Strategic Mediation of Information in Autocracies*

Aleksandr Levkun†

May 21, 2020

Click here for the latest version

Abstract

This paper presents the optimal editorial policy for state-owned media manipulating information flow from a strategic informed elite to an uninformed receiver. The receiver attempts to match the state of the ruler's competence with a binary action. If the elite's and audience's preferences are too distant from each other, then the editorial policy is uninformative. Otherwise, the media signal whether the state is higher or lower than a threshold which depends on the elite's preferences. The media benefit from a more lenient elite, as long as the elite is not too lenient. 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. When the receiver has private information about how critical he is, I characterize the lower bound on the media's payoff obtained within the class of restricted editorial policies. I identify a sufficient condition that implies the bound is achieved.

*I am grateful to my advisors Joel Sobel and Simone Galperti for their continuous guidance and support. The comments and suggestions from Renee Bowen, Songzi Du, Elliot Lipnowski, Konstantin Sonin, Denis Shishkin, Maria Titova, Joel Watson, Weijie Zhong, and participants of the UCSD Workshop in Theory and Behavioral Economics are greatly appreciated. All errors are mine.

†Department of Economics, University of California, San Diego. E-mail: alevkun@ucsd.edu.

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. 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). As claimed by Guriev and Treisman (2019), these states seek to convince the population in ruler's competence to lead the country into a prosperous future. One of the main instruments for achieving this goal is an information manipulation through multiple channels including state-owned media. That is why such authoritarian states are sometimes referred to as informational autocrats.[1] The state-owned media consistently manipulates facts and censors information to influence the population beliefs about the ruler's competence. These beliefs then play a role when the population is faced with a decision whether to adhere to the ruler's political target.

In many situations, the state-owned media does not have access to the facts. Instead, it has to rely on reports generated by a strategically-interested third party.[2] For example, reports could be research conducted by an independent statistical agency in hopes of providing the most accurate information to the general public. On the other hand, reports could surface from ruler's cabinet members acting in their own best interest. In the presence of a strategic source, how likely does the state-owned media sway citizens' decisions toward the ruler's favor? What is the optimal editorial policy for the media and how does the policy depend on the preferences of citizens and the source?

To answer these questions, I consider a model of optimal information disclosure by a state-owned media to a representative uninformed receiver. The media wants the receiver to choose the mobilizing action over the status-quo action. The examples of the pairs of the status-quo action and the mobilizing action faced by the receiver include voting against/for the ruler in the election, revolting/not revolting, participating/not participating in antigovernment protests, causing/not causing panic in the

[1] This term is used in Guriev and Treisman (2019) that provides an extensive overview of the inner processes in modern autocracies.

[2] Allgaier (2011) provides evidence that compared to journalists that specialize in education issues, science correspondents opt for a narrower scope of sources for the coverage of science in schools.

society (Gehlbach and Sonin, 2014). However, the receiver prefers the mobilizing action only if the ruler's competence is high enough. The media does 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 independent research, proximity to the ruler, or ability to understand political processes better than the receiver (Guriev and Treisman, 2018). The elite's ordinal preferences are such that if the elite observes that competence lies in the set Θ_1 (Θ_0), then she prefers the mobilizing action (the status-quo action). The elite's preferences is a central object in my analysis, and I allow it to be fairly general. The elite cannot communicate to the receiver directly. Instead, having learnt the competence, the elite sends a message to the media. The media then generates a report to the receiver. Finally, the receiver chooses an action based on the media's report. I 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. Independence of the state-owned media comes from the ruler's need to delegate responsibility for reporting news to designated institution.

I derive the media's optimal editorial policy. This 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.[3] In the more interesting case, the media reveals some information only about whether the ruler's competence lies in Θ_0 or Θ_1. By doing so, the media ensures that the elite truthfully reports information about the ruler's competence. If the ruler's competence is in Θ_1 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. I 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 is worse off. It becomes harder for the media to convince the receiver to choose the mobilizing action. I 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

[3]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.

critical toward their government. Thus, as the 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 benefits from a more lenient elite, as long as the elite is not too lenient.

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 largely heterogeneous in pickiness towards the ruler, even in authoritarian states. For example, 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.[4] 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. This observation leads me to consider the state-owned media that attempts to sway decisions of the population of receivers having private information about their pickiness. I characterize the solution to the media's problem choosing over the restricted simple class of editorial policies. The optimal simple policy provides the lower bound for the media's unconstrained problem. I show the sufficient condition on the distribution of receiver's private information, under which this lower bound is attained. I illustrate the lower bound for the unimodal distribution of receiver's types.

