OPTIMAL PERSUASION WITH AN APPLICATION TO MEDIA CENSORSHIP

Anton Kolotilin, Tymofiy Mylovanov, Andriy Zapechelnyuk

Abstract. A sender designs a signal about a state to persuade a receiver. Under standard assumptions, an optimal signal reveals the states below a cutoff and pools the states above the cutoff. This result holds in continuous and discrete environments with general and monotone partitional signals. We provide monotone comparative statics results on the informativeness of the optimal signal. We apply our results to the problem of media censorship by a government.

JEL Classification: D82, D83, L82

Keywords: Bayesian persuasion, information design, upper censorship, lower censorship, media censorship

---

Date: 24th April 2019.

Kolotilin: School of Economics, UNSW Australia, Sydney, NSW 2052, Australia. *E-mail:* akolotilin@gmail.com.
Mylovanov: University of Pittsburgh, Department of Economics, 4714 Posvar Hall, 230 South Bouquet Street, Pittsburgh, PA 15260, USA. *E-mail:* mylovanov@gmail.com.
Zapechelnyuk: School of Economics and Finance, University of St Andrews, Castlecliffe, the Scores, St Andrews KY16 9AR, UK. *E-mail:* az48@st-andrews.ac.uk.

We are grateful for discussions with Ming Li with whom we worked on the companion paper Kolotilin et al. (2017). An early version of the results in this paper and the results in the companion paper were presented in our joint working paper Kolotilin et al. (2015). We thank Ricardo Alonso, Dirk Bergemann, Patrick Bolton, Alessandro Bonatti, Steven Callander, Odilon Câmara, Rahul Deb, Péter Esö, Johannes Hörner, Florian Hoffman, Roman Inderst, Emir Kamenica, Navin Kartik, Daniel Kräehmer, Marco Ottaviani, Mallesh Pai, Andrea Prat, Ilya Segal, Larry Samuelson, Joel Sobel, Konstantin Sonin, as well as many conference and seminar participants, for helpful comments. Kolotilin gratefully acknowledges financial support from the Australian Research Council Discovery Early Career Research Award DE160100964. Zapechelnyuk gratefully acknowledges support from the Economic and Social Research Council Grant ES/N01829X/1.

1. Introduction

Consider a sender who seeks to influence beliefs and actions of a receiver by disclosing information. Specifically, suppose that the sender can design a procedure for obtaining and revealing information about a state of the world to the receiver. In many situations, the sender reveals the states in some interval and pools (censors) the states in the complementary interval. For simplicity, we focus on *upper-censorship* policies that reveal the states below a cutoff and pool the states above this cutoff. We derive simple conditions under which such upper-censorship policies are robustly optimal. We also provide monotone comparative statics results on the amount of information that is optimally revealed.

In our model, the receiver chooses whether to act or not. If the receiver does not act, the sender and receiver's utilities are normalized to zero. If the receiver acts, the sender and receiver's utilities depend on a state of the world and a type of the receiver. The state and type are independent random variables that represent, respectively, the receiver's benefit and cost from action. We assume that the type is a continuous random variable, but we allow the state to be either a continuous or discrete random variable.[1]

The sender wishes to persuade the receiver to act, but she may also care about the receiver. That is, the sender's expected utility is a weighted sum of the probability that the receiver acts and the receiver's expected utility.[2] The sender and receiver share a common prior about the state; the receiver privately knows his type. To influence the receiver's choice, the sender designs a signal that reveals information about the state. After observing the signal realization, the receiver updates his beliefs about the state and makes a choice that maximizes his expected utility.

The main result shows that if the probability density of the receiver's type is log-concave, then, and only then, upper-censorship is optimal for all prior distributions of the state and for all weights that the sender puts on the receiver's utility. Many commonly used probability densities are log-concave (see Table 1 in Bagnoli and Bergstrom 2005). Log-concave densities are well-behaved and exhibit nice properties, such as single-peakedness and the monotonicity of the hazard rates.

The main result is obtained under two scenarios. In the first scenario, the signal is a random variable with arbitrary correlation with the state. In the second scenario, the signal is constrained to be a monotone partition of the state space, so that each state is either revealed or pooled with some adjacent states.

To compare the main result under these two scenarios, consider the cutoff of an optimal upper-censorship signal. If the state is a continuous random variable, then it does not matter whether this cutoff state is pooled or revealed. Thus, an optimal

[1]Extending the analysis to the case where the state and type are general random variable presents technical difficulties but does not yield new economic insights.

[2]In the paper, we consider a more general utility function of the sender.

upper-censorship signal is, in fact, a monotone partition. However, if the state is a discrete random variable, then the optimal signal is stochastic upper censorship so that the cutoff state is pooled and revealed with some probabilities. In contrast, the optimal monotone partition is deterministic upper censorship so that the cutoff state is either completely pooled or fully revealed.

In many applications of interest, the state is discrete, and thus optimal signals and optimal monotone partitions generally differ. We find an optimal signal using tools in Kolotilin, Mylovanov, Zapechelnyuk, and Li (2017). This optimal signal provides an upper bound on the value of persuasion. To achieve this bound, the sender should be able to commit to randomize over two signal realizations conditional on the cutoff state, which may be hard to enforce in practice. In contrast, monotone partitions are relatively simple signals that can be ex post verified and enforced.

The problem of finding an optimal monotone partition, however, is a discrete optimization problem, which cannot be solved using existing tools from the persuasion literature. Nevertheless, if the density of the receiver's type is log-concave, we are able to show that upper censorship is an optimal monotone partition. Specifically, for any monotone partition that is not upper censorship, we construct another monotone partition that is preferred by the sender.

When upper-censorship is optimal, it is possible to perform the comparative statics analysis on the amount of information that is optimally revealed. We show that the sender reveals less information as the sender becomes more biased and each type of the receiver experiences a preference shock in favor of the sender.

We apply our results to the problem of media censorship by the government. We consider a stylized setting with a finite number of media outlets and a continuum of heterogeneous citizens (receivers). We extend the model to permit the aggregate action of the citizens to affect their utility but not their optimal actions. For example, an election outcome impacts all citizens but does not change their preferences over candidates. The government wishes to influence citizens' actions by deciding which media outlets are to be censored. We show that if the probability density of the citizens' types is log-concave, then the optimal media censorship policy prescribes to permit all sufficiently loyal media outlets and to censor the remaining outlets.

Related Literature. This paper presents a theory of optimal Bayesian persuasion in environments where payoffs are linear in the state. The literature on Bayesian persuasion was set in motion by the seminal papers of Rayo and Segal (2010) and Kamenica and Gentzkow (2011). Linear persuasion is a workhorse environment for Bayesian persuasion literature. The leading example in Kamenica and Gentzkow (2011) is linear with two states. Linear persuasion has been studied by Kamenica and Gentzkow (2011), Gentzkow and Kamenica (2016), Kolotilin, Mylovanov, Zapechelnyuk, and Li (2017), Kolotilin (2017), and Dworczak and Martini (2018). Linear persuasion has been applied, for example, to selection of projects in organizations in Boleslavsky and Cotton (2018), school grading policies in Ostrovsky and Schwarz (2010), trading mechanisms with resale opportunities in Dworczak (2017), macroprudential stress tests in Orlov, Zryumov, and Skrzypach (2018), transparency benchmarks in over-the-counter markets in Duffie, Dworczak, and Zhu (2017), clinical trials in Kolotilin (2015), media control in Gehlbach and Sonin (2014), and voter persuasion in Alonso and Câmara (2016a,b), stress tests for financial institutions in Goldstein and Leitner (2018).

Our paper provides necessary and sufficient *prior independent* conditions for randomized censorship to be optimal. Importantly, we show that whenever these conditions are satisfied deterministic censorship is optimal in the class of monotone partitional signals (simple signals). In addition, we establish monotone comparative statics results on the informativeness of the optimal signal. Kolotilin (2017) provides necessary and sufficient *prior dependent* conditions for optimality of randomized censorship. In various contexts, sufficient conditions for optimality of randomized censorship have been provided by Alonso and Câmara (2016b), Kolotilin, Mylovanov, Zapechelnyuk, and Li (2017), and Dworczak and Martini (2018). Alonso and Câmara (2016b) establish related comparative statics results for the case of a finite number of states. There are no counterparts in the literature to our results about optimality of deterministic censorship in the class of monotone partitional signals.

2. Model

2.1. **Setup.** There are two players: a sender (she) and a receiver (he). The receiver chooses one of two actions $a \in \{0,1\}$. The players' utilities depend on the action and two independent random variables: a state of nature $\omega \in [0,1]$ and the receiver's type $r \in [0,1]$. The distributions F of ω and G of r admit continuously differentiable densities f and g. Action $a=0$ is interpreted as a status quo and gives zero utility to both players. Action $a=1$ gives utility $\omega - r$ to the receiver and $v(\omega,r)$ to the sender, where $v(\omega,r)$ is linear in ω and continuously differentiable in r.

The receiver privately knows his type, but he does not observe the state. The sender can influence the action taken by the receiver by releasing a signal that reveals information about the state. A *signal* is a random variable $s \in [0,1]$ that is independent of r but, possibly, correlated with ω. For example, s is fully revealing if it is perfectly correlated with ω, and s is completely uninformative if it is independent of ω.

