UNIVERSITY OF CALIFORNIA SAN DIEGO

Essays on Strategic Mediation of Information

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Economics

by

Aleksandr Levkun

Committee in charge:

Professor Simone Galperti, Co-Chair
Professor Joel Sobel, Co-Chair
Professor Renee Bowen
Professor Songzi Du
Professor Isabel Trevino

2022

The Dissertation of Aleksandr Levkun is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2022

DEDICATION

To my mother.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I am indebted to Renee Bowen, Simone Galperti, and Joel Sobel for their patience, guidance, and unconditional support. I want to thank Renee, Simone, and Joel for invaluable time they put in mentoring me. They were very supportive during the most difficult times of my studies. I feel fortunate to have had the chance to work under supervision of these incredibly smart and wonderful people.

I am grateful to my committee members Songzi Du and Isabel Trevino for their encouragement and feedback. I want to thank Mark Machina, Denis Shishkin, Maria Titova, Joel Watson for engaging and thoughtful discussions. My coauthors Simone Galperti and Jacopo Perego taught me how to present a research paper and what questions and answers are important, and I am thankful for that.

I am grateful to my family and friends for encouraging me to never give up. I want to also thank the UCSD Department of Economics for the funding and hospitality during my graduate studies.

Finally, I would like to thank my mentors at the Higher School of Economics. I am particularly grateful to Boris Demeshev, Maarten Janssen, and Anton Suvorov.

Chapter 1 of this dissertation is currently being prepared for submission for publication of the material. The dissertation author, Aleksandr Levkun, is the sole author of this material. Chapter 2 of this dissertation is currently being prepared for submission for publication of the material. Chapter 2 is coauthored with Galperti, Simone and Perego, Jacopo. The dissertation author, Aleksandr Levkun, is one of primary authors of this chapter. Chapter 3 of this dissertation is currently being prepared for submission for publication of the material. The dissertation author, Aleksandr Levkun, is the sole author of this material.

VITA

2016	Bachelor, Higher School of Economics
2020	Master of Arts, University of California San Diego
2017–2022	Teaching Assistant, Department of Economics University of California San Diego
2022	Doctor of Philosophy, University of California San Diego

ABSTRACT OF THE DISSERTATION

Essays on Strategic Mediation of Information

by

Aleksandr Levkun

Doctor of Philosophy in Economics

University of California San Diego, 2022

Professor Simone Galperti, Co-Chair
Professor Joel Sobel, Co-Chair

This dissertation is a collection of three essays on the topic of strategic mediation of information. A recurring theme is a presence of the information intermediary influencing the interaction between informed and uninformed parties.

Chapter 1 studies communication between an informed sender and an uninformed receiver with a presence of a strategic fact-checker. I show that if the cost of checking is small, the optimal fact-checking policy is full fact-checking; otherwise, no fact-checking is optimal. 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.

Chapter 2 considers an online platform that intermediates trade between sellers and buyers using data records of the buyers' personal characteristics. An important component of the value of a data record for the platform is a novel externality that arises when a platform pools records to withhold information from the sellers. Ignoring this externality can significantly bias our understanding of the value of data records. Chapter 2 then characterizes a platform's willingness to pay for more data, thereby establishing a series of basic properties of the demand side of data markets.

Chapter 3 presents the optimal editorial policy for state-owned media manipulating information flow from a strategic informed elite to an uninformed receiver. I show conditions on players' preferences under which the media meaningfully communicate information on the ruler's competence. An elite that is more aligned with the media benefits the media, as long as the alignment is not too close. The media are worse off when the receiver is more critical of the ruler, whereas the elite generally is better off when the receiver is more critical. I characterize the lower bound on the media's payoff when the receiver has private information about how critical he is.

Chapter 1

Communication with Strategic Fact-checking

1.1 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,

[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).

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 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

[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.

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 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

[7] See also Scientific American (2020).

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 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,

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

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

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

or support for 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.

1.2 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 1.6 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-

[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

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*:

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 $\underset{m\in\mathcal{M}}{\operatorname{argmax}} \left\{\chi_p(m)\alpha(m, 1\{\theta = m\}) + (1 - \chi_p(m))\alpha(m, \varnothing)\right\}$.[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

will take χ_p into account.

[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:

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

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

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]

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).

1.3 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

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

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]

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

[19]Irrespective of the environment, some equilibria can still be informative, with some messages revealing the issue. However, the receiver's payoff is fixed across all χ_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.

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 1.1. *The feasible sender's payoffs are*

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

All proofs are in 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 1.1 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 1.1.

We now relate the result to the best possible communication outcome for the sender. In a setting without the fact-checker but with the sender's commitment power as in Kamenica and Gentzkow (2011), the sender can obtain the ex ante payoff of $\frac{\mu}{\omega}$ in SUE. To achieve this, 1-sender always sends a "winning" message $m_w \in \mathcal{M}$ and 0-sender sends m_w with probability $\frac{\mu}{1-\mu} \cdot \frac{1-\omega}{\omega}$ to make the receiver exactly indifferent between taking actions $a = 1$ and $a = 0$ upon observing m_w. The tie is broken in the sender's favor. In our setting, even when the fact-checking technology never fails, the maximum ex ante payoff is $U_S = \mu$ achieved by the full fact-checking policy. The sender's commitment payoff is not achievable, since it requires an undetectable randomization on the side of 0-sender. Our sender lacks commitment power. If θ-sender sends multiple messages, then he is indifferent between sending any one of them. Fact-checking cannot make 0-sender randomize without revealing him. We note that for large state space $\theta \in [0,1]$, this is no longer true. The reason is that the best communication outcome for the sender no longer requires randomization on his side.[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 1.6.

[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.

Proposition 1.1 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 1.2. *When the fact-checking policy is more aggressive:*

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

The key insight behind Proposition 1.2 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

[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 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 1.2 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 1.2 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 1.2 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 1.1 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 1.1. 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 1.1. 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

Table 1.2 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 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 1.1, 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 1.2. 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 1.2. 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:

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

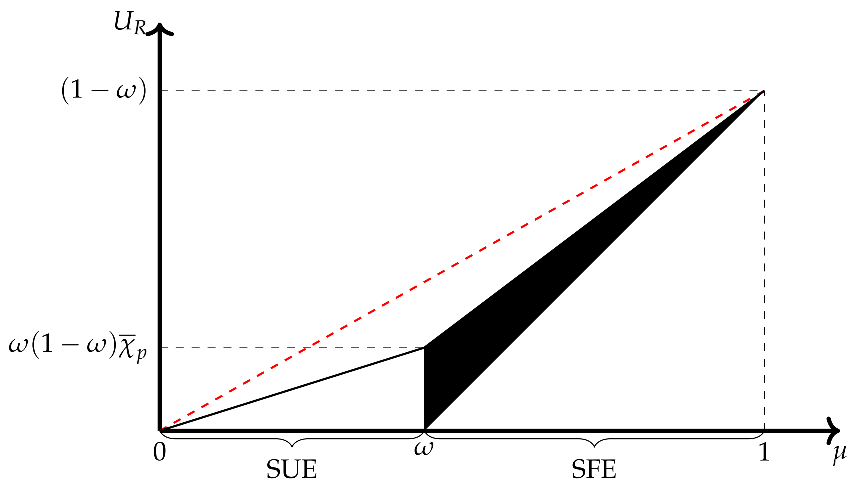


Figure 1.1. Feasible U_R depending on prior μ for fact-checking policies with fixed $\overline{\chi}_p$. The dashed red line corresponds to the receiver's payoff under complete information.

Figure 1.1 provides an illustration of the part of Proposition 1.2 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.

1.4 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 1.3. *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,*

- $\bar{c}(u_F) = \omega(1-p)\,[u_F(1,1) - u_F(0,1)]$ *in the sender-unfavorable environment,*
- $\bar{c}(u_F) = (1-\mu)(1-p)\,[u_F(0,0) - u_F(1,0)]$ *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 1.3 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

[23]Note that picking $\underline{\chi}_p = 1 - p$ is not necessary for optimality. However, when there are multiple

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 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 1.1, 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 $\lambda(a=1,\theta'|\varepsilon,\chi_p)$ in χ_p-equilibrium ε. Therefore, different fact-checking policies can only affect the fact-checker's payoff in issue θ.

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

The cost threshold depends on the prior only in SFE. Moreover, $\overline{c}(u_F)$ 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

equilibrium patterns, the fact-checker is able to steer players toward the preferred χ_p-equilibrium in which $\underline{m}$ is never played.

on an unchecked message, $\sigma(\overline{m}|0) = \frac{\mu}{1-\mu} \cdot \frac{1-\omega}{\omega}$ and $\sigma(m_s|0) = 1 - \sigma(\overline{m}|0)$, such that the receiver's incentive constraints are intact. As a consequence, the fact-checker's benefit and the minimal cost of fact-checking are linear in $\mu\overline{\chi}$, and $\overline{c}(u_F)$ is independent of the prior. In SFE, the minimal cost of implementing any equilibrium pattern from Table 2 is achieved by implementing χ_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 1.3 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 1.1. *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 1.2, the receiver prefers a more aggressive fact-checking policy. By Proposition 1.3, 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 1.1 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 1.1, 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 1.2 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.

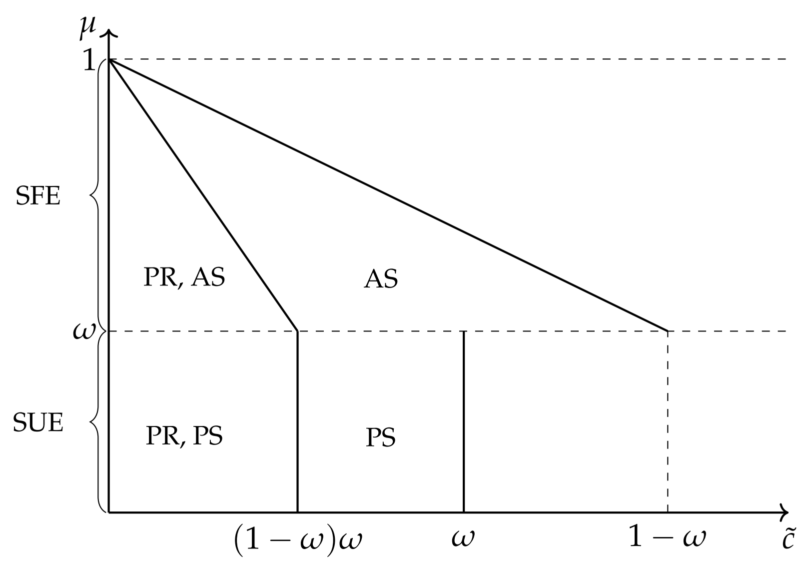


Figure 1.2. This figure shows the regions in the $(\tilde{c}, \mu)$ space, where $\tilde{c} = \frac{c}{1-p}$, for fixed $\omega < \frac{1}{2}$, where pro-receiver (PR), pro-sender (PS), and anti-sender (AS) fact-checkers choose the full fact-checking policy.

Figure 1.2 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 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 $\mu(1-p)[u_F(1,1)-u_F(0,1)]$. In SFE, the fact-checker that desires to implement a more aggressive fact-checking policy will not be able to sustain informative communication under the worst-equilibrium selection. Indeed, 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.

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

1.5 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 $\chi_p(m) := 1 - (1 - \chi_{p,1}(m))(1 - \chi_{p,2}(m))$. 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 1.3 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

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

[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.

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 $\bar{c}(\cdot)$ as given by Proposition 1.3 is fixed, that is, we fix parameters μ, ω, and p. First, note that either of conditions $\bar{c}(u_{F,i}) < 0$ or $c \geq \bar{c}(u_{F,i})$ imply that the optimal policy involves no fact-checking by fact-checker i. Indeed, if fact-checker i does not want to 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 1.3. 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.

[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.

Proposition 1.4. *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 1.4 holds for both SUE and SFE, with cost threshold $\bar{c}(\cdot)$ given by Proposition 1.3. 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 additional fact-checker does not assist in overcoming a failure of fact-checking technology. The payoffs of the sender and the receiver remain unaffected.

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

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.

1.6 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.

[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 1.5. *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,*

- $\overline{c}(u_F) = u_F(1,1) - u_F(0,1)$ *in the sender-unfavorable environment,*
- $\overline{c}(u_F) = \frac{1-\mu}{\mu} \cdot [u_F(0,0) - u_F(1,0)]$ *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 $a = 1$ and $a = 0$ upon observing m_w. The tie is broken in the sender's favor. In our setting, messages are cheap but the fact-checker can provide their verification. Thus, the pro-sender fact-checker is able to deliver the sender's commitment payoff in SUE, if the fact-checking cost is low enough. In particular, the fact-checker only checks m_w with probability one. The outcome does not rely on the selection and does not involve randomization on the sender's side. In our binary setting, the commitment payoff is not achievable, since it requires undetectable randomization by 0-sender which the fact-checker cannot sustain without revealing him. Note that the similar construction to Titova (2021) can show that the anti-sender fact-checker uses the same structure to implement the sender-worst outcome in SFE,

[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.

with a difference that the cutoff value θ^* for messages m_w and m_l is chosen to make the receiver exactly indifferent between taking actions $a = 1$ and $a = 0$ upon observing m_l and the tie is broken against the sender.

1.7 Conclusion

This paper examines 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.

Chapter 1 is currently being prepared for submission for publication of the material. The dissertation author, Aleksandr Levkun, is the sole author of this material.

Chapter 2

The Value of Data Records

2.1 Introduction

Personal data is the “new oil” of modern economies. Markets for data have been rapidly developing and have fueled major policy debates (Federal Trade Commission, 2014). These markets also have spurred intense research to understand their unique properties (Bergemann and Ottaviani, 2021). However, many critical questions remain. Among them, what is the value of an individual’s data for the firm using it? How does this value depend on the data’s content and the firm’s goals? Answering these questions can shed light on the demand side of data markets and on how people should be fairly compensated for their specific data (Lanier, 2013; Acquisti, Taylor, and Wagman, 2016; Arrieta-Ibarra et al., 2018).

Understanding the value of data in modern economies raises new challenges. First, in many markets data is traded based on its specific informational content (Bergemann and Bonatti, 2019). Yet, standard theories almost exclusively evaluate information before it realizes. Second, data is often used by firms that act as *intermediaries*—like e-commerce marketplaces, search engines, and matching platforms—to strategically direct interactions between agents with conflicting interests. Yet, standard theories mostly evaluate information for single decision makers. To overcome these challenges, our approach combines modern information design with classic duality methods. We find that the value of data for intermediation problems differs fundamentally from standard decision problems. This is because, to manage conflicting interests, an intermediary may tailor the information it conveys to the agents by pooling data records, thus creating complex externalities between them.[1]

Consider an example. An online platform mediates the interactions between a population of buyers and a monopolist, who produces a good at zero marginal cost. For each buyer, the platform owns a data *record*, which consists of a list of the buyer's personal characteristics (gender, age, etc.). There are different types of records depending on how much the platform knows about the buyer. For simplicity, suppose type ω_k reveals that her valuation for the seller's good is k for $k \in \{1,2\}$. The collection of buyers' records forms the platform's database. Suppose its composition q consists of 3 million records of type ω_1 and 6 million of type ω_2. The seller knows only q. For each interaction, the platform sends a signal about ω to the seller so as to influence the price he charges.[2] Concretely, the platform may divide the buyers into market segments based on their records and tell the seller to which segment each buyer belongs. Our main goal is to determine how much value the platform derives from each buyer's record. This is immediate if the platform maximizes the seller's profits (e.g., because it keeps a share of it). This case is effectively the same as if the platform itself were the seller and directly set a price for each buyer knowing ω. Since this is akin to a decision problem, the value of a record is equal to the payoff the platform directly obtains conditional on ω.

The answer is no longer immediate when we consider other objectives of the platform. To illustrate, suppose it maximizes the buyers' surplus (e.g., because it cares

[1] This practice is widespread in many digital platforms. For example, Google's "quality score" pools people's searches to increase competition among advertisers (see, e.g., Sayedi, Jerath, and Srinivasan, 2014); Uber conceals the riders' destinations from drivers to increase riders' welfare; and Airbnb withholds the host's profile picture to decrease discrimination.

[2] This is in the spirit of Bergemann, Brooks, and Morris (2015). Elliott, Galeotti, and Koh (2020) study a related problem with multiple horizontally differentiated sellers.

about their loyalty). One way to do this is to assign each buyer whose record is of type ω_2 to either a subprime segment $\underline{s}$ or to a prime segment $\bar{s}$, with equal probability; instead, it assigns all buyers whose record is of type ω_1 to $\underline{s}$. The seller optimally sets a price of 1 for segment $\underline{s}$ and a price of 2 for $\bar{s}$. The expected payoff that the platform directly obtains from a record of type ω_1 is 0, while it is $\frac{1}{2}$ for a record of type ω_2. Do these payoffs reflect the actual value the platform derives from each record? The answer is no. We will show that the actual values are $v^*(\omega_1) = 1$ and $v^*(\omega_2) = 0$. That is, the most valuable records for the platform are those that yield the lowest payoff. To see why, imagine two buyers, Ann and Bonnie, whose records are of type ω_1 and ω_2 respectively. Bonnie's record yields a positive payoff to the platform only when pooled with Ann's record through segment $\underline{s}$. In this case, Ann's record helps to persuade the seller to set a low price for Bonnie. Hence, Ann's record should not be worthless, even though Ann's interaction with the seller yields zero payoff to the platform. Indeed, $v^*(\omega_1) = 1$ reflects that Ann's record exerts a positive information externality on Bonnie's interaction. By contrast, $v^*(\omega_2) = 0$ reflects that we have to discount this externality from Bonnie's record.

Our main contribution is to characterize what determines the value of data records for intermediaries like the platform above. Our analysis delivers $v^*(\omega)$ as the *unit* value of every type-ω record in the database, leveraging the linear structure of intermediation problems. At the same time, $v^*(\omega)$ also equals the marginal effect on the platform's total payoff of adding type-ω records to the database. As the example showed, $v^*(\omega)$ can differ significantly from the payoff the platform *directly* obtains from a record because it pools records to tailor the information it conveys to the seller. We show that $v^*(\omega)$ is the sum of the platform's direct payoff from each type-ω record and the externalities caused by that record on other records and their interactions.[3] We relate

[3] Importantly, these externalities arise even when data records are statistically independent. As such, they differ fundamentally from "learning" externalities highlighted by the literature (see below), which depend on the correlation between records.

these externalities to how the platform exploits the seller's incentives across interactions. We explain when these externalities are positive and negative. For instance, in price-discrimination settings—which generalize our example—they satisfy a single-crossing property as long as the platform cares more about the buyers' surplus than the seller's profit: The externality is positive for buyers whose valuation for the seller's product is low and negative for those whose valuation is high. This means that ignoring such externalities could lead to overcompensating the latter for their data at the expense of the former.

This characterization of the value of data records is a necessary step to study an intermediary's willingness to pay for more data. In our context, acquiring more data can have two meanings. In our example, (i) the platform can obtain *more* records for its database and, hence, gain the ability to mediate more interactions between the corresponding buyers and the seller (e.g., because new buyers join it); or (ii) our platform can obtain *better* records by observing more informative characteristics about existing buyers (e.g., because they become more active online).[4]

With regard to obtaining more records, a key insight is that the platform's preference over databases is pinned down by v^* as a function of their composition q. In particular, v^* determines the platform's willingness to pay for more records and the substitutability between types of records. We find that this willingness to pay is stepwise diminishing. Moreover, record types are imperfect substitutes (or even complements) if and only if the platform withholds some information from the seller. These properties establish a "scarcity principle" for data: In any intermediation problem, scarcer types of records are more valuable both in absolute and in relative terms. They also enable us to infer how the platform uses its data from observable features of its demand function,

[4]The distinction between more and better records is consistent with that between *marketing lists* and *data appends*, the two main products traded in the data brokerage industry (Federal Trade Commission, 2014). The former allows companies to identify new customers who have specific characteristics. The latter allows companies to learn new characteristics about existing customers.

which can be derived using standard maximization subject to a budget constraint.

The platform's preference over databases is also useful to study its willingness to pay for better records. This is because obtaining more information about existing buyers changes their records' type and hence the database composition q. Imagine the platform refines the record of a buyer, called Cindy, by observing new characteristics about her. We show that such refinements have a positive direct effect: Cindy's record becomes more valuable in expectation. Because they change q, refinements also have indirect effects: Unrefined records can become more or less valuable. These effects are due to the aforementioned externalities and exist even if Cindy's new characteristics provide no information about other buyers (i.e., refinements are independent). We find that, despite their negative effects on the value of some records, independent refinements always benefit the platform overall, which therefore has a positive willingness to pay for them. This benefit is decreasing in the extensive margin—namely, how many records of a given type are refined. The benefit becomes zero under a precise condition, even if the platform would act on the new information it gets and use the refined and unrefined records differently. This is in sharp contrast with decision problems, where getting information is strictly beneficial if and only if it changes optimal behavior. Another difference is that the platform's willingness to pay can be *negative* for refinements that are correlated between records.

Our analysis applies to any setting where an intermediary (principal) mediates interactions between multiple agents by providing them with information or by affecting incentives with its actions. We can also let the agents have some payoff-relevant data, as long as the intermediary has direct access to everybody's data.[5] We view the intermediary as using each interaction's data as an input to produce information or choose its actions. Once we see intermediation problems through this lens, it becomes

[5] In a related project, we analyze the case where the intermediary has to first elicit the data from its sources.

natural to use linear-programming duality to characterize the value v^* of the data inputs, adapting the classic work of Dorfman, Samuelson, and Solow (1987) and Gale (1989).

This paper contributes to improving our understanding of the *demand side* of data markets. We believe that its insights are useful to informing empirical strategies for estimating the demand for data or to inferring from market observables how data is used. Progress in this area is essential for studying the welfare effects of critical policy interventions, such as new antitrust or privacy regulations.[6] Finally, a better understanding of the value of people's data may help improve on the status quo where they receive no compensation for it (Lanier, 2013; Arrieta-Ibarra et al., 2018; Jones and Tonetti, 2020).

Related Literature

This paper contributes to the burgeoning literature on data markets, comprehensively reviewed by Bergemann and Bonatti (2019) and Bergemann and Ottaviani (2021).

One of its strands studies the optimal "use" of a database. This often involves a single party—such as a platform or data broker—who owns a database and designs information products for some agents—such as sellers, advertisers, or decision makers—to either charge a price or influence their behavior (or both). In Admati and Pfleiderer (1986) and Admati and Pfleiderer (1990), a platform sells signals (i.e., Blackwell experiments) about an asset to market traders. In Bergemann and Bonatti (2015), a platform sells segments of buyers to advertisers and charges a linear price based on the segment size. In Bergemann, Bonatti, and Smolin (2018), a platform designs menus of signals to screen information buyers with heterogeneous priors. Yang (2020) studies a related problem in considerably richer settings. Our platform also owns a database and uses it to design information. However, our focus is not on the information products and

[6] See, e.g., Stigler Report (2019); Crémer, de Montjoye, and Schweitzer (2019); Goldberg, Johnson, and Shriver (2021).

their prices but on the value of their data inputs in the "upstream" market. These data records have two key features: Each record gives access to a buyer, on top of information about her, and each record can be valued ex post based on its specific content.

Another strand of the literature on data markets studies how to incentivize consumers to disclose their data. Choi, Jeon, and B.-C. Kim (2019), Acemoglu et al. (2021), and Ichihashi (2021) study the "learning" externalities that one consumer's disclosure has on others when their data is correlated. Bergemann and Morris (2019) examine how this correlation affects consumers' incentives to participate in data markets and other market observables. These papers differ from ours in two ways. First, our platform is assumed to already have the database.[7] This offers a useful benchmark to study the effects of privacy regulations. Second, we isolate a new data externality, which stems from the platform's pooling records to withhold information from the sellers and arises even if consumers' records are statistically independent.

Our work is related to the literature on data privacy, reviewed by Acquisti, Taylor, and Wagman (2016). Calzolari and Pavan (2006) analyze information externalities between sequential interactions. Nageeb, Lewis, and Vasserman (2022) examine when giving consumers control over their data can help them benefit from personalized pricing. Ichihashi (2020) finds that a multi-product platform can prefer not to use consumers' data for personalized pricing and maximizes profits via product recommendations.

Our methods build on the information-design literature, reviewed by Bergemann and Morris (2019). We formulate our "data-use" problem as a linear program, using standard arguments (Bergemann and Morris, 2016), and then consider its dual to obtain our "data-value" problem. Others have used duality to study information design (Kolotilin, 2018; Galperti and Perego, 2018; Dworczak and Martini, 2019; Dworczak and Kolotilin, 2019; Dizdar and Kováč, 2020). These papers exploit the dual to solve

[7] This may seem far-fetched but most data-brokers' transactions happen without the consumer's knowledge (Federal Trade Commission, 2014).

the primal design problem. We use the dual to address a distinct economic question of independent interest—what is the value of data? Unlike those papers, we also study problems with multiple agents through the notion of Bayes-correlated equilibrium. This links our work to an earlier literature on dual analysis of correlated equilibria (Nau and McCardle, 1990; Nau, 1992; Myerson, 1997). Finally, the mechanism-design literature has used duality methods at least since Myerson (1983; 1984), as well as more recently to study informationally robust mechanisms (e.g., Du, 2018; Brooks and Du, 2020; Brooks and Du, 2021).

2.2 Model

For ease of exposition, we present the model and analysis in a context similar to our example in the Introduction: An e-commerce platform mediates interactions between buyers and sellers. Our approach and results apply much more broadly to settings where a principal influences the behavior of multiple strategic agents with information, its actions, or both. Section 2.5 discusses this and other aspects of the model.

Let $i = 0$ denote the platform, which is the principal. Let $I = \{1, \ldots, n\}$ be a set of sellers, who are the strategic agents. Let A_i be the finite set of seller i's actions. We can interpret a_i as the price, quality, or other features of seller i's product. The platform is used by a continuum of buyers, each interested in buying a product from the sellers. Each buyer's preference over the sellers' products is pinned down by a random variable θ, which is independently and identically distributed across buyers over a finite set Θ. We use the pronoun 'it' for the platform, 'he' for each seller, and 'she' for each buyer.

The platform has access to some data about each buyer. We think of this data as a *record* of personal characteristics that is informative about her θ—perhaps only partially.

We assume that each buyer's record is uninformative about the other buyers' θ.[8] There are different *types* of records—denoted by ω in some finite set Ω—depending on what the platform knows about the buyer. Thus, the content of each buyer's record is analogous to the realization of an exogenous signal about her underlying preference. Only the platform observes ω, which gives it an informational advantage over the sellers. Let $q \in \mathbb{R}_+^{\Omega}$ denote the collection of buyers' records, where $q(\omega)$ are of type ω. We refer to q as the platform's *database*.

For each interaction between a buyer and the sellers, we leave her purchase decision given their actions and θ implicit and embed it in the payoff functions of the sellers and the platform. For every ω and $a = (a_1 \ldots, a_n)$, let $u_i(a, \omega)$ be i's expected payoff conditional on the buyer's record. Let $\Gamma_\omega = \{I, (A_i, u_i(\cdot, \omega))_{i=0}^n\}$, which defines a complete-information game between the sellers. We may also refer to Γ_ω as a buyer-sellers interaction of type ω. The primitives $\Gamma = \{\Gamma_\omega\}_{\omega \in \Omega}$ and q are common knowledge.

The platform mediates each interaction by privately conveying information about its type to each seller so as to influence their actions. The sellers combine this information with Γ and q to form beliefs and act. Our platform has full commitment power, similar to the omniscient information designer in Bergemann and Morris (2019). Formally, it publicly commits to an information structure that, for each interaction, produces a private signal about ω for each seller i. As is standard (Myerson, 1983; Myerson, 1984; Bergemann and Morris, 2016), we can focus on information structures in the form of recommendation mechanisms, where the platform privately recommends an action to each seller that he must find optimal to follow (obedience). A mechanism is then a function $x : \Omega \to \Delta(A)$, where $x(a|\omega)$ can be interpreted as the share of interactions of

[8]We make this assumption to emphasize the novel aspects of our results. Our model can accommodate correlation among records (see Section 2.5.1).

type ω that lead to recommendation profile a.[9] Formally, the problem is

$$\begin{aligned}
\mathcal{U}_q: \qquad & \max_x \sum_{\omega\in\Omega, a\in A} u_0(a,\omega)x(a|\omega)q(\omega) \\
& \text{s.t. for all } i\in I \text{ and } a_i, a_i' \in A_i, \\
& \sum_{\omega\in\Omega, a_{-i}\in A_{-i}} \left(u_i(a_i,a_{-i},\omega) - u_i(a_i',a_{-i},\omega)\right) x(a_i,a_{-i}|\omega)q(\omega) \geq 0. \qquad (2.1)
\end{aligned}$$

Constraint (2.1) is equivalent to requiring that a_i maximize seller i's expected utility conditional on the information conveyed by a_i given x and the database q. Denoting any optimal mechanism by x_q^*, we define the *direct payoff* generated by each record of type ω as

$$u_q^*(\omega) \triangleq \sum_{a\in A} u_0(a,\omega)x_q^*(a|\omega),$$

and the *total payoff* generated by the database as

$$U^*(q) \triangleq \sum_{\omega\in\Omega} u_q^*(\omega)q(\omega). \tag{2.2}$$

We assume that $\mathcal{U}_q$ satisfies a minor regularity property, which holds generically in the space of sellers' payoff functions: No more than $|A\times\Omega|$ of the constraints (2.1) are ever active at the same time (see Remark A.2.1 in Appendix A.2).