2 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. Closer to my paper, Guriev and Treisman (2018) present the model of the informational autocracy that yearns for staying in power. A competent ruler brings higher living standards. As in my model, there is an elite that is informed about the ruler's competence. The elite may send messages to the population. These messages can be censored by the ruler, or the ruler may buy silence off the elite. The ruler is also able to send propaganda directly to the population. The authors establish conditions under which the manipulation of information is more beneficial for the ruler than opting for repressions or improving living standards. In my paper, the only way for the elite to

[4] URL: https://www.levada.ru/2019/11/18/vladimir-putin-7/ (available in Russian only).

communicate with the public is through the state-owned media. The media does not generate information itself, and transfers are not allowed. The manipulation of information is the only available instrument to the media. I characterize the effectiveness of this instrument depending on the preferences of the elite, the media, and the population.

To date, most of the papers in this strand of 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. The challenger's policy is unknown to the population, whereas the incumbent's policy is known. As in my model, 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, in my model, I consider the general slanting technology for the one-dimensional state and allow 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 et al. (2015) adopt the assumption of the media's commitment power to a probabilistic information structure, as my 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 the communication between an informed sender and an uninformed receiver through the third-party called 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 et al. (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. They also derive the optimal level of noise that achieves the best sender's and receiver's payoffs across all communication devices. These papers concentrate on the neutral mediator and characterize the best mediator for the sender and the receiver. In my paper, the media playing the role of the mediator has its own objective, specifically, to increase the probability of the mobilizing action chosen by the receiver. I 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, I assume the media's ability to commit to the mediation plan. Therefore, the media in my setup will generally obtain higher payoff. The closest to my paper is Salamanca (2016), which studies the informed party that is able to choose and commit to the mediation plan. The author characterizes sender's value function as the concavification of sender's indirect virtual utility function over prior beliefs. This paper can be seen as the case of complete alignment between sender's and mediator's preferences. In my paper, if the information source and the mediator are aligned, then the only equilibrium is completely uninformative.

Lastly, this paper 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 Tsakas and Tsakas (2019) study the setup where 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).

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. The persuader's value function is then characterized as a concavification of Lagrangian. However, in my 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. 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 faces novel constraints capturing the media's inability to access the state directly. Instead, the media has to incentivize the source to supply information by carefully designing the information protocol.

Boleslavsky and Kim (2018) and Ball (2020) have similar setup to my paper in the sense that the persuading side does not have direct access to the state. Boleslavsky and 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. This paper has the additional constraints in the form of moral hazard, whereas in my paper the honesty constraints correspond to the adverse selection problem. Moreover, the agent's problem can be summarized by a single equation that can be written as the expectation over the distribution of posterior beliefs, so that the methodology of Doval and Skreta (2018) is applicable. Ball (2020) 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. The scoring agency is aligned in preferences with the receiver. The optimal scoring rule puts smaller weights on some features to defer an excessive distortion by the sender.

3 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 decision-maker, to whom I refer to as a *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.[5] The payoff of the receiver u_R depends on his action $a \in A = \{a_0, a_1\}$

[5]Some examples of mobilization in the literature include voting for the ruler in the election, voting

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 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, I assume the following *tension* condition

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

This condition indicates that under the prior the receiver opts for the status-quo action.[6]

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. For my purposes, the source's payoff function $u_S : A \times \Theta \to \mathbb{R}$ is summarized by the two 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.[7] 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 wishes to promote the ruler's interests. In particular, the media wants 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 produces a costless report $r \in \mathcal{R}$ observed by the receiver. The set of reports $\mathcal{R}$ has at

in favor of the ruler's proposal to change the Constitution, not revolting, not going to antigovernment protests, etc. See Gehlbach and Sonin (2014) and Shadmehr and Bernhardt (2015).

[6]The media prefers the mobilizing action over the status-quo action. Therefore, if the tension condition is not satisfied, then the media ensures that no information is transmitted and the receiver chooses the mobilizing action according to his prior.

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

least two elements.

I assume that the media has the commitment power in 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 chooses 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. I refer to the editorial policy π as a strategic dissemination protocol, or simply a *protocol*.

Timing.—I 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 $\operatorname{argmax}_{m \in \mathcal{M}} \int_\mathcal{R} [u_S(a_1, \theta)\alpha(\cdot) + u_S(a_0, \theta)(1 - \alpha(\cdot))] \, d\pi(\cdot|m)$ for every $\theta \in \Theta$.

Following the information design literature[8], the agent with a commitment power is assumed to be able to steer other agents toward her favorite PBE. Thus, for every π,

[8] See Kamenica and Gentzkow (2011), Alonso and Câmara (2016), and Bergemann and Morris

M chooses a PBE that gives her the best ex ante payoff denoted as $V(\pi)$. Finally, the equilibrium π is chosen to maximize $V(\pi)$. I denote the value function of the media as V. This completes the definition of the equilibrium.

4 Analysis

I 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. I characterize the set of the direct protocols that satisfy honesty and obedience conditions. Finally, I 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 makes 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 is honest and R is obedient. Specifically, S is honest given R's obedience if

$$u_S(a_1, \theta)\pi(a_1|\theta) + u_S(a_0, \theta)\pi(a_0|\theta) \geqslant u_S(a_1, \theta)\pi(a_1|\theta') + u_S(a_0, \theta)\pi(a_0|\theta') \tag{2}$$

for every $\theta, \theta' \in \Theta$. I call a direct protocol π *honest* if it satisfies (2).