The timing is as follows. First, the sender publicly chooses a signal s. Then, realizations of ω, r, and s are drawn. Finally, the receiver observes the realizations of his type r and the signal s, and then chooses between $a=0$ and $a=1$.

We consider two scenarios:

(i) the sender can choose any signal,

(ii) the sender can choose any monotone partitional signal.

Formally, a signal s is a *monotone partitional signal* (for short, monotone partition) if there exists a nondecreasing function $\sigma(\omega)$ such that $s = \sigma(\omega)$ for each ω. Every such signal induces a partition of the state space $[0, 1]$ into intervals and singletons, and the receiver observes the partition element that contains the state.

Under Scenario (i), we are interested in an *optimal signal* that maximizes the sender's expected utility among all signals. This is a standard persuasion problem.

Under Scenario (ii), we are interested in an *optimal monotone partition* that maximizes the sender's expected utility among all monotone partitional signals. This scenario incorporates constraints that information designers often face in practice. For example, a non-monotone grading policy that gives better grades to worse performing students will be perceived as unfair and will be manipulated by strategic students.

2.2. **Optimality of Upper-censorship.** In this section we provide necessary and sufficient conditions for optimality of a simple class of signals called upper-censorship.

Let m be the expected state conditional on observing a realization of a signal s. Because that the sender and receiver's utilities are linear in ω, they depend on the information about ω revealed by signal s only through the expected state m. In particular, the receiver chooses $a = 1$ if and only if $r \leq m$.

Let $V(m)$ denote the *indirect utility* of the sender conditional on m,

$$V(m) = \int_{r\leq m} \mathbb{E}\left[v(\omega, r) | \, m \right] g(r) \mathrm{d}r = \int_0^m v(m, r) g(r) \mathrm{d}r, \quad m \in [0, 1]. \tag{1}$$

where $\mathbb{E}\left[v(\omega, r) | \, m \right] = v(m, r)$ by the linearity of v in ω.

A function V is said to be *S-shaped* if it is convex below some threshold and concave above that threshold, or, equivalently, if V'' is single-crossing from above:

$$\text{there exists } m' \text{ such that } \; V''(m) \geq (\leq)\, 0 \; \text{ for all } \; m < (>)\, m'.$$

A signal s is *upper-censorship* if there exists a cutoff $\omega^* \in [0, 1]$ such that the states below ω^* are revealed and the states above ω^* are pooled. Note that it does not matter whether the cutoff state ω^* is revealed or pooled, because the state is a continuous random variable. For concreteness, suppose that ω^* is pooled, so that

$$s = \begin{cases} \omega, & \text{if } \omega < \omega^*, \\ m^*, & \text{if } \omega \geq \omega^*, \end{cases}$$

where

$$m^* = \mathbb{E}[\omega | \omega \geq \omega^*]$$

is the expected state conditional on being above the cutoff.

Remark 1. A *lower-censorship* signal is defined symmetrically: there exists a cutoff $\omega^* \in [0,1]$ such that the states below ω^* are pooled and the states above ω^* are revealed. For clarity of exposition, we focus on upper-censorship. Note that the same results will hold if we replace *upper-censorship* by *lower-censorship* and V by $-V$.

We now provide the criterion for the optimality of upper-censorship.

Theorem 1. *Let V be S-shaped. Then, and only then, an optimal signal is upper-censorship for all F.*

As follows from this theorem, optimal persuasion takes a very simple form of upper-censorship whenever V is S-shaped. In this case, the sender's optimization problem is reduced to finding an optimal censorship cutoff ω^*. Moreover, this result is tight, in the sense that if V is not S-shaped, then there exists a distribution F of the state such that no upper-censorship signal is optimal.

Note that every upper-censorship signal is a monotone partition. Therefore, Theorem 2 applies in both Scenario A, where the sender is free to choose any signal, and Scenario B, where the sender is constrained to choose monotone partitions.

The intuition behind Theorem 2 is as follows. Observe that when no information about the state is revealed, the receiver's best-response action does not change with the state. The more information is revealed, the more variable the receiver's behavior will be in response to this information. Consider an interval of states where $V(\omega)$ is concave. The sender would prefer not to reveal any information over this interval, since a certain outcome is preferred to any lottery with the same expected state. Conversely, consider an interval of states where $V(\omega)$ is convex. The sender would prefer to fully reveal the state in this interval, since now lotteries are preferred. If V is S-shaped, that is, it is convex below some threshold and concave above that threshold, the induced optimal persuasion takes the form of upper-censorship.

Let s be upper-censorship with a cutoff ω^*. If the realized state ω is below the cutoff, then it is revealed to the receiver, so the expected utility of the sender is $V(\omega)$. If the realized state ω is above the cutoff, then the posterior expected state is $m^* = \mathbb{E}[\omega|\omega \geq \omega^*]$, so the expected utility of the sender conditional on $\omega > \omega^*$ is $V(m^*)$. The sender thus needs to solve the problem

$$\max_{\omega^* \in [0,1]} \int_0^{\omega^*} V(\omega) f(\omega) \mathrm{d}\omega + \int_{\omega^*}^1 V(m^*) f(\omega) \mathrm{d}\omega. \tag{2}$$

Proposition 1. *Let V be S-shaped. There exists an optimal upper-censorship signal with a cutoff ω^* that satisfies*

$$V(\omega) - V(m^*) + V'(m^*)(m^* - \omega) \geq (\leq)\, 0 \quad \textit{for all } \omega \leq (\geq)\, \omega^*. \tag{3}$$

The expression in (7) represents the first-order condition and captures three possible cases: the boundary solutions $\omega^* = 0$ and $\omega^* = 1$ if the expression in (7) has the same

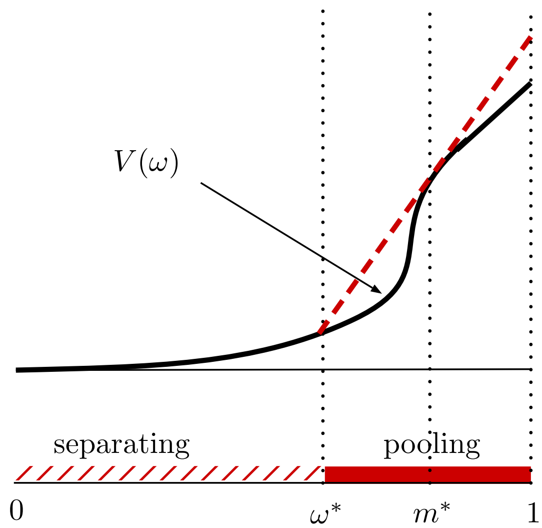

FIGURE 1. Optimal upper-censorship with cutoff ω^*.

sign for all ω, and an interior solution ω^* such that

$$V(\omega^*) - V(m^*) + V'(m^*)(m^* - \omega^*) = 0. \tag{4}$$

This first-order condition is illustrated by Fig. 2. The solid line is $V(\omega)$, and the dashed line is $V(m^*) - V'(m^*)(m^* - \omega)$, which is the tangent line to V at m^*.

The solution of the sender's problem (6) is particularly simple if V is either globally convex or globally concave. In this case, the first-order condition in (7) has a constant sign, so either $\omega^* = 0$ or $\omega^* = 1$ must be optimal. If V is convex, then $\omega^* = 1$ is optimal, which corresponds the fully informative signal. Similarly, if V is concave, then $\omega^* = 0$ is optimal, which corresponds the completely uninformative signal. This is summarized in the following corollary.

Corollary 1. *An optimal signal is*

(a) *fully informative for all F if and only if V is convex,*

(b) *completely uninformative for all F if and only if V is concave.*

3. MODEL

3.1. **Setup.** There are two players: a sender (she) and a receiver (he). The receiver chooses one of two actions $a \in \{0, 1\}$. The players' utilities depend on the action and two independent random variables: a state of nature $\omega \in [0, 1]$ and the receiver's type $r \in [0, 1]$. The distributions F of ω and G of r admit continuously differentiable densities f and g. Action $a = 0$ is interpreted as a status quo and gives zero utility to both players. Action $a = 1$ gives utility $\omega - r$ to the receiver and $v(\omega, r)$ to the sender, where $v(\omega, r)$ is linear in ω and continuously differentiable in r.

The receiver privately knows his type, but he does not observe the state. The sender can influence the action taken by the receiver by releasing a signal that reveals information about the state. A *signal* is a random variable $s \in [0, 1]$ that is independent of r but, possibly, correlated with ω. For example, s is fully revealing if it is perfectly correlated with ω, and s is completely uninformative if it is independent of ω.

The timing is as follows. First, the sender publicly chooses a signal s. Then, realizations of ω, r, and s are drawn. Finally, the receiver observes the realizations of his type r and the signal s, and then chooses between $a = 0$ and $a = 1$.

We consider two scenarios:

(i) the sender can choose any signal,

(ii) the sender can choose any monotone partitional signal.