2.3 The Unit Value of Data

This section addresses our main question: How much value does the platform derive from each buyer's record and what are its properties? To get a sense of why the answer is nontrivial, it is useful to compare our problem $\mathcal{U}_q$ with standard decision problems. We can interpret $\mathcal{U}_q$ as a collection of decisions: For each buyer-sellers interaction, the platform uses its record to decide what to disclose about the buyer so as

[9] Note that restricting $x(\cdot|\omega)$ to be the same between records of the same type ω is without loss of generality.

to influence the sellers' actions (i.e., $x(\cdot|\omega)$ for every ω).

To establish our benchmark, imagine that all parties have aligned interests: u_i is an affine transformation of u_0 for all $i = 1, \ldots, n$. Then, constraints (2.1) can be omitted and it is as if the platform *directly* controlled the sellers' actions. In this case, all the decisions in $\mathcal{U}_q$ are independent of one another. Indeed, $\mathcal{U}_q$ is separable across records: For each of them, the platform effectively faces a standard decision problem in which it chooses a to maximize $u_0(a, \omega)$ guided by the information in the record. For this reason, we will slightly abuse terminology and refer to our model with aligned interests as a *standard-decision problem*.

In this paper, however, we are mainly interested in instances of $\mathcal{U}_q$ where parties have conflicting interests, to which we will refer as *intermediation problems*. In this case, the platform can only influence the sellers' actions *indirectly*, subject to constraints (2.1). While it continues to face the collection of decisions in $\mathcal{U}_q$, these are no longer independent. That is, $\mathcal{U}_q$ is no longer separable across records because what information a signal conveys about one record depends on which other records lead to the same signal.[10] Consequently, while the value of each record continues to be determined by how it is used to guide decisions—like in standard-decision problems—this use is not confined to the interaction physically attached to that record—unlike in standard-decision problems. Thus, to answer our question, we need to systematically keep track of all the ways the platform uses each record to mediate all interactions and the resulting interdependencies.

This contrast between intermediation and standard-decision problems will be helpful to better understand our results and relate them to the classic work on the comparison of experiments under the decision-theoretic framework of Blackwell (1951, 1953) (see also Laffont, 1989). We will return to this point in Section 2.5.

[10]This dependence between signal decisions is orthogonal to our commitment assumption. It would arise even if our platform could not commit and we had to rely on some equilibrium notion.

2.3.1 The Data-Value Problem

Our approach builds on the observation that any information-design problem is a linear program. A standard economic interpretation is that linear programs describe the problem of optimally using some scarce inputs to produce some output (Dorfman, Samuelson, and Solow, 1987, p. 39). We think of information design as a "data-use" problem, where the inputs are the records in the database and the output is the information conveyed by each mechanism in the form of recommendations. Following Dorfman, Samuelson, and Solow (1987), we then exploit the dual of this data-use problem to evaluate each record.

We call this evaluation task the *data-value* problem. Let $\lambda = (\lambda_1, \ldots, \lambda_n)$ where $\lambda_i : A_i \times A_i \to \mathbb{R}_+$ for all $i \in I$. For each i and (a, ω) define

$$t_i(a,\omega) \triangleq \sum_{a'_i \in A_i} \left(u_i(a_i, a_{-i}, \omega) - u_i(a'_i, a_{-i}, \omega) \right) \lambda_i(a'_i | a_i), \tag{2.3}$$

and $t(a,\omega) \triangleq \sum_{i \in I} t_i(a,\omega)$. The data-value problem is

$$\begin{aligned} \mathcal{V}_q : \quad & \min_{v,\lambda} \sum_{\omega \in \Omega} v(\omega) q(\omega) \\ & \text{s.t. for all } \omega \in \Omega, \\ & v(\omega) = \max_{a \in A} \left\{ u_0(a,\omega) + t(a,\omega) \right\}. \end{aligned} \tag{2.4}$$

We denote any optimal solution by (v^*_q, λ^*_q) and the induced functions t by t^*_q. By standard linear-programming arguments v^*_q is unique generically with respect to q. Note that v^*_q can depend on q for intermediation problems but not for standard-decision problems, as in this case $v^*_q(\omega) = \max_{a \in A} u_0(a, \omega)$ for all ω.

We refer to equation (2.4) as the *value formula*, which defines our main object of interest. The reason hinges on the next relation between the data-use and data-value problems and on the following interpretation. All proofs are in the appendix.

Lemma 2.1. *For any q, $\mathcal{V}_q$ is equivalent to the dual of $\mathcal{U}_q$. Thus, for every x_q^* and (v_q^*, λ_q^*)*

$$\sum_{\omega \in \Omega} v_q^*(\omega) q(\omega) = U^*(q) \triangleq \sum_{\omega \in \Omega} u_q^*(\omega) q(\omega). \tag{2.5}$$

This duality relation follows from basic linear-programming results. When applied to our specific problem, it becomes the key to answering our economic questions. In $\mathcal{U}_q$, every x defines a joint measure $\chi \in \mathbb{R}_+^{\Omega \times A}$, which must satisfy $\sum_{a \in A} \chi(a, \omega) = q(\omega)$; that is, the use of type-ω records to produce recommendations must exhaust their stock $q(\omega)$ in the database. Formally, $v(\omega)$ is the multiplier of this constraint, which is usually interpreted as the shadow price of the corresponding input through the thought experiment of adding a marginal unit of it. In fact, $v_q^*(\omega)$ is equal to the derivative of $U^*(q)$ with respect to $q(\omega)$, as for any constrained optimization. However, it would be incorrect to think that $v_q^*(\omega)$ captures only the value of a marginal record of type ω. The linear structure of $\mathcal{V}_q$ and our value formula demonstrate that $v_q^*(\omega)$ is the value of *each* record of type ω in the database. Note that, by (2.4), $v_q^*(\omega)$ is measured in terms of the payoffs of the platform and the sellers. We will then call $v_q^*(\omega)$ the *unit value* of a record of type ω (see also Gale, 1989, p. 12). Note that $\mathcal{V}_q$ assigns such values simultaneously to all records and does not require finding x_q^*.

The rest of the paper characterizes the properties of v_q^* and their economic implications. Here, we begin with a useful lower bound. For $\omega \in \Omega$, let $CE(\Gamma_\omega)$ be the set of correlated equilibria of the game Γ_ω.

Lemma 2.2 (Lower Bound)**.** *For every q,*

$$v_q^*(\omega) \geq \bar{u}(\omega) \triangleq \max_{y \in CE(\Gamma_\omega)} \sum_{a \in A} u_0(a, \omega) y(a), \qquad \omega \in \Omega.$$

Lemma 2.1 and 2.2 imply that $v_q^*(\omega) = \bar{u}(\omega)$ for all ω if and only if there is an optimal x_q^* that satisfies $x_q^*(\cdot|\omega) \in CE(\Gamma_\omega)$ for all ω. In words, for such an x_q^* the platform fully discloses the buyer's record to the sellers for all interactions.

Unit Values and Individual Compensations

By quantifying how much a record contributes to the total payoff $U^*(q)$, v_q^* offers a *benchmark* for individually compensating each buyer as the "owner" of her record. In $\mathcal{V}_q$, we can view the platform as choosing v to minimize the total expenditure to compensate the buyers. However, the platform is constrained by equation (2.4), which imposes a lower bound for each buyer's compensation that takes into account how her record is used.[11] Paraphrasing Dorfman, Samuelson, and Solow (1987), p. 43, this interpretation is reminiscent of the operation of a competitive market where competition forces the platform to offer the "owner" of a record the full value to which her input gives rise, while competition among these "owners" drives down this value to the minimum consistent with this limitation. Gale (1989), Chapter 3.5, also shows how dual problems can deliver competitive prices of scarce inputs. In general, how much individuals will actually receive for their data can depend on the market structure, their bargaining power, and the need to incentivize them to disclose their data truthfully.

Nonetheless, to see the importance of guiding the compensation of data owners using v_q^* and not u_q^*, consider again our introduction example of a surplus-maximizing platform. Suppose it decides—perhaps forced by some regulation or court order—to compensate the buyers for their contribution to $U^*(q)$ by giving back some share δ to them. How δ is chosen and the compensations implemented is important in practice but irrelevant for the point we want to make here. The more fundamental question is how much each buyer should get. It seems that the answer should take into account each buyer's specific record. One could use u_q^*, which would result in incorrectly allocating $\delta U^*(q)$ only to the buyers with $\omega = \omega_2$ (each receiving $\delta u_q^*(\omega_2) = \delta 0.5$)

[11] In fact, by complementary slackness $v_q^*(\omega) = u_0(a,\omega) + t_q^*(a,\omega)$ if $x_q^*(a|\omega) > 0$. We provide another independent economic interpretation of the data-value problem in Appendix A.2.

because $u_q^*(\omega_1) = 0$. In fact, only the buyers with $\omega = \omega_1$ contribute to $U^*(q)$ because $v_q^*(\omega_2) = 0$. Thus, $\delta U^*(q)$ should be allocated entirely to these buyers (each receiving $\delta v_q^*(\omega_1) = \delta$).

2.3.2 Value Decomposition and Data Externalities

What determines the unit value of a record? Why and how are the direct payoffs u_q^* a biased measure of these values? We show next that the value of a record can be decomposed into two parts: its direct payoff and an additional component, which captures that record's effects on the information the platform discloses about other records and thus on their direct payoffs.

Proposition 2.1. *For all ω, $v_q^*(\omega) = u_q^*(\omega) + t_q^*(\omega)$ where*

$$t_q^*(\omega) \triangleq \sum_{a \in A} t_q^*(a, \omega) x_q^*(a|\omega) \overset{a.e.}{=} \sum_{\omega' \in \Omega} \frac{\partial u_q^*(\omega')}{\partial q(\omega)} q(\omega'). \tag{2.6}$$

This result highlights that the effects captured by $t_q^*(\omega)$ are akin to an externality. Consider a buyer called Ann. Simply by belonging to the database, her record affects how the platform mediates the interactions that other buyers have with the sellers. Formally, $t_q^*(\omega)$ summarizes the marginal effect that Ann's record has on the direct payoff of other records (equation (2.6)). This externality is purely informational: Ann's record contributes to the information advantage that the platform has for all interactions it mediates and hence affects its decisions with other records through x_q^*. In fact, $\frac{\partial}{\partial q(\omega)} u_q^*(\omega') = \sum_a u_0(a, \omega') \frac{\partial}{\partial q(\omega)} x_q^*(a|\omega')$. Adjustments in x_q^* can arise because changing $q(\omega)$ can render x_q^* no longer feasible (i.e., obedient) or optimal.

This externality is a hallmark of intermediation problems. It arises when an intermediary tailors the information for the agents by pooling data records so as to manage conflicts of interest. Indeed, the externality is absent in standard-decision problems, where it is optimal to fully disclose the type of each record.[12] It is worth emphasizing that this externality arises even if records are statistically independent. As such, it is distinct and complementary to the "learning" externalities discussed in Section 2.1, which arise because a buyer's record is informative about another buyer's preferences. This channel is intentionally switched off in our paper, which emphasizes externalities that arise endogenously from how data is used.

Which records generate positive and which records generate negative externalities?

Corollary 2.1. *$t_q^*(\omega) < 0$ for some ω if and only if $t_q^*(\omega') > 0$ for some ω'. Moreover, $t_q^*(\omega) < 0$ implies $u_q^*(\omega) > \bar{u}(\omega)$, while $u_q^*(\omega) < \bar{u}(\omega)$ implies $t_q^*(\omega) > 0$.*[13]

The externalities lead to cross-subsidization of value from records with $t_q^*(\omega) < 0$ to records with $t_q^*(\omega') > 0$. Since the total payoff is fixed, records with $v_q^*(\omega') > u_q^*(\omega')$ must take their extra value from some other records. The second part of the corollary explains this cross-subsidization. Records with $t_q^*(\omega) < 0$ generate a direct payoff that exceeds the full-disclosure payoff $\bar{u}(\omega)$, which requires that $u_0(a,\omega) > \bar{u}(\omega)$ and $x_q^*(a|\omega) > 0$ for some a. That is, the platform earns a payoff with type-ω records that would never be possible by fully disclosing them, so it relies on pooling them with records of a different type. In this case, type-ω records do not "deserve" the full $u_q^*(\omega)$ and their value discounts the help received from other records. Conversely, this help from type-ω' records justifies why $t_q^*(\omega') > 0$ and their value exceeds $u_q^*(\omega')$. Finally, we can interpret $u_q^*(\omega) < \bar{u}(\omega)$ as "sacrificing" type-ω records, as the platform could fully disclose them and get $\bar{u}(\omega)$. For this sacrifice to be worthwhile, such records must receive compensation, explaining $t_q^*(\omega) > 0$. This last part offers a simple sufficient

[12] Whenever full disclosure is optimal, $u_q^*(\omega) = \bar{u}(\omega)$ and $v_q^*(\omega) = \bar{u}(\omega)$ as discussed after Lemma 2.2, so $t_q^*(\omega) = 0$. Note that the converse is not true: There are examples where $t_q^*(\omega) = 0$, but $v_q^*(\omega) > \bar{u}(\omega)$ for all ω.

[13] The corollary follows because Lemma 2.1 implies $\sum_{\omega\in\Omega} t_q^*(\omega)q(\omega) = 0$, and Lemma 2.2 and Proposition 2.1 imply $t_q^*(\omega) \geq \bar{u}(\omega) - u_q^*(\omega)$ for all ω.

condition for $t_q^* \neq 0$. Appendix A.2 provides another condition based on primitives.

Proposition 2.1 also highlights that the externalities through u_q^* are tightly related to how the platform exploits the sellers' incentives with its information. By the first part of (2.6), $t_q^*(\omega)$ aggregates externalities that type-ω records generate by inducing specific actions a. These are inversely related to the platform's resulting payoff, in the following sense.

Corollary 2.2. *Suppose $x_q^*(a|\omega) > 0$ and $x_q^*(a'|\omega) > 0$. Then, $u_0(a,\omega) > u_0(a',\omega)$ if and only if $t_q^*(a,\omega) < t_q^*(a',\omega)$.*[14]

Thus, inducing actions whose payoff exceeds $\bar{u}$ by more, for instance, requires paying larger externalities to other records. Since $t_q^*(a,\omega) \triangleq \sum_{i \in I} t_{q,i}^*(a,\omega)$, we can view $t_{q,i}^*(a,\omega)$ as how much seller i contributes to the externality. Recall that $t_{q,i}^*(a,\omega)$ differs from zero only if $\lambda_{q,i}^*(a_i'|a_i) > 0$ for some a_i' (see (2.3)). By standard arguments (complementary slackness), $\lambda_{q,i}^*(a_i'|a_i) > 0$ only if

$$\sum_{\omega, a_{-i}} \Big(u_i(a_i, a_{-i}, \omega) - u_i(a_i', a_{-i}, \omega) \Big) x_q^*(a_i, a_{-i}|\omega) q(\omega) = 0; \tag{2.7}$$

the converse also holds generically in q. In words, $\lambda_{q,i}^*(a_i'|a_i) > 0$ if and only if seller i is indifferent between a_i and a_i' conditional on receiving recommendation a_i from x_q^*.

Corollary 2.3. *The sellers who contribute to the externality $t_q^*(\omega)$ are only those whom x_q^* renders indifferent with the actions it recommends using records of type ω (i.e., (2.7) holds).*

Thus, we can also interpret $t_q^*(\omega)$ as aggregating the "cost of incentives" for the actions that the platform recommends using type-ω records. This cost is positive for seller i if recommending a with type-ω records hinders satisfying (2.7) because $u_i(a_i, a_{-i}, \omega) < u_i(a_i', a_{-i}, \omega)$, which then lowers $t_q^*(a,\omega)$ and hence $v_q^*(\omega)$. The opposite happens if recommending a with type-ω records helps satisfying (2.7) because

[14] This follows from complementary slackness, namely $v_q^*(\omega) = u_0(a,\omega) + t_q^*(a,\omega)$ if $x_q^*(a|\omega) > 0$.

$u_i(a_i, a_{-i}, \omega) > u_i(a_i', a_{-i}, \omega)$. Appendix A.2 elaborates on this interpretation and how the platform exploits the sellers to determine their contribution to the externalities. Note that Corollary 2.3 differs from the immediate fact that optimal solutions of linear programs occur on the boundary of the feasible set, which here means that some obedience constraint must bind. Also, as q varies, x_q^* and hence t_q^* may change. However, as long as λ_q^* does not change, how each seller contributes to $t_q^*(\omega)$ does not change.

2.3.3 Application (Part I): Price Discrimination and the Externalities

To illustrate the importance of these data externalities, we consider a more general version of our example in the Introduction. There is only one seller ($n = 1$) who chooses the price a_1 for his product. For each buyer, θ is her valuation for the product. Let $\Omega = \{\omega_1, \ldots, \omega_K\} \subset \mathbb{R}_+$, $K \geq 2$, and ω_k be strictly increasing in k. Records of type ω_k fully reveal that $\theta = \omega_k$. Normalizing the seller's constant marginal cost to zero, his profit is a_1 if $\omega \geq a_1$ and zero otherwise: $u_1(a_1, \omega) = a_1 \mathbb{I}\{\omega \geq a_1\}$. The platform maximizes a weighted sum of profits and consumer surplus: $u_0(a_1, \omega) = \pi a_1 \mathbb{I}\{\omega \geq a_1\} + (1-\pi)\max\{\omega - a_1, 0\}$, where $\pi \in [0,1]$. Finally, let a_q be the optimal uniform monopoly price.

Proposition 2.2. *For $\pi \leq \frac{1}{2}$,*

$$v_q^*(\omega) = \begin{cases} (1-\pi)\omega & \text{if } \omega < a_q \\ \pi a_q + (1-\pi)(\omega - a_q) & \text{if } \omega \geq a_q; \end{cases}$$

moreover, $t_q^*(\omega) > 0$ for $\omega < a_q$ and $t_q^*(\omega) \leq 0$ for $\omega \geq a_q$. For $\pi \geq \frac{1}{2}$, $v_q^*(\omega) = u_q^*(\omega) = \pi\omega$ for all ω.

To understand this result, we note that x_q^* takes only two forms depending on π (see Appendix A.2). If $\pi \leq \frac{1}{2}$, the platform maximizes the buyers' surplus subject to holding the seller's expected profit at a_q, as when $\pi = 0$. Thus, it is as if trade happens for every interaction, generating total surplus equal to ω, and only the buyers with a product valuation of at least a_q contribute to guaranteeing this profit. If $\pi \geq \frac{1}{2}$, the platform fully discloses all records. This allows perfect price discrimination, so profits always equal the buyer's valuation and her surplus is zero.

Whenever the platform cares more about the buyers' surplus than the seller's profits, the direct payoff u_q^* provides a biased account of the value of each record. This bias has a specific structure: t_q^* satisfies a single-crossing property in ω and this holds generally across q. That is, u_q^* is biased downward for low-valuation buyers (i.e., $\omega < a_q$) and upward for high-valuation buyers (i.e., $\omega \geq a_q$). Thus, ignoring the externalities we highlight may lead to overcompensating high-valuation buyers for their data at the expense of low-valuation buyers.

How does caring more about the buyers' surplus affect the value of their records? By simple algebra, lowering $\pi \leq \frac{1}{2}$ decreases $v_q^*(\omega)$ if and only if the buyer has an intermediate valuation ($a_q \leq \omega < 2a_q$). Intuitively, for such records a larger share of the buyers' product valuation goes to fund the seller's guaranteed profits of a_q, which becomes more costly as their surplus becomes more important to the platform. By contrast, the records of buyers with low valuation help the platform achieve a positive surplus with other buyers, and the records of buyers with high valuation just yield a large surplus. For $\pi \geq \frac{1}{2}$, $v_q^*(\omega)$ increases in π independently of ω. This is because the platform helps the seller extract the full surplus from each interaction, and it cares more about doing so.

2.4 Willingness to Pay for Data

What is the platform's willingness to pay for "having more data"? This colloquial expression can have two meanings. The first—analyzed in Section 2.4.1—is that the platform obtains *more* records in the database and, hence, it mediates more interactions between the buyers and sellers. The second—analyzed in Section 2.4.3—is that the platform obtains *better* records; namely, it observes more informative characteristics about existing buyers. In either case, having more data ultimately changes the database q, which is the basis for the sellers' beliefs. Hereafter, we assume that how the platform changes q is publicly observed and hence q is always commonly known.[15] Building on Section 2.3, we can then study the platform's willingness to pay for more data by analyzing how the records' values v_q^* depend on q. Alternatively, we can interpret the the following analysis as comparative static exercises that show how the values of records vary between platforms which differ only in their databases.

2.4.1 More Records: Preferences Over Databases

Analyzing the platform's willingness to pay for more records can shed light on the properties of the demand for data records. For example, are demand curves downward sloping? Are data records complements or substitutes and, if so, why? We can view the platform as a "consumer" of records, whose utility function is U^*. The platform's preferences over databases are then fully characterized by v_q^*. Indeed, $v_q^*(\omega)$ is akin to the marginal utility of type-ω records at q, which determines the platform's willingness to pay. We can also measure the substitutability between records of type ω and ω' at q by computing their marginal rate of substitution as usual, which satisfies $MRS_q(\omega,\omega') \stackrel{\text{a.e.}}{=} -\frac{v_q^*(\omega)}{v_q^*(\omega')}$.

A classic property in standard consumer theory is that marginal utilities are diminishing. Does the same hold for the platform? More generally, how does v_q^* vary with q? We show that as records of a given type become more abundant, they become less valuable and do so stepwise. This follows from the next two results. The first

[15]Of course, in reality the platform may change its database privately without the sellers' knowing exactly how. Allowing for this introduces complications and requires enriching the model accordingly. We leave this for future research.

establishes a general "scarcity principle" for data. Given q, define the share of type-ω records by

$$\mu_q(\omega) \triangleq \frac{q(\omega)}{\sum_{\omega'} q(\omega')}, \qquad \omega \in \Omega.$$

Proposition 2.3 (Scarcity Principle)**.** *Consider databases q and q'. Fix ω. If $\mu_q(\omega) < \mu_{q'}(\omega)$, then $v^*_q(\omega) \geq v^*_{q'}(\omega)$. Moreover, there exists $\bar{\mu}(\omega) < 1$ such that, if $\mu_q(\omega) > \bar{\mu}(\omega)$, then $v^*_q(\omega) = \bar{u}(\omega)$.*

This property holds generally, irrespective of the details of the intermediation problem. It implies that $v^*_q(\omega)$ is weakly decreasing in $q(\omega)$. Hence, holding fixed the quantity of all other types of records, the platform's demand for type-ω records is downward sloping and converges to $\bar{u}(\omega)$ when $q(\omega)$ is sufficiently large. Equivalently, the individual contribution of type-ω records to the platform's payoff—hence, their owners' benchmark compensation—decreases as their quantity increases.

This decline in value is stepwise because v^*_q is locally constant in q.

Proposition 2.4 (Stability)**.** *There exists a finite collection $\{Q_1, \ldots, Q_M\}$ of open, convex, and disjoint subsets of $\mathbb{R}^{\Omega}_{+}$ such that $\cup_m Q_m$ has full measure and, for every m, v^*_q is unique and constant for $q \in Q_m$.*

Each Q_m is the interior of a cone in the space of databases $\mathbb{R}^{\Omega}_{+}$.[16] Importantly, v^*_q is constant even though the platform may adjust how it uses its data when q changes. We can show that within each cone, while $v^*_q(\omega)$ is constant, the optimal $x^*_q(\omega)$ changes as a function of q (see Remark A.2.1 in Appendix A.2). Intuitively, this is because x^*_q has to be fine-tuned to maximally exploit the sellers' incentives. By contrast, v^*_q depends only on which sellers' incentives are exploited, but not on how much (recall equation (2.3) and Corollary 2.3).

[16]It is easy to see that unit values are constant along the rays in the space of databases: If $q' = \alpha q$ for $\alpha > 0$, then $v^*_q = v^*_{q'}$. This is because only the frequency of record types matters for the sellers' incentives.

Returning to the platform's marginal rate of substitution between records, is it diminishing as in standard consumer theory? The answer is yes, at least weakly: The platform's preferences are always convex, because $U^*(q)$ is always a concave function of q.[17] However, in some cases records are perfect substitutes, namely $MRS_q(\omega,\omega')$ is constant. An example is when the platform faces a standard-decision problem, since then $v^*_q(\omega) = \bar{u}(\omega)$ for all ω. The next result characterizes which intermediation problems also lead to perfect substitutability between all records (i.e., $MRS_q(\omega,\omega')$ does not depend on q for all ω, ω').

Proposition 2.5. *All records are perfect substitutes if and only if there is some database $q \in \mathbb{R}^\Omega_{++}$ at which it is optimal for the platform to fully disclose every record. In this case, full disclosure is optimal for all $q \in \mathbb{R}^\Omega_+$.*

This result has several implications. First, suppose we can estimate the platform's demand functions by observing its transactions in the data market. Then, by detecting any imperfect substituatbility between record types, we can infer that the platform is withholding information from the sellers. We can do so even if we know nothing about the intermediation problem it faces (i.e., Γ). Indeed, by Proposition 2.5 some types of records are imperfect substitutes if and only if it is never optimal to fully disclose all records. More generally, the intermediary's transactions in a data market reveal properties of how it uses its database, which may be harder to observe.

Another implication is that the optimality of withholding some information does not depend on the database composition. This simplifies assessing whether an intermediary will withhold information based on the primitives of a specific application (i.e., Γ). One way is to start from full disclosure and show that for some $q \in \mathbb{R}^\Omega_{++}$ we can do strictly better by sometimes concealing any type of records. Since the answer

[17] Concavity follows because, by (2.5), we can view U^* as the result of minimizing a family of functions that are linear in q (Rockafellar, 1970, Theorem 5.5). It is related directly to the concavification results in Mathevet, Perego, and Taneva (2020) and indirectly to the individual-sufficiency results in Bergemann and Morris (2016a).

does not depend on q, we can pick it in any convenient way (e.g., uniform quantities). Alternatively, we can identify conditions for the optimality of withholding information directly in terms of Γ. Appendix A.2 provides such a condition.

Last but not least, convexity of the platform's preferences leads to standard demand analysis. In particular, choosing an optimal database subject to a budget constraint is a well-behaved problem. Given market price $p(\omega) > 0$ for every ω, the optimal q is characterized by a generalized version of the usual tangency condition that deals with kinks in indifference curves:

$$\max_{v \in v_q^*} \frac{v(\omega)}{v(\omega')} \geq \frac{p(\omega)}{p(\omega')} \geq \min_{v \in v_q^*} \frac{v(\omega)}{v(\omega')}, \qquad \omega, \omega' \in \Omega.^{18}$$

It is easy to see that if withholding information is optimal (i.e., $v_q^* \neq \bar{u}$ for some q), then there is an open set of prices for which the platform chooses a nontrivial database containing multiple types of records. More generally, we can use v_q^* to characterize the platform's demand functions for records, thus enabling a general study of the demand side of data markets. These functions satisfy some simple properties that may be useful for empirical analysis: Since $U^*(q)$ is homothetic, data records are normal goods and the optimal database composition depends only on price ratios, not on the platform's budget. Which prices will prevail in the market is of course determined by the interplay of demand and supply. Under perfect competition, Dorfman, Samuelson, and Solow (1987) and Gale (1989) provide arguments for equilibrium prices to equal v_q^*.

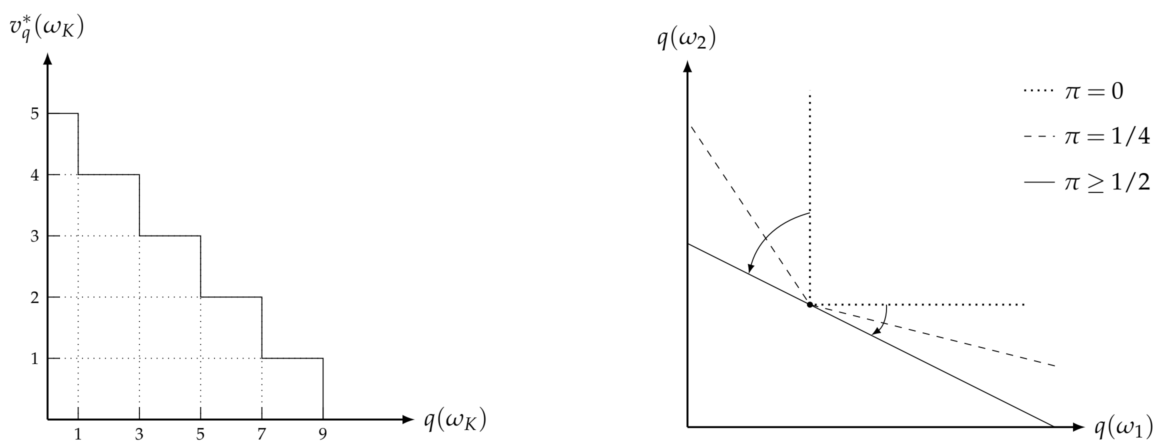
2.4.2 Application (Part II): Demand Curve and Substitutability

Returning to the setting of Section 2.3.3, recall that the platform maximizes a weighted sum of the buyers' surplus and the profits of a single price-setting seller,

[18] We slightly abuse notation by letting v_q^* stand for the *set* of optimal solutions at q. This condition is equivalent to $p \in \partial U^*(q)$, where $\partial U^*(q)$ is the superdifferential of U^* at q. Note that in the special case with a unitary budget and $p(\omega) = 1$ for all ω, choosing q is isomorphic to choosing an optimal prior in $\Delta(\Omega)$.

where the latter receives weight $\pi \in [0,1]$.