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

$$\int_0^1 \delta_R(\theta)\pi(a_1|\theta)d\mu_0 \geqslant 0, \tag{3}$$

and following recommendation a_0 is optimal, that is,

$$-\int_0^1 \delta_R(\theta)\pi(a_0|\theta)d\mu_0 \geqslant 0. \tag{4}$$

Note that the tension condition (1) and the inequality (3) imply the inequality (4), since by the tension condition

$$\int_0^1 \delta_R(\theta)d\mu_0 = \int_0^1 \delta_R(\theta)\pi(a_1|\theta)d\mu_0 + \int_0^1 \delta_R(\theta)\pi(a_0|\theta)d\mu_0 < 0.$$

(2016) among others. Mathevet et al. (2020) develop the methodology of analyzing persuasion problems for various equilibrium selection rules.

I call a direct protocol π *obedient* if it satisfies (3).

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

Lemma 1. Given any PBE in the original game followed by π, there exists an incentive-compatible direct protocol π^* in which the media gets the same expected utility when S is honest and R is obedient as in this PBE.

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

By Lemma 1 the optimal incentive-compatible direct protocol is also an equilibrium in the class of all possible protocols. From now on I 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 2 summarizes this observation.

Lemma 2. A direct protocol π is honest if and only if there exist π_0 and π_1, with $\pi_0 \leqslant \pi_1$, such that $\pi(a_1|\theta) = \pi_0$ for every $\theta \in \Theta_0$, and $\pi(a_1|\theta) = \pi_1$ for every $\theta \in \Theta_1$.

Proof. The honesty constraint (2) can be rewritten as follows: for every $\theta, \theta' \in \Theta$,

$$(u_S(a_1, \theta) - u_S(a_0, \theta))\,(\pi(a_1|\theta) - \pi(a_1|\theta')) \geqslant 0.$$

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, $\pi(a_1|\theta') \geqslant \pi(a_1|\theta'')$ and $\pi(a_1|\theta'') \geqslant \pi(a_1|\theta')$, 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 $\theta_1 \in \Theta_1$,

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

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

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

For the types of the source that have the same preferred action, the probability of this action being implemented has to be the same. Intuitively, the source will always attempt to induce highest probability of a_i, $i \in \{0, 1\}$, when the type is in Θ_i.

Thus, within Θ_i, the probability of a_i prescribed by protocol π has to be identical. 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.

The honesty constraint simplifies the search for the set of incentive-compatible protocols, since all the honest protocols are characterized by the pair of probabilities π_0 and π_1, with $\pi_0 \leqslant \pi_1$. To find the set of incentive-compatible protocols, I combine the insight of Lemma 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 $\mu_0(\Theta_i) \cdot I_i$, $i \in \{0, 1\}$, 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 1 characterizes the set of incentive-compatible 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 1. The set of incentive-compatible direct protocols $\mathcal{I}$ is characterized by the pair (π_0, π_1), such that $\pi_0 \leqslant \pi_1$, $\pi(a_1|\theta) = \pi_0$ for every $\theta \in \Theta_0$, and $\pi(a_1|\theta) = \pi_1$ for every $\theta \in \Theta_1$. Furthermore,

- if $I_0 < 0$ and $I_1 > 0$, then $\mathcal{I} = \left\{ (\pi_0, \pi_1) \in [0,1]^2 : \pi_1 \geqslant \frac{-I_0}{I_1} \cdot \pi_0 \right\}$;
- if $I_0 < 0$ and $I_1 = 0$, then $\mathcal{I} = \{ (\pi_0, \pi_1) : \pi_0 = 0, \pi_1 \in [0, 1] \}$;
- in all other cases, $\mathcal{I} = \{(0, 0)\}$.

Proof. By Lemma 2, the obedience constraint (3) can be written as follows:

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

or $\pi_0 I_0 + \pi_1 I_1 \geqslant 0$. There are three cases to consider, depending on the signs of I_0 and I_1. The case in which $I_0 \geqslant 0$ and $I_1 \geqslant 0$ is ruled out by the tension condition (1).

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

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

Case 3. Suppose $I_0 < 0$ and $I_1 \geqslant 0$. 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

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

where the second inequality is implied by the tension condition (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 $\pi_1 \geqslant \frac{-I_0}{I_1} \cdot \pi_0$.

□

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. By the standard intuition from sender-receiver games, if the conflict between sender's and receiver's preferences is too large, then no information can be transmitted.[9] Thus, in this case the only incentive-compatible protocol is complete pooling: $\pi_0 = \pi_1 = 0$. 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 1 shows that in this case the obedience constraint implies the requirement $\pi_0 \leqslant \pi_1$ of the honesty constraint.

