m) = 1$ or $\alpha(m,\varnothing) = 0$. If $\overline{\chi}_p > 0$ and $\overline{\chi}_p > \underline{\chi}_p$, then the optimal behavior for 1-sender prescribes $\sigma(\overline{m}|1) = 1$. If $\overline{\chi}_p = 1$, 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 $\overline{\chi}_p < 1$, 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 $\overline{\chi}_p = \underline{\chi}_p > 0$, 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 $\overline{\chi}_p = 1$, 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 $\overline{\chi}_p < 1$, 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 $\overline{\chi}_p = 0$, 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_p-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 $\chi_p(m) = 1$ 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 $m \in \arg\min \chi_p(\cdot)$.

Suppose $\sigma(m_s|1) > 0$. First, consider $\underline{\chi}_p > 0$. 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 $\overline{\chi}_p > \underline{\chi}_p = 0$. 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 $\overline{\chi}_p = 0$. 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 $\underline{\chi}_p = 1$. Then any $\sigma(\cdot|0)$ is an equilibrium strategy of 0-sender, since any strategy brings him the payoff of zero. Now suppose that $\underline{\chi}_p \in [0,1)$ and $\overline{\chi}_p > \underline{\chi}_p$. 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 $\overline{\chi}_p = \underline{\chi}_p \in [0,1)$, 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 $\overline{\chi}_p = 1$, then any $\sigma(\cdot|0)$ is an equilibrium strategy of 0-sender, since any strategy brings him the payoff of zero. If $\overline{\chi}_p < 1$, 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_p-equilibria, since we exhausted all possibilities.

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

In SUE, $U_S(1) = \overline{\chi}_p$, 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 $U_S = \mu\overline{\chi}_p$. The receiver's payoff is $U_R = \mu(1-\omega)\overline{\chi}_p$. 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 no-communication payoff $U_R = \mu - \omega$. Instead, suppose that the support of $\sigma(\cdot|1)$ does not include m_s but includes $\underline{m}$ that is checked with probability $\underline{\chi}_p \in [0, \overline{\chi}_p]$. Then the support of $\sigma(\cdot|0)$ only includes messages that are checked with probability $\underline{\chi}_p$. The payoff of 0-sender is $U_S(0) = 1 - \underline{\chi}_p$ and the sender's ex ante payoff is $U_S = 1 - (1-\mu)\underline{\chi}_p$. The receiver's payoff is $U_R = \mu(1-\omega) + (1-\mu)(1-\underline{\chi}_p)(-\omega) = \mu - \omega + (1-\mu)\omega\underline{\chi}_p$. Finally, suppose that $\sigma(\overline{m}|1) = 1$. Then either $\overline{\chi}_p = 1$ or 0-sender pools on $\overline{m}$, $\sigma(\overline{m}|0) = 1$. In either case, the payoff of 0-sender can be summarized by $U_S(0) = 1 - \overline{\chi}_p$. The sender's ex ante payoff is $U_S = 1 - (1-\mu)\overline{\chi}_p$. A similar calculation as above demonstrates that $U_R = \mu - \omega + (1-\mu)\omega\overline{\chi}_p$.

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

Proof of Proposition 3 and 5

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

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

For a fact-checking policy with fixed $\overline{\chi}_p$, the cheapest equilibrium to implement for the fact-checker depends on whether $\overline{\chi}_p = 1$ or $\overline{\chi}_p = 0$. If $\overline{\chi}_p = 1$, then an equilibrium that is associated with the minimal cost of fact-checking has $\sigma(\overline{m}|1) = 1$ and $\sigma(m_s|0) = 1$.

Table 3: The joint distribution of actions and issues in the sender-unfavorable environment for a fixed fact-checking policy χ_p for any χ_p-equilibrium ε.

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

Indeed, condition $\sigma(\overline{m}|1)=1$ has to hold. Thus, if 0-sender is never checked, then the fact-checker minimizes cost of fact-checking. If $\overline{\chi}_p<1$, then $\sigma(m_s|0)=1$ is not available anymore. Indeed, in any χ_p-equilibrium, we need to have $\alpha(m,\varnothing)=0$ for any m that is checked with probability $\overline{\chi}_p$. 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 total probability of initiating a fact-check is then $\frac{\mu\chi(\overline{m})}{\omega}$. Since $\overline{\chi}_p=(1-p)\chi(\overline{m})$, the implied minimal cost of implementing a fact-checking policy with $\overline{\chi}_p$ is

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

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

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

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

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

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

Table 4: The joint distribution of actions and issues in the sender-favorable environment for a fixed fact-checking policy χ_p for any χ_p-equilibrium ε.