We first show an example of a downward-sloping demand curve. Figure 2.1(a) shows the value of records of type ω_K calculated using Proposition 2.2. This value is stepwise diminishing as these records become more abundant (Propositions 2.4 and 2.3). The figure also shows that as $q(\omega_K)$ becomes sufficiently large, $v_q^*(\omega_K)$ reaches a lower bound, which in this case is 0.



(a) Example of a demand curve: $\pi = 0$, $K = 10$, $\theta_k = k$ ($\forall k$), $q(\omega_k) = 1$ ($\forall k < K$).

(b) Example of indifference curves becoming less convex: $K = 2$, $\theta_k = k$ ($\forall k$)

Figure 2.1. Platform's demand and indifference curves.

Next, we explore how the substitutability between records depends on π. When $\pi \geq \frac{1}{2}$, all types of records are perfect substitutes and $MRS_q(\omega,\omega') = -\frac{\omega}{\omega'}$ for all q, which is thus constant in π. When $\pi < \frac{1}{2}$ instead, records become more substitutable as the platform cares less about the buyers' surplus (i.e., π increases). Recall that a_q is the seller's optimal price if he knows only the database composition q.

Corollary 2.4. *Fix q and increase $\pi < \frac{1}{2}$. If $\omega,\omega' < a_q$, $MRS_q(\omega,\omega')$ is constant at $-\frac{\omega}{\omega'}$. If $\omega < a_q \leq \omega'$, $MRS_q(\omega,\omega')$ increases monotonically toward $-\frac{\omega}{\omega'}$ from below. If $\omega' > \omega \geq a_q$, $MRS_q(\omega,\omega')$ decreases monotonically toward $-\frac{\omega}{\omega'}$ from above.*

In words, as π increases toward $\frac{1}{2}$, for record types on the opposite side of a_q the platform's indifference curves rotate counterclockwise in the direction of perfect substitutability. For records on the same side of a_q, its indifference curves rotate clockwise in the direction of perfect substitutes. Thus, the indifference curves become "less convex" around the dimension $\omega = a_q$. In particular, at $\pi = 0$ records of type $\omega = a_q$ are perfect complements with every other type. These patterns are illustrated in the right panel of Figure 2.1, which shows the platform's indifference curves in the case with two types of records.

2.4.3 Better Records and Willingness to Pay for Information

A platform can also change its database by refining some of its existing records with better information. For example, this could involve observing new personal characteristics about a subset of buyers. Intuitively, refining a record changes its type according to what the platform learns. How do such refinements change the value derived from each record? Do they always improve the platform's total payoff and consequently command a positive willingness to pay?

We first need to formalize what a refinement is. Recall that every buyer's record of type $\omega \in \Omega$ is informative about her θ, so it induces a belief $\beta_\omega \in \Delta(\Theta)$. A refinement of a record of type ω is a distribution $\sigma_\omega \in \Delta(\Omega)$ that satisfies the usual Bayes' consistency condition $\beta_\omega = \sum_{\omega' \in \Omega} \sigma_\omega(\omega')\beta_{\omega'}$. That is, any such refinement is equivalent to observing an exogenous signal that transforms a record of type ω into a record of type ω' with probability $\sigma_\omega(\omega')$. For instance, the original record may contain only the buyer's age, while the refined record may also contain her gender. When refining multiple records of type ω—in particular, a *share* $\alpha \in [0,1]$ of $q(\omega)$—each record is refined independently according to σ_ω.[19] Importantly, implicit in the definition there is the assumption that Ω is "rich" in the sense that it already contains all record types that can result from σ_ω. This allows us to use the platform's preferences over databases in $\mathbb{R}_+^\Omega$ characterized by v_q^* to assess the consequences of refinement σ_ω.

[19]We discuss refinements that are correlated among records in Section 2.5.1.

Consider refining a share α of type-ω records according to σ_ω. How does this change the unit value that the platform derives from its records? Such a refinement has both direct and indirect effects, as it affects both the records that are being refined and those that are not. The root of these interdependencies is the externality discussed in Section 2.3.2. Refining α of type-ω records changes the original database q into a new one, denoted by q_α, which contains fewer records of type ω and more records of the types ω' that result from the refinement (i.e., $\omega' \in \operatorname{supp} \sigma_\omega$). Thus, the unit value of the former records may increase and that of the latter decrease by the scarcity principle (Proposition 2.3). Formally, given $\alpha \in [0,1]$ and σ_ω, by the Law of Large Numbers $q_\alpha(\omega) = (1-\alpha)q(\omega)$ and $q_\alpha(\omega') = q(\omega') + \alpha\sigma_\omega(\omega')q(\omega)$ (where we can interpret $\alpha = 0$ as refining only one record since it is infinitesimal and $q_0 = q$). Note that the composition q_α of the new database is certain, even though it is uncertain which records of type ω become of type ω'. Thus, it suffices that the sellers know that a database q has been refined according to σ_ω and α for them to know the resulting composition q_α.

Corollary 2.5. *Fix q. Suppose a share α of type-ω records is refined according to σ_ω.*

> Direct Effects: *The value of* refined *records increases in expectation. That is, we have $\sum_{\omega' \in \Omega} v^*_{q_\alpha}(\omega')\sigma_\omega(\omega') - v^*_q(\omega) \geq 0$. This increase shrinks as α gets larger.*
>
> Indirect Effects: *The value of* unrefined *records of type ω increases: $v^*_{q_\alpha}(\omega) \geq v^*_q(\omega)$. The value of* unrefined *records of type $\omega' \in \operatorname{supp} \sigma_\omega$ decreases: $v^*_{q_\alpha}(\omega') \leq v^*_q(\omega')$. Both these effects are larger as α gets larger.*

With regard to the refined records, the expected gain in their value can shrink but never turn into a loss, even if refining more records lowers the value of the record types that result from it. Intuitively, for each refined record the platform knows more about the corresponding interaction, so it can better tailor its signals for the sellers and achieve more with that record. However, the externalities that contribute to its value may now be smaller. The former positive aspect dominates the latter because we are considering independent refinements. This is no longer true if refinements are correlated between records (see Section 2.5.1). Finally, note that the direct effects of a refinement depend on q, so the net value of the information it adds to a record cannot be quantified in absolute terms (unlike for standard-decision problems).

Corollary 2.5 highlights a novel implication of people's decisions to disclose their data. Recall that v_q^* is a benchmark for compensating buyers for their specific record. We can interpret a refinement as a buyer's decision of whether to disclose more of her personal characteristics. Imagine a group of similar buyers—that is, whose records are of the same type—which includes Ann and Bonnie. Ann decides to disclose, expecting that her record will become more valuable and hence may result in a higher compensation (direct effect). Bonnie instead decides *not* to disclose, yet her record may also become more valuable but for different reasons (indirect effect). Moreover, a larger group of disclosing buyers decreases Ann's expected gain in value, but increases Bonnie's. The disclosing buyers can also cause the value of, say, Cindy's record to fall—hence, lower her compensation—if her record is of one of the types that can result from the refinement (indirect effect). Importantly, these effects happen even if the platform does not learn anything new about Bonnie and Cindy from what Ann and the other buyers disclosed—in contrast to the learning externalities discussed in the literature (see Section 2.3.2).

Given these mixed effects of refinements on the unit value of all records, it is unclear whether they benefit the platform overall. In fact, those effects reflect a fundamental trade-off that refining records can generate in intermediation problems (but not in standard-decision problems). On the one hand, knowing more about each refined record allows the platform to better tailor its signals for the sellers and possibly achieve more in those interactions. On the other hand, changing the database q can change the sellers' beliefs about each buyer, which in turn can weaken the platform's informational advantage and hence its ability to influence the sellers' actions. A mechanism x may be obedient before the refinement but not after it, which can hurt the platform. Nonetheless, we obtain the following.

Proposition 2.6. *Fix q. Suppose a share α of type-ω records is refined according to σ_ω. The platform weakly benefits from this refinement: $U^*(q_\alpha) - U^*(q) \geq 0$. Moreover, the benefit is zero for all $\alpha \in [0,1]$ if (and only if generically in q) there exists $a \in supp\, x^*_q(\cdot|\omega'')$ for $\omega'' = \omega$ and all $\omega'' \in supp\, \sigma_\omega$. Finally, the refinement's marginal benefit decreases in α.*

This implies that the platform's willingness to pay for a refinement is always weakly positive, so the positive effects on refined records always dominate the negative effects on other records. The platform's willingness to pay can be strictly negative for refinements that are correlated between records (see Section 2.5.1).

Proposition 2.6 provides a sharp condition for the willingness to pay for refinements to be zero, which depends only on the initial q. Given this q, there must be a common action profile that the platform induces with positive probability both for the original record to be refined and for every type that it can turn into when refined. Intuitively, this means that the platform is exploiting its information advantage to sometimes use the original record as if it was already refined, so refining it does not make it more valuable. In fact, under this condition all direct and indirect effects in Corollary 2.5 are zero. Importantly, note that the direct effect of refining a record can be zero even if the platform uses it differently after the refinement (i.e., even if $x^*_q(\cdot|\omega) \neq x^*_q(\cdot|\omega')$ and $u^*_q(\omega) \neq u^*_q(\omega')$ for some $\omega' \in \operatorname{supp} \sigma_\omega$). Overall, the platform may be unwilling to pay a strictly positive price for refining its records, despite acting on the information it receives (i.e., changing x^*_q). This is different from standard-decision problems, for which a key insight is that more information is strictly beneficial if it changes the optimal choices.

Finally, Proposition 2.6 shows that the marginal benefit of a refinement is diminishing in the share of refined records. This may be reminiscent of classic results in standard decision problems where information has decreasing marginal returns (see, e.g., Moscarini and Smith, 2002; Varian, 2019). However, there is an important difference. Our exercise is not to gradually give the platform more information about one fixed interaction so that it can better learn the buyer's preferences. Focusing on this intensive margin is perhaps the most typical way of studying returns from information, especially in standard decision problems (see Bergemann and Ottaviani, 2021, Section 2.5). We instead fix the amount of information we give the platform for each interaction (i.e., σ_ω) and vary how many interactions we *independently* refine in this way (i.e., α). As such, this extensive-margin exercise has constant returns for standard-decision problems but not for intermediation problems—again, due to the externalities documented in Section 2.3.2.[20]

2.4.4 Application (Part III): Refinements

We illustrate some of these points using the setting of Section 2.3.3 with a single price-setting seller. Suppose the platform maximizes the buyers' surplus ($\pi = 0$). As before, ω_1 and ω_2 are the record types that correspond to buyers whose valuation θ is 1 and 2; instead, ω° corresponds to buyers' whose valuation is believed to be $\theta = 2$ with probability $h > \frac{1}{2}$ and $\theta = 1$ otherwise. Fix any q that satisfies $q(\omega^\circ) < q(\omega_1) < q(\omega_2)$ in which case we have that $v_q^*(\omega_1) = 1$, $v_q^*(\omega_2) = 0$, and $v_q^*(\omega^\circ) = 1 - h$. Now, suppose we refine a share α of type-ω° records with a refinement σ_{ω° such that $\sigma_{\omega^\circ}(\omega_2) = h$ and $\sigma_{\omega^\circ}(\omega_1) = 1 - h$. As shown in Table 2.1, the platform changes how it uses the refined records—compare $x_q^*(\cdot|\omega^\circ)$ and $x_q^*(\cdot|\omega_2)$—as well as the unrefined records of type ω_2—note that $x_q^*(\cdot|\omega_2)$ depends on q. Nonetheless, the "if" condition in Proposition 2.6 holds. Therefore, for any $\alpha \in [0,1]$ the platform's willingness to pay for the refinement as well

[20] In fact, given σ_ω the marginal effect of changing α on $U^*(q_\alpha)$ equals $\sum_{\omega' \in \Omega} v_{q_\alpha}^*(\omega')\sigma_\omega(\omega') - v_{q_\alpha}^*(\omega)$ (see the proof of Proposition 2.6 for details).

as the expected increase in unit value of each refined record are zero: $U^*(q) = U^*(q_\alpha)$ and $v^*(\omega_1)\sigma_{\omega^\circ}(\omega_1) + v^*(\omega_2)\sigma_{\omega^\circ}(\omega_2) = v^*(\omega^\circ)$. Both are instead strictly positive if $q(\omega_1) < q(\omega^\circ)$ and $\alpha > 0$ is sufficiently small. See Appendix A.2 for more details.

Table 2.1. Optimal x_q^*.

$x_q^*(a\|\omega)$	ω_1	ω_2	$\omega = \omega^\circ$
$a = 1$	1	$\frac{q(\omega_1)-(2h-1)q(\omega^\circ)}{q(\omega_2)}$	1
$a = 2$	0	$1 - \frac{q(\omega_1)-(2h-1)q(\omega^\circ)}{q(\omega_2)}$	0

2.5 Discussion

2.5.1 Correlation between Records and General Refinements

Throughout the paper we assumed that each buyer's record is uninformative about other buyers' preferences (i.e., records are independent). We did so to clarify that the interdependencies between the values of data records arise not from exogenous correlation, but endogenously from how the data is used. While this assumption may seem restrictive, it can be easily relaxed. For a fixed database, each buyer's record should already contain all the observations available to the platform that are relevant to that buyer, which may include variables that refer to other individuals. For example, if the average income in Ann's neighborhood is predictive of Ann's income, then it should be listed in Ann's record. Once this assignment is done for each buyer, conditional on her record any other record adds no information about her θ by construction. We can then replace our original independence assumption with this *conditional* independence assumption, and nothing changes in our analysis for a fixed database.

The possibility that one buyer's data is informative about other buyers has deeper implications with regard to refinements. For example, observing Bonnie's income may require updating her record as well as the record of her neighbor Ann. This introduces correlation in how records are updated, so it leads to more general refinements than those analyzed in Section 2.4.3. Nonetheless, we can continue to view such refinements as changing the platform's database and analyze their consequences using our tools (i.e., v_q^* over $\mathbb{R}_+^\Omega$). Ultimately, a refinement transforms the type of each affected record into a new one, so it changes the original q to another q'. This change can exhibit correlation between records; yet, for each of them the Bayes' consistency condition $\beta_\omega = \sum_{\omega' \in \Omega} \hat{\sigma}_\omega(\omega')\beta_{\omega'}$ must hold, where $\hat{\sigma}_\omega$ is the marginal distribution of the type changes for records of type ω.

Interestingly, correlated refinements can overturn some of the results from Section 2.4.3. For standard-decision problems, they always weakly increase both the records' unit values and the platform's total payoff. By contrast, for intermediation problems there are refinements that *decrease* the unit value of the refined records as well as the platform's total payoff. This is because they change more drastically not only what the platform knows about the buyers, but also the degree and nature of its information advantage over the sellers (recall that q is always commonly known).

Table 2.2. Value of records and total payoffs for specific databases ($\pi = 0$).

	ω_1	ω_2	ω°	
v_q^*	1	0	$1-h$	$U^*(q) = q(\omega_1) + (1-h)q(\omega_\circ)$
$v_{q'}^*$	0	1	–	$U^*(q') = q'(\omega_2)$
$v_{q''}^*$	1	0	–	$U^*(q'') = q''(\omega_1)$

We illustrate this possibility with an example. We use again the setting of Section 2.4.4 and assume that the platform maximizes buyers' surplus ($\pi = 0$). Let the initial q satisfy $q(\omega_\circ) < q(\omega_1) < q(\omega_2)$ and $q(\omega_2) < q(\omega_1) + (1-h)q(\omega^\circ)$. Consider the following refinement, which is arguably extreme but serves to make our point as clearly as possible. Suppose the platform learns that all its type-ω° records involve buyers with the same valuation. Thus, if refined, with probability $1-h$ they *all* become records of type ω_1 and with probability h they *all* become records of type ω_2. Thus, this refinement transforms the original database q into a new one: With probability $1-h$, the new database is q' and satisfies $q'(\omega_1) > q'(\omega_2) > q'(\omega^\circ) = 0$; with probability h, the new database is q'' and satisfies $q''(\omega_2) > q''(\omega_1) > q''(\omega^\circ) = 0$. Table 2.2 reports the value of each record for these databases (see Appendix A.2 for details). We find that the refinement has a strictly negative effect on both the unit value of the refined records and the platform's total payoff. Indeed, note that $v^*_q(\omega^\circ) > (1-h)v^*_{q'}(\omega_1) + hv^*_{q''}(\omega_2) = 0$ and $U^*(q) > U^*(q') > U^*(q'')$ because $q'(\omega_2) = q(\omega_2)$ and $q''(\omega_1) = q(\omega_1)$. By contrast, if the platform maximizes the seller's profits ($\pi = 1$), we have $v^*_{\hat{q}}(\omega_1) = 1$, $v^*_{\hat{q}}(\omega_2) = 2$, and $v^*_{\hat{q}}(\omega^\circ) = 2h$ for all $\hat{q}$. Thus, the same refinement has a strictly positive effect on both the unit value of the refined records and the platform's total payoff. The key is that a profit-maximizing platform treats each buyer-seller interaction as an independent decision problem, so it does not care about correlation in how it learns about records. Instead, a surplus-maximizing platform cares about such correlation, because it can have profound consequences on its information advantage through the composition of its database.

2.5.2 Standard-Decision Versus Intermediation Problems

It is instructive to briefly explain how our setting relates to the classic decision-theoretic framework of Blackwell (1951) and Blackwell (1953) and the ensuing literature (see, e.g., Laffont, 1989). In a standard decision problem there is an unknown state of nature $\theta \in \Theta$ that is drawn according to some distribution $\psi \in \Delta(\Theta)$. The decision maker observes an exogenous signal $\omega \in \Omega$ from a known experiment $e : \Theta \to \Omega$. Let $\tilde{u}_0(a,\theta)$ be the ex-post utility of the decision maker from the payoff-relevant action $a \in A$. Then, conditional on the signal realization ω, the expected payoff of the decision maker is $u_0(a,\omega) = \mathbb{E}_{\psi,e}(\tilde{u}_0(a,\theta)|\omega)$. Last but not least, the decision maker directly chooses a.

Our setting shares this framework's basic elements. Our intermediary is the analogue of the decision-maker, but faces a fundamentally different decision. Instead of directly choosing a, the intermediary has to decide what to disclose about the exogenous signal ω so as to influence another agent's choice of a. The capacity to influence this choice depends on the intermediary's information advantage over the agent and is constrained by the agent's incentives.

Moreover, our intermediary mediates a *collection* of such problems whose respective θ and ω have already realized. While the frequency of problems whose signal is ω is common knowledge, the agent cannot identify which ones these are. Only the intermediary can: A data record of type ω allows the intermediary to identify a problem whose signal realization was ω.[21] While mathematically this collection of problems is the analogue of the usual prior distribution, this "frequentist" approach has two advantages. First, it is more descriptive: It allows us to think of data records as physical objects rather than mutually exclusive possibilities. This is important if we want to think about data records as being traded based on their specific content, as it is often the case in data markets (Bergemann and Bonatti, 2019). Second, this approach allows us to ask natural and practical questions—such as the effects of adding more records or refining existing records—which would be artificial with the standard approach. We can think of refining an existing record as observing the realization of an experiment, in line with the tradition following Blackwell. Hence, we can view Proposition 2.6 as studying the "value of information" in intermediation problems.

2.5.3 General Intermediation Problems

Our framework and results apply more broadly to any setting where a principal mediates interactions between multiple agents using data. For ease of exposition, we simplified the model in several ways. Neither changes the analysis or its interpretations. First, we can allow the principal to also choose an action $a_0 \in A_0$ for each mediated interaction. In this case, a mechanism x also has to specify a_0 for each ω. Second, we

[21] Rather than simultaneously mediating all problems in the collection, an equally valid interpretation is that the intermediary commits to a mechanism for the whole collection and then problems are drawn independently and mediated one at a time.

can allow each agent i to also privately observe some data about the interaction he is in. For example, in our leading e-commerce example the seller can observe the quality or the history of customer reviews of his product. We can again model the realizations of such data with some finite set Ω_i, where each ω_i is ultimately an exogenous signal about some underlying payoff-relevant θ. Let $\Omega = \Omega_0 \times \ldots \times \Omega_n$ with typical element $\omega = (\omega_0, \ldots, \omega_n)$. The key assumption is that the principal also observes the private data of each agent—i.e., the entire $\omega = (\omega_0, \ldots, \omega_n)$—as does the omniscient designer in Bergemann and Morris (2016a). Thus, now the whole vector ω defines a type of data record in the principal's database and characterizes each interaction that it mediates. Our proofs already take into account this more general setting.

2.5.4 Additional Examples

We sketch other possible applications of our model.

Navigation Services

A navigation app uses data about routes' conditions to direct traffic by providing drivers with information—such as recommended routes and travel times. Das, Kamenica, and Mirka (2017) propose a simple way to model this complex problem. Suppose the app (principal) seeks to minimize congestion. We can think of an interaction as consisting of a group of drivers (agents) in some city who simultaneously choose, say, one of two routes between the residential and business district. For each route, the travel time increases in how many drivers choose it but at different rates (e.g., because one is a highway and one is surface streets); travel times also depend on some uncertain event (e.g., construction work), which is observed only by the app. For each city served by the app, the realization of this event defines its data record.

Ridesharing

A ridesharing platform mediates the interactions between n drivers and a population of potential riders who just landed at an airport. Each rider seeks to reach her final destination $\vartheta_d \in [0,1]$ and values a ride $\vartheta_v \geq 0$. Her preference is then pinned down by $\theta = (\vartheta_d, \vartheta_v)$. The platform knows ϑ_d and some personal characteristics of the rider, which are informative about ϑ_v. The drivers do not know anything about the riders and compete for them by posting a price a_i. Drivers have known preferences over final destinations—for instance, they are increasing in ϑ_d. Once an offer is accepted, the driver must honor it regardless of the final destination. The platform chooses what information about a rider's θ to disclose to the drivers so as to maximize a combination of the rider's and drivers' payoffs.

Online Advertisement

A population of individuals uses a search engine run by a platform. For each individual, it keeps a record that includes the searched keywords and some personal characteristics that are informative about her tastes in an horizontally differentiated product market, summarized by $\theta \in [0,1]$. There is a finite set $I \subseteq [0,1]$ of advertisers. The index $i \in I$ captures the advertiser's exogenous position in the product market—e.g., whether he advertises men's or women's apparel. The advertisers compete in a second-price auction to display an ad to each individual. Advertiser's i expected profits from winning access to an individual decreases in the distance between θ and i. The platform chooses which information about each individual's θ to disclose to the advertisers so as to maximize its total payoff.

2.6 Conclusion

This paper explains what determines the individual value of specific data records and what its properties are. In doing so, it advances our understanding of the demand side of data markets, thus shedding light on how they work and how they may be affected by regulatory interventions. This can provide insights into a key part of the digital economy. To the best of our knowledge, our approach to assessing the value of data has not been used before, it is broadly applicable, and it lays the foundations on which more questions can be tackled by future research. One direction is to fully analyze specific applications, such as e-commerce or the settings sketched in Section 2.5.4. Another is to explicitly model some of the privacy regulations discussed in policy circles, or to consider richer and possibly private ways in which intermediaries can change their databases.

Chapter 2 is currently being prepared for submission for publication of the material. Chapter 2 is coauthored with Galperti, Simone and Perego, Jacopo. The dissertation author, Aleksandr Levkun, is one of primary authors of this chapter.

Chapter 3

Strategic Mediation of Information in Autocracies

3.1 Introduction

The dominant model of dictatorship has evolved over the course of the twentieth century. The authoritarian states rely less on terror and ideology to make the citizens abide by ruler's political objectives than before.[1] Softer autocracies have emerged, including Russia, Venezuela, Ecuador, Turkey before the coup attempt in 2016, among others. These states no longer practice massive repressions. Instead, they hold elections and allow legal opposition in the attempts to imitate democracy (Gandhi and Lust-Okar, 2009).[2] These states seek to convince the population in ruler's competence to lead the country into a prosperous future (Guriev and Treisman, 2019). One of the main instruments for achieving this goal is an information manipulation through multiple channels including state-owned media.[3] The state-owned media consistently manipulate facts

[1]Potential reasons for this transformation are improved personal freedom, means of mobility, and the declined appeal of authoritarian ideologies since the end of the Cold War (Naim, 2022; Guriev and Treisman, 2019).

[2]As Brownlee (2007), p. 6, writes, "regimes have permitted opposition movements to contest elections but have stopped short of rotating power or allowing fair elections that would have risked their secure tenure in office".

[3]That is why such authoritarian states are sometimes referred to as informational autocrats. This term is used in Guriev and Treisman (2019) that provides an extensive overview of the inner processes in modern autocracies.

and censor information to influence citizens' beliefs about the ruler's competence. However, the state-owned media do not necessarily have direct access to the facts.[4] Instead, they have to rely on reports generated by a strategically-interested information source.[5] For example, reports could be research conducted by an independent statistical agency in hopes of providing the most accurate information to the general public. Reports could also surface from ruler's cabinet members acting in their own best interest. This paper seeks to pin down information sources that are useful to the media in their persuasive attempts. In the presence of a strategic source, how likely is swaying citizens' decisions toward the ruler's favor? What is the optimal editorial policy for the state-owned media and how does the policy depend on the preferences of citizens and the source?

To answer these questions, we consider a model of the optimal information disclosure by a state-owned media to a representative uninformed receiver. The receiver must decide between taking a mobilizing or the status-quo actions. Some examples of mobilization include voting for the ruler in the election, voting in favor of the ruler's proposal to change the Constitution, not revolting, not going to anti-government protests (Gehlbach and Sonin, 2014; Shadmehr and Bernhardt, 2015). The media want the receiver to take the mobilizing action.[6] However, the receiver prefers the mobilizing action only if the ruler's competence is high enough. The media do not have access to the ruler's competence and such information has to be supplied by the informed elite. The elite knows the state of the ruler's competence. This knowledge can come from an independent research, a proximity to the ruler, or an ability to understand political processes better than the receiver (Guriev and Treisman, 2018). The elite's ordinal

[4]This can happen due to asymmetric information, so that journalists require sources to gain access to information or help to interpret it. Egorov, Guriev, and Sonin (2009) and Lorentzen (2014) also indicate the need of authoritarian regimes for "watchdog reporting" to govern more efficiently.

[5]Schudson (2002) and Waisbord (2000), p. 108, point out the constraints that the media face due to the source's incentives to advance their own interests.

[6]The ruler needs to delegate the responsibility for reporting news to a designated institution. The ruler's interests are presented by the state-owned media with the same objective, even though the ruler does not participate in the game.

preferences are such that only if the elite observes that the ruler's competence is above a certain threshold, then she prefers the mobilizing action. The elite cannot communicate to the receiver directly. Instead, having learnt the competence, the elite sends a message to the media. The media then generate a report to the receiver. Finally, the receiver chooses an action based on the media's report. We study the media's problem under the commitment assumption. That is, at the beginning of the interaction, the media's editorial policy on how reports are generated from elite's messages is announced.

The media's optimal editorial policy is simple to describe. If the elite and the receiver disagree on the favorable action for a sufficiently large set of the ruler's competence levels, then the media cannot do better than providing no information and the receiver opts for the status-quo action.[7] Otherwise, the media signal whether the ruler's competence is higher or lower than a threshold which depends on the elite's preferences. By doing so, the media ensure that the elite reports information truthfully. If the ruler's competence is high enough then the editorial policy suggests the mobilizing action to the receiver. Otherwise, the status-quo action is suggested with some probability. This probability is calibrated to make the receiver indifferent between two actions. We show that if the receiver becomes more critical of the ruler, that is, he requires a higher ruler's competence to oblige with choosing the mobilizing action, then the media are worse off. It becomes harder for the media to convince the receiver to choose the mobilizing action. We also show that the elite generally benefits when the receiver becomes more critical. As discussed in Guriev and Treisman (2019), highly educated citizens in the authoritarian states tend to be more critical toward their government. Thus, as a prediction of this model, the spread of higher education would make it harder for the media to sway the receiver toward the mobilizing action and make the informed elite better off. The media benefit from the elite that is more aligned with the media as long as the

[7] The intuition is similar to the one in sender-receiver games, where if the sender's bias is too large, then the equilibrium is necessarily uninformative.

alignment is not too close: reports generated from the source having a similar objective to the media's lead to the receiver's skepticism.