Figure 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 my problem lies in Θ_0 or Θ_1.

[9] 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.

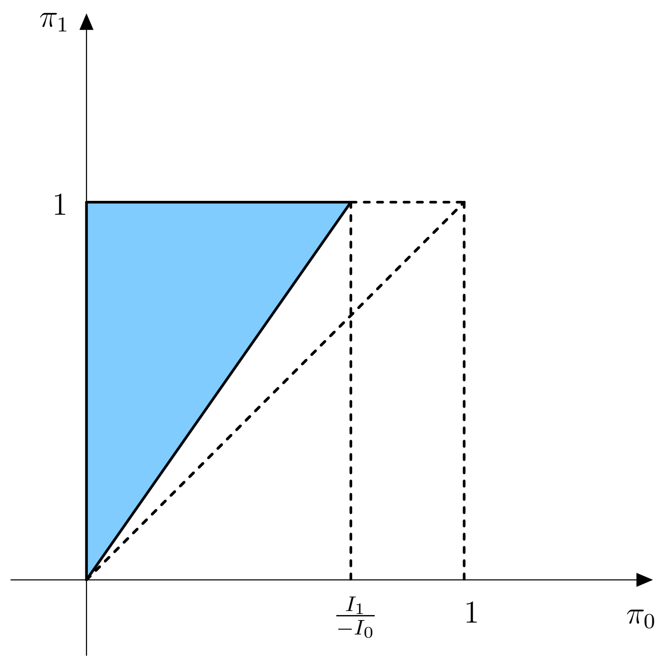


Figure 1. The set of incentive-compatible direct protocols, when $\mathbb{E}[\delta_R(\theta)|\theta \in \Theta_0] < 0$ and $\mathbb{E}[\delta_R(\theta)|\theta \in \Theta_1] > 0$.

Proposition 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:

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

Thus, the media maximizes the linear function over the set of incentive-compatible protocols $\mathcal{I}$ defined by the linear inequalities. Then the solution is necessarily an extreme point of $\mathcal{I}$. Proposition 2 finds this extreme point.

Proposition 2. If $I_0 < 0 \leqslant I_1$, 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 $I_0 < 0 \leqslant I_1$ is not satisfied then $\pi_0 = \pi_1 = 0$ and the media's payoff is 0.

Proof. 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 chooses 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 wants 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

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

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

□

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 makes the protocol informative about whether the ruler's competence is in Θ_0 or Θ_1. From the receiver's point of view, if the receiver gets the recommendation of the status-quo action, then he knows for certain 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.

4.1 Example

In order to gain insight about properties of the media-optimal protocol, I impose the functional forms for the players' payoffs. I assume that the receiver's payoff function is $u_R(\theta, a) = 1\{a = a_1\} \cdot (\theta - \omega)$, 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 (1), $\omega > \frac{1}{2}$. If $\omega \leqslant \frac{1}{2}$, then the media makes the protocol completely uninformative and extracts the payoff of 1.

Furthermore, I 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. I 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. Thus, $\bar{\theta}$ can be interpreted as the source's pickiness level that is potentially different from the receiver's ω. 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 $[0, 1]$.

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 2 readily establishes the optimal protocol and the media's ex ante payoff from this protocol.

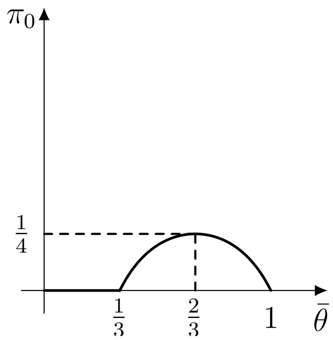
Claim 1. If $\bar{\theta} < 2\omega - 1$, then the optimal protocol is complete pooling $\pi_0 = \pi_1 = 0$, resulting in the media's payoff of 0. If $\bar{\theta} \geqslant 2\omega - 1$, then the optimal protocol is

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

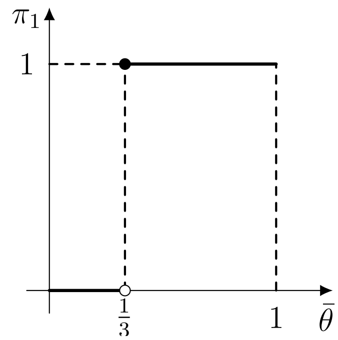
and the media's payoff is

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

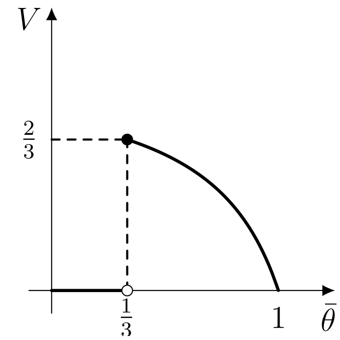
Figure 2 shows the media's optimal protocol and the value function for $\omega = \frac{2}{3}$ for various source's thresholds $\bar{\theta}$.