Formally, a signal s is a *monotone partitional signal* (for short, monotone partition) if there exists a nondecreasing function $\sigma(\omega)$ such that $s = \sigma(\omega)$ for each ω. Every such signal induces a partition of the state space $[0, 1]$ into intervals and singletons, and the receiver observes the partition element that contains the state.

Under Scenario (i), we are interested in an *optimal signal* that maximizes the sender's expected utility among all signals. This is a standard persuasion problem.

Under Scenario (ii), we are interested in an *optimal monotone partition* that maximizes the sender's expected utility among all monotone partitional signals. This scenario incorporates constraints that information designers often face in practice. For example, a non-monotone grading policy that gives better grades to worse performing students will be perceived as unfair and will be manipulated by strategic students.

3.2. **Optimality of Upper-censorship.** In this section we provide necessary and sufficient conditions for optimality of a simple class of signals called upper-censorship.

Let m be the expected state conditional on observing a realization of a signal s. Because that the sender and receiver's utilities are linear in ω, they depend on the information about ω revealed by signal s only through the expected state m. In particular, the receiver chooses $a = 1$ if and only if $r \leq m$.

Let $V(m)$ denote the *indirect utility* of the sender conditional on m,

$$V(m) = \int_{r \leq m} \mathbb{E}\left[v(\omega, r) | \, m \right] g(r) \mathrm{d}r = \int_0^m v(m, r) g(r) \mathrm{d}r, \quad m \in [0, 1]. \tag{5}$$

where $\mathbb{E}\left[v(\omega, r) | \, m \right] = v(m, r)$ by the linearity of v in ω.

A function V is said to be *S-shaped* if it is convex below some threshold and concave above that threshold, or, equivalently, if V'' is single-crossing from above:

$$\text{there exists } m' \text{ such that } \; V''(m) \geq (\leq)\, 0 \;\text{ for all }\; m < (>)\, m'.$$

A signal s is *upper-censorship* if there exists a cutoff $\omega^* \in [0,1]$ such that the states below ω^* are revealed and the states above ω^* are pooled. Note that it does not matter whether the cutoff state ω^* is revealed or pooled, because the state is a continuous random variable. For concreteness, suppose that ω^* is pooled, so that

$$s = \begin{cases} \omega, & \text{if } \omega < \omega^*, \\ m^*, & \text{if } \omega \geq \omega^*, \end{cases}$$

where

$$m^* = \mathbb{E}[\omega | \omega \geq \omega^*]$$

is the expected state conditional on being above the cutoff.

Remark 2. A *lower-censorship* signal is defined symmetrically: there exists a cutoff $\omega^* \in [0,1]$ such that the states below ω^* are pooled and the states above ω^* are revealed. For clarity of exposition, we focus on upper-censorship. Note that the same results will hold if we replace *upper-censorship* by *lower-censorship* and V by $-V$.

We now provide the criterion for the optimality of upper-censorship.

Theorem 2. *Let V be S-shaped. Then, and only then, an optimal signal is upper-censorship for all F.*

As follows from this theorem, optimal persuasion takes a very simple form of upper-censorship whenever V is S-shaped. In this case, the sender's optimization problem is reduced to finding an optimal censorship cutoff ω^*. Moreover, this result is tight, in the sense that if V is not S-shaped, then there exists a distribution F of the state such that no upper-censorship signal is optimal.

Note that every upper-censorship signal is a monotone partition. Therefore, Theorem 2 applies in both Scenario A, where the sender is free to choose any signal, and Scenario B, where the sender is constrained to choose monotone partitions.

The intuition behind Theorem 2 is as follows. Observe that when no information about the state is revealed, the receiver's best-response action does not change with the state. The more information is revealed, the more variable the receiver's behavior will be in response to this information. Consider an interval of states where $V(\omega)$ is concave. The sender would prefer not to reveal any information over this interval, since a certain outcome is preferred to any lottery with the same expected state. Conversely, consider an interval of states where $V(\omega)$ is convex. The sender would prefer to fully reveal the state in this interval, since now lotteries are preferred. If V is S-shaped, that is, it is convex below some threshold and concave above that threshold, the induced optimal persuasion takes the form of upper-censorship.

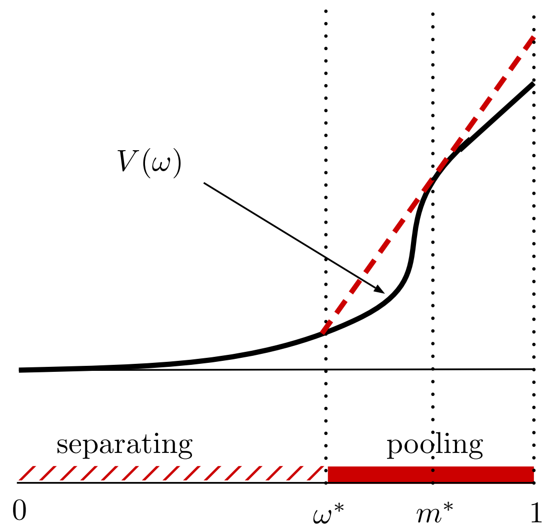

FIGURE 2. Optimal upper-censorship with cutoff ω^*.

Let s be upper-censorship with a cutoff ω^*. If the realized state ω is below the cutoff, then it is revealed to the receiver, so the expected utility of the sender is $V(\omega)$. If the realized state ω is above the cutoff, then the posterior expected state is $m^* = \mathbb{E}[\omega|\omega \geq \omega^*]$, so the expected utility of the sender conditional on $\omega > \omega^*$ is $V(m^*)$. The sender thus needs to solve the problem

$$\max_{\omega^* \in [0,1]} \int_0^{\omega^*} V(\omega) f(\omega) \mathrm{d}\omega + \int_{\omega^*}^1 V(m^*) f(\omega) \mathrm{d}\omega. \tag{6}$$

Proposition 2. *Let V be S-shaped. There exists an optimal upper-censorship signal with a cutoff ω^* that satisfies*

$$V(\omega) - V(m^*) + V'(m^*)(m^* - \omega) \geq (\leq)\, 0 \quad \textit{for all } \omega \leq (\geq)\, \omega^*. \tag{7}$$

The expression in (7) represents the first-order condition and captures three possible cases: the boundary solutions $\omega^* = 0$ and $\omega^* = 1$ if the expression in (7) has the same sign for all ω, and an interior solution ω^* such that

$$V(\omega^*) - V(m^*) + V'(m^*)(m^* - \omega^*) = 0. \tag{8}$$

This first-order condition is illustrated by Fig. 2. The solid line is $V(\omega)$, and the dashed line is $V(m^*) - V'(m^*)(m^* - \omega)$, which is the tangent line to V at m^*.

The solution of the sender's problem (6) is particularly simple if V is either globally convex or globally concave. In this case, the first-order condition in (7) has a constant sign, so either $\omega^* = 0$ or $\omega^* = 1$ must be optimal. If V is convex, then $\omega^* = 1$ is optimal, which corresponds the fully informative signal. Similarly, if V is concave, then $\omega^* = 0$ is optimal, which corresponds the completely uninformative signal. This is summarized in the following corollary.

Corollary 2. *An optimal signal is*

(a) *fully informative for all F if and only if V is convex,*

(b) *completely uninformative for all F if and only if V is concave.*

3.3. **Constant Bias.** We now impose more structure on the sender's utility, which is relevant in many applications. We assume that

$$\begin{aligned} &v(\omega, r) = 1 + \rho(\omega - r), \quad \rho \in \mathbb{R}, \\ &g \text{ is strictly positive.} \end{aligned} \tag{A$_1$}$$

That is, the sender's utility is a weighted sum of the receiver's utility and action. The sender is biased towards $a = 1$ but also puts a weight ρ on the receiver's utility. In particular, if the weight ρ is large, then the sender's and receiver's interests are aligned, whereas if the weight is zero, then the sender cares only about a.

The density g of the receiver's type r is said to be *log-concave* if $\ln g(r)$ is concave in r. Note that $\ln g(r)$ is well defined by Assumption (A_1).

Under Assumption (A_1), the shape of the sender's indirect utility V defined by (5) is connected to the shape of the density g of the receiver's type as follows.

Lemma 1. *Let (A_1) hold. Then V is S-shaped for all ρ if and only if g is log-concave.*

That is, if the density of the receiver's type g is log-concave, then the sender's indirect utility V is S-shaped. Moreover, this result is tight, in the sense that if g is not log-concave, then there exists ρ such that V is not S-shaped.

By Theorem 2 and Lemma 1, we obtain a criterion for the optimality of upper-censorship with the condition on the primitive of the model, the density g.

Theorem 3. *Let (A_1) hold and let the density g be log-concave. Then, and only then, an optimal signal is upper-censorship for all F and all ρ.*

Following Remark 2, the symmetric statement is also true: an optimal signal is *lower-censorship* for all F and all ρ if and only if $-g$ is log-concave. Consequently, if both g and $-g$ are log-concave (that is, g is exponential), then there exists an optimal signal which is both upper- and lower-censorship. There are only two signals with this property: fully informative and completely uninformative. This allows to obtain conditions on the distribution of the receiver's type under which the optimal signal polarizes between fully informative and completely uninformative signals, as, for example, in Lewis and Sappington (1994) and Johnson and Myatt (2006).