The media need to incentivize the elite to supply information by designing the editorial policy tailored to the elite's preferences. Such honesty constraints discipline the editorial policy and render some information sources to be futile to the media. Our assumptions on players' payoffs describe a range of situations in which the media attempt to influence citizens' beliefs, e.g., a voting application.[8] The elites are directly interested in the election outcome, and their preferences determine informativeness of communication. This paper helps to understand the relationships between the informed elites and the state-owned media in modern autocracies. In particular, we show that the elite closely aligned to the ruler's objective is fruitless to the state-owned media if the receiver is able to correctly make inferences. Importantly, even if the elite is heavily critical of the ruler, the state-owned media can still weaponize elite's information to persuade the receiver.

The initial analysis assumes that the population is identical in how critical they are, and hence it can be represented by a single representative receiver. However, in reality, the population is heterogeneous in pickiness towards the ruler, even in authoritarian states.[9] This observation leads us to consider the state-owned media attempting to sway decisions of the population of receivers having private information about their pickiness. Without honesty constraints, the state-owned media implement the upper censorship policy revealing low levels of the ruler's competence and pooling high levels (Kolotilin et al., 2017). Honesty constraints reduce the media's welfare, since the upper censorship policy cannot be implemented. Revealing all low states of the

[8]For example, see Alonso and Câmara (2016).

[9]For an example of such heterogeneity in citizens' preferences, Russian independent pollster Levada center reports respondents' answers to the question "Do you approve the decisions of Vladimir Putin as the president of Russia?" in October 2019. 26 % of the respondents answered "Yes, absolutely", 44 % answered "Rather yes, than no", 18 % answered "Rather no, than yes", 10 % answered "Absolutely not", 2 % abstained (Levada Center, 2019). See also Neundorf, Gerschewski, and Olar (2020).

ruler's competence is not available due to incentives of the low types of the source. We characterize the solution to the media's problem choosing over the restricted simple class of editorial policies with a support having at most two elements. The optimal simple policy provides the lower bound for the media's unconstrained problem. We show the sufficient condition on the distribution of receiver's private information, under which this lower bound is attained. We illustrate the lower bound for the unimodal distribution of receiver's types.

Related literature

This paper is a part of an active literature that studies *strategic information dissemination decisions by the media* concerned with swaying the beliefs of its audience. Guriev and Treisman (2018) present the model of the informational autocracy that yearns for staying in power. There is an elite that is informed about the ruler's competence. The authors establish conditions under which manipulation of information provided by the elite is more beneficial for the ruler than opting for repressions or improving living standards. Relative to Guriev and Treisman (2018), in our model the only way for the elite to communicate with the public is through the state-owned media. The media do not generate information itself, and transfers are not allowed. The manipulation of information is the only available instrument to the media. We characterize the effectiveness of this instrument depending on the preferences of the elite, the media, and the population.

Most of the papers in the literature assume that news is exogenous in the sense that it is a realization of the payoff-relevant random variable for the audience. Shadmehr and Bernhardt (2015) explore a ruler's decision of whether to censor information available to citizens to avoid a revolution. The citizen's net payoff from successful revolution depends on the news that can be censored by the ruler. The authors characterize the ruler's censorship strategy and show that the ruler is better off by committing to censor less than he does in the equilibrium. Duggan and Martinelli (2011) consider the election model with an incumbent and a challenger in which the media can affect the public opinion. The state of the world is the challenger's policy on a level of public good provision and an income tax rate. As in this paper, the media can commit to how it systematically distorts the information about the challenger's policy. The authors characterize the choice of the media slant for pro-incumbent and pro-challenger media. The slanting technology is fixed and represented by the projection of the two-dimensional policy on a straight line with a slope representing the media slant. Instead, this model considers a general slanting technology for a one-dimensional state and allows the information supplier to be strategic.

Chiang and Knight (2011) and Gehlbach and Sonin (2014) provide empirical evidence that voters take media's bias into account when forming beliefs about political candidates. Gehlbach and Sonin (2014), Duggan and Martinelli (2011), and Gentzkow, Shapiro, and Stone (2015) adopt the assumption of the media's commitment power to a probabilistic information structure, as this paper does. This assumption captures the government's need to delegate responsibility for reporting news to correspondents, reporters, and editors who make frequent decisions about the framing of the news they decide to cover.

This paper also contributes to the literature on *mediated cheap talk*, which studies communication between an informed sender and an uninformed receiver through the mediator. The informed party makes a report to the mediator, who then makes a non-binding recommendation to the receiver. The literature focuses on the optimal mediation for the sender and the receiver. The optimal mediation generally adds noise to communication. Goltsman et al. (2009) characterize the optimal mediation for the uniform-quadratic setup of the cheap-talk game of Crawford and Sobel (1982). Blume, Board, and Kawamura (2007) analyze the special case of the mediation protocol: with some probability, the sender's message is transmitted perfectly to the receiver; with the remaining probability, the noisy message is generated. The authors show that noise generally improves welfare. These papers concentrate on the neutral mediator and characterize the best mediator for the sender and the receiver. In this paper, the media play the role of the mediator and have a strategic objective, specifically, to increase the probability of the mobilizing action chosen by the receiver. We analyze the optimal mediation plan for different assumptions on the sender's and receiver's preferences. Within the uniform-quadratic setup, Ivanov (2010) shows that that there is no welfare loss if the strategic mediator is chosen properly. Compared to this paper, we assume the media's ability to commit to the mediation plan. Therefore, the media in this paper will generally obtain a higher payoff. Salamanca (2021) studies the informed party that is able to choose and commit to the mediation plan. The author characterizes sender's value function under the assumption of aligned sender's and mediator's preferences. In this paper, if the information source and the mediator are aligned, then the only equilibrium is completely uninformative. Skreta and Perez-Richet (2021) consider a game with a similar extensive form to this paper. The mediator designs a test to detect the costly falsification of the state of the world by the sender.[10] The main difference of Skreta and Perez-Richet (2021) and this paper is the payoff structure: in Skreta and Perez-Richet (2021) the mediator designs the test to maximize the receiver's payoff while the sender wishes the receiver to take a certain action.[11]

This paper also contributes to the *constrained information design* literature. This literature seeks to extend the standard Bayesian persuasion framework of Kamenica and Gentzkow (2011) by adding meaningful constraints the persuading side has to face. Le Treust and Tomala (2019) and E. Tsakas and N. Tsakas (2021) study the setup where

[10] Skreta and Perez-Richet (2021) cover the case of the costless falsification that corresponds to cheap-talk messages.

[11] Ball (2019) introduces the model of predictive scoring. The scoring agency with commitment power aggregates multiple features of the sender into a score. The sender's features are correlated with the state that the receiver wishes to match. The sender is able to distort her features at a cost. Similarly to Skreta and Perez-Richet (2021), the scoring agency is aligned in preferences with the receiver.

the persuading side communicates with the receiver through a channel that is subject to exogenous noise. The optimal payoff is characterized as a function of the Shannon channel capacity in Le Treust and Tomala (2019).[12] Lipnowski and Mathevet (2018) impose the behavioral assumptions on the receiver that leads to non-Bayesian updating and analyze the optimal information disclosure. Compared to these papers, the media in the role of a persuading side face novel constraints capturing the media's inability to access the state directly. Instead, the media incentivize the source to supply information by carefully designing the information protocol. Boleslavsky and K. Kim (2018) consider the setup where the agent exerts a privately observed effort that determines the state distribution. Thus, the persuader has to not only persuade the receiver to take some action, but also incentivize the agent's effort. Boleslavsky and K. Kim (2018) has the additional constraints in the form of moral hazard, whereas in this paper the honesty constraints correspond to the adverse selection problem.

3.2 Model

This section introduces a game between an informed elite, which I call a *source* (S, she), a state-owned *media* (M, it), and an uninformed *receiver* (R, he).

The receiver has to decide whether to undertake the status-quo action a_0, or the mobilizing action a_1. The mobilizing action corresponds to some political objective of the ruler. The payoff of the receiver $u_R(a,\theta)$ depends on his action $a \in A = \{a_0, a_1\}$ and the ruler's competence $\theta \in \Theta = [0,1]$. The ruler's competence is unknown to the receiver but he holds a prior μ_0 on Θ that is common to all the players. The receiver's preference parameter $\delta_R(\theta) = u_R(a_1,\theta) - u_R(a_0,\theta)$ captures the receiver's net payoff

[12] The technique developed in Le Treust and Tomala (2019) and Doval and Skreta (2018) corresponds to rewriting the additional constraints as a function of receiver's posterior beliefs distribution. However, in our problem, the state space is continuous and honesty constraints have to be satisfied for every pair of states. Even though the inequalities corresponding to honesty conditions can be written in accordance with Doval and Skreta (2018), their method does not make the problem tractable.

from the mobilizing action, and the function $\delta_R(\theta)$ is assumed to be strictly increasing. That is, the receiver prefers the mobilizing action if the ruler's competence is high enough. Moreover, we assume the following *tension* condition:

$$\int_0^1 \delta_R(\theta) d\mu_0 < 0. \tag{3.1}$$

This condition indicates that under the prior the receiver opts for the status-quo action.

A source perfectly learns the ruler's competence and cares about the receiver's action. The source is referred to as type-θ source if she learns that the ruler's competence is θ. The source's payoff function $u_S(a,\theta)$ is summarized by two measurable sets representing source's ordinal preferences, $\Theta_0 = \{\theta \in [0,1] : u_S(a_0,\theta) > u_S(a_1,\theta)\}$ and $\Theta_1 = \{\theta \in [0,1] : u_S(a_1,\theta) > u_S(a_0,\theta)\}$. In words, Θ_0 captures the source types that strictly prefer the status-quo action, whereas Θ_1 captures the source types that strictly prefer the mobilizing action. The measure of types that are indifferent between a_0 and a_1 is assumed to be zero.[13] The source can only communicate with the receiver indirectly, by sending a costless message $m \in \mathcal{M}$ to the state-owned media. The set of messages $\mathcal{M}$ has at least as many elements as Θ.

The state-owned media wish to promote the ruler's interests. In particular, the media want the receiver to undertake the mobilizing action irrespective of the ruler's competence. The media's payoffs of the status-quo action and the mobilizing action are normalized to 0 and 1, respectively. Therefore, the media's expected payoff is simply the probability of the mobilizing action being chosen. The media can communicate with the receiver but cannot generate information itself. Instead, information has to be provided to the media by the source in the form of message $m \in \mathcal{M}$. The media then produce a costless report $r \in \mathcal{R}$ observed by the receiver. The set of reports $\mathcal{R}$ has at

[13] If the measure of indifferent source types is nonzero, the media's problem is relaxed in the sense that the media have to satisfy fewer incentive constraints on the source's side.

least two elements.

We assume that the media can commit to how the reports are generated based on the messages provided by the source. In particular, at the beginning of the game, the state-owned media publicly choose an editorial policy $\pi : \mathcal{M} \to \Delta(\mathcal{R})$, where $\pi(r|m)$ is the probability of generating report r after observing the source's message m. We refer to the editorial policy π as a strategic dissemination protocol, or simply a *protocol*.

Timing

We summarize the timing of the game. The game starts with the state-owned media committing to the strategic dissemination protocol, $\pi : \mathcal{M} \to \Delta(\mathcal{R})$, observed by all players. The ruler's competence $\theta \in [0,1]$ then realizes as the draw from the distribution μ_0. The source observes θ and π and sends a costless message $m \in \mathcal{M}$ to the media. The report $r \in \mathcal{R}$ is then generated by the media as the draw from the distribution $\pi(\cdot|m)$. Finally, the receiver observes the protocol π and the report r, forms the posterior belief μ, and decides whether to undertake the status-quo action a_0 or the mobilizing action a_1. The payoffs then are realized.

Equilibrium

An *equilibrium* consists of four measurable maps: a messaging strategy $\rho : \Theta \to \Delta(\mathcal{M})$ for S, an information dissemination protocol $\pi : \mathcal{M} \to \Delta(\mathcal{R})$ for M, a probability of choosing the mobilizing action $\alpha : \mathcal{R} \to [0,1]$ for R, and a belief mapping $\mu : \mathcal{R} \to \Delta(\Theta)$ for R. An equilibrium is the protocol π chosen by M and a perfect Bayesian equilibrium of the subgame that follows the M's choice. Specifically, given π, a perfect Bayesian equilibrium (PBE) is a tuple (ρ, α, μ) that satisfies

1. (belief formation)

   μ is obtained from μ_0 via Bayes' rule, given ρ, whenever well-defined;

2. (receiver's best-response)

$\alpha(r) = 1$ if $\int_\Theta \delta_R(\cdot) d\mu(\cdot|r) > 0$, and $\alpha(r) = 0$ if $\int_\Theta \delta_R(\cdot) d\mu(\cdot|r) < 0$;

3. (sender's best-response)

   $\rho(\theta)$ is supported on $\underset{m \in \mathcal{M}}{\operatorname{argmax}} \int_\mathcal{R} [u_S(a_1,\theta)\alpha(\cdot) + u_S(a_0,\theta)(1-\alpha(\cdot))] \, d\pi(\cdot|m)$ for every $\theta \in \Theta$.

Following the information design literature[14], the agent with a commitment power is assumed to be able to steer other agents toward her favorite PBE. Thus, for every π, M chooses a PBE that generates the highest media's ex ante payoff denoted as $V(\pi)$. Finally, the equilibrium π is chosen to maximize $V(\pi)$. We denote the value function of the media as V. This completes the definition of the equilibrium.

3.3 Equilibrium Analysis

We analyze the model using revelation principle. It is without loss to focus on the direct protocols where the source truthfully reports the ruler's competence and the receiver obediently follows an action recommendation. We characterize the set of the direct protocols that satisfy honesty and obedience conditions. Finally, we solve for the optimal protocol for the media and discuss its properties.

A protocol π is said to be direct if $\mathcal{M} = \Theta$ and $\mathcal{R} = A$. That is, for a direct protocol π, the source is asked to report a competence level θ and the media make a binary action recommendation to the receiver.

In a direct protocol, S is said to be *honest* if it is optimal for her to report the ruler's competence truthfully. R is said to be *obedient* if it is optimal for him to follow a recommendation. A direct protocol $\pi : \Theta \to \Delta(A)$ is Bayesian incentive-compatible if S

[14] See Kamenica and Gentzkow (2011), Alonso and Câmara (2016), and Bergemann and Morris (2016b) among others. Mathevet, Perego, and Taneva (2020) develop the methodology of analyzing persuasion problems for various equilibrium selection rules.

is honest and R is obedient. Specifically, S is honest given R's obedience if

$$u_S(a_1,\theta)\pi(a_1|\theta) + u_S(a_0,\theta)\pi(a_0|\theta) \geq u_S(a_1,\theta)\pi(a_1|\theta') + u_S(a_0,\theta)\pi(a_0|\theta') \qquad (3.2)$$

for every $\theta,\theta' \in \Theta$. We call a direct protocol π *honest* if it satisfies (3.2).

R is obedient given S's honesty if following recommendation a_1 is optimal, that is,

$$\int_0^1 \delta_R(\theta)\pi(a_1|\theta)d\mu_0 \geq 0, \qquad (3.3)$$

and following recommendation a_0 is optimal, that is,

$$-\int_0^1 \delta_R(\theta)\pi(a_0|\theta)d\mu_0 \geq 0. \qquad (3.4)$$

Note that the tension condition (3.1) and the inequality (3.3) imply the inequality (3.4). Then we call a direct protocol π *obedient* if it satisfies (3.3).

The revelation principle states that without loss the media can focus on the direct protocols that are honest and obedient.

Lemma 3.1. *Given any PBE in the original game followed by π, there exists an incentive-compatible direct protocol π^* under which the media get the same expected utility when S is honest and R is obedient as in this PBE.*

All proofs are in the appendix. By Lemma 3.1 an optimal incentive-compatible direct protocol is also an equilibrium in the class of all possible protocols. From now on we focus on the characterization of the optimal incentive compatible direct protocol.

It turns out that the honesty constraint significantly simplifies the problem by disciplining any incentive-compatible direct protocol. Lemma 3.2 summarizes this observation.

Lemma 3.2. *A direct protocol π is honest if and only if there exist π_0 and π_1, with $\pi_0 \leq \pi_1$, such that $\pi(a_1|\theta) = \pi_0$ for every $\theta \in \Theta_0$, and $\pi(a_1|\theta) = \pi_1$ for every $\theta \in \Theta_1$.*

Thus, an honest direct protocol is characterized by a pair of numbers π_i, $i \in \{0,1\}$, capturing the probability that the receiver takes the mobilizing action when the ruler's competence is in Θ_i. Intuitively, the source will always attempt to induce the highest probability of her preferred action. Therefore, the probability of a_i prescribed by protocol π has to be identical across all $\theta \in \Theta_i$. Furthermore, the source types that prefer action a_1 have to be provided with a higher probability π_1 of implementing this action compared to probability π_0 provided to the types that prefer action a_0.

To find the set of incentive-compatible protocols, we combine the insight of Lemma 3.2 with the obedience condition. To this end, define $I_0 = \int_{\Theta_0} \delta_R(\theta) d\mu_0$ and $I_1 = \int_{\Theta_1} \delta_R(\theta) d\mu_0$. In words, I_0 and I_1 capture receiver's preferences in conjunction with source's preferences. It turns out that these statistics are sufficient to pin down the set of incentive-compatible protocols. Note that $\frac{I_i}{\mu_0(\Theta_i)}$, $i \in \{0,1\}$, is the expectation of the receiver's net payoff from the mobilizing action conditional on the competence being in Θ_i, $\mathbb{E}[\delta_R(\theta)|\theta \in \Theta_i]$. Proposition 3.1 characterizes the set of incentive-compatible direct protocols depending on the signs of the conditional expectations $\mathbb{E}[\delta_R(\theta)|\theta \in \Theta_0]$ and $\mathbb{E}[\delta_R(\theta)|\theta \in \Theta_1]$.

Proposition 3.1. *The set of incentive-compatible direct protocols $\mathcal{I}$ is characterized by (π_0, π_1), $\pi_0 \leq \pi_1$, where π_i is the probability of implementing action a_1 following any $\theta \in \Theta_i$. Furthermore,*

- *if $I_0 < 0$ and $I_1 \geq 0$, then $\mathcal{I} = \{(\pi_0, \pi_1) \in [0,1]^2 : \pi_1 I_1 + \pi_0 I_0 \geq 0\}$;*
- *otherwise, $\mathcal{I} = \{(0,0)\}$.*

To get the intuition behind this result, first, consider the case of $\mathbb{E}[\delta_R(\theta)|\theta \in \Theta_1] < 0$. In words, this inequality means that there is a sufficiently large portion of the source types who disagree with the receiver on the preferable action: they prefer a_1, whereas the receiver would choose a_0 if the ruler's competence was known. The receiver is too skeptical of the statements of the source closely aligned with the media in trying to assure the mobilizing action.[15] Thus, in this case the only incentive-compatible protocol involves pooling: $\pi_0 = \pi_1 = 0$.[16] Otherwise, the information about whether the ruler's competence is in Θ_0 or Θ_1 can be meaningfully transmitted to the receiver. Furthermore, Proposition 3.1 shows that in this case the obedience constraint implies the requirement $\pi_0 \leq \pi_1$ of the honesty constraint.

Figure 3.1 depicts the set of incentive-compatible protocols for the case when complete pooling is not the only incentive-compatible protocol. This figure is similar to the set of obedient protocols in the standard Bayesian persuasion problem with a binary state (see, for example, Bergemann and Morris, 2019). Here, however, the state space is continuous. The honesty constraints discipline the protocol over the states in Θ_0 and Θ_1. Thus, the binary state in the standard Bayesian persuasion problem can be seen as whether the state in our problem lies in Θ_0 or Θ_1.

Proposition 3.1 paves the way to finding the optimal media's protocol. Indeed, given the previous insights, the media's problem can be written as follows:

$$V = \max_{(\pi_0, \pi_1) \in \mathcal{I}} \{\mu_0(\Theta_0) \cdot \pi_0 + \mu_0(\Theta_1) \cdot \pi_1\}.$$

Thus, the media maximize the linear function over the set of incentive-compatible protocols $\mathcal{I}$ defined by the linear inequalities and given by Proposition 3.1. Then there necessarily is an extreme point solution. Proposition 3.2 finds this extreme point.

[15] By the standard intuition from sender-receiver games, if the conflict between sender's and receiver's preferences is too large, then no meaningful information can be transmitted. For example, in the uniform-quadratic setup of the cheap-talk game of Crawford and Sobel (1982), if the sender's bias is too large, then the equilibrium is necessarily completely uninformative.

[16] Note that even though $\pi_0 = \pi_1 = 0$, communication is not necessarily uninformative. However, all receiver's posterior beliefs generated by this protocol are low enough so that the status-quo action is always taken.

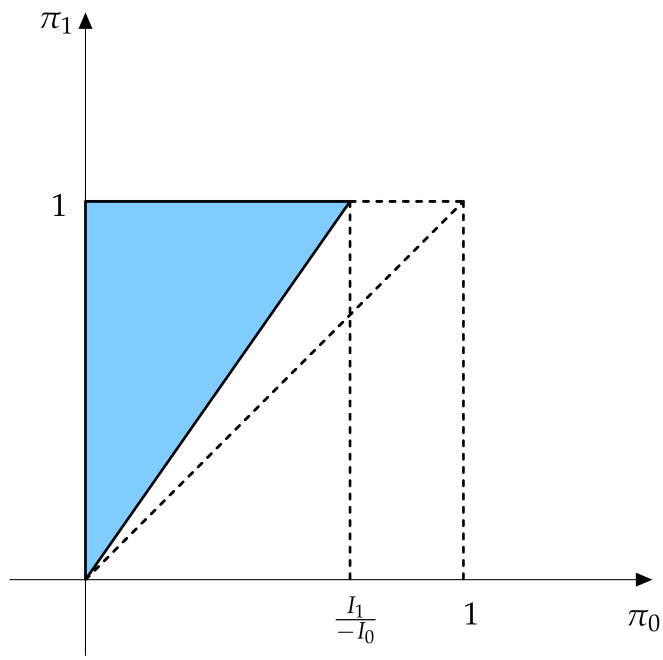


Figure 3.1. The set of incentive-compatible direct protocols, when $\mathbb{E}[\delta_R(\theta)|\theta \in \Theta_0] < 0$ and $\mathbb{E}[\delta_R(\theta)|\theta \in \Theta_1] > 0$.

Proposition 3.2. *If $I_0 < 0 \leq I_1$, then the solution to the media problem is the pair (π_0, π_1) such that $\pi_0 = \pi(a_1|\theta)$ for every $\theta \in \Theta_0$, $\pi_1 = \pi(a_1|\theta)$ for every $\theta \in \Theta_1$, and $(\pi_0, \pi_1) = \left(\frac{I_1}{-I_0}, 1\right)$. The ex ante media's payoff is*

$$V = \mu_0(\Theta_1) \cdot \frac{\mathbb{E}[\delta_R(\theta)|\theta \in \Theta_1] - \mathbb{E}[\delta_R(\theta)|\theta \in \Theta_0]}{-\mathbb{E}[\delta_R(\theta)|\theta \in \Theta_0]}.$$

If $I_0 < 0 \leq I_1$ is not satisfied then $\pi_0 = \pi_1 = 0$ and the media's payoff is 0.

When the conflict between source's and the receiver's preferences is too large, the media cannot do better than making the protocol completely uninformative. Otherwise, the media signal whether the ruler's competence lies in Θ_0 or Θ_1. From the receiver's perspective, if the receiver gets the recommendation of the status-quo action, then he knows with certainty that the ruler's competence is in Θ_0. On the other hand, the recommendation of the mobilizing action can come from any competence level, but the protocol renders the probabilities in the exact way to make the receiver indifferent between two actions.

3.3.1 Example

In order to gain insight about properties of the media-optimal protocol, we impose the functional forms for the players' payoffs. We assume that the receiver's payoff function is $u_R(a,\theta) = 1\{a = a_1\} \cdot (\theta - \omega)$, where $\omega \in [0,1]$ is interpreted as receiver's pickiness commonly known to the players. The pickiness level corresponds to how critical the receiver is of the government. A pickier receiver would require a higher ruler's competence to oblige with choosing the mobilizing action. If the ruler's competence surpasses ω, then the receiver prefers the mobilizing action. The receiver's net payoff from the mobilizing action is then simply a linear function $\delta_R(\theta) = \theta - \omega$. By the tension condition (3.1), $\omega > \frac{1}{2}$. If $\omega \leq \frac{1}{2}$, then the media make the protocol completely uninformative and extracts the payoff of 1.

Furthermore, we assume that Θ_0 and Θ_1 are half-intervals: $\Theta_0 = [0,\bar{\theta})$ and $\Theta_1 = (\bar{\theta},1]$, where $\bar{\theta} \in [0,1]$ is the relevant source's payoff parameter commonly known to the players. We refer to $\bar{\theta}$ as a source's *threshold*. That is, all the source types above $\bar{\theta}$ strictly prefer the mobilizing action, whereas all the source types below $\bar{\theta}$ strictly prefer the status-quo action. The source with type $\bar{\theta}$ is indifferent between a_0 and a_1. If $\bar{\theta} = 0$, then the source is aligned in preferences with the media. If $\bar{\theta} = \omega$, then the source is aligned in preferences with the receiver. Finally, for the sake of exposition, let the prior μ_0 be the uniform distribution on $[0,1]$.

Note that in the general case covered before, there may coexist two sets of the source types that do not want their information to be revealed: types in Θ_0 that know that the ruler's competence exceeds ω and types in Θ_1 that know that the ruler's competence is below ω. The example leaves only one of these sets present depending on the relation between $\bar{\theta}$ and ω. Threshold $\bar{\theta}$ can be interpreted as the source's pickiness level that is potentially different from the receiver's ω.

For this example, the summary statistics I_0 and I_1 of source's and the receiver's preferences can be directly calculated as

$$I_0 = \int_0^{\bar{\theta}} (\theta - \omega) d\theta = \frac{\bar{\theta}(\bar{\theta} - 2\omega)}{2},$$

$$I_1 = \int_{\bar{\theta}}^1 (\theta - \omega) d\theta = \frac{(1 - \bar{\theta})(1 + \bar{\theta} - 2\omega)}{2}.$$

Then Proposition 3.2 readily establishes the optimal protocol and the media's ex ante payoff from this protocol.

Claim 3.1. *If* $\bar{\theta} < 2\omega - 1$*, then the optimal protocol is* $\pi_0 = \pi_1 = 0$*, resulting in the media's payoff of* 0*. If* $\bar{\theta} \geq 2\omega - 1$*, then the optimal protocol is*

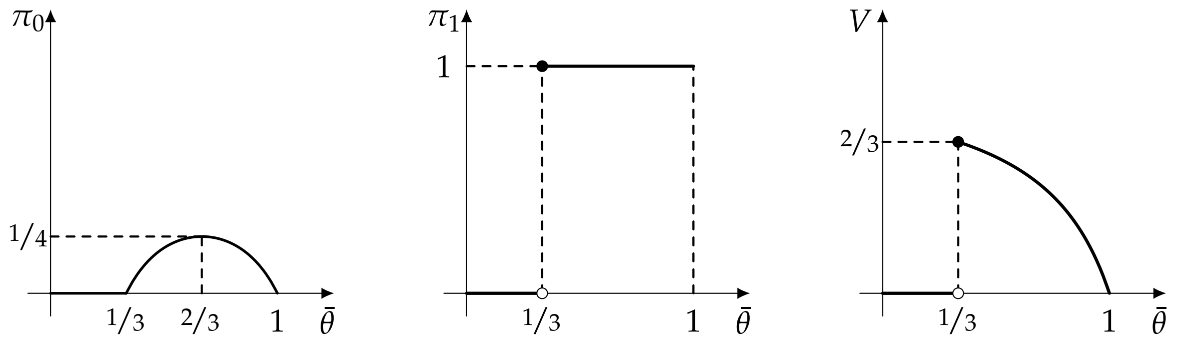
$$(\pi_0, \pi_1) = \left(\frac{(1 - \bar{\theta})(1 + \bar{\theta} - 2\omega)}{\bar{\theta}(2\omega - \bar{\theta})}, 1 \right),$$

and the media's payoff is

$$V = \frac{1 - \bar{\theta}}{2\omega - \bar{\theta}}.$$

Figure 3.2 shows the media's optimal protocol and the value function for $\omega = \frac{2}{3}$ for various source's thresholds $\bar{\theta}$.

Given Claim 3.1, it is straightforward to derive the relevant comparative statics with respect to $\bar{\theta}$ and ω. A receiver with pickiness ω' is said to be more aligned with the media than a receiver with pickiness ω'' if $\omega' < \omega''$. Similarly, a source with threshold



(a) The probability of implementing a_1 if $\theta \in \Theta_0$.
(b) The probability of implementing a_1 if $\theta \in \Theta_1$.
(c) The media's ex ante payoff.

Figure 3.2. The solution to the media's problem for $\omega = 2/3$ as a function of $\bar{\theta}$.

$\bar{\theta}'$ is said to be more aligned with the media than a source with threshold $\bar{\theta}''$ if $\bar{\theta}' < \bar{\theta}''$.