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

Given Claim 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 *pickier* (*more lenient*) than a receiver with pickiness ω'' if $\omega' > \omega''$ ($\omega' < \omega''$). Similarly, a source with threshold $\bar{\theta}'$ is said to be *pickier* (*more lenient*) than a source with threshold $\bar{\theta}''$ if $\bar{\theta}' > \bar{\theta}''$ ($\bar{\theta}' > \bar{\theta}''$).

Claim 2. The media gets the ex ante payoff of 0 if $\bar{\theta} < 2\omega - 1$. As long as $\bar{\theta} \geqslant 2\omega - 1$, the media's ex ante payoff decreases if the receiver becomes pickier or the source becomes pickier.

It is harder for the media to persuade a pickier receiver to undertake the mobilizing action. For a pickier source, there are two effects. First, the measure of types Θ_1 that guarantees the implementation of action a_1 goes down. 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 $\bar{\theta} \geqslant 2\omega - 1$. To sum up, for a sufficiently lenient source, the media's payoff stays at zero. When the source's pickiness reaches the level of $2\omega - 1$, the media's payoff jumps to $2 - 2\omega$ (calculated in Claim 1) and then starts to decrease to 0 with increasing $\bar{\theta}$. This effect is illustrated in Figure 2(c).

Observe that the receiver's ex ante payoff is always zero. Indeed, when the protocol corresponds to complete pooling, the receiver undertakes the status-quo action and gets zero payoff. If the protocol is informative, the media ensures that the receiver is kept indifferent between the status-quo and mobilizing actions, so that the ex ante payoff of the receiver is again zero.

Claim 3 establishes a comparative statics of the source's payoff as a function of the receiver's pickiness.

Claim 3. Suppose that $\bar{\theta} \geqslant 2\omega - 1$. 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 better off with a pickier receiver.

Hence, by Claim 3, ex ante (before learning the ruler's competence) the source is better off with a pickier receiver. Indeed, the source types from Θ_1 always get their favorite action. The source types from Θ_0 benefit from a pickier receiver as it becomes harder for the media to persuade the receiver to undertake the mobilizing action.

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

Claim 4. The solution to the media's problem facing no honesty constraints is

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

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

Proof. 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) tailored to the example:

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

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

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

But then the solution is $\pi(a_1|\theta) = 1$ for $2\omega - 1 \leqslant \theta < \omega$, 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 $2 - 2\omega$. □

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 implication of the discontinuity of the set of incentive-compatible protocols in $\bar{\theta}$. When $\bar{\theta} < 2\omega - 1$, only the uninformative protocol 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 has access to a larger set of incentive-compatible protocols. However, in this case the solution to my problem and Bayesian persuasion problem coincide.

5 Persuading the Public

This section allows the receiver to have private information. The media attempts 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 $a_i \in \{a_0, a_1\}$.[10] I impose the same assumptions on the payoff functions as in Section 4.1. That is, $\Theta_0 = [0, \bar{\theta})$ and $\Theta_1 = (\bar{\theta}, 1]$. The 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

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

the measure of a set $\Omega \subseteq [0,1]$ generated by H as $\eta(\Omega)$. The timing is unchanged, with an exception that 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.

Clearly, this setup is isomorphic to the problem with a single receiver having private information about ω. The common prior on ω is captured by the distribution H, and θ and ω are assumed to be independent. I assume that the media cannot elicit private information from the receiver.[11] Then the analogue of the revelation principle for this case can be shown. Instead of an unconditional action recommendation, the media now offers 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, I focus on a protocol $\phi : \Theta \to \Delta(\mathcal{P}([0,1])$, where $\mathcal{P}(\cdot)$ is the power set. Let suppϕ denote the set of contingent recommendations that appear in the protocol ϕ after some reported competence level θ, $\phi(\cdot|\theta) > 0$. For the sake of exposition, suppose that suppϕ is finite.[12]

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

$$\sum_{\Omega_1 \in \text{supp}\phi} \left[\eta(\Omega_1) u_S(a_1, \theta) + (1 - \eta(\Omega_1)) u_S(a_0, \theta))\right] (\phi(\Omega_1|\theta) - \phi(\Omega_1|\theta')) \geqslant 0 \tag{5}$$

for every $\theta, \theta' \in \Theta$, that is, it is optimal for the source to report the ruler's competence. I call ϕ honest if it satisfies (5).

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

$$\int_0^1 \delta_R(\cdot, \omega) d\phi(\Omega_1|\cdot) \geqslant 0, \tag{6}$$

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

$$\int_0^1 \delta_R(\cdot, \omega) d\phi(\Omega_1|\cdot) \leqslant 0. \tag{7}$$

[11] In the setup with no honesty concerns, Kolotilin et al. (2017) show that there is no value from elicitation. However, it is unclear whether the media gains from elicitation when it has to satisfy additional incentive constraints.