Corollary 3. *Let (A_1) hold. An optimal signal is either fully informative or completely uninformative for all F and all ρ if and only if there exist $\lambda \in \mathbb{R}$ and $c > 0$ such that $g(r) = ce^{-\lambda r}$ for $r \in [0, 1]$.*

If $g(r) = ce^{-\lambda r}$, then the fully informative signal is optimal whenever $\rho \leq -\lambda$ and the completely uninformative signal is optimal whenever $\rho \geq -\lambda$ (and any signal is optimal when $\rho = -\lambda$). In particular, if $\rho = 0$, the optimal signal is fully determined by the sign of λ, which is in turn determined by whether the mean of r is greater or smaller than 1/2.

3.4. **Comparative Statics.** Theorem 3 allows for a sharp comparative statics analysis on the amount of information that is optimally disclosed by the sender.

We compare signals by their Blackwell informativeness (Blackwell, 1953). To compare upper-censorship signals s_1 and s_2, we only need to compare their cutoffs ω_1^* and ω_2^*. Signal s_1 is more informative than signal s_2 if $\omega_1^* \geq \omega_2^*$. Indeed, state $\omega \in [0, \omega_2^*)$ is fully revealed by both s_1 and s_2, and state $\omega \in [\omega_2^*, 1]$ is partially revealed by s_1 and not revealed at all by s_2, so s_1 is more Blackwell informative than s_2.

For the purpose of comparison, we extend the definition of density function g to the real line and assume that g is log-concave. Consider a family of densities g_t of the receiver's type

$$g_t(r) = g(r - t),$$

where $t \in \mathbb{R}$ is a parameter. Because g_t is log-concave on $[0, 1]$ for every t, an upper-censorship mechanism is optimal by Theorem 3. Let $\omega^*(\rho, t) \in [0, 1]$ be the optimal upper-censorship cutoff as given by Proposition 2.

We now show that the sender optimally discloses more information when she is less biased relative to the receiver (the bias parameter ρ is greater), and when the receiver is more reluctant to act (the shift parameter t is greater).

Theorem 4. *Let (A_1) hold. For all ρ and t*

(a) $\omega^*(\rho, t)$ *is increasing in* ρ,

(b) $\omega^*(\rho, t)$ *is increasing in* t.

The intuition for part (a) is that for a higher ρ, the sender puts more weight on the receiver's utility, so she optimally endows the receiver with a higher utility by providing more information.

The intuition for part (b) is that for a higher t, each type of the receiver has a greater cost of action, so to persuade the same type of the receiver, the sender needs to increase $\mathbb{E}[\omega | \omega \geq \omega^*]$ by expanding the full disclosure interval $[0, \omega^*]$.

4. Discrete State

In many applications, the state is either a continuous random variable or a discrete random variable. Section 3 covers the continuous case. In this section we assume that the state is a discrete random variable; that is, it can take only a finite number of values.

In Section 3, an upper-censorship signal is defined to reveal the states below some cutoff state ω^* and pool the remaining states, where the cutoff state is assumed to be pooled. However, when the state is discrete, the sender will generally prefer the cutoff state to be revealed with some probability. Therefore, an optimal signal is generally not a monotone partition. We thus provide separate results for optimal signals and optimal monotone partitions.

A signal is *stochastic upper-censorship* if there exists a cutoff $\omega^* \in [0,1]$ and a probability $q^* \in [0,1]$ such that the states below ω^* are revealed, the states above ω^* are pooled, and the state ω^* is revealed with probability q^* and pooled with probability $1-q^*$. For example,

$$s = \begin{cases} \omega \text{ with probability one,} & \text{if } \omega < \omega^*, \\ \omega^* \text{ and } \bar{m}(\omega^*, q^*) \text{ with probabilities } q^* \text{ and } 1-q^*, & \text{if } \omega = \omega^*, \\ \bar{m}(\omega^*, q^*) \text{ with probability one,} & \text{if } \omega > \omega^*, \end{cases}$$

where

$$\bar{m}(\omega^*, q^*) = \frac{\int_{(\omega^*,1]} \omega \mathrm{d}F(\omega) + \omega^*(1-q^*)\Pr(\omega = \omega^*)}{\int_{(\omega^*,1]} \mathrm{d}F(\omega) + (1-q^*)\Pr(\omega = \omega^*)}$$

is the posterior expected state induced by the pooling signal.

A signal is *deterministic upper-censorship* if it is stochastic upper-censorship with $q^* \in \{0,1\}$. Note that deterministic upper-censorship is a monotone partitional signal, but stochastic upper-censorship with $q^* \in (0,1)$ is not.

Theorem 2′. *Let V be S-shaped. Then, and only then, for all discrete F*

(i) an optimal signal is stochastic upper-censorship,

(ii) an optimal monotone partition is deterministic upper-censorship.

Under Assumption (A_1) that imposes an additional structure on the sender's utility, we obtain the result analogous to Theorem 3.

Theorem 3′. *Let (A_1) hold and let the density g be log-concave. Then, and only then, for all discrete F and all ρ*

(i) an optimal signal is stochastic upper-censorship,

(ii) an optimal monotone partition is deterministic upper-censorship.

Finally, we discuss how the comparative statics result (Theorem 4) changes. In Section 3.4, we argue that upper-censorship signals can be compared by the comparison of their censorship cutoffs. However, when the state is discrete, a censorship cutoff is described by a pair (ω^*, q^*) where ω^* is a cutoff state and q^* is the probability of revealing this cutoff state when it realizes.

Consider two stochastic upper-censorship signals s_1 and s_2 with cutoffs (ω_1^*, q_1^*) and (ω_2^*, q_2^*). Denote $(\omega_1^*, q_1^*) \succeq (\omega_2^*, q_2^*)$ if $\omega_1^* > \omega_2^*$, or $\omega_1^* = \omega_2^*$ and $q_1^* \geq q_2^*$. Observe that s_1 is more (Blackwell) informative than s_2 if and only if $(\omega_1^*, q_1^*) \succeq (\omega_2^*, q_2^*)$. This comparison also applies to deterministic upper-censorship signals, with the constraint that q_1^* and q_2^* are in $\{0, 1\}$.

The statement of Theorem 4 remains same, with respect to the above order on censorship cutoffs. That is, the sender optimally discloses more information when she is less biased relative to the receiver (the bias parameter ρ is greater), and when the receiver is more reluctant to act (the shift parameter t is greater).

Remark 3. The above results show that a form upper-censorship emerges as an optimal persuasion strategy under the same condition (V is S-shaped in Theorem 2′ and g is log-concave in Theorem 3′), whether the sender chooses an arbitrary signal or is restricted to monotone partitions. It is tempting to conclude that if the sender was allowed to choose an arbitrary partition (not necessarily monotone), the same result would hold. However, this is not true.

For illustration, suppose that the sender wishes to maximize the probability of $a = 1$, and the receiver has type $r = 1/2$ with certainty.[3] Let the state ω take values in $\{0, \frac{1}{4}, 1\}$ with probabilities $(\frac{1}{6}, \frac{2}{3}, \frac{1}{6})$, so the prior mean is $\frac{1}{3}$.

An optimal signal is stochastic upper-censorship $(\omega^*, q^*) = (\frac{1}{4}, \frac{1}{2})$ that sends the pooling message with certainty when $\omega = 1$ and with probability $1 - q^* = \frac{1}{2}$ when $\omega = \frac{1}{4}$, and reveals the state otherwise. The pooling message induces the posterior mean $m = \frac{1}{2}$, leading to $a = 1$; all other messages induce posterior means below $\frac{1}{2}$, leading to $a = 0$. Under the optimal signal, action $a = 1$ is chosen with probability $\frac{2}{3} \cdot \frac{1}{2} + \frac{1}{6} = \frac{1}{2}$.

An optimal *monotone* partition is deterministic upper-censorship that fully reveals the state, where only $\omega = 1$ leads to $a = 1$. Under this partition, action $a = 1$ is chosen with probability $\frac{1}{6}$.

An optimal partition pools the extreme states $\omega = 0$ and $\omega = 1$, and reveals the intermediate state $\omega = \frac{1}{4}$. The pooling message induces the posterior mean $m = \frac{1}{2}$, leading to $a = 1$. This optimal partition is nonmonotone. Under this partition, action $a = 1$ is chosen with probability $\frac{1}{6} + \frac{1}{6} = \frac{1}{3}$, which is strictly better than the probability of $\frac{1}{6}$ obtained under the optimal monotone partition.

5. Application to Media Censorship

In this section, we apply our results to the problem of media censorship by the government. In the contemporary world, people obtain information about the government's state through various media sources such as television, newspapers, and internet blogs. Without the media, most people would not know what policies and reforms

[3]In this case, $V(m) = 1$ if $m \geq 1/2$ and $V(m) = 0$ if $m < 1/2$. Observe that the graph of V is S-shaped, and it can be approximated arbitrarily closely by a continuously differentiable S-shaped function.

the government pursues and how effective they are. Media outlets have different positions on the political spectrum and differ substantially in how they select and present facts to cover the same news. People choose their sources of information based on their political ideology and socioeconomic status. This information is valuable for significant individual decisions in migration, investment, and voting, to name a few. Individuals do not fully internalize externalities that their decisions impose on the society. Likewise, the government may not have the society's best interest at heart. To further its goals, the government then wishes to influence individual decisions by manipulating the information through media. In autocracies and countries with weak checks and balances, the government has power to censor the media content.