Claim 3.2. *The media get the ex ante payoff of* 0 *if* $\bar{\theta} < 2\omega - 1$*. As long as* $\bar{\theta} \geq 2\omega - 1$*, the media's ex ante payoff increases if the receiver or the source become more aligned with the media.*

It is easier for the media to persuade a more aligned receiver to undertake the mobilizing action. For a more aligned source, there are two effects. First, the measure of types Θ_1 that guarantees the implementation of action a_1 goes up. Second, for the source types from Θ_0, the probability of implementing the mobilizing action is increasing in $\bar{\theta}$ for $\bar{\theta} < \omega$ and decreasing for $\bar{\theta} > \omega$. When π_0 is increasing in $\bar{\theta}$, the first effect outweighs the second effect. Therefore, the media's payoff is lower for the pickier source as long as $\bar{\theta} \geq 2\omega - 1$. To summarize, if the source is closely aligned with the media, $\bar{\theta} < 2\omega - 1$, the media's payoff stays at zero. When $\bar{\theta}$ reaches the level of $2\omega - 1$, the media's payoff jumps to $2 - 2\omega$ and then starts to decrease to 0 with increasing $\bar{\theta}$. This effect is illustrated in Figure 3.2c.

Observe that the receiver's ex ante payoff is always zero irrespective of the source's preferences. Indeed, when the protocol never recommends the mobilizing action, the receiver undertakes the status-quo action and gets zero payoff. When the source is not too closely aligned with the media, the media ensure that the receiver is kept indifferent between the status-quo and mobilizing actions upon receiving the recommendation of the mobilizing action, so that the ex ante payoff of the receiver is again zero.

Claim 3.3 establishes a comparative statics of the source's payoff as a function of the receiver's pickiness.

Claim 3.3. *Suppose that $\bar{\theta} \geq 2\omega - 1$. The source types from Θ_1 get their favorite action a_1 with probability 1 irrespective of ω. The source types from Θ_0 get their favorite action a_0 with probability $1 - \pi_0$ that is increasing in ω. That is, the source types from Θ_0 are worse off when the receiver is more aligned with the media.*

Hence, by Claim 3.3, ex ante (before learning the ruler's competence) the source is worse off when the receiver is more aligned with the media. Indeed, the source types from Θ_1 always get their favorite action. The source types from Θ_0 benefit from a receiver with greater ω as it becomes harder for the media to persuade the receiver to undertake the mobilizing action.

Finally, we compare the media's optimal protocol to the protocol in the standard Bayesian persuasion problem, that is, the problem with no honesty constraints.

Claim 3.4. *The solution to the media's problem facing no honesty constraints is*

$$\pi(a_1|\theta) = \begin{cases} 1, & \text{if } \theta \geq 2\omega - 1, \\ 0, & \text{otherwise.} \end{cases}$$

The media's ex ante payoff is equal to $2 - 2\omega$.

The optimal protocol with honesty concerns achieves the payoff of the media facing no honesty constraints when $\bar{\theta} = 2\omega - 1$. In other words, the media-optimal source's threshold corresponds to the threshold on the Bayesian persuasion protocol that renders the receiver to be indifferent between a_0 and a_1. If the source's threshold $\bar{\theta}$ falls below $2\omega - 1$, the media's payoff drops to zero. This discontinuity is the consequence of the discontinuity of the set of incentive-compatible protocols in $\bar{\theta}$. When $\bar{\theta} < 2\omega - 1$, only the protocol that never recommends a_1 is available. At $\bar{\theta} = 2\omega - 1$, the Bayesian persuasion protocol becomes available and is employed by the media. Importantly, even when $\bar{\theta} = 2\omega - 1$, the media facing no honesty constraints have access to a larger set of incentive-compatible protocols. However, in this case the solution to our problem and Bayesian persuasion problem coincide.

3.4 Persuading the Public

This section allows the receiver to have private information. The media attempt to persuade the population of receivers to choose the mobilizing action. The media's report r is publicly revealed to the unit mass of receivers. Each receiver cares only about his own action $a_i \in \{a_0, a_1\}$.[17] We impose the same assumptions on the payoff functions as in Section 3.3.1. That is, $\Theta_0 = [0, \bar{\theta})$ and $\Theta_1 = (\bar{\theta}, 1]$. The ruler's competence level θ is assumed to be a draw from the uniform distribution on $[0,1]$. The receiver i's net payoff from the mobilizing action is a linear function $\delta_R(\theta, \omega_i) = \theta - \omega_i$, where the receiver's pickiness $\omega_i \in [0,1]$ is the receiver i's private information. The mass of the receivers with the pickiness below or equal to ω is captured by an absolutely continuous cumulative distribution function H, with a strictly positive on $(0,1)$ density h. Denote the measure of a set $\Omega \subseteq [0,1]$ generated by H as $\eta(\Omega)$. Type-θ source extracts $u_S(a_1, \theta)$ from each receiver taking the mobilizing action and $u_S(a_0, \theta)$ from each receiver taking the status-quo action. The timing is unchanged. However, the source and the media have to evaluate their payoffs as the expectation over the receiver's types. The media's goal is to maximize the proportion of the receivers that choose the mobilizing action.

This setup is isomorphic to the problem with a single receiver having private

[17] In a voting application, this assumption corresponds to the sincere voting paradigm. For example, see Alonso and Câmara (2016).

information about ω. The common prior on ω is captured by the distribution H, and θ and ω are assumed to be independent. We assume that the media cannot elicit private information from the receiver.Then the analogue of the revelation principle for this case can be shown. Instead of an unconditional action recommendation, the media now offer a contingent recommendation, i.e., an action recommendation for each receiver's type. A typical contingent recommendation has the form of $\omega \mapsto \{a_0, a_1\}$. Thus, without loss, it can be seen as the subset Ω_1 of receivers that are recommended to choose a_1. To this end, we focus on a protocol $\phi : \Theta \to \Delta(\mathcal{P}([0,1]))$, where $\mathcal{P}([0,1])$ is the set of measurable subsets of the unit interval. We will show that only recommendations of the form $[0, b]$, $b \in [0,1]$, are sent. Let suppϕ denote the set of contingent recommendations that appear in the protocol ϕ after some reported competence level θ, $\phi(\cdot|\theta) > 0$, and assume that suppϕ is finite.[18]

A contingent protocol $\phi : \Theta \to \Delta(\mathcal{P}([0,1])$ is Bayesian incentive-compatible if S's honesty and R's obedience form an equilibrium. Specifically, S is honest given R's obedience if

$$\sum_{\Omega_1 \in \text{supp}\phi} [\eta(\Omega_1) u_S(a_1, \theta) + (1 - \eta(\Omega_1)) u_S(a_0, \theta))] (\phi(\Omega_1|\theta) - \phi(\Omega_1|\theta')) \geq 0 \quad (3.5)$$

for every $\theta, \theta' \in \Theta$, that is, it is optimal for the source to report the ruler's competence. We call ϕ honest if it satisfies (3.5).

R is obedient given S's honesty if, for every $\Omega_1 \in \text{supp}\phi$,

for every $\omega \in \Omega_1$,

$$\int_0^1 \delta_R(\cdot, \omega) d\phi(\Omega_1|\cdot) \geq 0, \quad (3.6)$$

for every $\omega \in [0,1] \setminus \Omega_1$,

$$\int_0^1 \delta_R(\cdot, \omega) d\phi(\Omega_1|\cdot) \leq 0. \quad (3.7)$$

[18]In principle, a protocol ϕ can have an infinite support. Then the summation in the honesty constraint (3.5) has to be substituted by appropriate integration.

We call ϕ obedient if it satisfies (3.6) and (3.7).

The following lemma shows that an optimal incentive-compatible contingent protocol is also an equilibrium in the class of all possible protocols.

Lemma 3.3. *Given any PBE in the original game followed by π, there exists an incentive-compatible contingent protocol $\phi : \Theta \to \Delta(\mathcal{P}([0,1])$ under which the media get the same expected utility when S is honest and R is obedient as in this PBE.*

As in the setup with a known receiver's type, the incentive constraints reduce the dimension of the search for the optimal protocol. We start by simplifying the obedience constraint. Lemma 3.4 shows that any contingent recommendation is necessarily an interval. The obedience constraint can be reduced to a single equation for each recommendation in the support. This equation requires that the receiver with the highest pickiness level among the receivers that are recommended a_1 has to be indifferent between the mobilizing and status-quo actions.

Lemma 3.4. *Every obedient protocol $\phi : \Theta \to \Delta(\mathcal{P}([0,1])$ exclusively sends contingent recommendations of the following form: a_1 is recommended to receivers with $\omega \in [0,b]$, $b \in [0,1]$; otherwise, a_0 is recommended. For every $b \in supp\phi$, the obedience constraint is summarized by*

$$\int_0^1 (\theta - b) d\phi([0,b]|\theta) = 0.$$

Lemma 3.4 illustrates that the contingent recommendations are intuitive: only receivers that are sufficiently aligned with the media are recommended to take the mobilizing action. Lemma 3.5 provides the further simplification that mirrors Lemma 3.2 and characterizes honest protocols.

Lemma 3.5. *A protocol $\phi : \Theta \to \Delta(\mathcal{P}([0,1])$ is honest if and only if there exist s_0 and s_1, with*

$s_0 \leq s_1$*, such that for every* $\theta \in \Theta_i$*,* $i \in \{0,1\}$*,*

$$s_i = \sum_{[0,b] \in supp\phi} \phi([0,b]|\theta) H(b).$$

As before, the source types within sets Θ_0 and Θ_1 have to be provided the same probability of their preferred action being implemented. This probability has to be evaluated as an expectation over the receiver's private information. The source types from Θ_1 get the favorite mobilizing action with probability s_1, and the source types from Θ_0 get the undesirable mobilizing action with probability s_0. Naturally, for the honesty constraint to hold, the media ensure that $s_1 \geq s_0$.

Combining the results of Lemma 3.3, 3.4, and 3.5, the media's problem can be written as the following Lemma 3.6 prescribes.

Lemma 3.6. *The media's problem is*

$$V = \max_{\phi, s_0, s_1} \left\{ \bar{\theta} \cdot s_0 + (1 - \bar{\theta}) \cdot s_1 \right\},$$

subject to

$$s_0 = \sum_{[0,b] \in supp\phi} \phi([0,b]|\theta) H(b)$$

for every $\theta \in [0, \bar{\theta})$*,*

$$s_1 = \sum_{[0,b] \in supp\phi} \phi([0,b]|\theta) H(b)$$

for every $\theta \in [\bar{\theta}, 1]$*;* $s_0 \leq s_1$*; and*

$$\int_0^1 (\theta - b) d\phi([0,b]|\theta) = 0$$

for every $[0,b] \in supp\phi$*.*

This problem is still complicated, since the optimal protocol can potentially provide multiple contingent recommendations given each reported competence level θ. We do not describe the solution to the problem stated in Lemma 3.6 explicitly. Instead, we deliver the lower bound on the media's payoff by finding the optimal protocol within the restricted class of protocols. Define a *simple protocol* as a protocol with a support having at most two elements, $[0,\bar{b}]$ and $[0,\underline{b}]$, with $\bar{b} \geq \underline{b}$. In what follows, we characterize the optimal simple protocol. We will show the condition on the distribution of receiver's types H, under which the media's payoff achieves this lower bound.

For a simple protocol, the honesty constraints in Lemma 3.6 can be written as follows:

$$s_0 = H(\underline{b}) + \phi_0(H(\bar{b}) - H(\underline{b})),$$

where $\phi_0 = \phi([0,\bar{b}]|\theta)$ for every $\theta \in \Theta_0$, and

$$s_1 = H(\underline{b}) + \phi_1(H(\bar{b}) - H(\underline{b})),$$

where $\phi_1 = \phi([0,\bar{b}]|\theta)$ for every $\theta \in \Theta_1$. It has to be the case that $\phi_0 \leq \phi_1$. For a simple protocol, the honesty constraints specify that the probability of generating the "larger" contingent recommendation $[0,\bar{b}]$ has to be constant within the competence levels in Θ_0 and Θ_1 and it has to be higher for the sources in Θ_1.

The obedience constraint then pins down $\bar{b}$ and $\underline{b}$ as a function of ϕ_0 and ϕ_1. Lemma 3.7 establishes the bounds on $\bar{b}$ and $\underline{b}$ and shows that, for any pair $(\underline{b},\bar{b})$ within these bounds, there exists a simple protocol with the support on $[0,\underline{b}]$ and $[0,\bar{b}]$. The operator $\mathbb{E}$ is the expectation with respect to the prior distribution on Θ.

Lemma 3.7. *Every simple incentive-compatible protocol* $\phi : [0,1] \to \Delta(\{[0,\underline{b}],[0,\bar{b}]\})$ *has to satisfy* $\bar{b} \in [\mathbb{E}[\theta], \mathbb{E}[\theta|\theta \in \Theta_1]]$ *and* $\underline{b} \in [\mathbb{E}[\theta|\theta \in \Theta_0], \mathbb{E}[\theta]]$. *For every pair* $(\underline{b},\bar{b})$ *within these bounds, there exists an incentive-compatible simple protocol with the support on* $[0,\underline{b}]$ *and* $[0,\bar{b}]$, *such that* $\phi_0 = \phi([0,\bar{b}]|\theta)$ *for every* $\theta \in \Theta_0$, $\phi_1 = \phi([0,\bar{b}]|\theta)$ *for every* $\theta \in \Theta_1$, *with* $\phi_0 \leq \phi_1$.

From Lemma 3.7, every simple protocol is characterized by the pair of numbers $\bar{b}$ and $\underline{b}$. This observation paves the way to finding the optimal simple protocol. Proposition 3.3 provides the geometric characterization of the solution to the media's problem. Let cav$H$ be the concavification of H, that is, the smallest concave function that majorizes H. Let $\hat{H}$ be the function H reduced to the domain $[\mathbb{E}[\theta|\theta \in \Theta_0], \mathbb{E}[\theta|\theta \in \Theta_1]]$. Proposition 3.3 shows that the optimal simple protocol delivers the media the payoff equal to the concavification of $\hat{H}$ evaluated at $\mathbb{E}[\theta]$.

Proposition 3.3. *The media's payoff from the optimal simple protocol is equal to cav*$\hat{H}[\mathbb{E}[\theta]]$.

Proposition 3.3 obtains the lower bound on the media's equilibrium payoff. The upper bound on the media's payoff is given by cav$H[\mathbb{E}[\theta]]$. Indeed, the receiver ultimately bases his decision on the posterior mean of the ruler's competence. If it was possible for the media to induce every distribution of posterior means whose expectation is the prior mean, then the solution to the media's problem would correspond to cav$H[\mathbb{E}[\theta]]$. However, this is not always feasible (see, for example, Gentzkow and Kamenica, 2016). If those bounds are equal to each other, then the simple protocol is optimal across all incentive-compatible protocols. Corollary 3.1 summarizes this observation.

Corollary 3.1. *The simple protocol is optimal for the media if cav*$\hat{H}[\mathbb{E}[\theta]] =$ *cav*$H[\mathbb{E}[\theta]]$.

Note that this sufficient condition can be checked just by knowing primitives of the model: distribution H, prior distribution on Θ, and source's preferences. The immediate consequence of Corollary 3.1 is that for concave H, the simple protocol is optimal. Moreover, $\bar{b} = \underline{b}$, that is, the optimal protocol is uninformative.

Finally, to illustrate the bound provided in Proposition 3.3, we consider a unimodal distribution of receiver's types that recently gained a lot of attention in the literature, namely, Kolotilin et al. (2017), Lipnowski, Ravid, and Shishkin (2019), and Shishkin (2021).

The unimodal distribution corresponds to the density strictly increasing until reaching mode m and then strictly decreasing. As a result, the corresponding cumulative distribution function is convex-concave. The black line in Figure 3 illustrates the example of such distribution.

With no honesty concerns, Kolotilin et al. (2017) show that the optimal policy is upper censorship, when the distribution of receiver's types is unimodal: it reveals all states below and pools all states above a threshold.[19] The upper-censorship policy in our setup corresponds to the protocol ϕ, with $\phi([0,\theta]|\theta) = 1$ for $\theta < t$ and $\phi([0,b]|\theta) = 1$ for $\theta \geq t$, where $t > 0$ is a threshold and $b = \mathbb{E}[\theta|\theta \geq t]$. However, the upper-censorship policy is not incentive-compatible, when the honesty constraints are present. Indeed, pick two types $\theta', \theta'' \in \Theta_0 \cap [0,t)$, $\theta' < \theta''$ and consider the incentives of type-θ'' source under the upper-censorship. Reporting θ' gives the source probability $1 - H(\theta')$ of the status-quo action chosen, whereas reporting θ'' produces probability $1 - H(\theta'')$. Type-θ'' source prefers to misreport, and the honesty constraint is not satisfied. Indeed, by Lemma 3.5 the sources preferring the status-quo action have to be provided with the same probability of this action implemented. Under the upper-censorship policy, each source type from Θ_0 wants to mimic the lowest possible type, so that the status-quo action is always chosen. This observation leads me to expect some pooling for the low states in the optimal protocol.

Figure 3.3 illustrates the bounds on the media's payoff for the case of unimodal distribution. The media's payoff from the optimal protocol has to lie in the interval $[\underline{V}, \bar{V}]$. The payoff $\underline{V}$ can be achieved with an optimal simple protocol. This protocol has a support having two elements, $[0,\bar{b}]$, $[0,\underline{b}]$, where $\underline{b} = \mathbb{E}[\theta|\theta \in \Theta_0]$. To maximize the probability of the mobilizing action, the media sometimes reveal that the ruler's competence is low: the contingent recommendation $[0,\mathbb{E}[\theta|\theta \in \Theta_0]]$ reveals that the

[19] The mode of the distribution has to be sufficiently large. Otherwise, the uninformative policy is optimal as most of the receivers are closely aligned with the media.

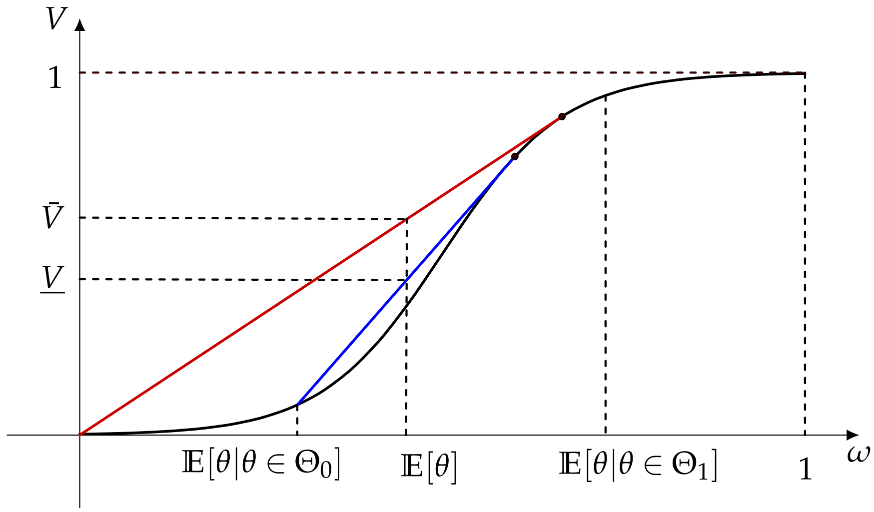


Figure 3.3. Bounds on the media's payoff. The black line corresponds to H. The red line corresponds to cav$H$. The blue line corresponds to cav$\hat{H}$.

ruler's competence is in Θ_0. However, the contingent recommendation $[0,\bar{b}]$ can come from any θ: $\phi([0,\bar{b}]|\theta) = \phi_0 \in (0,1)$ for every $\theta \in \Theta_0$, $\phi([0,\bar{b}]|\theta) = 1$ for every $\theta \in \Theta_1$.

Honesty constraints reduce the media's welfare, since the upper censorship policy cannot be implemented by the media. Revealing all low states of the ruler's competence is not available due to incentives of the low types of the source. The optimal simple protocol delivers the lower bound on the media's payoff. This protocol sometimes reveals that the source is in Θ_0 but does not given information beyond that. The optimal protocol in the unrestricted class will possess the same feature: low states need to be pooled together to satisfy honesty constraints.

3.5 Conclusion

This paper presents a model of information disclosure by a state-owned media to an uninformed receiver choosing between two actions. The problem is that the media do not have direct access to relevant information. Instead, it has to be supplied by the informed elite having interests in the receiver's action. Therefore, the optimal media's editorial policy has to not only convince the receiver to undertake the media-favorite action, but also cater to the elite's preferences to incentivize the information supply. This paper shows how these additional incentive constraints shape the optimal editorial policy and outlines the welfare implications of this policy. We show the conditions under which the honesty constraints are binding and there is a meaningful communication depending on the preference parameters. We close with the discussion of assumptions that are substantial for our results.

Discussion of assumptions

We assume that there are no transfers between players. In this sense, we study the purely informational model of the interaction between the source, the media, and the receiver. In reality, the source may be paid for promoting the ruler's competence or the receiver may be paid by the ruler for undertaking the mobilizing action. We leave this possibility out of the model.

We assume that the media have commitment power. As explained in Gehlbach and Sonin (2014) and Gentzkow, Shapiro, and Stone (2015), the editorial policy cannot be easily changed and consistent bias in reporting is detected by receivers. This commitment assumption may be relaxed in the fashion of Lipnowski, Ravid, and Shishkin (2019), where with some probability the media may secretly change the editorial policy after observing the source's message. The analysis provided here can be seen as the best the media can achieve over different possible communication protocols and equilibrium selection rules.

The messages produced by the source and the reports published by the media are assumed to be costless. In reality of the authoritarian states, messages that suggest the ruler's incompetence may be associated with the consequential punishment. The introduction of cost associated with specific messages imposes modeling challenges and makes the methodology developed in this paper futile. Instead, one would need to make use of, for example, the methodology of the papers that study strategic communication with lying costs as in Kartik (2009).

Chapter 3 is currently being prepared for submission for publication of the material. The dissertation author, Aleksandr Levkun, is the sole author of this material.

Appendix A

Supplemental Material

A.1 Omitted Proofs for Chapter 1

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

Proof of Proposition 1.1

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 1.2

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 1.1, 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 1.1, 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 1.3 and 1.5

Fix a fact-checking policy χ_p. The characterization of χ_p-equilibria provided in the proof of Proposition 1.2 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

Table A.1. 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	$\mu\overline{\chi}_p$

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 $\sigma(m_s|0)=1$. 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

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

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

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

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 A.2. 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 1.4

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 $\bar{c}(u_F) = \omega(1-p)(u_F(1,1) - u_F(0,1))$ by Proposition 1.3. 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 1.3. 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 $\overline{c}(u_F) = (1-\mu)(1-p)(u_F(0,0) - u_F(1,0))$ by Proposition 1.3. 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 $m \in \arg\max \overline{\chi}_{p,2}(\cdot)$ 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

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

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 1.3. 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.

A.2 Omitted Proofs and Supplemental Material for Chapter 2

All proofs in this appendix are for the general case where the agents (sellers) observe private data in the form of $\omega_i \in \Omega_i$ and hence $\omega = (\omega_0, \omega_1, \ldots, \omega_n)$ (see Section 2.5). The special case where only the principal (platform) observes data obtains by having $|\Omega_i| = 1$ for all $i \in I$.

Proof of Lemma 2.1

We will formulate $\mathcal{U}_q$ in terms of choosing a measure $\chi \in \mathbb{R}_+^{A\times\Omega}$:

$$\begin{aligned}
\mathcal{U}_q : \qquad & \max_{\chi} \sum_{\omega\in\Omega, a\in A} u_0(a,\omega)\chi(a,\omega), \\
& \text{subject to for all } i \in I, \omega_i \in \Omega_i, \text{ and } a_i, a_i' \in A_i, \\
& \sum_{\omega_{-i}\in\Omega_{-i}, a_{-i}\in A_{-i}} \left(u_i(a_i,a_{-i},\omega) - u_i(a_i',a_{-i},\omega)\right)\chi(a_i,a_{-i},\omega) \geq 0, \qquad \text{(A.1)} \\
& \text{and for all } \omega \in \Omega, \\
& \sum_{a\in A} \chi(a,\omega) = q(\omega).
\end{aligned}$$

It is convenient to express this problem in matrix form. Fix an arbitrary total ordering of the set $A \times \Omega$. We denote by $\mathbf{u}_0 \in \mathbb{R}^{A\times\Omega}$ the vector whose entry corresponding to (a,ω) is $u_0(a,\omega)$. For every player i, let $\mathbf{U}_i \in \mathbb{R}^{(A_i\times A_i\times\Omega_i)\times(A\times\Omega)}$ be a matrix thus defined: For each row $(a_i', a_i'', \omega_i') \in A_i \times A_i \times \Omega_i$ and column $(a,\omega) \in A \times \Omega$, let the corresponding entry be

$$\mathbf{U}_i((a_i',a_i'',\omega_i'),(a,\omega)) = \begin{cases} u_i(a_i',a_{-i},\omega) - u_i(a_i'',a_{-i},\omega) & \text{if } a_i' = a_i, \omega_i' = \omega_i, \\ 0 & \text{else.} \end{cases}$$

Thus, $\mathbf{U}_i(a_i',a_i'',\omega_i')$ denotes the row labeled by (a_i',a_i'',ω_i') (which defines the corresponding obedience constraint) and $\mathbf{U}_i(a,\omega)$ denotes the column labeled by (a,ω). Define the matrix $\mathbf{U}$ by stacking all the matrices $\{\mathbf{U}_i\}_{i\in I}$ on top each other. Finally, define the indicator matrix $\mathbf{I} \in \{0,1\}^{\Omega\times(A\times\Omega)}$ such that, for each row ω' and column (a,ω'),

$$\mathbf{I}(\omega',(a,\omega)) := \begin{cases} 1 & \text{if } \omega' = \omega, \\ 0 & \text{else.} \end{cases}$$

With this notation and treating q as a vector, $\mathcal{U}_q$ can be written as follows:

$$\begin{aligned} \max_{\chi} \quad & \mathbf{u}_0^T \chi, \\ \text{s.t.} \quad & \mathbf{U}\chi \geq \mathbf{0}, \\ & \mathbf{I}\chi = q, \\ & \chi \geq \mathbf{0}. \end{aligned} \tag{A.2}$$

By standard linear-programming arguments (Bertsimas and Tsitsiklis, 1997) the dual of $\mathcal{U}_q$ can be written as

$$\min_{\lambda,v} \mathbf{0}^T \lambda + q^T v,$$

subject to, for all $i = 1, \ldots, n$, $a_i, a_i' \in A_i$, and $\omega_i \in \Omega_i$,

$$\lambda_i(a_i'|a_i,\omega_i) \geq 0,$$

$v(\omega) \in \mathbb{R}$ for all $\omega \in \Omega$, and for all $(a,\omega) \in A \times \Omega$

$$u_0(a,\omega) \leq v(\omega) - \sum_{i \in I} \left\{ \sum_{a_i' \in A_i} \left(u_i(a_i, a_{-i}, \omega) - u_i(a_i', a_{-i}, \omega) \right) \lambda_i(a_i'|a_i,\omega_i) \right\}.$$

The objective simplifies to

$$\min_{\lambda,v} \sum_{\omega \in \Omega} v(\omega) q(\omega).$$

The second set of constraints can be written as

$$v(\omega) \geq u_0(a,\omega) + \sum_{i \in I} \left\{ \sum_{a_i' \in A_i} \left(u_i(a_i, a_{-i}, \omega) - u_i(a_i', a_{-i}, \omega) \right) \lambda_i(a_i'|a_i,\omega_i) \right\}.$$

Define the sum in this expression by $t(a,\omega)$ for all (a,ω). Since for every $\omega \in \Omega$ this constraint has to hold for all $a \in A$ and we have a minimization problem, we conclude that each $v(\omega)$ has to satisfy

$$v(\omega) = \max_{a \in A} \{u_0(a,\omega) + t(a,\omega)\}.$$

Thus, we obtain our data-value problem $\mathcal{V}_q$.

Non-degeneracy and a Remark on the Structure of Solutions

In Section 2.2, we assumed that no more than $|A \times \Omega|$ of the constraints (2.1) are ever active at the same time. We now formalize that assumption following Bertsimas and Tsitsiklis (1997). Consider the polyhedron defined by the constraints in (A.2) and recall that $\chi \in \mathbb{R}_+^{A \times \Omega}$, which has dimension $|A \times \Omega|$. A basic feasible solution of $\mathcal{U}_q$ is a χ such that (*i*) all equality constraints are active, (*ii*) $|A \times \Omega|$ of the constraints active at χ are linearly independent, and (*iii*) all constraints are satisfied. Formally, we assume the following.

Assumption A.2.1 (Non-degeneracy)**.** At every basic feasible solution χ of problem $\mathcal{U}_q$ there are only $|A \times \Omega|$ active constraints.

The next remark describes the structure of optimal solutions of $\mathcal{U}_q$ and $\mathcal{V}_q$.