[12] In principle, a protocol ϕ can have an infinite support. Then the summation in the honesty constraint (5) has to be substituted by appropriate integration.

I call ϕ obedient if it satisfies (6) and (7).

Lemma 3. It is without loss for the media's objective to focus on the protocols $\phi : \Theta \to \Delta(\mathcal{P}([0,1])$ that are honest and obedient.

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

As in the setup with a known receiver's type, the incentive constraints reduce the dimension of the search for the optimal protocol. I start by simplifying the obedience constraint. Lemma 4 shows that the contingent recommendation is necessarily an interval $[0,b]$ (with a slight abuse of notation). The obedience constraint can be reduced to a single equation for each recommendation in the support.

Lemma 4. Every obedient protocol $\phi : \Theta \to \Delta(\mathcal{P}([0,1])$ exclusively sends contingent recommendations of the following form: a_1 is recommended for receivers with $\omega \in [0,b]$, $b \in [0,1]$; otherwise, a_0 is recommended. For every $b \in \text{supp}\phi$, the obedience constraint is summarized by

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

Proof. Suppose $\Omega_1 \in \text{supp}\phi$. Let $b = \sup(\Omega_1) \in [0,1]$. Note that the left-hand side of inequalities (6) and (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 (6). Similarly, any $\omega > b$ has to lie in $[0,1] \setminus \Omega_1$ by (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.

□

Lemma 4 illustrates that the contingent recommendations are very intuitive: sufficiently lenient receivers are recommended to take the mobilizing action, whereas sufficiently picky receivers are recommended to opt for the status-quo action. Lemma 5 provides the further simplification that mirrors Lemma 2 and characterizes honest protocols.

Lemma 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 $s_0 \leqslant s_1$, such that

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

for every $\theta \in \Theta_0$,

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

for every $\theta \in \Theta_1$.

Proof. Using Lemma 4, the honesty constraint (5) can be written as follows:

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

for every $\theta, \theta' \in \Theta$. Thus, by the same argument as in Lemma 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 $\theta \in \Theta_1$. Similarly, $s_0 = \sum_{[0,b]\in \text{supp}\phi} \phi([0,b]|\theta)H(b)$ 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 $s_1 \geqslant s_0$. □

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. However, now 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 has to ensure that $s_1 \geqslant s_0$.

Combining the results of Lemmas 3, 4, and 5, the media's problem can be written as the following Proposition 3 prescribes.

Proposition 3. 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

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

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

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

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

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

for every $[0,b] \in \text{supp}\phi$.

This problem is still complicated, since the optimal protocol can potentially provide multiple contingent recommendations given each reported competence level θ. To this end, I restrict the analysis to what I refer to as simple protocols. This restriction is inspired by the results of the previous section. A *simple* protocol is a protocol with a support having at most two elements, $[0, \bar{b}]$ and $[0, \underline{b}]$, with $\bar{b} \geqslant \underline{b}$. In what follows, I characterize the optimal simple protocol. This optimal simple protocol provides the lower bound on the media's payoff. I 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 Proposition 3 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 $\phi_0 \leqslant \phi_1$. Thus, 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 the constant has to be higher for the ruler's competence in Θ_1.

The obedience constraint then pins down $\bar{b}$ and $\underline{b}$ as a function of ϕ_0 and ϕ_1. Lemma 6 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 6. 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 $\phi_0 \leqslant \phi_1$.

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

$$\bar{b} = f_1(\phi_0, \phi_1) = \frac{\phi_0 \int_{\Theta_0} \theta d\mu_0 + \phi_1 \int_{\Theta_1} \theta d\mu_0}{\phi_0 \mu_0(\Theta_0) + \phi_1 \mu_0(\Theta_1)}, \tag{8}$$

$$\underline{b} = f_2(\phi_0, \phi_1) = \frac{(1 - \phi_0) \int_{\Theta_0} \theta d\mu_0 + (1 - \phi_1) \int_{\Theta_1} \theta d\mu_0}{(1 - \phi_0)\mu_0(\Theta_0) + (1 - \phi_1)\mu_0(\Theta_1)}. \tag{9}$$

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:

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

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

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

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

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 picky receivers 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 $\phi_1 \geqslant \phi_0$, such that $\bar{b} = f_1(\phi_0, \phi_1)$ and $\underline{b} = f_2(\phi_0, \phi_1)$. This is a consequence of 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

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

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 $\phi_1 \geqslant \phi_0$. If $\bar{b} = \underline{b}$, then $\bar{b} = \underline{b} = \mathbb{E}[\theta]$ and this happens as long as $\phi_0 = \phi_1$. $\square$

From Lemma 6, 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 4 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 4 shows that the optimal simple protocol delivers the media the payoff equal to the concavification of $\hat{H}$ evaluated at $\mathbb{E}[\theta]$.