The government's problem of media censorship can be represented as the persuasion problem in Section 3. We apply our results to provide conditions for the optimality of upper-censorship policies that censor all media outlets except the most supportive ones. An interpretation of our comparative statics results is as follows. First, the government increases censorship if influencing society decisions becomes relatively more important than maximizing individual welfare. Second, the government increases censorship if the society experiences an ideology shock in favor of the government.

5.1. **Setup.** There is a continuum of heterogeneous citizens indexed by $r \in [0,1]$ distributed with G. Each citizen chooses between $a = 0$ and $a = 1$. The utility of a citizen of type r is given by

$$u(\theta, r, a_r, \bar{a}) = (\theta - r)a_r + \xi(\theta, r, \bar{a}),$$

where $a_r \in \{0,1\}$ denotes the citizen's own action, $\bar{a} = \int a_r \mathrm{d}G(r)$ denotes the aggregate action in the society, $\theta \in [0,1]$ captures an unobserved benefit from action 1 as compared to action 0, and ξ captures the impact of the aggregate action $\bar{a}$ on the citizen's utility. The term $(\theta - r)a_r$ is a private surplus of a citizen of type r. The term $\xi(\theta, r, \bar{a})$ is an externality, because for a citizen of type r it is optimal to ignore this term and choose $a_r = 1$ if and only if $\theta \geq r$.

There is a government which is concerned with a weighted average of the social utility and the government's intrinsic benefit from the aggregate action. For a given θ, the government's utility is given by

$$(1-\delta)\int_0^1 \nu(\theta, r, a_r, \bar{a})\mathrm{d}G(r) + \delta\gamma(\theta, \bar{a}).$$

The term ν captures a citizen's utility from the government's perspective. We allow ν to be different from u to reflect paternalistic or other concerns. The term γ captures the government's intrinsic benefit from the aggregate action. The parameter $\delta \in [0,1]$ captures a relative weight of the aggregate action in the government's utility.

Let T be a distribution of the random variable θ. We assume that distributions G and T are independent and admit continuously differentiable and strictly positive densities g and τ. We also assume that β, γ, and ν are linear in θ and continuously differentiable in r and $\bar{a}$. Furthermore, to simplify interpretations, we assume that ν

and γ are non-decreasing in θ and $\bar{a}$, so that for the government, a high θ is a good news, and a higher aggregate action is preferable.

Citizens obtain information about the unobservable benefit θ through media outlets. Each media outlet is identified by its editorial policy $c \in [0,1]$, and it endorses action $a = 1$ if $\theta \geq c$ and criticizes it if $\theta < c$.[4] A set C of media outlets is either the entire interval $[0,1]$, or a finite subset of $[0,1]$.

The government's censorship policy is a measurable set of the media outlets $X \subset C$ that are permitted to broadcast; so the rest of the media outlets are censored.

The timing is as follows. First, the government chooses a set $X \subset C$ of permitted media outlets. Second, state θ is realized, and each permitted media outlet endorses or criticizes action $a = 1$ according to its editorial policy. Finally, each citizen observes messages from all permitted media outlets, updates his beliefs about θ, and chooses an action.

5.2. **Discussion.** We now discuss interpretations of the key components of the media censorship application. As in Gehlbach and Sonin (2014), there can be various interpretations of the citizen's action $a = 1$, such as voting for the government, supporting a government's policy, or taking an individual decision that benefits the government. A citizen's type r can be interpreted as his ideological position or preference parameter. A citizen who is more supportive of the government has a smaller r.

A media outlet with a higher editorial policy c can be interpreted as more disloyal to the government because it criticizes the government on a larger set of states. An editorial policy $c \in C$ can therefore represent a slant or political bias of the outlet against the government and can be empirically measured as the frequency with which the outlet uses anti-government language. Gentzkow and Shapiro (2010) construct such a slant index for U.S. newspapers. Empirical findings of their paper suggest that the editorial policies of media outlets are driven by reader preferences, justifying our assumption of the existence of a large variety of editorial policies.[5] As in Suen (2004), Chan and Suen (2008), and Chiang and Knight (2011), the assumption of the binary media reports that communicate only whether the state θ is above some standard c can be justified by a cursory reader's preference for simple messages such as positive or negative opinions and yes or no recommendations.

The government's censorship of media outlets can take different forms. For example, the government can ban access to internet sites, withdraw licenses, disrupt financing, confiscate print materials and equipment, and discredit, arrest, or even murder editors and journalists. In some countries, the government can exercise direct control over media editorial policies either through state ownership or administrative pressure.

[4] The tie-breaking in the event of $\theta = c$ is unimportant, as θ is a continuous random variable.

[5] Theoretical literature has explored the determinants of media slant of an outlet driven by its citizens (Mullainathan and Shleifer, 2005, Gentzkow and Shapiro, 2006, and Chan and Suen, 2008) and its owners (Baron, 2006, and Besley and Prat, 2006).

5.3. **Formulating Media Censorship as Persuasion.** We now show that the media censorship problem can be formulated as a linear persuasion problem, in which the government is a sender and a representative citizen is a receiver.

Observe that the citizens' and the government's utilities are linear in θ. Therefore, given any information from the media outlets, the utilities depend only on the posterior mean of θ.

Consider an arbitrary posterior mean of θ, denoted by $m \in [0, 1]$. Each citizen of type r chooses $a_r = 1$ if and only if $r \le m$. The externality term β plays no role in this decision. Therefore, the aggregate action $\bar{a}$ is simply the mass of all citizens whose types do not exceed m, so $\bar{a} = G(m)$.

Next, using the citizens' optimal behavior, we derive the government's expected utility conditional on the posterior mean m:

$$\begin{aligned} V(m) &= \mathbb{E}\left[(1-\delta) \int_0^1 \nu(\theta, r, a_r, G(m)) \mathrm{d}G(r) + \delta\gamma(\theta, G(m)) \middle| \, m \right] \\ &= (1-\delta) \int_0^1 \nu(m, r, \mathbf{1}_{\{r \le m\}}, G(m)) \mathrm{d}G(r) + \delta\gamma(m, G(m)). \end{aligned} \tag{9}$$

This is the government's *indirect utility.* As in Section 3, the government now chooses a signal which is informative about the state to maximize its expected utility. However, in contrast to Section 3, here the government is restricted to signals that are implementable by a subset of a given set of media outlets.

Next, we show that, with an appropriate definition of the state, the restriction to signals implementable by a subset of a given set of media outlets can be formulated as a restriction to monotone partitional signals, as in Scenario (ii) in Section 3.1. Thus we will be able to use our results in Sections 3 and 4 to find the optimal censorship policies.

Recall that the information about θ is only available through the media outlets. Let ω denote a random variable equal to the conditional expectation of θ given the messages of all media outlets. Note that if the set C of media outlets is the whole interval, $C = [0, 1]$, then ω coincides with θ; if C is a finite subset of $[0, 1]$, then ω is a discrete random variable. Let F denote the distribution of ω. From now on, we treat the random variable ω as the state, and denote by s a signal as defined in Section 3.1.

Let $\mathcal{H}_M$ denote the set of *all* distributions of the posterior mean state induced by all monotone partitional signals, and let $\mathcal{H}_C$ denote the set of *all* distributions of the posterior mean state induced by all finite subsets of the set C of media outlets. Observe that any set of media outlets induces a monotone partitional signal, that is, $\mathcal{H}_C \subset \mathcal{H}_M$. We now show that every outcome implementable by an arbitrary monotone partition in $\mathcal{H}_M$ is also implementable by a subset of media outlets.

Lemma 2. *$\mathcal{H}_C = \mathcal{H}_M$.*

For illustration, suppose that there is only one media outlet with the editorial policy $c \in (0,1)$. There are two observable events: $\theta \geq c$ and $\theta < c$. In each realized event, the citizens compute the posterior expectation of θ, which we will call ω. So, in this example, the distribution F of ω has two mass points, one for each of the two events.

Given these two mass points, every monotone partition of $[0,1]$ induces one of two possible distributions of the posterior mean state, (i) the mass points separated, and (ii) the mass points pooled. Observe, however, that these distributions are implementable by permitting and censoring the media outlet, respectively.

A censorship policy is *upper censorship* if it censors all sufficiently disloyal media outlets. Specifically, there exists a cutoff $c^* \in C$ and an indicator $q^* \in \{0,1\}$ such that all media outlets whose editorial policies are below c^* are permitted, all media outlets whose editorial policies are above c^* are censored, and c^* is permitted if and only if $q^* = 1$. That is, $X = \{c \in C : c \leq c^*\}$ if $q^* = 1$ and $X = \{c \in C : c < c^*\}$ if $q^* = 0$. Note that the *full censorship* policy $c^* = 0$ and $q^* = 0$ (where all media outlets are censored) and the *free media* policy $c^* = 1$ and $q^* = 1$ (where all media outlets are permitted) are the two extreme upper censorship policies.