Remark A.2.1. We can transform $\mathcal{U}_q$ to the standard form $\mathcal{U}_q^S$ which can be written as follows:

$$\begin{aligned} \max_{\chi, s} \quad & \mathbf{u}_0 \chi, \\ \text{s.t.} \quad \mathbf{U}\chi - s &= \mathbf{0}, \\ I\chi &= q, \\ \chi, s &\geq \mathbf{0}, \end{aligned} \tag{A.3}$$

where each $s_i(a_i'|a_i,\omega_i)$ is a nonnegative slack variable. The dual of $\mathcal{U}_q^S$ coincides with the data-value problem $\mathcal{V}_q$. Note that $\mathcal{U}_q$ always has an optimal solution χ_q^*, which is generically unique and hence corresponds to an extreme point of the polyhedron of feasible χ. Moreover, this χ_q^* is an optimal solution of $\mathcal{U}_q^S$ as well. The extreme point χ_q^* is nondegenerate by Assumption A.2.1 and characterized by a square, nonsingular, active-constraint submatrix $\mathbf{B}$ consisting of linearly independent rows of the stacked matrix $\left[\frac{\mathbf{U}}{I}\middle|\frac{-\mathbf{1}}{\mathbf{0}}\right]$, where $\mathbf{1}$ is the identity matrix. As illustrated in Bertsimas and Tsitsiklis (1997, Chapter 4), given $\mathbf{B}$, we have

$$\left[\frac{\chi_q^*}{s_q^*}\right] = \mathbf{B}^{-1}\left[\frac{\mathbf{0}}{q}\right], \tag{A.4}$$

where s_q^* is the vector of optimal slack variables in $\mathcal{U}_q^S$. A corresponding solution of $\mathcal{V}_q$ is given by

$$\left[\frac{v_q^*}{\lambda_q^*}\right] = \mathbf{u}_0\mathbf{B}^{-1}. \tag{A.5}$$

It follows that as long as the optimal solutions of $\mathcal{U}_q$ and $\mathcal{V}_q$ are defined by the same extreme point given by $\mathbf{B}$, χ_q^* varies with q, but (v_q^*, λ_q^*) does not.

Proof of Lemma 2.2

Fix an optimal solution (v_q^*, λ_q^*) of $\mathcal{V}_q$. For every q, $\omega \in \Omega$, and $x(\cdot|\omega) \in CE(\Gamma_\omega)$, by (2.4) we have

$$\begin{aligned}
v_q^*(\omega) &\geq \sum_{a\in A} u_0(a,\omega)x(a|\omega) + \sum_{a\in A} t(a,\omega)x(a|\omega) \\
&= \sum_{a\in A} u_0(a,\omega)x(a|\omega) \\
&\quad + \sum_{a\in A}\left\{\sum_{i\in I}\sum_{\hat{a}_i\in A_i}\left(u_i(a_i,a_{-i},\omega) - u_i(\hat{a}_i,a_{-i},\omega)\right)\lambda_i^*(\hat{a}_i|a_i,\omega_i)\right\}x(a|\omega) \\
&= \sum_{a\in A} u_0(a,\omega)x(a|\omega) \\
&\quad + \sum_{i\in I}\sum_{a_i,\hat{a}_i\in A_i}\lambda_i^*(\hat{a}_i|a_i,\omega_i)\left\{\sum_{a_{-i}\in A_{-i}}\left(u_i(a_i,a_{-i},\omega) - u_i(\hat{a}_i,a_{-i},\omega)\right)x(a|\omega)\right\} \\
&\geq \sum_{a\in A} u_0(a,\omega)x(a|\omega),
\end{aligned}$$

where the last inequality follows because any $x(\cdot|\omega) \in CE(\Gamma_\omega)$ is defined by the property that, for all $i \in I$ and $a_i, a_i' \in A_i$,

$$\sum_{a_{-i} \in A_{-i}} \left(u_i(a_i, a_{-i}, \omega) - u_i(a_i', a_{-i}, \omega) \right) x(a_i, a_{-i}|\omega) \geq 0.$$

Since $x(\cdot|\omega)$ is an arbitrary element of $CE(\Gamma_\omega)$, we conclude that $v_q^*(\omega) \geq \bar{u}(\omega)$.

Proof of Proposition 2.1

By complementary slackness, $x_q^*(a,\omega) > 0$ implies $v_q^*(\omega) = u_0(a,\omega) + t_q^*(a,\omega)$. Hence,

$$v_q^*(\omega) = \sum_{a \in A} u_0(a,\omega) x_q^*(a|\omega) + \sum_{a \in A} t_q^*(a,\omega) x_q^*(a|\omega) = u_q^*(\omega) + t_q^*(\omega).$$

Suppose we start from database q, with $q(\omega) > 0$, and we increase the quantity of ω-datapoints from $q(\omega)$ to $\hat{q}(\omega)$, thus obtaining the database $\hat{q}$. We can write

$$U^*(\hat{q}) - U^*(q) = u_{\hat{q}}^*(\omega)[\hat{q}(\omega) - q(\omega)] + \sum_{\omega' \in \Omega} [u_{\hat{q}}^*(\omega') - u_q^*(\omega')]\hat{q}(\omega')$$

Dividing both sides by $\hat{q}(\omega) - q(\omega)$, taking limits as $\hat{q}(\omega) \to q(\omega)$, and using Lemma 2.1, we obtain that

$$\begin{aligned}
t_q^*(\omega) &= v_q^*(\omega) - u_q^*(\omega) = \frac{\partial U^*(q)}{\partial q(\omega)} - u_q^*(\omega) \\
&= \lim_{\hat{q}(\omega) \to q(\omega)} \frac{\sum_{\omega' \in \Omega} [u_{\hat{q}}^*(\omega') - u_q^*(\omega')]\hat{q}(\omega')}{\hat{q}(\omega) - q(\omega)} = \sum_{\omega' \in \Omega} \frac{\partial u_q^*(\omega')}{\partial q(\omega)} q(\omega') \\
&= \sum_{\omega' \in \Omega, a \in A} u_0(a,\omega') \left(\lim_{\hat{q}(\omega) \to q(\omega)} \frac{[x_{\hat{q}}^*(a|\omega') - x_q^*(a|\omega')]}{\hat{q}(\omega) - q(\omega)} \right) \hat{q}(\omega') = \\
&= \sum_{\omega' \in \Omega, a \in A} u_0(a,\omega') \frac{\partial x_q^*(a|\omega')}{\partial q(\omega)} q(\omega'),
\end{aligned}$$

where the existence of the derivative $\frac{\partial x_q^*(a|\omega')}{\partial q(\omega)}$ almost everywhere follows from (A.4).

Proof Proposition 2.2

First, note that we can write $u_0(a_1,\omega) = \pi a_1 \mathbb{I}\{\omega \geq a_1\} + (1-\pi)\max\{\omega - a_1, 0\}$ as

$$[a(2\pi - 1) + (1-\pi)\omega]\mathbb{I}\{\omega \geq a\},$$

which is strictly increasing in a if and only if $\pi > \frac{1}{2}$. Let $\bar{x}^*$ be the profit-maximizing solution (i.e., for $\pi = 1$) and $\underline{x}^*$ be the surplus-maximizing solution (i.e., for $\pi = 0$).

Lemma A.1. *$\bar{x}^*$ is optimal for all $\pi \geq \frac{1}{2}$ and $\underline{x}^*$ is optimal for all $\pi \leq \frac{1}{2}$.*

Proof. Fix any (non-trivial) q and $\pi \in (0,1)$. Problem $\mathcal{U}_q$ involves maximizing

$$\sum_{\omega,a} u_\pi(a,\omega)x(a|\omega)q(\omega) = \sum_{\omega \geq a}[a(2\pi-1) + (1-\pi)\omega]x(a|\omega)q(\omega),$$

subject to constraints (2.1).

Suppose that $\pi > \frac{1}{2}$. Note that $\bar{x}^*$ is feasible and maximizes the objective function pointwise for every ω. Indeed, since $\bar{x}^*(\omega|\omega) = 1$, for every ω we have that $\bar{x}^*$ selects the highest $a \leq \omega$ for every ω, thereby maximizing $a(2\pi-1)\mathbb{I}\{\omega \geq a\}$; it also maximizes $\sum_{a \leq \omega} \omega x(a|\omega)$ for every ω. We can invoke the Theorem of the Maximum to extend the optimality of $\bar{x}^*$ at $\pi = \frac{1}{2}$.

Suppose now that $\pi < \frac{1}{2}$. Now for each ω the objective is to pair ω with the smallest possible a and do so with the highest probability allowed by (2.1). This is what $\underline{x}^*$ essentially does. We can again invoke the Theorem of the Maximum to extend the optimality of $\underline{x}^*$ at $\pi = \frac{1}{2}$. □

We now derive the expression of $v_q^*(\omega)$ in the statement of the proposition. The case of $\pi \geq \frac{1}{2}$ follows immediately from the fact that $\bar{x}^*$ is full disclosure. Now suppose

$\pi < \frac{1}{2}$. We will construct a candidate v_q^* and prove it solves $\mathcal{V}_q$ using strong duality. First, under $\underline{x}^*$ we have

$$\begin{aligned} U^*(q) &= \sum_{\omega,a}[\pi u_1(a,\omega) + (1-\pi)u_0(a,\omega)]\underline{x}^*(a|\omega)q(\omega) \\ &= \pi\sum_{\omega,a} a\mathbb{I}\{\omega \geq a\}\underline{x}^*(a|\omega)q(\omega) \\ &\quad +(1-\pi)\left[\sum_{\omega<a_q}\omega q(\omega) + \sum_{\omega \geq a_q}(\omega - a_q)q(\omega)\right]. \end{aligned}$$

Note that

$$\sum_{\omega,a} a\mathbb{I}\{\omega \geq a\}\underline{x}^*(a|\omega)q(\omega) = a_q \sum_{\omega \geq a_q} q(\omega),$$

because the left-hand side is the seller's expected profits under $\underline{x}^*$, which by construction equal to the expected profit from the fixed uninformed price a_q. Therefore, we can write

$$\begin{aligned} U^*(q) &= \pi a_q \sum_{\omega \geq a_q} q(\omega) + (1-\pi)\left[\sum_{\omega<a_q}\omega q(\omega) + \sum_{\omega \geq a_q}(\omega - a_q)q(\omega)\right] \\ &= (2\pi - 1)a_q \sum_{\omega \geq a_q} q(\omega) + (1-\pi)\sum_{\omega}\omega q(\omega). \end{aligned}$$

Now we construct (v_q^*, λ_q^*) and show that it satisfies all dual constraints and yields $\sum_\omega v_q^*(\omega)q(\omega) = U^*(q)$, which proves that (v_q^*, λ_q^*) is optimal by strong duality. Recall that, in general, for all (a,ω) the dual constraint reads as

$$v(\omega) \geq u_\pi(a,\omega) + \sum_{a'}[u_1(a,\omega) - u_1(a',\omega)]\lambda(a'|a).$$

Let $\lambda_q^*(a'|a) = 0$ for all $a' \neq a_q$. Let $\lambda_q^*(a_q|a) = 1 - 2\pi$ for all $a \in \operatorname{supp}\underline{x}(\cdot|\omega)$ for some ω and $\lambda_q^*(a_q|a) = 0$ otherwise. Given this, for $\omega < a_q$, the right-hand side of the dual constraint equals

$$\begin{cases} \pi a + (1-\pi)(\omega - a) + a\lambda_q^*(a_q|a) & \text{if } a \leq \omega, \\ 0 & \text{if } a > \omega. \end{cases}$$

Given $\lambda_q^*(a_q|a)$, the first line always equals $(1-\pi)\omega > 0$. Therefore, for $\omega < a_q$ define

$$v_q^*(\omega) = (1-\pi)\omega.$$

For $\omega \geq a_q$, the right-hand side of the dual constraint equals

$$\begin{cases} \pi a + (1-\pi)(\omega - a) + (a - a_q)\lambda_q^*(a_q|a) & \text{if } a \leq \omega \\ -a_q\lambda_q^*(a_q|a) & \text{if } a > \omega. \end{cases}$$

Given $\lambda_q^*(a_q|a)$, the first line always equals

$$(2\pi - 1)a_q + (1-\pi)\omega = \pi a_q + (1-\pi)(\omega - a_q) > 0.$$

Therefore, for $\omega \geq a_q$ define

$$v_q^*(\omega) = (2\pi - 1)a_q + (1-\pi)\omega.$$

Note that by construction v_q^* satisfies all dual constraint and $\sum_\omega v_q^*(\omega)q(\omega) = U^*(q)$, as desired.

It follows immediately that for $\pi < \frac{1}{2}$ we have $t_q^*(\omega) > 0$ for $\omega < a_q$ and $t_q^*(\omega) \leq 0$ for $\omega \geq a_q$.

Proof Proposition 2.4

By the formulation of $\mathcal{V}_q$ and Lemma 2.2, the polyhedron of feasible solutions of $\mathcal{V}_q$, denoted by $F(\mathcal{V}_q)$ does not contain a line because all dual variables are bounded from below. By Theorem 2.6 in Bertsimas and Tsitsiklis (1997), $F(\mathcal{V}_q)$ has at least one extreme point and at most finitely many of them by Corollary 2.1 in Bertsimas and Tsitsiklis (1997). By Theorem 4.4 in Bertsimas and Tsitsiklis (1997), $\mathcal{V}_q$ has at least one optimal solution. By Theorem 2.7 in Bertsimas and Tsitsiklis (1997), we can focus on solutions that are extreme points of $F(\mathcal{V}_q)$.

Fix q and suppose that the optimal solution (v_q^*, λ_q^*) of the dual of $\mathcal{U}_q$ is unique. As explained in Remark A.2.1, there exists a submatrix $\mathbf{B}$ such that (v_q^*, λ_q^*) satisfies (A.5). Given Assumption A.2.1, Theorem 3.1 and Exercise 3.6 in Bertsimas and Tsitsiklis (1997) imply that

$$\left[\begin{array}{c|c} \mathbf{U} & -\mathbf{1} \\ \hline I & \mathbf{0} \end{array}\right] \mathbf{B}^{-1} \begin{bmatrix} \mathbf{0} \\ \hline q \end{bmatrix} \geq \begin{bmatrix} \mathbf{0} \\ \hline q \end{bmatrix}.$$

The inequality is strict for each row of $\mathbf{U}$ that corresponds to $\lambda_{q,i}^*(a_i'|a_i,\omega_i) = 0$:

$$\left[\mathbf{U}_i(a_i, a_i', \omega_i) \mid -\mathbf{1}_i(a_i, a_i', \omega_i)\right] \mathbf{B}^{-1} \begin{bmatrix} \mathbf{0} \\ \hline q \end{bmatrix} > 0, \tag{A.6}$$

where $\mathbf{1}_i(a_i, a_i', \omega_i)$ is the row of the identity matrix $\mathbf{1}$ that corresponds to (i, a_i, a_i', ω_i). Note that for each row ω of the indicator matrix I (i.e., $I(\omega)$), which corresponds to variable $v_q^*(\omega)$, it automatically holds that $\left[I(\omega) \mid \mathbf{0}\right] \mathbf{B}^{-1} \left[\frac{\mathbf{0}}{q}\right] = q(\omega)$. Similarly, for each row of $\mathbf{U}$ that corresponds to $\lambda_{q,i}^*(a_i'|a_i,\omega_i) > 0$, it holds that $\left[\mathbf{U}_i(a_i, a_i', \omega_i) \mid -\mathbf{1}_i(a_i, a_i', \omega_i)\right] \mathbf{B}^{-1} \left[\frac{\mathbf{0}}{q}\right] = 0$ as long as $\mathbf{B}$ identifies the optimal extreme point.

Now consider changes in q and note that it only enters the objective of $\mathcal{V}_q$. Each condition (A.6) defines an open set of q's in $\mathbb{R}_+^\Omega$ that satisfy it. Define $(v_{\mathbf{B}}^*, \lambda_{\mathbf{B}}^*)$ identified by $\mathbf{B}$ as in (A.5) and

$$Q(\mathbf{B}) = \{q : (A.6) \text{ holds for all } i \in I \text{ and } (a_i, a_i', \omega_i) \text{ s.t. } \lambda^*_{\mathbf{B},i}(a_i|a_i', \omega_i) = 0\}.$$

Note that $Q(\mathbf{B})$ is an open set because it is the intersection of finitely many open sets.

Now recall that there are only finitely many extreme points of the dual polyhedron of feasible solutions. Therefore, there are finitely many submatrices $\{\mathbf{B}_1, \ldots, \mathbf{B}_M\}$ such that each identifies an optimal $(v^*_{\mathbf{B}_m}, \lambda^*_{\mathbf{B}_m})$, where $v^*_{\mathbf{B}_m}$ is unique for all $q \in Q(\mathbf{B}_m)$. For all $m = 1, \ldots, M$, define $Q_m = Q(\mathbf{B}_m)$. By construction, each Q_m is open and $q, q' \in Q_m$ implies that $v^*_q = v^*_{q'}$. Since v^*_q is generically unique with respect to q, it follows that $\mathbb{R}^\Omega_+ \setminus \cup_m Q_m$ has Lebesgue measure zero.

Proof of Proposition 2.3

Fix $\mu_1, \mu_2 \in \Delta(\Omega)$. Let $\Omega^i = \{\omega \in \Omega : \mu_i(\omega) > \mu_j(\omega), j \neq i\}$, $i \in \{1,2\}$, and $\Omega^3 = \Omega \setminus \{\Omega^1 \cup \Omega^2\}$.

Let $Y = \mathbb{R}^\Omega \times \mathbb{R}^{A_1 \times A_1}_+ \times \ldots \times \mathbb{R}^{A_n \times A_n}_+$. Associate the canonical component-wise order with Y, with an exception that the order is reversed for $\omega \in \Omega^1$. Y is a lattice, with a typical element (v, λ), where $v \in \mathbb{R}^\Omega$ and $\lambda \in \mathbb{R}^{A_1 \times A_1}_+ \times \ldots \times \mathbb{R}^{A_n \times A_n}_+$.

The data-value problem is equivalent to the problem $\max_{(v,\lambda) \in S} f(v, \lambda; \mu)$, where $f(v, \lambda; \mu) = -\sum_{\omega \in \Omega} v(\omega)\mu(\omega)$ and the feasible set $S \subset Y$ is given by the inequalities

$$v(\omega) \geq u_0(a, \omega) + \sum_{i \in I} \sum_{a_i' \in A_i} (u_i(a_i, a_{-i}, \omega) - u_i(a_i', a_{-i}, \omega))\lambda_i(a_i'|a_i, \omega_i).$$

We treat μ as a parameter. Note that S does not depend on μ. Furthermore, μ is an element of $(|\Omega| - 1)$-dimensional simplex, with which we associate the following partial order: $\mu' \geq \mu$ if $\mu'(\omega) \geq \mu(\omega)$ for $\omega \in \Omega^1$, $\mu'(\omega) \leq \mu(\omega)$ for $\omega \in \Omega^2$, and $\mu'(\omega) = \mu(\omega)$ for $\omega \in \Omega^3$. Note that $\mu_1 \geq \mu_2$ in accordance with this partial order.

We want to show that f is supermodular in (v, λ) and has increasing differences in $(v, \lambda; \mu)$. Observe that

$$\begin{aligned} f(v', \lambda'; \mu) + f(v'', \lambda''; \mu) &= -\sum_{\omega \in \Omega} v'(\omega)\mu(\omega) - \sum_{\omega \in \Omega} v''(\omega)\mu(\omega) \\ &= -\sum_{\omega \in \Omega} (v'(\omega) + v''(\omega))\mu(\omega) \\ &= -\sum_{\omega \in \Omega} (\max\{v'(\omega), v''(\omega)\} + \min\{v'(\omega), v''(\omega)\})\mu(\omega) \\ &= f((v', \lambda') \wedge (v'', \lambda''); \mu) + f((v', \lambda') \vee (v'', \lambda''); \mu). \end{aligned}$$

Then f is supermodular in (v, λ).

Fix $(v', \lambda') \geq (v, \lambda)$ and $\mu' \geq \mu$. Observe that

$$\begin{aligned} & (f(v', \lambda', \mu') - f(v, \lambda, \mu')) - (f(v', \lambda', \mu) - f(v, \lambda, \mu)) \\ = & \sum_{\omega \in \Omega} (v(\omega) - v'(\omega))(\mu'(\omega) - \mu(\omega)) \\ = & \sum_{\omega \in \Omega^1} (v(\omega) - v'(\omega))(\mu'(\omega) - \mu(\omega)) + \sum_{\omega \in \Omega^2} (v(\omega) - v'(\omega))(\mu'(\omega) - \mu(\omega)) \geq 0, \end{aligned}$$

where the inequality follows from the adapted partial orders. Then, f has increasing differences in $(v, \lambda; \mu)$.

Finally, by Theorem 5 in Milgrom and Shannon (1994), $\arg\max_{(v,\lambda) \in S} f(v, \lambda; \mu)$ is monotone nondecreasing in μ. This monotone comparative statics coupled with generic uniqueness of v_q^* with respect to q imply that if $\mu_q(\omega) > \mu_{q'}(\omega)$ for two databases q and q' then $v_q^*(\omega) \leq v_{q'}^*(\omega)$. That is, this monotonicity of $v_q^*(\omega)$ holds for any selection v_q^* from the optimal solution correspondence of $\mathcal{V}_q$.

We now prove the second part of the proposition. When only records of type ω are present in the database (i.e., $\mu_q(\omega) = 1$), we have $v_q^*(\omega) = \bar{u}(\omega)$. Indeed, the definition of $\bar{u}(\omega)$ implies that it can be written as

$$\bar{u}(\omega) = \min_{b_\omega, \ell_\omega} \max_{a \in A} \left\{ u_0(a,\omega) + t_{\lambda_\omega}(a,\omega) \right\},$$

where $\lambda_\omega = (\lambda_{1,\omega}, \ldots, \lambda_{n,\omega})$, with $\lambda_{i,\omega} : A_i \to \Delta(A_i)$, and

$$t_{\lambda_\omega}(a,\omega) = \sum_{i \in I} \sum_{a'_i \in A_i} (u_i(a_i, a_{-i}, \omega) - u_i(a'_i, a_{-i}, \omega)) \lambda_{i,\omega}(a'_i | a_i, \omega_i).$$

For $\varepsilon > 0$, consider a set $M_\varepsilon(\omega) = \{\mu \in \Delta(\Omega) : \mu(\omega') \in (0,\varepsilon) \text{ for } \omega \neq \omega', \mu(\omega) < 1\}$. By Proposition 2.4, there exists a finite collection $\{\mathcal{P}_1, \ldots, \mathcal{P}_K\}$ of open, convex, and disjoint subsets of $\Delta(\Omega)$ such that $\cup_k \mathcal{P}_k$ has measure one and, for every k, v_q^* is unique and constant for q, with $\mu_q \in \mathcal{P}_k$. Therefore, we can always find $\mathcal{P}_m \in \{\mathcal{P}_1, \ldots, \mathcal{P}_K\}$, such that $\mathcal{P}_m \cap M_\varepsilon(\omega)$ is nonempty, open, and convex for all $0 < \varepsilon \leq \delta$, where $\delta > 0$. Then $v_q^*(\omega)$ is unique and constant for all $q \in \mathbb{R}^\Omega_{++}$, with $\mu_q \in \mathcal{P}_m \cap M_\delta(\omega)$. Let us refer to this constant as $\hat{u}(\omega)$. If $\hat{u}(\omega) = \bar{u}(\omega)$, then the result follows. Suppose, on the contrary, that $\hat{u}(\omega) \neq \bar{u}(\omega)$. We can always pick a sequence μ^n, $n \in \mathbb{N}$, from $\mathcal{P}_m \cap M_\delta(\omega)$ that converges to $\tilde{\mu}$, with $\tilde{\mu}(\omega) = 1$. Then for every $n \in \mathbb{N}$, $v_q^*(\omega) = \hat{u}(\omega)$ for every q, such that $\mu_q = \mu^n$. By the Berge's maximum theorem, (v_q^*, λ_q^*) is an upper-hemicontinuous correspondence and therefore has a closed graph. Hence, $\hat{u}(\omega) \in v_q^*(\omega)$ for every q, with $\mu_q = \tilde{\mu}$. We obtain the desired contradiction, since $v_q^*(\omega) = \bar{u}(\omega)$ for such q.

Proof of Proposition 2.5

If all types of records are perfect substitutes, $MRS_q(\omega, \omega') = -\frac{v_q^*(\omega)}{v_q^*(\omega')}$ must be constant for all (ω, ω') and q. By Lemma 2.2 and Proposition 2.3, it follows that $v_q^*(\omega) = \bar{u}(\omega)$ for all ω and q. It follows that it is optimal to always fully disclose every record.

Fix $q \in \mathbb{R}^\Omega_{++}$. Suppose that an optimal mechanism x_q^* involves full disclosure.

Then, we have

$$v_q^*(\omega) = u_q^*(\omega) + \sum_{a \in A} t_q^*(a,\omega) x_q^*(a|\omega) \geq u_q^*(\omega),$$

where the inequality follows from $x_q^*(\cdot|\omega) \in CE(\Gamma_\omega)$ for all ω. Since by Lemma 2.1 we must have $\sum_\omega v_q^*(\omega)q(\omega) = \sum_\omega u_q^*(\omega)q(\omega)$, it follows that $v_q^*(\omega) = u_q^*(\omega)$ for all ω. Finally, since x_q^* is optimal, it must be that $u_q^*(\omega) = \bar{u}(\omega)$ for all ω. Now, note that v_q^* defines a supporting hyperplane of the iso-payoff line of level $U^*(q)$ at q. The intercept of such an hyperplane on each ω-axis is $\hat{q}_\omega(\omega) = \frac{U^*(q)}{\bar{u}(\omega)}$ and $\hat{q}_\omega(\omega') = 0$ for $\omega' \neq \omega$. By definition, each $\hat{q}_\omega$ also belongs to the iso-payoff line of level $U^*(q)$ and therefore $U^*(q) = U^*(\hat{q}_\omega)$ for all ω. In other words, the intercepts of the hyperplane and the iso-payoff line coincide for all ω.

Now consider any $q' \in \mathbb{R}_{++}$, $q' \neq q$, that belongs to the supporting hyperplane of level $U^*(q)$ at q. By definition, we can obtain q' as a convex combination of intercepts $\hat{q}_\omega$ on each axis. Specifically, there exists $\beta \in \Delta(\Omega)$ such that $q'(\omega) = \beta(\omega)\hat{q}_\omega(\omega)$ for all ω. By concavity of $U^*(q)$ (Footnote 17), we must have that

$$U^*(q') = \sum_{\omega \in \Omega} v_{q'}^*(\omega) q'(\omega) \leq U^*(q) = \sum_{\omega \in \Omega} \beta(\omega) U^*(\hat{q}_\omega) = \sum_{\omega \in \Omega} \bar{u}(\omega) q'(\omega).$$

But since $v_{q'}^*(\omega) \geq \bar{u}(\omega)$ for all ω by Lemma 2.2, we must have $v_{q'}^*(\omega) = \bar{u}(\omega)$ for all ω. Then $v_{q''}^*(\omega) = \bar{u}(\omega)$ for all q'' that belong to the supporting hyperplane of level $U^*(q)$ at q. Finally, since v_q^* is invariant to scaling of q, it follows that $v_q^*(\omega) = \bar{u}(\omega)$ for all ω and all $q \in \mathbb{R}_+^\Omega$.

Proof of Corollary 2.4

For $\pi > \frac{1}{2}$, we have $MRS_q(\omega,\omega') = -\frac{\omega}{\omega'}$ and all ω,ω'. Consider now $\pi < \frac{1}{2}$:

$$MRS_q(\omega,\omega') = \begin{cases} -\frac{\omega}{\omega'} & \text{if } \omega,\omega' < a_q \\ -\frac{(1-\pi)\omega}{\pi a_q + (1-\pi)(\omega' - a_q)} & \text{if } \omega < a_q \leq \omega' \\ -\frac{\pi a_q + (1-\pi)(\omega - a_q)}{\pi a_q + (1-\pi)(\omega' - a_q)} & \text{if } \omega,\omega' \geq a_q. \end{cases}$$

Thus, we have

$$\frac{\partial MRS_q(\omega,\omega')}{\partial \pi} = \begin{cases} 0 & \text{if } \omega,\omega' < a_q \\ \frac{\omega a_q}{[\pi a_q + (1-\pi)(\omega' - a_q)]^2} & \text{if } \omega < a_q \leq \omega' \\ -\frac{a_q(\omega' - \omega)}{[\pi a_q + (1-\pi)(\omega' - a_q)]^2} & \text{if } \omega,\omega' \geq a_q. \end{cases}$$

Finally, it is easy to see that $MRS_q(\omega,\omega') < -\frac{\omega}{\omega'}$ for $\omega < a_q \leq \omega'$ and that $MRS_q(\omega,\omega') > -\frac{\omega}{\omega'}$ for $\omega' > \omega \geq a_q$.