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

We now fix g and discuss an equilibrium that minimizes the cost of fact-checking for fixed fact-checking policy with $\overline{\chi}_p$. When $\overline{\chi}_p = 1$ 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 $g(\overline{\chi}_p) = 0$, 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 $g(\overline{\chi}_p) = \underline{\chi}_p \in (0, \overline{\chi}_p)$, 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 $g(\overline{\chi}_p) = \overline{\chi}_p$ and $\overline{\chi}_p < 1$, both 0-sender and 1-sender send only messages that are checked with probability $\overline{\chi}_p$. We conclude that the minimal cost of implementing a fact-checking policy with $\overline{\chi}_p$ is

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

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

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

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

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

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

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

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

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

This solution can be achieved by choosing $g(\chi_p) = \chi_p$ for all χ_p. Note that $g(\overline{\chi}_p) = 1$ 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_p-equilibrium guarantees that for any χ_p, 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 $\overline{c}(u_F) = \omega(1-p)(u_F(1,1) - u_F(0,1))$ by Proposition 3. Fix the strategy of the second fact-checker $\chi_{p,2}$. Note that $\overline{\chi}_p$ is bounded below by $\overline{\chi}_{p,2} := \max\{\chi_{p,2}(1), \chi_{p,2}(0)\}$. Then the cheapest way to generate $\overline{\chi}_p \in [\overline{\chi}_{p,2}, 1-p^2]$ is to check the message $m \in \arg\max \overline{\chi}_{p,2}(\cdot)$ with probability $\overline{\chi}_{p,1} = \frac{\overline{\chi}_p - \overline{\chi}_{p,2}}{1-\overline{\chi}_{p,2}}$. The problem of the first fact-checker is

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

subject to $\overline{\chi}_p = 1 - (1-\overline{\chi}_{p,1})(1-\overline{\chi}_{p,2})$. 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

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

Note that if $c < p\overline{c}(u_{F,1})$, then $\overline{\chi}_{p,1}(\cdot) = 1-p$ 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 $\overline{\chi}_{p,2}(\cdot) = 0$. Otherwise,

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

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 $\overline{\chi}_{p,1} = \overline{\chi}_{p,2} = 0$ 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 $\overline{\chi}_{p,i} = 0$ and $\overline{\chi}_{p,j} = 1-p$. 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 $\overline{\chi}_{p,1} = \overline{\chi}_{p,2} = 1-p$. If $c < p\overline{c}(u_{F,i})$ is true for one $i \in \{1,2\}$, but not for $j \neq i$, then $\overline{\chi}_{p,i} = 1-p$ and $\overline{\chi}_{p,j} = 0$ (when $c = p\overline{c}(u_{F,j})$, $\overline{\chi}_{p,j} \in [0, 1-p]$). Finally, suppose that $c < p\overline{c}(u_{F,i})$ is false for both $i \in \{1,2\}$. Then there are three equilibria: (1) $\overline{\chi}_{p,1} = 0, \overline{\chi}_{p,2} = 1-p$; (2) $\overline{\chi}_{p,1} = 1-p, \overline{\chi}_{p,2} = 0$; (3) $\overline{\chi}_{p,1} = 1 - \frac{c}{\overline{c}(u_{F,2})}$, $\overline{\chi}_{p,2} = 1 - \frac{c}{\overline{c}(u_{F,1})}$.

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 $\overline{\chi}_{p,2} = 1$, the best response for the first fact-checker is $\overline{\chi}_{p,1} = 0$. As before, the first fact-checker can generate $\overline{\chi}_p \in [\overline{\chi}_{p,2}, 1)$ by checking message $m \in \arg\max \overline{\chi}_{p,2}(\cdot)$ with probability $\overline{\chi}_{p,1} = \frac{\overline{\chi}_p - \overline{\chi}_{p,2}}{1 - \overline{\chi}_{p,2}} \in [0, 1)$. The cost of doing so is $C_{\text{SUE}}(\overline{\chi}_{p,1}) = \frac{\mu \overline{\chi}_{p,1}}{\omega} \cdot c$. Alternatively, the fact-checker can generate $\overline{\chi}_p = 1$ by selecting $\overline{\chi}_{p,1} = 1$ at a cost of μc. Note that if $\overline{\chi}_{p,1} > \omega$, then the latter option is cheaper. The problem of the first fact-checker is find a maximum between

$$\sup_{\overline{\chi}_{p,1} \in [0,1)} \left\{ \mu \overline{\chi}_p (u_{F,1}(1,1) - u_{F,1}(0,1)) - C_{\text{SUE}}(\overline{\chi}_{p,1}) \right\}$$

and

$$\mu(u_{F,1}(1,1) - u_{F,1}(0,1)) - \mu c,$$

subject to $\overline{\chi}_p = 1 - (1 - \overline{\chi}_{p,1})(1 - \overline{\chi}_{p,2})$. There cannot be an interior solution. Indeed, the objective in the inner problem is linear in $\overline{\chi}_{p,1}$. Thus, the supremum is achieved on either $\overline{\chi}_{p,1} = 0$ or $\overline{\chi}_{p,1} = 1$. If the supremum is achieved on $\overline{\chi}_{p,1} = 1$, 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