As shown above, the media censorship problem is equivalent to the linear persuasion problem in which the sender is restricted to monotone partitional signals and the sender's indirect utility V is given by (9). We thus apply Theorem $2'$(ii) to obtain the following result.

Theorem $2''$. *If V is S-shaped, then an optimal censorship policy is upper-censorship.*

To illustrate Theorem $2''$, suppose the government is interested only in the aggregate action, so that $v = 0$ and γ depends only on the aggregate action $\bar{a}$; so $V(m) = \gamma(G(m))$. Thus, an upper-censorship policy is optimal if the composition function $\gamma(G(\cdot))$ is S-shaped. For example, this condition holds if the government is interested in reaching a certain approval threshold, so that γ is a step function. This condition also holds if γ is S-shaped and G is uniform or S-shaped with the same inflection point as γ.[6]

5.4. **Constant Bias.** We now impose more structure on the utilities to obtain a sharper result for optimality of upper censorship and to perform a comparative statics analysis. Similarly to Section 3.3, we assume that the government's utility is a weighted average of the citizens' utility and their aggregate action:

$$\begin{aligned}&(1-\delta)\int_0^1 u(\theta, r, a_r, \bar{a})\mathrm{d}G(r) + \delta\bar{a},\\ &\text{where } u(\theta, r, a_r, \bar{a}) = (\theta - r)a_r + \zeta(r)\bar{a}\end{aligned} \tag{A$_2$}$$

[6]This condition holds in many special cases where γ and G are S-shaped even when γ and G have different inflection points.

for some continuously differentiable function ζ. Define

$$\beta = (1-\delta)\int_0^1 \zeta(r)\mathrm{d}G(r) + \delta.$$

The term $(1-\delta)\int_0^1 \zeta(r)\mathrm{d}G(r)$ is the government's expected bias towards the citizens' action $a = 1$ due to the citizens' externality, and the term δ is the government's intrinsic bias towards a greater aggregate action $\bar{a}$. Thus, we interpret β as the government's aggregate bias. This decomposition of the bias allows for different interpretations why the government is biased. So, β can be high because the government is too self-serving (high δ), or because the government is benevolent (low δ) but wishes to internalize strong positive externalities of the citizens (high $\zeta(r)$). For the ease of interpretation, we assume that $\beta > 0$.[7]

Under Assumption (A_2), the government's indirect utility V given by (9) becomes

$$V(m) = (1-\delta)\int_0^m (m-r)\mathrm{d}G(r) + \beta\bar{a} = \int_0^m v(m,r)\mathrm{d}G(r),$$

where

$$v(m,r) = \beta\left(1 + \frac{1-\delta}{\beta}(m-r)\right).$$

This is the same as v given by Assumption (A_1) with $\rho = (1-\delta)/\beta$, up to rescaling by a positive constant. Consequently, we can apply Theorem 3′(ii) for V to obtain the following result.

Theorem 3″. *Let (A_2) hold. If the density g of citizens' types is log-concave, then an optimal censorship policy is upper-censorship.*

We now apply the comparative statics analysis presented in Section 3.4. Note that the upper censorship policies are ordered according to the amount of information transmitted to the citizens, in the sense of Blackwell (1953). A greater censorship threshold c^* means that the interval of states below c^* where the readers receive best possible information is greater, and the pooling interval of states above c^* is smaller. With this order in mind, we will make a comparative statics analysis on the amount of information that is optimally disclosed by the government.

First, Theorem 4(a) states that the censorship cutoff is weakly increasing in ρ. Recall that this is a reciprocal of the government's bias, $\rho = (1-\delta)/\beta$. This means that the government optimally discloses more information (the censorship cutoff is greater) when it is less biased (β is smaller). Intuitively, as β decreases, the government puts more weight on the citizens' utility, so it optimally endows the population with a higher utility by censoring fewer media outlets and disclosing more information.

[7] If $\beta < 0$, then swapping the roles of $a = 0$ and $a = 1$ reverses the sign of the bias. If $\beta = 0$, then the government's utility and the citizens' private interests coincide, so it is trivially optimal to disclose maximum information.

Second, Theorem 4(b) states that the censorship cutoff is weakly increasing in t. Recall that parameter t is the magnitude of the horizontal shift of the density g (see Section 3.4), so a greater t is a greater opportunity cost of action $a = 1$ for each citizen. This means that the government optimally discloses more information (the censorship cutoff is greater) when the citizens' are more difficult to persuade to take action $a = 1$ (parameter t is greater). Informally speaking, to persuade the same type of the citizen, the government needs to increase the posterior mean. But the expected posterior mean must be equal to the prior mean, so it is not possible to increase all posterior means. Due to the log-concave shape of the density of the citizens' types, this tradeoff is resolved by increasing the posterior mean of the pooling interval, $\mathbb{E}[\omega|\omega \geq c^*]$. This is done by shrinking the pooling interval $[c^*, 1]$, that is, increasing the censorship cutoff c^*.

5.5. **Extensions and Open Questions.** Let us now consider a few extensions of our model of censorship.

In our model, the set of media outlets is exogenous, and the government's only instrument is censorship. We now consider three alternative ways of expanding the government's instruments of influence.

First, suppose that the government can not only censor existing media outlets, but also introduce new media outlets with chosen editorial policies. This is equivalent to our censorship model where all media outlets in $[0, 1]$ are initially available, and the government can censor any subset of them. So, this extension is, in fact, as special case of our model.

Second, suppose that the government can garble information available from media outlets. That is, the government is not restricted to monotone partitions, it can create arbitrary signals about state ω for the citizens to observe. This becomes a general persuasion problem, and our Theorem 2′(i) applies.

Third, suppose that the government can restrict not only which media outlets are permitted, but also how many media outlets each citizen can choose to observe. In this extension the citizens are not allowed to communicate with one another, as otherwise they would have shared the information, thus observing all permitted media outlets indirectly. This extension does not affect our results, as long as each citizen is allowed to access at least one media outlet of his choice, as in Chan and Suen (2008). Intuitively, this is because each citizen categorizes the information from the media outlets into "good news" where $a = 1$ is optimal and "bad news" where $a = 0$ is optimal. Because the information from the media outlets induces a monotone partition, it means that "good news" is separated from "bad news" by a threshold media outlet that depends on the citizen's type. That is, it is sufficient to observe a single threshold media outlet to distinguish "good news" from "bad news".

In our model, each citizen's utility depends on the aggregate action $\bar{a}$ through the externality term $\beta(\theta, r, \bar{a})$ which does not affect the chosen action. Let us relax this assumption, so that a citizen's optimal choice can depend on $\bar{a}$. We can still write the sender's indirect utility V as a function of the posterior mean state and apply our results. However, now V is not uniquely determined by the primitives of the model. It is endogenous and depends on an equilibrium the citizens play, as each citizen's optimal action now depends on what all citizens do in equilibrium. For example, given the same information about the state, a citizen could prefer to choose $a = 1$ if and only if many enough citizens choose the same, so $\bar{a}$ is large enough. This creates the problem of multiplicity of equilibria and, as a consequence, the dependence of optimal censorship on equilibrium selection.

Our media censorship problem can be applied to spatial voting models, as in Chiang and Knight (2011). Consider a government party ($p = G$) and an opposition party ($p = O$) competing in an election. If party p wins, a voter with ideological position r gets utility $w_p - (r - r_p)^2$, where w_p is the quality or valence of party p, and r_p is the ideology or policy platform of party p. Voters know the parties' ideologies and obtain information about the parties' qualities from all available media outlets. Each voter supports the party that maximizes his expected utility. Our analysis applies, because the voter's utility difference between the government and opposition parties is proportional to $\theta - r$, where $\theta = \left(w_G - w_O + r_O^2 - r_G^2\right)/2(r_O - r_G)$.

There are a few more extensions that can be practically relevant. First, instead of complete censorship of a media outlet, there could be a cost of accessing it. For example, an international news channel could be censored by a local government, but citizens could still to access it through VPN at some cost. Second, it can be costly for the government to censor media outlets, so an important question is how to prioritize censoring. Finally, citizens could incur some cost of following each media outlet. We already mentioned that citizens have no benefit in following more than one outlet. But they might stop watching news altogether if the news is sufficiently uninformative. These extensions are nontrivial and left for future research.

Appendix

Proof of Theorem 2. Theorem 2 is a special case of Theorem 2′(i).

Proof of Theorem 2′(i). Each signal s induces a random variable $m = \mathbb{E}[\omega|s]$, called the posterior mean. Let H be the distribution of m induced by a signal. Because the sender's and receiver's utilities are linear in the state, H summarizes all relevant information about a signal.

The distribution H implements the receiver's interim utility U given by

$$U(r) = \int_r^1 (m - r)\mathrm{d}H(m) = \int_r^1 (1 - H(m))\mathrm{d}m \text{ for } r \in [0, 1],$$

where the first equality holds because the receiver acts iff $m \geq r$, and the second equality holds by integration by parts. Notice that each such function U uniquely determines H, so that $H(m) = 1 + U'(m)$ where $U'(m)$ is the right derivative of U at m. Thus, we can represent each signal by U.