Proof of Corollary 2.5

Fix ω, q, and a refinement σ_ω. Since $u_i(a,\omega) = \mathbb{E}_{\sigma_\omega}[u_i(a,\omega')|\omega]$ for all i, by (2.4) we have

$$\begin{aligned} v_q^*(\omega) &= \max_{a \in A} \sum_{\omega' \in \Omega} [u_0(a,\omega') + t_q^*(a,\omega')]\sigma_\omega(\omega') \\ &\leq \sum_{\omega' \in \Omega} \max_{a \in A} [u_0(a,\omega') + t_q^*(a,\omega')]\sigma_\omega(\omega') = \sum_{\omega' \in \Omega} v_q^*(\omega')\sigma_\omega(\omega'). \end{aligned} \tag{A.7}$$

Thus, if refining $\alpha q(\omega)$ of the original records of type ω according to σ_ω does not change the value of any record, then (A.7) implies the desired inequality. Now consider the other case: There exists a share $\alpha > 0$ such that refining $\alpha q(\omega)$ of the current records of type ω according to σ_ω leads to a new database q_α such that $v_{q_\alpha}^*(\omega') \neq v_q^*(\omega')$ for some $\omega' \in \operatorname{supp} \sigma_\omega$ or $\omega' = \omega$. Since the total quantity of records does not change, we have that $\mu_{q_\alpha}(\omega) < \mu_q(\omega)$ and $\mu_{q_\alpha}(\omega') > \mu_q(\omega')$ for all $\omega' \in \operatorname{supp} \sigma_\omega$. By Proposition 2.3, it follows that $v^*_{q_\alpha}(\omega) \geq v^*_q(\omega)$ and $v^*_{q_\alpha}(\omega') \leq v^*_q(\omega')$ for all $\omega' \in \operatorname{supp} \sigma_\omega$ and that the indirect effects are increasing in α. Now, note that for all α,

$$\sum_{\omega' \in \Omega} v^*_{q_\alpha}(\omega')\sigma_\omega(\omega') \geq v^*_{q_\alpha}(\omega) \geq v^*_q(\omega), \tag{A.8}$$

where the first inequality follows from (A.7). This implies that the direct effect of a refinement is always non-negative and decreasing in α.

Proof of Proposition 2.6

The directional derivative of U^* at any $\hat{q}$ in the direction σ_ω is equal to

$$\sum_{\omega' \in \Omega} v^*_{\hat{q}}(\omega')\sigma_\omega(\omega') - v^*_{\hat{q}}(\omega).$$

The linear path from q to q_α can be parametrized as follows: for $t \in [0,1]$, define $q_t(\omega) = q(\omega) - t\alpha q(\omega)$, $q_t(\omega') = q(\omega') + t\alpha\sigma_\omega(\omega')q(\omega)$ for $\omega' \in \operatorname{supp} \sigma_\omega$, and $q_t(\omega'') = q(\omega'')$ for remaining ω''. Note that $\sum_{\omega' \in \Omega} v^*_{q_t}(\omega')\sigma_\omega(\omega') - v^*_{q_t}(\omega)$ is non-negative by (A.7) and decreasing in t by the scarcity principle. Finally, by the gradient theorem,

$$U^*(q_\alpha) - U^*(q) = \int_0^1 v^*_{q_t} \cdot \nabla q_t \, dt = \alpha q(\omega) \int_0^1 \left[\sum_{\omega' \in \Omega} v^*_{q_t}(\omega')\sigma_\omega(\omega') - v^*_{q_t}(\omega) \right] dt \geq 0,$$

where ∇q_t is the gradient of q_t with respect to t.

Suppose that there exists a common $\tilde{a} \in \operatorname{supp} x^*_q(\cdot|\omega)$ that satisfies $x^*_q(\tilde{a}|\omega'') > 0$ for all $\omega'' \in \operatorname{supp} \sigma_\omega$. By complementary slackness, it follows that for all $\omega'' \in \operatorname{supp} \sigma_\omega$, we have $v_q^*(\omega'') = u_0(\tilde{a},\omega'') + t_q^*(\tilde{a},\omega'')$. Therefore, by the scarcity principle,

$$\sum_{\omega''\in\Omega} v_{q_\alpha}^*(\omega'')\sigma_\omega(\omega'') \le \sum_{\omega''\in\Omega} v_q^*(\omega'')\sigma_\omega(\omega'') = v_q^*(\omega) \le v_{q_\alpha}^*(\omega),$$

which, combined with (A.8), implies that $\sum_{\omega''\in\Omega} v_{q_\alpha}^*(\omega'')\sigma_\omega(\omega'') = v_{q_\alpha}^*(\omega)$ for all $\alpha \in [0,1]$. In turn, this implies that $U^*(q_\alpha) = U^*(q)$ for all $\alpha \in [0,1]$.

Conversely, suppose that for every $\hat{a} \in \operatorname{supp} x_q^*(\cdot|\omega)$ there exists $\omega' \in \operatorname{supp}\sigma_\omega$ that satisfies $x_q^*(\hat{a}|\omega') = 0$. If the solution to the data-value problem is unique for database q—which is the case generically—then $x_q^*(\hat{a}|\omega') = 0$ implies $v_q^*(\omega') > u_0(\hat{a},\omega') + t_q^*(\hat{a},\omega')$ by strict complementary slackness. This and Proposition 2.4 imply that there exists $t' > 0$ such that $\sum_{\omega'\in\Omega} v_{q_t}^*(\omega')\sigma_\omega(\omega') > v_{q_t}^*(\omega)$ for all $t \in [0,1]$. It follows that $U^*(q_\alpha) > U^*(q)$.

Interpreting the Data-Value Problem

To further understand the value of data records and the externalities between them, we provide a stand-alone interpretation of the data-value problem $\mathcal{V}_q$. With minor adjustments, this extends to the problems described in Section 2.5.3. We fix $q \in \mathbb{R}_{++}^\Omega$ and so drop it from notation.

We first rewrite $\mathcal{V}$ in the following equivalent way by exploiting the structure of the specific problem at hand. For every i, we can set $\lambda_i(a_i|a_i) = 1$ (or any strictly positive number) for all $a_i \in A_i$. Given this, for every i and $(a_i) \in A_i$, define

$$b_i(a_i) = \sum_{a_i'\in A_i} \lambda_i(a_i'|a_i),$$

which is strictly positive by construction. Also, for every i and $(a_i', a_i) \in A_i \times A_i$ define

$$\ell_i(a_i'|a_i) = \frac{\lambda_i(a_i'|a_i)}{b_i(a_i)},$$

which implies that $\ell_i(\cdot|a_i) \in \Delta(A_i)$. After constructing $b = (b_1, \ldots, b_n)$ and $\ell = (\ell_1, \ldots, \ell_n)$

in this way, for each $i \in I$ and (a, ω) define

$$t_i(a,\omega) = b_i(a_i) \sum_{a_i' \in A_i} \left(u_i(a_i, a_{-i}, \omega) - u_i(a_i', a_{-i}, \omega) \right) \ell_i(a_i' | a_i)$$

and $t(a,\omega) = \sum_{i \in I} t_i(a,\omega)$. The data-value problem can be written as

$$\begin{aligned} \mathcal{V}: \quad & \min_{v,b,\ell} \sum_{\omega \in \Omega} v(\omega) q(\omega) \\ & \text{s.t. for all } \omega \in \Omega, \\ & v(\omega) = \max_{a \in A} \left\{ u_0(a,\omega) + t(a,\omega) \right\}, \end{aligned} \tag{A.9}$$

Gambles Against the Agents

Our interpretation hinges on unpacking how the platform determines the sellers' contributions to the externalities between records. By (A.9), it does so by choosing b and ℓ, which fully pin down $t(a,\omega)$ and hence $v(\omega)$. Recall that the platform wants to *minimize* the values of its records, so it would like to lower $t(a,\omega) = \sum_{i \in I} t_i(a,\omega)$ as much as possible for all (a,ω). Each term of $t_i(a,\omega)$ takes the form

$$b_i(a_i) \ell_i(a_i' | a_i) \left(u_i(a_i, a_{-i}, \omega) - u_i(a_i', a_{-i}, \omega) \right),$$

which contributes to lowering $t_i(a,\omega)$ if and only if $\ell_i(a_i'|a_i) > 0$ and $u_i(a_i, a_{-i}, \omega) < u_i(a_i', a_{-i}, \omega)$. That is, if seller i knew ω and his opponents' offers a_{-i}, he would strictly prefer a_i' to a_i. In this case, offering a_i amounts to making a mistake from an ex-post viewpoint. We will say that seller i regrets offering a_i.

Thus, inducing sellers to make offers they will regret emerges as an intrinsic goal of the platform's problem—together with maximizing u_0. In this view, (b_i, ℓ_i) becomes an exploitation strategy on the part of the platform against seller i. Inducing regrettable actions requires withholding information from seller i about ω or a_{-i}. This explains why the platform may prefer partial disclosure, but from the perspective of the data-value problem. In the end, $v(\omega)$ results from a trade-off between $u_0(a,\omega)$ and the return from inducing actions the sellers regret.

This return depends on the structure of b and ℓ, which define a family of gambles against the sellers. To see this, fix (a,ω) and seller i. Then, $\ell_i(\cdot|a_i) \in \Delta(A_i)$ defines a lottery whose prize for the platform is $u_i(a_i,a_{-i},\omega) - u_i(a_i',a_{-i},\omega)$ for each a_i'; the scaling term $b_i(a_i)$ defines the stakes that it bets on this lottery. The platform "wins" when $u_i(a_i,a_{-i},\omega) < u_i(a_i',a_{-i},\omega)$ and "loses" otherwise. Thus, $t(a,\omega)$ is the overall expected prize from (b,ℓ). We can then think of $\mathcal{V}$ as a fictitious environment where money is a medium of exchange and the platform can write monetary gambling contracts with each seller. Such contracts are enforced through contingent-claim markets that determine prizes based on the interaction's type ω and outcome a.[1]

We can then link how the platform chooses these gambles in $\mathcal{V}$ with the externalities between records. Negative externalities $t^*(\omega) < 0$ correspond to favorable gambles, in the sense that the platform wins in expectation. This requires the help of other records to withhold information and induce the sellers to make offers they will regret. Conversely, positive externalities $t^*(\omega) > 0$ correspond to unfavorable gambles. Corollary 2.1 implies that, at the optimum, the platform chooses gambles that favor it for some records, but not for others. In fact, this stems from deeper constraints and trade-offs in the use of such gambles against the sellers.

Feasible Gambles and Trade-offs

The feasible gambles in $\mathcal{V}$ have specific features that shed light on the data-value problem.

Some features reflect structural properties of $\mathcal{V}$. While the prizes of each gamble are contingent on ω and the entire a, for each seller i both b_i and ℓ_i can depend only on a_i.

[1]See Nau (1992) for a related interpretation.

This limits the platform's ability to tailor its gambles across records and sellers. These properties reflect in $\mathcal{V}$ key interdependences in $\mathcal{U}$: The independence of (b_i, ℓ_i) from a_{-i} reflects the interdependence in $\mathcal{U}$ between sellers' incentives; the independence of (b_i, ℓ_i) from ω reflects the non-separability of $\mathcal{U}$ across data records. To see this, suppose $\ell_i(\hat{a}_i|a_i) > 0$. Then, (b_i, ℓ_i) links the value formula (A.9) for (a_i, a_{-i}, ω) and (a_i, a'_{-i}, ω'). In particular, if $u_i(a_i, a_{-i}, \omega) < u_i(\hat{a}_i, a_{-i}, \omega)$ but $u_i(a_i, a'_{-i}, \omega') > u_i(\hat{a}_i, a'_{-i}, \omega')$, the platform faces a trade-off because it may not be possible to use (b_i, ℓ_i) to lower $v(\omega)$ without also raising $v(\omega')$. This is another way to see why and how externalities arise between records. When committing to (b, ℓ) the platform has to take into account these effects of each (b_i, ℓ_i) across records.

How it solves the trade-offs depends on the relative frequency of records in the database (hence q). Importantly, this transformation of non-separabilities in $\mathcal{U}$ into independence properties of (b, ℓ) is what enables $\mathcal{V}$ to assign values individually to each record.

The platform faces other constraints in its ability to *jointly* exploit the sellers. Given $\mathcal{V}$, it is clear that it would want to choose (b, ℓ) so that $t(a, \omega) \leq 0$ for all (a, ω) with some strict inequality. Such gambles would guarantee a sure arbitrage against the sellers, but are infeasible in the following sense. By complementary slackness $x^*(a|\omega) > 0$ implies $v^*(\omega) = u_0(a, \omega) + t^*(a, \omega)$. Thus, since every ω must induce some a for every x, action profiles that cannot be in the support of any obedient $x(\cdot|\omega)$ are irrelevant for determining $v^*(\omega)$. Given this, define

$$\mathbf{X} = \{(a, \omega) \in A \times \Omega : x(a|\omega) > 0 \text{ for some obedient } x\}.$$

Let $G(\mathbf{X})$ be the set of gambles that can be contingent only on $(a, \omega) \in \mathbf{X}$ (formally, we restrict the functions b and ℓ to the subdomain $\mathbf{X}$). Note that restricting the platform to choosing from $G(\mathbf{X})$ in $\mathcal{V}$ is immaterial, as restricting x to domain $\mathbf{X}$ is immaterial in $\mathcal{U}$.

Proposition A.2.1. For every gamble $(b,\ell) \in G(\mathbf{X})$, if $t(a,\omega) < 0$ for some (a,ω), there must exist (a',ω') such that $t(a',\omega') > 0$.

This property is closely related to a similar result in Nau (1992). For completeness we provide a proof below, which relies on a dual characterization of $\mathbf{X}$ using Farkas' lemma.

The economic takeaway is that in the attempt to minimize values v by exploiting the sellers with (b,ℓ), the platform faces a fundamental trade-off that is a hallmark of $\mathcal{V}$. Successfully exploiting the sellers for records of type ω with some outcome a requires paying the cost of losing against them for records of some other type ω' or outcome a'. This result sheds light on how and how much the platform can actually manipulate sellers by conveying information.

Proof of Proposition A.2.1 This proof is for the general case where the principal can choose $a_0 \in A_0$ and each agent i can privately observe some own data $\omega_i \in \Omega_i$ about the interaction he is in. Fix $(a^*,\omega^*) \in \mathbf{X}$ and introduce $\mathbf{1}_{a^*,\omega^*}$ as a vector of size $|\mathbf{X}|$ with $\varepsilon > 0$ in the position indexed by (a^*,ω^*) and 0 in all other positions. Constitute a matrix $\mathbf{W}$ such that its rows are indexed by $(a,\omega) \in \mathbf{X}$, its columns are indexed by (i,a_i',a_i,ω_i), $i \in I$, and its entries are as follows:

$$\mathbf{W}\left((\tilde{a},\tilde{\omega}),(i,a_i',a_i,\omega_i)\right) = 1\{a_i = \tilde{a}_i, \omega_i = \tilde{\omega}_i\}\left(u_i(a_i,\tilde{a}_{-i},\omega_i,\tilde{\omega}_{-i}) - u_i(a_i',\tilde{a}_{-i},\omega_i,\tilde{\omega}_{-i})\right).$$

By a variant of the Farkas' lemma, either there exists $\lambda \geq 0$, such that $\mathbf{W}\lambda \leq -\mathbf{1}_{a^*,\omega^*}$, or else there exists $\chi \geq 0$, such that $\mathbf{W}^T\chi \geq 0$, with $\chi^T\mathbf{1}_{a^*,\omega^*} > 0$. Now we show that the latter is true. Indeed, we can pick $\chi(a,\omega) = q(\omega)x(a|\omega)$, where x is obedient and satisfies $x(a^*|\omega^*) > 0$. We can find such x, since $(a^*,\omega^*) \in \mathbf{X}$. Then $\chi \geq 0$ and $\chi^T\mathbf{1}_{a^*,\omega^*} > 0$ are satisfied automatically. Finally, $\mathbf{W}^T\chi \geq 0$ corresponds exactly to the set of obedience constraints in $\mathcal{U}_q$ restricted to the subdomain $\mathbf{X}$.

Since any λ can be decomposed as $\lambda_i(a_i'|a_i,\omega_i) = b_i(a_i,\omega_i)\ell_i(a_i'|a_i,\omega_i)$, we conclude that there is no $(b,\ell) \in G(\mathbf{X})$ that satisfies $t(a,\omega) \leq 0$ for every $(a,\omega) \in \mathbf{X}$ and $t(a^*,\omega^*) < -\varepsilon$. The result then follows, since the choice of $(a^*,\omega^*) \in \mathbf{X}$ and $\varepsilon > 0$ was arbitrary.

A Sufficient Condition for Optimality of Withholding Information

We provide a sufficient condition on Γ for optimality of withholding information for the general case where the principal can choose $a_0 \in A_0$ and each agent i can privately observe some own data $\omega_i \in \Omega_i$. Recall that if the principal always fully disclose all ω, then its must be implementing a correlated equilibrium of the complete-information game Γ_ω for all ω (i.e., $x_q^*(\cdot|\omega) \in CE(\Gamma_\omega)$). The definition of CE in terms of inequalities can be adjusted to incorporate the principal's a_0.

Proposition A.2.2. Fix Γ. Suppose there exists (a,ω) that satisfies:

(1) $u_0(a,\omega) > \bar{u}(\omega)$,

(2) for every agent i and action $\hat{a}_i$, such that $u_i(a_i,a_{-i},\omega) < u_i(\hat{a}_i,a_{-i},\omega)$, there exists an $x(\cdot|\omega') \in CE(\Gamma_{\omega'})$ for some ω', with $\omega'_i = \omega_i$, that satisfies

$$\sum_{a \in A} u_0(a,\omega')x(a|\omega') = \bar{u}(\omega'),$$

$$\sum_{a_{-i} \in A_{-i}} \left(u_i(a_i,a_{-i},\omega') - u_i(\hat{a}_i,a_{-i},\omega')\right) x(a_i,a_{-i}|\omega') > 0.$$

Then it is not optimal in $\mathcal{U}_q$ to always fully disclose all records for any $q \in \mathbb{R}^{\Omega}_{++}$.

Condition (1) is clearly necessary: If for every records of type ω every action profile a cannot deliver a payoff higher than the full-information payoff $\bar{u}(\omega)$, then it is clearly optimal for the principal to fully reveal every ω. Given an outcome (a,ω) with $u_0(a,\omega) > \bar{u}(\omega)$, there must be an agent who would have a profitable deviation from a_i to $\hat{a}_i$ if he knew (a_{-i},ω_{-i}). Otherwise, given a_0, the profile a_{-0} is a Nash Equilibrium of

Γ_ω and hence $a_{-0} \in CE(\Gamma_\omega)$, which would imply $u_0(a,\omega) \leq \bar{u}(\omega)$. Then condition (2) requires that agent i's data ω_i is consistent with another record ω'—so that he cannot tell ω and ω' apart based on his own data only—which admits a principal-preferred correlated equilibrium that also recommends i to play a_i and renders the deviation to $\hat{a}_i$ strictly suboptimal. Note that this condition is easy to check in applications starting from the best full-disclosure mechanism x.

Proof of Proposition A.2.2 We will argue by contradiction. Suppose $q \in \mathbb{R}^{\Omega}_{++}$ and $\mathcal{U}_q$ admits a full-disclosure solution x^{**}_q and hence $x^{**}_q(\cdot|\tilde{\omega}) \in CE(\Gamma_{\tilde{\omega}})$ and $u^{**}_q(\tilde{\omega}) = \bar{u}(\tilde{\omega})$ for all $\tilde{\omega} \in \Omega$. Then $v^{**}_q(\tilde{\omega}) = u^{**}_q(\tilde{\omega}) = \bar{u}(\tilde{\omega})$ for all $\tilde{\omega} \in \Omega$ by Proposition 2.5.

Now suppose that (a,ω) satisfies both conditions in the statement of the proposition. For $(v^{**}_q, \lambda^{**}_q)$ to be feasible for $\mathcal{V}_q$, we must have for all $\tilde{\omega} \in \Omega$,

$$v^{**}_q(\tilde{\omega}) \geq u_0(a,\tilde{\omega}) + t^{**}_q(a,\tilde{\omega}).$$

Since $u_0(a,\omega) > \bar{u}(\omega) = v^{**}_q(\omega)$, we must have $t^{**}_q(a,\omega) < 0$. Therefore, there exists a pair $(i,\hat{a}_i)$ that satisfies $u_i(a_i,a_{-i},\omega) < u_i(\hat{a}_i,a_{-i},\omega)$ and $\lambda^{**}_{q,i}(\hat{a}_i|a_i,\omega_i) > 0$. For such a pair $(i,\hat{a}_i)$, there exists $x(\cdot|\omega') \in CE(\Gamma_{\omega'})$ with the properties listed in the proposition. Then, since $\lambda^{**}_{q,i}(\hat{a}_i|a_i,\omega_i) > 0$,

$$\begin{aligned}
& \sum_{\tilde{a}\in A} u_0(\tilde{a},\omega')x(\tilde{a}|\omega') + \sum_{\tilde{a}\in A} t^{**}_q(\tilde{a},\omega')x(\tilde{a}|\omega') \\
\geq\ & \sum_{\tilde{a}\in A} u_0(\tilde{a},\omega')x(\tilde{a}|\omega') \\
& + \lambda^{**}_{q,i}(\hat{a}_i|a_i,\omega_i)\left\{ \sum_{\tilde{a}_{-i}\in A_{-i}} \left(u_i(a_i,\tilde{a}_{-i},\omega') - u_i(\hat{a}_i,\tilde{a}_{-i},\omega')\right) x(a_i,\tilde{a}_{-i}|\omega') \right\} \\
>\ & \sum_{\tilde{a}\in A} u_0(\tilde{a},\omega')x(\tilde{a}|\omega') = v^{**}_q(\omega'),
\end{aligned}$$

where the first inequality follows because $x(\cdot|\omega') \in CE(\Gamma_{\omega'})$. The strict inequality is incompatible with constraint (2.4) and delivers the desired contradiction.

Data and Price Discrimination: Analysis

This section provides the calculations for Section 2.4.4 and the example in Section 2.5.1. Recall that $a \in \{1,2\}$ and that $u_0(a,\omega) = \max\{\omega - a, 0\}$ and $u_1(a,\omega) = a\mathbb{I}\{\omega \geq a\}$ for $\omega \in \{\omega_1, \omega_2\}$. For ω°, we have $u_i(a,\omega^\circ) = hu_i(a,\omega_2) + (1-h)u_i(a,\omega_1)$ for $i = 0,1$. For completeness, we solve both the information-design problem and the data-value problem.

Information Design

The objective function is

$$(\omega_2 - \omega_1)x(1|\omega_2)q(\omega_2) + h(\omega_2 - \omega_1)x(1|\omega^\circ)q(\omega^\circ) = x(1|\omega_2)q(\omega_2) + hx(1|\omega^\circ)q(\omega^\circ).$$

The obedience constraints are

$$-x(2|\omega_1)q(\omega_1) + x(2|\omega_2)q(\omega_2) + (2h-1)x(2|\omega^\circ)q(\omega^\circ) \geq 0,$$
$$x(1|\omega_1)q(\omega_1) - x(1|\omega_2)q(\omega_2) - (2h-1)x(1|\omega^\circ)q(\omega^\circ) \geq 0.$$

Consider first the case of $h > \frac{1}{2}$. From the second constraint we get $x_q^*(1|\omega_1) = 1$. The first constraint is then automatically satisfied. Since $h \in (0,1)$, it is always true that $2h - 1 < h$. The solution satisfies $x_q^*(1|\omega_2) = 0$ and $x_q^*(1|\omega^\circ) = \frac{1}{2h-1}\frac{q(\omega_1)}{q(\omega^\circ)}$, as long as $\frac{1}{2h-1}\frac{q(\omega_1)}{q(\omega^\circ)} \leq 1$.

Now consider the case of $h \leq \frac{1}{2}$. Combining obedience constraints, we get

$$x(1|\omega_1)q(\omega_1) - x(1|\omega_2)q(\omega_2) - (2h-1)x(1|\omega^\circ)q(\omega^\circ) \geq$$

$$\max\{2q(\omega_1) + (1-h)2q(\omega^\circ) - 1, 0\}.$$

It is immediate that $x_q^*(1|\omega^\circ) = x_q^*(1|\omega_1) = 1$, since this relaxes the constraint as much as possible. The constraint then becomes

$$q(\omega_1) - (2h-1)q(\omega^\circ) - \max\{2q(\omega_1) + (1-h)2q(\omega^\circ) - 1, 0\} \geq x(1|\omega_2)q(\omega_2).$$

Data Value

The data-value problem is

$$\min_{v,\lambda} q(\omega_1)v(\omega_1) + q(\omega_2)v(\omega_2) + q(\omega^\circ)v(\omega^\circ),$$

subject to $\lambda(2|1), \lambda(1|2) \geq 0$,

$$\begin{aligned} v(\omega_1) &= \max\{\lambda(2|1), -\lambda(1|2)\} = \lambda(2|1), \\ v(\omega_2) &= \max\{1 - \lambda(2|1), \lambda(1|2)\}, \\ v(\omega^\circ) &= \max\{h + (1-2h)\lambda(2|1), (2h-1)\lambda(1|2)\} \\ &= h \max\left\{1 - \frac{2h-1}{h}\lambda(2|1), \frac{2h-1}{h}\lambda(1|2)\right\}. \end{aligned}$$

As we noted before, $\frac{2h-1}{h} < 1$. Suppose that $h > \frac{1}{2}$. Then, it is optimal to set $\lambda_q^*(1|2) = 0$ to relax the problem as much as possible. We then have

$$\begin{aligned} v(\omega_1) &= \lambda(2|1), \\ v(\omega_2) &= \max\{1 - \lambda(2|1), 0\}, \\ v(\omega^\circ) &= h \max\left\{1 - \frac{2h-1}{h}\lambda(2|1), 0\right\}. \end{aligned}$$

There are three candidates for optimal $\lambda(2|1)$. When $\lambda(2|1) = 0$, the objective is $S_0 \triangleq q(\omega_2) + hq(\omega^\circ)$. When $\lambda(2|1) = 1$, the objective is $S_1 \triangleq q(\omega_1) + q(\omega^\circ)(1-h)$. When $\lambda(2|1) = \frac{h}{2h-1}$, the objective is $S_f \triangleq q(\omega_1)\frac{h}{2h-1}$. The following claims are true. First, $S_0 \leq S_1$ if and only if $q(\omega_1) \geq q(\omega_2) + (2h-1)q(\omega^\circ)$. Second, $S_0 \leq S_f$ if and only if

$q(\omega_1) \geq q(\omega_2)\frac{2h-1}{h} + (2h-1)q(\omega^\circ)$. Third, $S_1 \leq S_f$ if and only if $q(\omega_1) \geq (2h-1)q(\omega^\circ)$.

Suppose now that $h \leq \frac{\omega_1}{\omega_2}$. Then $v(\omega^\circ) = h - (2h-1)\lambda(2|1)$ and $\lambda_q^*(1|2) = 0$ is again optimal. There are only two candidates for optimal $\lambda(2|1)$, specifically, 0 and 1.

Summary

All these cases lead to three scenarios in terms of q.

Scenario 1: $q(\omega_1) \leq (2h-1)q(\omega^\circ)$. Note that this requires $h > \frac{1}{2}$. Table A.3 presents the optimal x_q^*.

Table A.3. Platform example, x_q^* for Scenario 1.

$x_q^*(a\|\omega)$		ω		
		ω_1	ω_2	ω°
a	1	1	0	$\frac{1}{2h-1}\frac{q(\omega_1)}{q(\omega^\circ)}$
	2	0	1	$1 - \frac{1}{2h-1}\frac{q(\omega_1)}{q(\omega^\circ)}$

The solution to the data-value problem is $\lambda_q^*(1|2) = 0$, $\lambda_q^*(2|1) = \frac{h}{2h-1}$ and the unit values are $v_q^*(\omega_1) = \frac{h}{2h-1}$, $v_q^*(\omega_2) = 0$, and $v_q^*(\omega^\circ) = 0$.

Scenario 2: $(2h-1)q(\omega^\circ) \leq q(\omega_1) \leq q(\omega_2) + (2h-1)q(\omega^\circ)$. Note that the lower bound on $q(\omega_1)$ is meaningful only if $h > \frac{1}{2}$. Table A.4 presents the optimal x_q^*.

Table A.4. Platform example, x_q^* for Scenario 2.

$x_q^*(a\|\omega)$		ω		
		ω_1	ω_2	ω°
a	1	1	$\frac{q(\omega_1)-(2h-1)q(\omega^\circ)}{q(\omega_2)}$	1
	2	0	$1 - \frac{q(\omega_1)-(2h-1)q(\omega^\circ)}{q(\omega_2)}$	0

The solution to the data-value problem is $\lambda_q^*(1|2) = 0$, $\lambda_q^*(2|1) = 1$, and the unit values are $v_q^*(\omega_1) = 1$, $v_q^*(\omega_2) = 0$, and $v_q^*(\omega^\circ) = 1 - h$.

Scenario 3: $q(\omega_1) \geq q(\omega_2) + (2h-1)q(\omega^\circ)$. Table A.5 presents the optimal x_q^*.

The solution to the data-value problem is $\lambda_q^*(1|2) = \lambda_q^*(2|1) = 0$ and the unit values are $v_q^*(\omega_1) = 0$, $v_q^*(\omega_2) = 1$, and $v_q^*(\omega^\circ) = h$.

Table A.5. Platform example, x_q^* for Scenario 3.

$x_q^*(a\|\omega)$		ω_1	ω_2	ω°
a	1	1	1	1
	2	0	0	0

A.3 Omitted Proofs for Chapter 3

Proof of Lemma 3.1

The proof is standard and thence omitted (see, for example, Myerson, 1982).