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

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 $\overline{\chi}_{p,2} = 0$. Otherwise, the best response of the second fact-checker is

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

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 $\overline{\chi}_{p,1} = \overline{\chi}_{p,2} = 0$ 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 $\overline{\chi}_{p,i} = 0$ and $\overline{\chi}_{p,j} = 1$. 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) $\overline{\chi}_{p,1} = 0$, $\overline{\chi}_{p,2} = 1$; (2) $\overline{\chi}_{p,1} = 1$, $\overline{\chi}_{p,2} = 0$.

Consider SFE. In this case, the cost threshold in the case of one fact-checker is given by $\bar{c}(u_F) = (1-\mu)(1-p)(u_F(0,0) - u_F(1,0))$ by Proposition 3. When $p > 0$, in any χ_p-equilibrium, $\sigma(\overline{m}|1) = \sigma(\overline{m}|0) = 1$. When $p = 0$ and $\overline{\chi}_p = 1$, 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 $\overline{\chi}_p$ is bounded below by $\overline{\chi}_{p,2} := \max\{\chi_{p,2}(1), \chi_{p,2}(0)\}$. To generate $\overline{\chi}_p \in [\overline{\chi}_{p,2}, 1-p^2]$, the first fact-checker checks the message $m \in \arg\max \chi_{p,2}(\cdot)$ with probability $\overline{\chi}_{p,1} = \frac{\overline{\chi}_p - \overline{\chi}_{p,2}}{1-\overline{\chi}_{p,2}}$. The problem of the first fact-checker is

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

subject to $\overline{\chi}_p = 1 - (1-\overline{\chi}_{p,1})(1-\overline{\chi}_{p,2})$, where $g(\cdot)$ is defined as follows. Let $g(\overline{\chi}_p) \in [0, \overline{\chi}_p]$ be a function that tracks what type of χ_p-equilibrium is played. Specifically, when $g(\overline{\chi}_p) = 0$, 1-sender's strategy has a message checked with zero probability in its support. When $g(\overline{\chi}_p) = \underline{\chi}_p \in (0, \overline{\chi}_p)$, 1-sender's strategy has a message checked with probability $\underline{\chi}_p$ in its support and $\sigma(m_s|1) = 0$. Finally, when $g(\overline{\chi}_p) = \overline{\chi}_p$, 1-sender's strategy only has messages checked with probability $\overline{\chi}_p$ in its support.

Under our selection, $g(\overline{\chi}_p) = \overline{\chi}_p$, and

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

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

$$\max_{\overline{\chi}_{p,1} \in [0,1-p]} \left\{ \overline{\chi}_{p,1} \left((1-\overline{\chi}_{p,2}) \bar{c}(u_{F,1}) - c \right) \right\}.$$

Then the best responses are the same as in SUE, subject to a changed cost threshold $\bar{c}(\cdot)$.

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

$$\max \left\{ \sup_{\overline{\chi}_{p,1} \in [0,1)} \left\{ \mu \bar{c}(u_{F,1}) \overline{\chi}_p - \frac{\overline{\chi}_{p,1}}{1-p} \cdot c \right\}, \mu \bar{c}(u_{F,1}) - \mu c \right\},$$

subject to $\overline{\chi}_p = 1 - (1-\overline{\chi}_{p,1})(1-\overline{\chi}_{p,2})$. There cannot be an interior solution for the same reason as in the problem in SUE under the perfect fact-checking technology. Then $\overline{\chi}_{p,1} = 1$

is optimal when $\mu\bar{c}(u_{F,1}) - \mu c \geq \mu\bar{c}(u_{F,1})\overline{\chi}_{p,2}$, or $c \leq (1 - \overline{\chi}_{p,2})\bar{c}(u_{F,1})$. When $c \geq (1 - \overline{\chi}_{p,2})\bar{c}(u_{F,1})$, $\overline{\chi}_{p,1} = 0$ is optimal. Then the best responses are the same as in SUE, subject to a changed cost threshold $\bar{c}(\cdot)$. This completes the proof.