A fully informative signal induces the distribution $\overline{H}$ of m equal to F, and thus implements the interim utility given by

$$\overline{U}(r) = \int_r^1 (1 - F(m))\mathrm{d}m \text{ for } r \in [0,1].$$

A completely uninformative signal induces the distribution $\underline{H}$ of m that assigns probability 1 to $m = \mathbb{E}[\omega]$, and thus implements the interim utility given by

$$\underline{U}(r) = \max\{\mathbb{E}[\omega] - r, 0\} \text{ for } r \in [0,1].$$

As follows from Theorem 1 in Kolotilin et al. (2017), there exists a signal that implements U if and only if $U(r)$ is convex on $[0,1]$ and $\underline{U}(r) \leq U(r) \leq \overline{U}(r)$ for all $r \in [0,1]$ (see Figure XXX).

The sender's expected utility given the posterior mean m is

$$V(m) = \int_0^m v(m,r)\mathrm{d}G(r) \text{ for } m \in [0,1].$$

Using integration by parts, as in Lemma 2 in Kolotilin et al. (2017), we can rewrite the sender's expected utility to obtain the following result.

Lemma 3 (Kolotilin et al. 2017)**.** *An optimal signal implements*

$$\begin{aligned} &U^* \in \arg\max_U \int_0^1 U(r)V''(r)\mathrm{d}r \\ &\text{subject to } U \text{ is convex and } \underline{U} \leq U \leq \overline{U}, \end{aligned} \tag{10}$$

where V'' is the second derivative of V.

The interim utility U^* under an upper-censorship signal is shown as a black kinked curve in Figure XXX. A receiver of type $r \leq \omega^*$ knows whether $\omega \geq r$, and hence gets the highest feasible utility $U^*(r) = \overline{U}(r)$, by choosing $a = 1$ whenever $\omega \geq r$. A receiver of type $r > \omega^*$ knows whether $\omega \geq \omega^*$, and hence gets the utility

$$U^*(r) = \max\{m^* - r, 0\} \cdot (1 - F(\omega^*) + \Pr(\omega = \omega^*)(1 - q^*)),$$

by choosing $a = 1$ whenever $\omega \geq \omega^*$ and $m^* \geq r$.

Consider an S-shaped function V, so that there exists $r^* \in [0,1]$ such that $V''(r) \geq 0$ for $r < r^*$ and $V''(r) \leq 0$ for $r > r^*$. Fix a value $y^* \in [\underline{U}(r^*), \overline{U}(r^*)]$. Define $\omega^* \in [0,1]$ and $q^* \in [0,1]$ such that the interim utility U^* satisfies $U^*(r^*) = y^*$ under the corresponding upper-censorship signal. It is easy to see from Figure XXX that for any convex U such that $\underline{U} \leq U \leq \overline{U}$ and $U(r^*) = y^*$, we have $U^*(r) \leq U(r)$ for $r < r^*$ and $U^*(r) \geq U(r)$ for $r > r^*$. Thus, by Lemma 3, an upper-censorship signal is optimal.

Conversely, suppose that V is not S-shaped. Then there exist $0 \leq r_1 < r_2 < r_3 \leq 1$ such that $V''(r) < 0$ for $r \in (r_1, r_2)$ and $V''(r) > 0$ for $r \in (r_2, r_3)$, because V'' is continuous by assumption. Thus, $-V$ is S-shaped on the interval $[r_1, r_3]$. Consider any F with the support equal to $[r_1, r_3]$ and any upper-censorship signal. Let y^* be the receiver's utility of type r^* under this signal. It is easy to see from Figure XXX that the sender strictly prefers a lower-censorship signal that gives the same utility y^* to the receiver.

Proof of Proposition 2. Observe that

$$(1 - F(\omega^*))\frac{\mathrm{d}m^*}{\mathrm{d}\omega^*} = f(\omega^*)(m^* - \omega^*).$$

Using the above, we find the first-order condition of the problem (6): $\omega^* \in [0,1]$ is a solution of (6) if it satisfies

$$\big[V(\omega) - V(m^*) + V'(m^*)(m^* - \omega)\big] f(\omega) \geq (\leq)\, 0 \quad \text{for all } \omega \leq (\geq)\, \omega^*. \tag{11}$$

As $f(\omega) \geq 0$ for all ω, the above condition is weaker than (7). So ω^* satisfying (7) must be a solution of (6).

It remains to show that there exists $\omega^* \in [0,1]$ that satisfies (7). This is true, because V is S-shaped, so the expression in (7) is single-crossing from above. To see this, consider Fig. 2. Observe that (8) can hold only if ω^* is on the convex part of V and m^* on the concave part of V. As the censorship cutoff ω^* increases, the posterior mean state of the pooling message m^* also increases. But because V is concave at m^*, the dashed tangent line becomes flatter, so it crosses the solid line at a smaller ω. It follows that the expression in (8) is nonpositive for all $\omega > \omega^*$ and nonnegative for all $\omega < \omega^*$. □

Proof of Lemma 1. Notice that, under assumption (A_1), we have

$$V''(r) = g'(r) + \rho g(r) \text{ for } r \in [0,1].$$

Recall that V is S-shaped iff $V''(r) = g'(r) + \rho g(r)$ is single-crossing from above. By Proposition 1 in Quah and Strulovici (2012), this holds for all $\rho \in \mathbb{R}$ if and only if $g'(r)/g(r)$ is nonincreasing in r (that is, $\ln g(r)$ is concave). □

Proof of Theorem 3. Immediate by Theorem 2 and Lemma 1. □

Proof of Theorem 4. Consider a signal that implements the interim utility U (equivalently, the distribution $H = 1 + U'$ of the posterior mean m). The sender's expected utility is then

$$\begin{aligned}\int_0^1 V(m)\mathrm{d}H(m) &= \int_0^1 \int_0^m (1+\rho(m-r))\mathrm{d}G(r)\mathrm{d}H(m) \\ &= \int_0^1 \int_r^1 (1+\rho(m-r))\mathrm{d}H(m)\mathrm{d}G(r) \\ &= \int_0^1 (-U'(r)+\rho U(r))\mathrm{d}G(r).\end{aligned}$$

Part (a). Consider $\rho_2 > \rho_1$. Suppose to get a contradiction that the cutoffs of the corresponding optimal upper-censorship signals are such that $F(\omega_2^*) < F(\omega_1^*)$. Since the sender prefers a signal that induces U_2 under ρ_2 and a signal that induces U_1 under ρ_1, we have

$$\begin{aligned}\int (-U_2'(r)+\rho_2 U_2(r))\,\mathrm{d}G(r) &\geq \int (-U_1'(r)+\rho_2 U_1(r))\,\mathrm{d}G(r), \\ \int (-U_1'(r)+\rho_1 U_1(r))\,\mathrm{d}G(r) &\geq \int (-U_2'(r)+\rho_1 U_2(r))\,\mathrm{d}G(r).\end{aligned}$$

Summing up these conditions gives:

$$(\rho_2-\rho_1)\int (U_2(r)-U_1(r))\,\mathrm{d}G(r) \geq 0,$$

leading to a contradiction to the fact that $U_2(r) \geq U_1(r)$ for all r with strict inequality for some r, which follows from $F(\omega_2^*) < F(\omega_1^*)$.

Part (b). It is easy to prove this part in the case where F admits a density or in scenario (i). Indeed, we want to show that $\omega_1^* \leq \omega_2^*$ for $t_1 < t_2$. Since ω_2^* is optimal under t_2, we have

$$V(\omega-t_2) \leq V(m_2^*-t_2)+V'(m_2^*-t_2)(m_2^*-\omega) \text{ for } \omega\in[\omega_2^*,1]. \tag{12}$$

For an S-shaped V, it is easy to see from a graph that the following inequality holds

$$V(\omega-t_1) \leq V(m_2^*-t_1)+V'(m_2^*-t_1)(m_2^*-\omega) \text{ for } \omega\in[\omega_2^*,1], \tag{13}$$

which implies that $\omega_1^* \leq \omega_2^*$.

Now consider the case where the sender chooses a monotone partition. Then the difficulty arises because (12) does not have to hold if there is an atom at ω_2^*. Let $\bar{m}(\omega^*, q^*)$ be the mean state of the pooling message in a given upper-censorship signal (ω^*, q^*). To prove the result, we first need to notice that

$$\begin{aligned}V_t^*(\omega^*,q^*) = &\int_{[0,\omega^*)} V(\omega-t)\mathrm{d}F(\omega) + V(\omega^*-t)q^*\Pr(\omega=\omega^*) \\ &+ V(\bar{m}(\omega^*,q^*))((1-q^*)\Pr(\omega=\omega^*)+1-F(\omega^*)),\end{aligned}$$

is single-peaked in (ω^*, q^*) in the lexicographic order (Blackwell informativeness order).