Proof of Lemma 3.2

The honesty constraint (3.2) can be rewritten as follows: for every $\theta, \theta' \in \Theta$,

$$(u_S(a_1,\theta) - u_S(a_0,\theta))\left(\pi(a_1|\theta) - \pi(a_1|\theta')\right) \geq 0.$$

Pick two source types θ' and θ'' that prefer a_1 over a_0, that is, $\theta', \theta'' \in \Theta_1$. Then from the inequality above, $\pi(a_1|\theta') \geq \pi(a_1|\theta'')$ and $\pi(a_1|\theta'') \geq \pi(a_1|\theta')$, and consequently $\pi(a_1|\theta') = \pi(a_1|\theta'')$. Since θ' and θ'' were chosen arbitrarily from Θ_1, $\pi(a_1|\theta)$ is constant across $\theta \in \Theta_1$. Call this constant π_1. Similarly, $\pi(a_1|\theta)$ is constant across $\theta \in \Theta_0$. Call this constant π_0. Thus, the honesty constraint is equivalent to the following condition: for every $\theta_0 \in \Theta_0$ and $\theta_1 \in \Theta_1$,

$$(u_S(a_1,\theta_0) - u_S(a_0,\theta_0))(\pi_0 - \pi_1) \geq 0,$$

$$(u_S(a_1,\theta_1) - u_S(a_0,\theta_1))(\pi_1 - \pi_0) \geq 0.$$

Both of these inequalities are equivalent to $\pi_1 \geq \pi_0$.

Proof of Proposition 3.1

By Lemma 3.2, the obedience constraint (3.3) can be written as follows:

$$\pi_0 \int_{\Theta_0} \delta_R(\theta) d\mu_0 + \pi_1 \int_{\Theta_1} \delta_R(\theta) d\mu_0 \geq 0,$$

or $\pi_0 I_0 + \pi_1 I_1 \geq 0$. There are three cases to consider, depending on the signs of I_0 and I_1. The case in which $I_0 \geq 0$ and $I_1 \geq 0$ is ruled out by the tension condition (3.1).

1. If $I_0 < 0$ and $I_1 < 0$, then the only way to satisfy the obedience constraint is to set $\pi_0 = \pi_1 = 0$.

2. If $I_0 \geq 0$ and $I_1 < 0$, then $\pi_1 \leq \frac{I_0}{-I_1} \cdot \pi_0$. By Lemma 3.2, $\pi_1 \geq \pi_0$. Finally, $\frac{I_0}{-I_1} < 1$, since by the tension condition (3.1), $I_0 + I_1 < 0$. The only way to satisfy these inequalities is again to set $\pi_0 = \pi_1 = 0$.

3. Suppose $I_0 < 0$ and $I_1 \geq 0$. If $I_1 = 0$, then in the honest and obedient protocol, $\pi_0 = 0$ and $\pi_1 \in [0,1]$. If $I_1 > 0$, then

$$\pi_1 \geq \frac{-I_0}{I_1} \cdot \pi_0 \geq \pi_0,$$

where the second inequality is implied by the tension condition (3.1). Thus, the set of honest and obedient protocols for this case is the set of $\pi_0, \pi_1 \in [0,1]$, such that $\pi_1 \geq \frac{-I_0}{I_1} \cdot \pi_0$.

Proof of Proposition 3.2

If $I_1 < 0$, then the only incentive-compatible protocol is $\pi_0 = \pi_1 = 0$. The associated media's payoff is then 0.

If $I_1 = 0$, then $\pi_0 = 0$. The media choose π_1 as high as possible, that is, $\pi_1 = 1$.

Finally, if $I_1 > 0$, then the set of incentive-compatible protocols is the triangle depicted in Figure 1. According to the media's objective, the media want the pair (π_0, π_1) to be as high as possible. The solution then is the extreme point $\left(\frac{I_1}{-I_0}, 1\right)$. The corresponding payoff is then

$$\mu_0(\Theta_0) \cdot \frac{I_1}{-I_0} + \mu_0(\Theta_1) = \mu_0(\Theta_0) \cdot \frac{\int_{\Theta_1} \delta_R(\theta) d\mu_0}{-\int_{\Theta_0} \delta_R(\theta) d\mu_0} + \mu_0(\Theta_1) =$$

$$\mu_0(\Theta_1) \cdot \frac{\mathbb{E}[\delta_R(\theta)|\theta \in \Theta_1] - \mathbb{E}[\delta_R(\theta)|\theta \in \Theta_0]}{-\mathbb{E}[\delta_R(\theta)|\theta \in \Theta_0]}.$$

Proof of Claims 3.1, 3.2, and 3.3

These results directly follow from Proposition 3.2.

Proof of Claim 3.4

The media's problem facing no honesty constraints can be written as follows:

$$\max_{\pi(a_1|\theta) \in [0,1]^{[0,1]}} \int_0^1 \pi(a_1|\theta) d\theta,$$

subject to the obedience constraint (3.3) tailored to the example:

$$\int_0^1 (\theta - \omega)\pi(a_1|\theta) d\theta \geq 0.$$

First, note that $\pi(a_1|\theta) = 1$ for $\theta \geq \omega$. Indeed, this choice relaxes the obedience constraint as much as possible and provides the maximal payoff to the media for $\theta \geq \omega$. Thus, the problem is reduced to

$$\max_{\pi(a_1|\theta) \in [0,1]^{[0,\omega]}} \int_0^\omega \pi(a_1|\theta) d\theta + 1 - \omega,$$

subject to

$$\int_0^{\omega} (\omega - \theta)\pi(a_1|\theta)d\theta \le \frac{(1-\omega)^2}{2}.$$

But then the solution is $\pi(a_1|\theta) = 1$ for $2\omega - 1 \le \theta < \omega$, since those types are associated with "cheaper" cost of persuasion, namely, $\omega - \theta$. The optimal information structure then follows. The media's payoff is the length of the interval $[2\omega - 1, 1]$, which is $2 - 2\omega$.

Proof of Lemma 3.3

The argument is standard and can be found in Proposition 3 in Bergemann and Morris (2019).

Proof of Lemma 3.4

Suppose $\Omega_1 \in \text{supp}\phi$. Let $b = \sup(\Omega_1) \in [0,1]$. Note that the left-hand side of inequalities (3.6) and (3.7) is continuous and strictly decreasing in ω, as $\delta_R(\theta,\omega) = \theta - \omega$. Hence, any $\omega < b$ has to lie in Ω_1 by (3.6). Similarly, any $\omega > b$ has to lie in $[0,1] \setminus \Omega_1$ by (3.7). Thus, $\Omega_1 = [0,b]$. Finally, since δ is continuous in ω,

$$\int_0^1 \delta_R(\cdot,b)d\phi([0,b]|\cdot) = 0$$

has to be satisfied.

Proof of Lemma 3.5

Using Lemma 3.4, the honesty constraint (3.5) can be written as follows:

$$(u_S(a_1,\theta) - u_S(a_0,\theta)) \sum_{[0,b]\in\text{supp}\phi} (\phi([0,b]|\theta) - \phi([0,b]|\theta'))H(b) \ge 0$$

for every $\theta,\theta' \in \Theta$. Thus, by the same argument as in Lemma 3.2, for every $\theta',\theta'' \in \Theta_1$, $\sum_{[0,b]\in\text{supp}\phi} \phi([0,b]|\theta')H(b) = \sum_{[0,b]\in\text{supp}\phi} \phi([0,b]|\theta'')H(b)$. Then there exists $s_1 \in [0,1]$, such that $s_1 = \sum_{[0,b]\in \text{supp}\phi} \phi([0,b]|\theta)H(b)$ for every $\theta \in \Theta_1$. The same argument shows that $s_0 = \sum_{[0,b]\in \text{supp}\phi} \phi([0,b]|\theta)H(b)$ for every $\theta \in \Theta_0$. Finally, the honesty constraint for some type $\theta_1 \in \Theta_1$ deliberating a misreport $\theta_0 \in \Theta_0$ pins down that $s_1 \geq s_0$.

Proof of Lemma 3.6

Lemma 3.3, 3.4, and 3.5 imply the constraints of the reduced problem. The specification of sets Θ_0 and Θ_1 implies the media's objective function.

Proof of Lemma 3.7

The obedience constraints pin down $\underline{b}$ and $\bar{b}$ as functions of ϕ_0 and ϕ_1:

$$\bar{b} = f_1(\phi_0, \phi_1) = \frac{\phi_0 \int_{\Theta_0} \theta d\mu_0 + \phi_1 \int_{\Theta_1} \theta d\mu_0}{\phi_0 \mu_0(\Theta_0) + \phi_1 \mu_0(\Theta_1)}, \tag{A.10}$$

$$\underline{b} = f_2(\phi_0, \phi_1) = \frac{(1-\phi_0) \int_{\Theta_0} \theta d\mu_0 + (1-\phi_1) \int_{\Theta_1} \theta d\mu_0}{(1-\phi_0)\mu_0(\Theta_0) + (1-\phi_1)\mu_0(\Theta_1)}. \tag{A.11}$$

Then the boundary conditions are established. If $\phi_0 = \phi_1$, then $\underline{b} = \bar{b} = \mathbb{E}[\theta]$. If $\phi_0 = 0$, then $\bar{b} = \mathbb{E}[\theta|\theta \in \Theta_1]$. If $\phi_1 = 1$, then $\underline{b} = \mathbb{E}[\theta|\theta \in \Theta_0]$. The derivatives of $\bar{b}$ and $\underline{b}$ can be directly calculated:

$$\frac{\partial \bar{b}}{\partial \phi_0} = \frac{\int_{\Theta_0} \theta d\mu_0 \cdot (\phi_0 \mu_0(\Theta_0) + \phi_1 \mu_0(\Theta_1)) - (\phi_0 \int_{\Theta_0} \theta d\mu_0 + \phi_1 \int_{\Theta_1} \theta d\mu_0) \cdot \mu_0(\Theta_0)}{(\phi_0 \mu_0(\Theta_0) + \phi_1 \mu_0(\Theta_1))^2}$$

$$= \frac{\phi_1 \mu_0(\Theta_1) \int_{\Theta_0} \theta d\mu_0 - \phi_1 \mu_0(\Theta_0) \int_{\Theta_1} \theta d\mu_0}{(\phi_0 \mu_0(\Theta_0) + \phi_1 \mu_0(\Theta_1))^2}$$

$$= \mu_0(\Theta_0)\mu_0(\Theta_1) \cdot \frac{\phi_1(\mathbb{E}[\theta|\theta \in \Theta_0] - \mathbb{E}[\theta|\theta \in \Theta_1])}{(\phi_0 \mu_0(\Theta_0) + \phi_1 \mu_0(\Theta_1))^2} < 0.$$

$$\frac{\partial \bar{b}}{\partial \phi_1} = \frac{\int_{\Theta_1} \theta d\mu_0 \cdot (\phi_0 \mu_0(\Theta_0) + \phi_1 \mu_0(\Theta_1)) - (\phi_0 \int_{\Theta_0} \theta d\mu_0 + \phi_1 \int_{\Theta_1} \theta d\mu_0) \cdot \mu_0(\Theta_1)}{(\phi_0 \mu_0(\Theta_0) + \phi_1 \mu_0(\Theta_1))^2}$$

$$= \frac{\phi_0 \mu_0(\Theta_0) \int_{\Theta_1} \theta d\mu_0 - \phi_0 \mu_0(\Theta_1) \int_{\Theta_0} \theta d\mu_0}{(\phi_0 \mu_0(\Theta_0) + \phi_1 \mu_0(\Theta_1))^2}$$

$$= \mu_0(\Theta_0)\mu_0(\Theta_1) \cdot \frac{\phi_0(\mathbb{E}[\theta|\theta \in \Theta_1] - \mathbb{E}[\theta|\theta \in \Theta_0])}{(\phi_0 \mu_0(\Theta_0) + \phi_1 \mu_0(\Theta_1))^2} > 0.$$

Similarly, it can be shown that $\frac{\partial \underline{b}}{\partial \phi_0} > 0$ and $\frac{\partial \underline{b}}{\partial \phi_1} < 0$. The bounds on $\bar{b}$ and $\underline{b}$ in the statement of the lemma are then implied.

The signs of these derivatives are intuitive. For example, if ϕ_0 increases, then the probability of getting the recommendation $[0, \bar{b}]$ goes up for the lower competence levels $\theta \in \Theta_0 = [0, \bar{\theta})$. Thus, some receivers with high ω will find it optimal to switch from the mobilizing action to the status-quo action. That is, the obedience constraint will make $\bar{b}$ lower.

Given the boundary conditions and the signs of the derivatives above, there always exist $\phi_0, \phi_1 \in [0,1]$, with $\phi_1 \geq \phi_0$, such that $\bar{b} = f_1(\phi_0, \phi_1)$ and $\underline{b} = f_2(\phi_0, \phi_1)$ by a multivariate version of the mean value theorem.

It is worth mentioning that under our assumptions on μ_0, Θ_0, and Θ_1, the system of equations $\bar{b} = f_1(\phi_0, \phi_1)$, $\underline{b} = f_2(\phi_0, \phi_1)$ can be solved directly. It is a matter of algebra to show that

$$\phi_0 = \frac{1}{\bar{\theta}} \cdot \frac{(1 - 2\bar{b} + \bar{\theta})\left(\frac{1}{2} - \underline{b}\right)}{\bar{b} - \underline{b}},$$

and

$$\phi_1 = \frac{2}{1 - \bar{\theta}} \cdot \frac{\left(\bar{b} - \frac{\bar{\theta}}{2}\right)\left(\frac{1}{2} - \underline{b}\right)}{\bar{b} - \underline{b}},$$

if $\bar{b} > \underline{b}$. These ϕ_0 and ϕ_1 can be readily checked to satisfy $\phi_0, \phi_1 \in [0,1]$ and $\phi_1 \geq \phi_0$. If $\bar{b} = \underline{b}$, then $\bar{b} = \underline{b} = \mathbb{E}[\theta]$ which occurs as long as $\phi_0 = \phi_1$.

Proof of Proposition 3.3

Equations (A.10) and (A.11) can be combined to get

$$\underline{b} + (\bar{b} - \underline{b})(\phi_0\mu_0(\Theta_0) + \phi_1\mu_0(\Theta_1)) = \mathbb{E}[\theta].$$

By Lemma 3.6, the objective of the media is

$$\mu_0(\Theta_0)s_0 + \mu_0(\Theta_1)s_1 = H(\underline{b}) + (H(\bar{b}) - H(\underline{b}))(\phi_0\mu_0(\Theta_0) + \phi_1\mu_0(\Theta_1))$$

$$= H(\underline{b}) + \frac{\mathbb{E}[\theta] - \underline{b}}{\bar{b} - \underline{b}}(H(\bar{b}) - H(\underline{b})) = \frac{\bar{b} - \mathbb{E}[\theta]}{\bar{b} - \underline{b}}H(\underline{b}) + \frac{\mathbb{E}[\theta] - \underline{b}}{\bar{b} - \underline{b}}H(\bar{b}),$$

as long as $\bar{b} > \underline{b}$. If $\bar{b} = \underline{b}$, then the objective of the media is $H(\mathbb{E}[\theta])$. Note that

$$\frac{\bar{b} - \mathbb{E}[\theta]}{\bar{b} - \underline{b}} \cdot \underline{b} + \frac{\mathbb{E}[\theta] - \underline{b}}{\bar{b} - \underline{b}} \cdot \bar{b} = \mathbb{E}[\theta].$$

By Lemma 3.7, any $\bar{b} \in [\mathbb{E}[\theta], \mathbb{E}[\theta|\theta \in \Theta_1]]$ and $\bar{b} \in [\mathbb{E}[\theta|\theta \in \Theta_0], \mathbb{E}[\theta]]$ can be achieved by some simple protocol. Therefore, the media's problem is a splitting problem with the value function $\text{cav}\hat{H}[E[\theta]]$ (Le Treust and Tomala, 2019). The corresponding $\bar{b}$ and $\underline{b}$ can then be established as the supporting points of this object.

Bibliography

ACEMOGLU, DARON, ALI MAKHDOUMI, AZARAKHSH MALEKIAN, and ASUMAN OZDAGLAR (2021), "Too Much Data: Prices and Inefficiencies in Data Markets", *Americal Economic Journal: Microeconomics*, Forthcoming.

ACQUISTI, ALESSANDRO, CURTIS TAYLOR, and LIAD WAGMAN (2016), "The Economics of Privacy", *Journal of Economic Literature*, 54, 2, pp. 442-92.

ADMATI, ANAT and PAUL PFLEIDERER (1986), "A Monopolistic Market for Information", *Journal of Economic Theory*, 39, 2, pp. 400-438.

— (1990), "Direct and Indirect Sale of Information", *Econometrica*, 58, 4, pp. 901-28.

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

ALONSO, RICARDO and ODILON CÂMARA (2016), "Persuading Voters", *American Economic Review*, 106, 11, pp. 3590-3605.

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.

ARRIETA-IBARRA, IMANOL, LEONARD GOFF, DIEGO JIMÉNEZ-HERNÁNDEZ, JARON LANIER, and E. GLEN WEYL (2018), "Should We Treat Data as Labor? Moving beyond "Free"", *AEA Papers and Proceedings*, 108, pp. 38-42.

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

BALL, IAN (2019), "Scoring Strategic Agents", *Working Paper*.

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.

BERGEMANN, DIRK and ALESSANDRO BONATTI (2015), "Selling Cookies", *American Economic Journal: Microeconomics*, 7, 3, pp. 259-94.

— (2019), "Markets for Information: An Introduction", *Annual Review of Economics*, 11, 1, pp. 85-107.

BERGEMANN, DIRK, ALESSANDRO BONATTI, and ALEX SMOLIN (2018), "The Design and Price of Information", *American Economic Review*, 108, 1, pp. 1-48.

BERGEMANN, DIRK, BENJAMIN BROOKS, and STEPHEN MORRIS (2015), "The Limits of Price Discrimination", *American Economic Review*, 105, 3, pp. 921-57.

BERGEMANN, DIRK and STEPHEN MORRIS (2016a), "Bayes Correlated Equilibrium and the Comparison of Information Structures in Games", *Theoretical Economics*, 11, pp. 487-522.

— (2016b), "Information design, Bayesian persuasion, and Bayes correlated equilibrium", *American Economic Review*, 106, 5, pp. 586-91.

— (2019), "Information Design: A Unified Perspective", *Journal of Economic Literature*, 57(1), pp. 44-95).

BERGEMANN, DIRK and MARCO OTTAVIANI (2021), "Information Markets and Nonmarkets", *Handbook of Industrial Organization, forthcoming*, 4.

BERTSIMAS, DIMITRIS and JOHN N. TSITSIKLIS (1997), *Introduction to Linear Optimization*, Athena Scientific.

BLACKWELL, DAVID (1951), "Comparison of Experiments", *Proc. Second Berkeley Sympos. on Mathematical Statistics and Probability*.

— (1953), "Equivalent Comparisons of Experiments", *Annals of Mathematical Statistics*, pp. 265-272.

BLUME, ANDREAS, OLIVER J. BOARD, and KOHEI KAWAMURA (2007), "Noisy Talk", *Theoretical Economics*, 2, 4, pp. 395-440.

BOLESLAVSKY, RAPHAEL and KYUNGMIN KIM (2018), "Bayesian Persuasion and Moral Hazard", *Working Paper*.

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.

BROOKS, BENJAMIN and SONGZI DU (2020), “A Strong Minimax Theorem for Informationally-Robust Auction Design”, *Working Paper*.

— (2021), “Optimal auction design with common values: An informationally robust approach”, *Econometrica*, 89, 3, pp. 1313-1360.

BROWNLEE, JASON (2007), *Authoritarianism in an Age of Democratization*, Cambridge University Press Cambridge.

CALZOLARI, GIACOMO and ALESSANDRO PAVAN (2006), “On the Optimality of Privacy in Sequential Contracting”, *Journal of Economic Theory*, 130, 1, pp. 168-204.

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.

CHIANG, CHUN-FANG and BRIAN KNIGHT (2011), “Media Bias and Influence: Evidence from Newspaper Endorsements”, *The Review of Economic Studies*, 78, 3, pp. 795-820.

CHOI, JAY PIL, DOH-SHIN JEON, and BYUNG-CHEOL KIM (2019), “Privacy and Personal Data Collection with Information Externalities”, *Journal of Public Economics*, 173, pp. 113-124.

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

CRÉMER, JACQUES, YVES-ALEXANDRE DE MONTJOYE, and HEIKE SCHWEITZER (2019), “Competition Policy for the Digital Era”, *European Commission*.

DAS, SANMAY, EMIR KAMENICA, and RENEE MIRKA (2017), “Reducing Congestion through Information Design”, *2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, IEEE, pp. 1279-1284.

DIZDAR, DENIZ and EUGEN KOVÁČ (2020), “A Simple Proof of Strong Duality in the Linear Persuasion Problem”, *Games and Economic Behavior*, 122, pp. 407-412.

DORFMAN, ROBERT, PAUL ANTHONY SAMUELSON, and ROBERT M. SOLOW (1987), *Linear Programming and Economic Analysis*, Courier Corporation.

DOVAL, LAURA and VASILIKI SKRETA (2018), “Mechanism Design with Limited Commitment”, *Working Paper*.

DU, SONGZI (2018), “Robust Mechanisms Under Common Valuation”, *Econometrica*, 86(5), pp. 1569-1588.

DUGGAN, JOHN and CESAR MARTINELLI (2011), “A Spatial Theory of Media Slant and Voter Choice”, *The Review of Economic Studies*, 78, 2, pp. 640-666.

DWORCZAK, PIOTR and ANTON KOLOTILIN (2019), “The Persuasion Duality”, *Working Paper*.

DWORCZAK, PIOTR and GIORGIO MARTINI (2019), “The Simple Economics of Optimal Persuasion”, *Journal of Political Economy*, 127, 5, pp. 1993-2048.

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*.

EGOROV, GEORGY, SERGEI GURIEV, and KONSTANTIN SONIN (2009), “Why Resource-Poor Dictators Allow Freer Media: A Theory and Evidence from Panel Data”, *American Political Science Review*, 103, 4, pp. 645-668.

ELLIOTT, MATTHEW, ANDREA GALEOTTI, and ANDREW KOH (2020), “Market Segmentation through Information”, *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.

FEDERAL TRADE COMMISSION (2014), *Data Brokers: A Call for Transparency and Accountability*, A Report by the Federal Trade Commission, May.

GALE, DAVID (1989), *The Theory of Linear Economic Models*, University of Chicago Press.

GALPERTI, SIMONE and JACOPO PEREGO (2018), “A Dual Perspective on Information Design”, *Available at SSRN 3297406*.

GANDHI, JENNIFER and ELLEN LUST-OKAR (2009), “Elections under Authoritarianism”, *Annual Review of Political Science*, 12, pp. 403-422.

GEHLBACH, SCOTT and KONSTANTIN SONIN (2014), “Government Control of the Media”, *Journal of Public Economics*, 118, pp. 163-171.

GENTZKOW, MATTHEW and EMIR KAMENICA (2016), “A Rothschild-Stiglitz approach to Bayesian persuasion”, *American Economic Review*, 106, 5, pp. 597-601.

GENTZKOW, MATTHEW, JESSE M. SHAPIRO, and DANIEL F. STONE (2015), "Media Bias in the Marketplace: Theory", *Handbook of media economics*, Elsevier, vol. 1, pp. 623-645.

GOLDBERG, SAMUEL, GARRETT JOHNSON, and SCOTT SHRIVER (2021), "Regulating Privacy Online: An Economic Evaluation of the GDPR", *Working Paper*.

GOLTSMAN, MARIA, JOHANNES HÖRNER, GREGORY PAVLOV, and FRANCESCO SQUINTANI (2009), "Mediation, Arbitration and Negotiation", *Journal of Economic Theory*, 144, pp. 1397-1420.

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.

GURIEV, SERGEI and DANIEL TREISMAN (2018), "Informational Autocracy: Theory and Empirics of Modern Authoritarianism." *Working Paper*.

— (2019), "Informational Autocrats", *Journal of Economic Perspectives*, 33, 4, pp. 100-127.

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.

ICHIHASHI, SHOTA (2020), "Online Privacy and Information Disclosure by Consumers", *American Economic Review*, 110, 2, pp. 569-95.

— (2021), "The Economics of Data Externalities", *Journal of Economic Theory*, 196, p. 105316.

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.

JONES, CHARLES I. and CHRISTOPHER TONETTI (2020), "Nonrivalry and the Economics of Data", *American Economic Review*, 110, 9, pp. 2819-58.

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

KARTIK, NAVIN (2009), "Strategic Communication with Lying Costs", *The Review of Economic Studies*, 76, 4, pp. 1359-1395.

KOLOTILIN, ANTON (2018), "Optimal Information Disclosure: A Linear Programming Approach", *Theoretical Economics*, 13, pp. 607-635.

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 (1989), *The Economics of Uncertainty and Information*, MIT Press.

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

LANIER, JARON (2013), *Who Owns the Future?*, Simon & Schuster.

LE TREUST, MAËL and TRISTAN TOMALA (2019), "Persuasion with Limited Communication Capacity", *Journal of Economic Theory*, 184, p. 104940.

LEVADA CENTER (2019), *Do you approve the decisions of Vladimir Putin as the president of Russia?*, https://www.levada.ru/2019/11/18/vladimir-putin-7, Last accessed on 2021-12-30.

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

LIPNOWSKI, ELLIOT and LAURENT MATHEVET (2018), "Disclosure to a Psychological Audience", *American Economic Journal: Microeconomics*, 10, 4, pp. 67-93.

LIPNOWSKI, ELLIOT, DORON RAVID, and DENIS SHISHKIN (2019), "Persuasion via Weak Institutions", *Working Paper*.

LORENTZEN, PETER (2014), "China's Strategic Censorship", *American Journal of Political Science*, 58, 2, pp. 402-414.

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.

MILGROM, PAUL and CHRIS SHANNON (1994), "Monotone Comparative Statics", *Econometrica*, pp. 157-180.

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

MOSCARINI, GIUSEPPE and LONES SMITH (2002), "The Law of Large Demand for Information", *Econometrica*, 70, 6, pp. 2351-2366.

MYERSON, ROGER B. (1982), "Optimal Coordination Mechanisms in Generalized Principal–Agent Problems", *Journal of Mathematical Economics*, 10, 1, pp. 67-81.

— (1983), "Mechanism Design by an Informed Principal", *Econometrica*, 51, 6, pp. 1767-1797.

— (1984), "Two-Person Bargaining Problems with Incomplete Information", *Econometrica*, 52, 2, pp. 461-488.

— (1997), "Dual Reduction and Elementary Games", *Games and Economic Behavior*, 21, 1-2, pp. 183-202.

NAGEEB, ALI, GREG LEWIS, and SHOSHANA VASSERMAN (2022), "Voluntary Disclosure and Personalized Pricing", *Working Paper*.

NAIM, MOISES (2022), *The Revenge of Power: How Autocrats Are Reinventing Politics for the 21st Century*, St. Martin's Press.

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.

NAU, ROBERT F. (1992), "Joint Coherence in Games of Incomplete Information", *Management Science*, 38, 3, pp. 374-387.

NAU, ROBERT F. and KEVIN F. MCCARDLE (1990), "Coherent Behavior in Noncooperative Games", *Journal of Economic Theory*, 50, 2, pp. 424-444.

NEUNDORF, ANJA, JOHANNES GERSCHEWSKI, and ROMAN-GABRIEL OLAR (2020), "How Do Inclusionary and Exclusionary Autocracies Affect Ordinary People?", *Comparative Political Studies*, 53, 12, pp. 1890-1925.

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.

ROCKAFELLAR, TYRRELL R. (1970), *Convex Analysis*, Princeton University Press.

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

SAYEDI, AMIN, KINSHUK JERATH, and KANNAN SRINIVASAN (2014), “Competitive Poaching in Sponsored Search Advertising and Its Strategic Impact on Traditional Advertising”, *Marketing Science*, 33 (4), pp. 586-608.

SCHUDSON, MICHAEL (2002), “The News Media as Political Institutions”, *Annual Review of Political Science*, 5, 1, pp. 249-269.

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

SHADMEHR, MEHDI and DAN BERNHARDT (2015), “State Censorship”, *American Economic Journal: Microeconomics*, 7, 2, pp. 280-307.

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

SKRETA, VASILIKI and EDUARDO PEREZ-RICHET (2021), “Test Design under Falsification”, *Working Paper*.

STIGLER REPORT (2019), "Stigler Committee on Digital Platforms", *Final Report*, available at https://research.chicagobooth.edu/stigler/media/news/committee-on-digitalplatforms-final-report, September.

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.

TSAKAS, ELIAS and NIKOLAS TSAKAS (2021), "Noisy persuasion", *Games and Economic Behavior*, 130, pp. 44-61.

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.

VARIAN, HAL (2019), *16. Artificial Intelligence, Economics, and Industrial Organization*, University of Chicago Press.

WAISBORD, SILVIO (2000), *Watchdog Journalism in South America: News, Accountability, and Democracy*, Columbia University Press.

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.

YANG, KAI HAO (2020), "Selling Consumer Data for Profit: Optimal Market-Segmentation Design and its Consequences", *Working Paper*.