Proposition 4. The media's payoff from the optimal simple protocol is equal to cav$\hat{H}[\mathbb{E}[\theta]]$.

Proof. Equations (8) and (9) 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 Proposition 3, the objective of the media is

$$\mu_0(\Theta_0)s_0 + \mu_0(\Theta_1)s_1 = H(\underline{b}) + (\phi_0\mu_0(\Theta_0) + \phi_1\mu_0(\Theta_1))(H(\bar{b}) - H(\underline{b}))$$
$$= H(\underline{b}) + \frac{\mathbb{E}[\theta] - \underline{b}}{\bar{b} - \underline{b}}(H(\bar{b}) - H(\underline{b})) = \frac{\bar{b} - \mathbb{E}[\theta]}{\bar{b} - \underline{b}}H(\underline{b}) + \frac{\mathbb{E}[\theta] - \underline{b}}{\bar{b} - \underline{b}}H(\bar{b}),$$

as long as $\bar{b} > \underline{b}$. If $\bar{b} = \underline{b}$, then the objective of the mediator 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 6, 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 cav$\hat{H}[E[\theta]]$.[13] The corresponding $\bar{b}$ and $\underline{b}$ can then be established as the supporting points of this object.

□

Proposition 4 obtains the lower bound on the media's payoff. The upper bound on the media's payoff is given by cav$H[E[\theta]]$. 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$H[E[\theta]]$.

[13]See, for example, Le Treust and Tomala (2019).

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 1 summarizes this observation.

Corollary 1. The simple protocol is optimal for the media if $\text{cav}\hat{H}[E[\theta]] = \text{cav}H[E[\theta]]$.

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 1 is that for concave H, the simple protocol is optimal. Moreover, $\bar{b} = \underline{b}$, that is, the solution corresponds to complete pooling.

Finally, to illustrate the bound provided in Proposition 4, I consider a unimodal distribution of receiver's types that recently gained a lot of attention in the literature, namely, Kolotilin et al. (2017), Lipnowski et al. (2019), and Shishkin (2019).

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 some threshold.[14] The upper-censorship policy in my setup corresponds to the protocol ϕ, with $\phi([0,\theta]|\theta) = 1$ for $\theta < t$ and $\phi([0,b]|\theta) = 1$ for $\theta \geqslant t$, where $t > 0$ is a threshold and $[0, b]$ is a pooling recommendation. However, the upper-censorship policy can be readily seen to be not incentive-compatible, when the honesty constraint are present. Indeed, pick two types $\theta', \theta'' \in \Theta_0 \cap [0, t)$, $\theta' < \theta''$. Reporting θ' gives the source probability $1 - H(\theta')$ of the status-quo action chosen, whereas reporting θ'' produces probability $1 - H(\theta'')$. Thus, the source of type θ'' prefers to misreport, and the honesty constraint is not satisfied. This observation leads me to expect some pooling for the low states in the optimal protocol.

[14] The mode of the distribution has to be sufficiently large. Otherwise, the uninformative policy is optimal as most of the receivers are lenient.

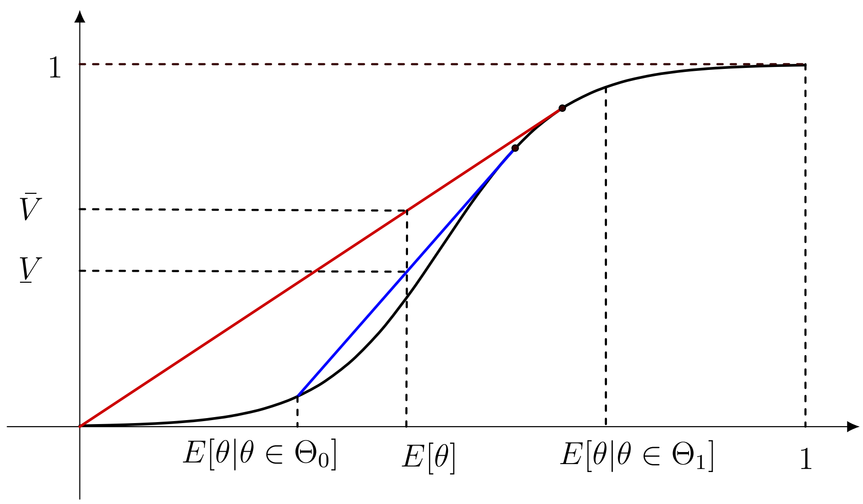


Figure 3. Bounds on the media's payoff. The black line corresponds to H. The red line corresponds to cav$H$. The blue line corresponds to cav$\hat{H}$.