If (12) holds, then the previous proof goes through. If (12) is violated, then $\Pr(\omega = \omega_2^*) \neq 0$ and $q_2^* = 0$, by single-peakedness of V_t^*. Moreover,

$$V(\bar{m}(\omega_2^*, 0) - t_2)(\Pr(\omega = \omega_2^*) + 1 - F(\omega_2^*)) \geq V(\omega_2^* - t_2)\Pr(\omega = \omega_2^*) + V(\bar{m}(\omega_2^*, 1) - t_2)(1 - F(\omega_2^*))$$

and

$$V(\omega - t_2) \leq V(\bar{m}(\omega_2^*, 1) - t_2) + V'(\bar{m}(\omega_2^*, 1) - t_2)(\bar{m}(\omega_2^*, 1) - \omega) \text{ for } \omega \in [\omega_2^*, 1],$$

again by single-peakedness of V_t^*.

Therefore, it suffices to show that

$$V(\bar{m}(\omega_2^*, 0) - t_1)(\Pr(\omega = \omega_2^*) + 1 - F(\omega_2^*)) \geq V(\omega_2^* - t_1)\Pr(\omega = \omega_2^*) + V(\bar{m}(\omega_2^*, 1) - t_1)(1 - F(\omega_2^*)).$$

Moreover, it seems to suffice to show that if the inequality holds with equality for t_2, then the inequality holds for $t_1 = t_2 - \mathrm{d}t$. This property is equivalent to

$$V'(\bar{m}(\omega_2^*, 0) - t_2)(\Pr(\omega = \omega_2^*) + 1 - F(\omega_2^*)) \geq V'(\omega_2^* - t_2)\Pr(\omega = \omega_2^*) + V'(\bar{m}(\omega_2^*, 1) - t_2)(1 - F(\omega_2^*)),$$

which holds as can be seen from the graph representing the condition

$$V(\bar{m}(\omega_2^*, 0) - t_2)(\Pr(\omega = \omega_2^*) + 1 - F(\omega_2^*)) = V(\omega_2^* - t_2)\Pr(\omega = \omega_2^*) + V(\bar{m}(\omega_2^*, 1) - t_2)(1 - F(\omega_2^*)).$$

As a side, a clean proof of Alonso and Câmara (2016b) does not seem to be applicable, as they have $\rho = 0$ and assume log-concavity of g. For their proof to be applicable, we would need to assume that $V'(r) = g(r) + \rho G(r)$ is log-concave, which cannot hold for all ρ.

Proof of Theorem 3′. Immediate by Theorem 2′ and Lemma 1. □

Proof of Lemma 4.

Lemma 4. *For each $s_1 \in S_M$ there exists $s_2 \in S_C$ such that $H_{s_1} = H_{s_2}$. Conversely, for each $s_1 \in S_C$ there exists $s_2 \in S_M$ such that $H_{s_1} = H_{s_2}$.*

First, observe that a signal induced by any set $X \subset C$ of permitted media outlets induces a monotone partition of $[0, 1]$, and thus $S_C \subset S_M$. Indeed, let $\underline{c}_X(\omega)$ and $\overline{c}_X(\omega)$ be the media outlets in X that are the closest to ω from above and from below:

$$\underline{c}_X(\omega) = \sup\Big(\{c \in X : c < \omega\} \cup \{0\}\Big) \quad \text{and} \quad \overline{c}_X(\omega) = \inf\Big(\{c \in X : c \geq \omega\} \cup \{1\}\Big).$$

By observing the media outlets in X, each citizen is either exactly informed about the state when $\underline{c}_X(\omega) = \overline{c}_X(\omega)$), or is informed that $\omega \in [\underline{c}_X(\omega), \overline{c}_X(\omega))$.

We prove that each monotone partitional signal in S_M induces the same distribution of posterior means as some set $X \subset C$ of media outlets. Consider a monotone partitional signal $s \in S_M$, and let H_s be the induced distribution of the posterior mean state. Observe that $H_s(\omega)$ can be increasing at ω, that is, $\lim_{\varepsilon\to 0}(H_s(\omega)-H_s(\omega-\varepsilon))/\varepsilon > 0$, only if $\omega \in C$, so there is a media outlet with the editorial policy ω.

Second, observe that if F has an atom on some $\omega \in (0,1)$, then

$$\underline{c}_X(\omega) < \omega < \overline{c}_X(\omega) \quad \text{for all } X \subset C, \tag{14}$$

in particular, for $X = C$. This is because the distribution T of θ has no atoms, so the only way ω has an atom is that ω is the posterior mean of θ on an interval that pools a positive measure of realizations of θ.

Equipped with the above observation, we prove that each monotone partition Π of $[0,1]$ induces the same distribution of posterior means as some set $X \subset C$ of media outlets. Consider a monotone partition Π. Assume that each nonsingleton element $\pi \in \Pi$ has a strictly positive measure under F. This is without loss of generality, because any nonsingleton element $\pi \in \Pi$ that has a measure zero under F can be merged with a neighboring element without affecting the posterior of that element.

For each element $\pi \in \Pi$ with a strictly positive measure, let $\chi(\pi)$ be the smallest convex set that is contained in π and has the same measure as π. That is, $\chi(\pi)$ is obtained from π by truncating measure zero tails on the left and on the right of π (if any). Define

$$\underline{x}(\pi) = \sup\{c \in C \cup \{0\} : c \leq \inf \chi(\pi)\big\}.$$

Let

$$X = \{\underline{x}(\pi)\}$$

For each element $\pi \in \Pi$, with abuse of notation let $F(\pi)$ denote the measure of π under

that has a positive measure under F, denote by $F(\pi)$

$\varphi(\pi)$ the posterior distribution of ω conditional on $\omega \in \pi$.

□

References

Alonso, R., and O. Câmara (2016a): “Persuading Voters,” *American Economic Review*, 106, 3590–3605.

——— (2016b): “Political disagreement and information in elections,” *Games and Economic Behavior*, 100, 390–412.

Bagnoli, M., and T. Bergstrom (2005): “Log-concave Probability and Its Applications,” *Economic Theory*, 26, 445–469.

Baron, D. P. (2006): “Persistent Media Bias,” *Journal of Public Economics*, 90(1), 1–36.

Besley, T., and A. Prat (2006): "Handcuffs for the Grabbing Hand? Media Capture and Government Accountability," *American Economic Review*, 96(3), 720–736.

Blackwell, D. (1953): "Equivalent Comparisons of Experiments," *Annals of Mathematical Statistics*, 24, 265–272.

Boleslavsky, R., and C. Cotton (2018): "Limited capacity in project selection: competition through evidence production," *Economic Theory*, 65(2), 385–421.

Chan, J., and W. Suen (2008): "A Spatial Theory of News Consumption and Electoral Competition," *Review of Economic Studies*, 75, 699–728.

Chiang, C.-F., and B. Knight (2011): "Media Bias and Influence: Evidence from Newspaper Endorsements," *Review of Economic Studies*, 78, 795–820.

Duffie, D., P. Dworczak, and H. Zhu (2017): "Benchmarks in Search Markets," *Journal of Finance*, 72, 1983–2044.

Dworczak, P. (2017): "Mechanism Design with Aftermarkets: Cutoff Mechanisms," *mimeo.*

Dworczak, P., and G. Martini (2018): "The Simple Economics of Optimal Persuasion," *Journal of Political Economy*, forthcoming.

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, Papers & Proceedings*, 106, 597–601.

Gentzkow, M., and J. Shapiro (2006): "Media Bias and Reputation," *Journal of Political Economy*, 114(2), 280–316.

——— (2010): "What Drives Media Slant? Evidence from US Daily Newspapers," *Econometrica*, 78(1), 35–71.

Goldstein, I., and Y. Leitner (2018): "Stress Tests and Information Disclosure," *Journal of Economic Theory*, 177, 34–69.

Johnson, J. P., and D. P. Myatt (2006): "On the Simple Economics of Advertising, Marketing, and Product Design," *American Economic Review*, 96, 756–784.

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

Kolotilin, A. (2015): "Experimental Design to Persuade," *Games and Economic Behavior*, 90, 215–226.

——— (2017): "Optimal Information Disclosure: A Linear Programming Approach," *Theoretical Economics*, forthcoming.

Kolotilin, A., M. Li, T. Mylovanov, and A. Zapechelnyuk (2015): "Persuasion of a privately informed receiver," *mimeo.*

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

Lewis, T. R., and D. Sappington (1994): "Supplying Information to Facilitate Price Discrimination," *International Economic Review*, 35, 309–327.

Mullainathan, S., and A. Shleifer (2005): "The Market for News," *American Economic Review*, 95(4), 1031–1053.

Orlov, D., P. Zryumov, and A. Skrzypach (2018): "Design of Macro-prudential Stress Tests," *mimeo*.
Ostrovsky, M., and M. Schwarz (2010): "Information Disclosure and Unraveling in Matching Markets," *American Economic Journal: Microeconomics*, 2, 34–63.
Quah, J., and B. Strulovici (2012): "Aggregating the Single Crossing Property," *Econometrica*, 80, 2333–2348.
Rayo, L., and I. Segal (2010): "Optimal Information Disclosure," *Journal of Political Economy*, 118, 949 – 987.
Suen, W. (2004): "The Self-Perpetuation of Biased Beliefs," *Economic Journal*, 114, 377–396.