Figure 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 $[\underline{V}, \bar{V}]$. The payoff $\underline{V}$ can be achieved with a simple protocol. This protocol has a support having two elements, $[0, \bar{b}]$, $[0, \underline{b}]$, where $\underline{b} = \mathbb{E}[\theta|\theta \in \Theta_0]$. The probabilities assigned to the contingent recommendation $[0, \bar{b}]$ are as follows: $\phi([0, \bar{b}]|\theta) = \phi_0 \in (0, 1)$ for every $\theta \in \Theta_0$, $\phi([0, \bar{b}]|\theta) = 1$ for every $\theta \in \Theta_1$. Thus, the contingent recommendation $[0, \mathbb{E}[\theta|\theta \in \Theta_0]]$ reveals that the ruler's competence is in Θ_0. In this sense, this simple protocol is similar to the optimal protocol in the case of known receiver's type.

6 Concluding Remarks

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 does 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. I show how these additional incentive constraints shape the optimal editorial policy and outline the welfare implications of this policy. I show when the honesty constraints are binding and when there is a meaningful communication depending on the preference parameters. I close with the discussion of assumptions that are substantial for my results.

Discussion of assumptions.— I assume that there are no transfers between players. In this sense, I 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. I leave this possibility out of the model.

I assume that the media has commitment power. As explained in Gehlbach and Sonin (2014) and Gentzkow et al. (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 et al. (2019), where with some probability the media may inaudibly change the editorial policy after observing the source's message. The analysis provided here can be seen as the best the media can potentially 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).

References

Allgaier, J. (2011). Who is having a voice? Journalists' selection of sources in a creationism controversy in the UK press. *Cultural Studies of Science Education 6*(2), 445–467.

Alonso, R. and O. Câmara (2016). Persuading voters. *American Economic Review 106*(11), 3590–3605.

Ball, I. (2020). Scoring strategic agents. Technical report, Yale University.

Bergemann, D. and S. Morris (2016). Information design, Bayesian persuasion, and Bayes correlated equilibrium. *American Economic Review 106*(5), 586–91.

Bergemann, D. and S. Morris (2019). Information design: A unified perspective. *Journal of Economic Literature 57*(1), 44–95.

Blume, A., O. J. Board, and K. Kawamura (2007). Noisy talk. *Theoretical Economics 2*(4), 395–440.

Boleslavsky, R. and K. Kim (2018). Bayesian persuasion and moral hazard. *Available at SSRN 2913669*.

Chiang, C.-F. and B. Knight (2011). Media bias and influence: Evidence from newspaper endorsements. *The Review of Economic Studies 78*(3), 795–820.

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

Doval, L. and V. Skreta (2018). Mechanism design with limited commitment. *arXiv preprint arXiv:1811.03579*.

Duggan, J. and C. Martinelli (2011). A spatial theory of media slant and voter choice. *The Review of Economic Studies 78*(2), 640–666.

Gandhi, J. and E. Lust-Okar (2009). Elections under authoritarianism. *Annual Review of Political Science 12*, 403–422.

Gehlbach, S. and K. Sonin (2014). Government control of the media. *Journal of Public Economics 118*, 163–171.

Gentzkow, M. and E. Kamenica (2016). A Rothschild-Stiglitz approach to Bayesian persuasion. *American Economic Review 106*(5), 597–601.

Gentzkow, M., J. M. Shapiro, and D. F. Stone (2015). Media bias in the marketplace: Theory. In *Handbook of media economics*, Volume 1, pp. 623–645. Elsevier.

Goltsman, M., J. Hörner, G. Pavlov, and F. Squintani (2009). Mediation, arbitration and negotiation. *Journal of Economic Theory 144*, 1397–1420.

Guriev, S. and D. Treisman (2018). Informational autocracy: Theory and empirics of modern authoritarianism. *Available at SSRN 2571905*.

Guriev, S. and D. Treisman (2019). Informational autocrats. *Journal of Economic Perspectives 33*(4), 100–127.

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

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

Kartik, N. (2009). Strategic communication with lying costs. *The Review of Economic Studies 76*(4), 1359–1395.

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

Le Treust, M. and T. Tomala (2019). Persuasion with limited communication capacity. *Journal of Economic Theory 184*, 104940.

Lipnowski, E. and L. Mathevet (2018). Disclosure to a psychological audience. *American Economic Journal: Microeconomics 10*(4), 67–93.

Lipnowski, E., D. Ravid, and D. Shishkin (2019). Persuasion via weak institutions. *Available at SSRN 3168103*.

Mathevet, L., J. Perego, and I. Taneva (2020). On information design in games. *Journal of Political Economy 128*(4), 000–000.

Myerson, R. B. (1982). Optimal coordination mechanisms in generalized principal–agent problems. *Journal of Mathematical Economics 10*(1), 67–81.

Salamanca, A. (2016). The value of mediated communication. Technical report, Toulouse School of Economics.

Shadmehr, M. and D. Bernhardt (2015). State censorship. *American Economic Journal: Microeconomics 7*(2), 280–307.

Shishkin, D. (2019). Evidence acquisition and voluntary disclosure. Technical report, Princeton University.

Tsakas, E. and N. Tsakas (2019). Noisy persuasion. *Available at SSRN 2940681*.