The Extensive Margin of Bayesian Persuasion

Pietro Dall'Ara*
Boston College

version: May 21, 2023

preliminary and incomplete

link to latest version

Abstract

I study the persuasion of a receiver who accesses information only if she exerts attention effort. The sender uses information to incentivize the receiver to pay attention. I show that persuasion mechanisms are equivalent to signals when the receiver's private information includes the cost of her attention effort. In a model of media capture, the sender finds it optimal to make the receiver distinguish between high states and intermediate states.

Keywords: Persuasion, Inattention, Cost of Information, Information Design.
JEL Codes: D82, D83, D91.

*I am grateful for helpful comments from Mehmet Ekmekci, Laurent Mathevet, Ryan Charour, Teddy Mekonnen, Bumin Yenmez, Utku Ünver, Chiara Margaria, Hideo Konishi, Manshu Khanna, Giacomo Rubbini, Marco Errico, Stefano Piasenti. First draft: December 2021.

1 Introduction

In the "information age," consumers of information decide whether an information source deserves attention because information acquisition is costly (Floridi, 2014; Simon, 1996). The information-design literature studies a *sender* who supplies information to a *receiver*, to persuade the receiver to take a certain action (Bergemann and Morris, 2019; Kamenica, 2019). When attention is costly, the sender faces the dual problem of (i) persuading the receiver to take a certain action and (ii) inducing her to pay attention. In this paper, I study the persuasion of a receiver who is privately informed about her cost and benefit of information, in which the sender uses information to reward the receiver for her effort.

The intensive margin of persuasion captures intensity of the sender's persuasion on the receiver's *action* decision, while the extensive margin of persuasion refers to whether or not the receiver pays *attention* to the sender's information. The study of the extensive margin of persuasion is important to determine which consumers have access to information. In a persuasion game, the sender effectively allocates information to a heterogeneous audience. For instance, today's central banks use "layered communication" to reach the general public, characterized by heterogeneous and limited information-processing ability, as discussed in Section 5. In my model, I investigate the following questions: Who accesses information? Does the receiver benefit from the limit to her information-processing ability?

In order to study the *extensive* and *intensive* margin of persuasion, I model the persuasion of an inattentive receiver who takes a binary action, 1 or 0. There is a state θ unknown to two players: Sender (he) and Receiver (she). Receiver chooses 1 only if she expects the state θ to exceed her outside option. Sender wants Receiver to choose 1 regardless of the state. In the baseline model, Sender designs a random variable S correlated with the state θ, called *signal*. Knowing the signal S, but not its realization, Receiver chooses her attention effort e: high effort is costly and increases the probability of observing the signal realization. The choice of effort captures the choice of acquiring information about the state, and the cost of effort may be monetary or psychological. The timing is as follows.

(1) Sender chooses signal S, without knowing the Receiver's *type*, which includes her effort cost and outside option.

(2) Receiver chooses her effort e;

(3) Receiver observes the realization of S with probability e, and observes an uninformative signal with the remaining probability. She chooses action 1 or 0 given her posterior belief.

For instance, let's suppose that a university (Sender) wants its graduates to find employment at a renowned firm (Receiver), regardless of their skills (state), while the firm finds it profitable only to hire high-skill graduates. The university decides how to best advertise its graduates to maximize the probability that the graduates are hired by the renowned firm. The university's marketing policy includes: grading policy, social-media presence, advertisement of graduates' achievements, and so on. There are two main forces that determine the optimal marketing: the university wants the firm to (i) pay attention to the marketing campaign, and (ii) hire the graduates. Paying attention refers to the extensive margin of persuasion: is the firm reached by the marketing efforts? The hiring decision refers to the intensive margin of persuasion: does the firm hire the graduates, given the information acquired from the marketing campaign? The private information of Receiver, in this example, captures the fact that the university is not fully informed about (i) the extent to which the firm is hiring, and the platforms where firms seek job candidates (cost of effort); and (ii) the firm's hiring process, including intervew questions and tests (outside option).

The extensive margin of persuasion arises because Receiver is privately informed about her type. In particular, the sender takes into account that increasing the correlation between the state and the signal has two effects: on Receiver's attention effort e — the extensive margin of persuasion —, and on Receiver's action if she observes the realization of S — the intensive margin of persuasion.

I show the equivalence between persuasion mechanisms and signals. Let's suppose that Sender commits to a persuasion mechanism, which is a menu of signals $S_\bullet$, as opposed to a single signal. Under a persuasion mechanism, Receiver reports a type and chooses an effort level. In particular, Receiver chooses the probability with which to observe the signal from the menu that corresponds to her reported type. A mechanism is incentive-compatible if Receiver finds it optimal to report her type truthfully. For every incentive-compatible persuasion mechanism $S_\bullet$, I construct a signal S that induces the same action and effort distributions over Receiver types (Theorem 1). The key

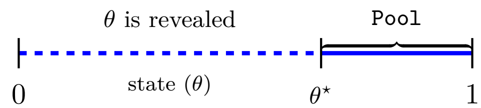

Figure 1: An upper censorship is a signal that reveals states below a cutoff state $\theta^\star$, and sends a single realization, `Pool`, if the state is above the cutoff.

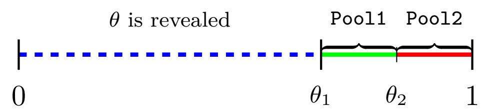

Figure 2: A bi-upper censorship is a signal that reveals low states and separates high from very high states.

is to establish a supermodularity property of type-t Receiver's expected utility: the return from effort is increasing in a t-specific informativeness order, which agrees with Blackwell's order whenever possible. I construct a single signal S that attaches to each Receiver's type the same t informativeness as the incentive-compatible mechanism $S_\bullet$. This result shows that Sender does not need to offer a fine collection of information structures, and allows the study of persuasion to focus on single signals.

I characterize the optimal information structure in commonly-studied applications, which censors high states. An upper censorship is a signal that reveals low states, and censors high states, as in Figure 1. Upper censorships are optimal signals if the Receiver's outside option admits a single-peaked distribution (Theorem 2). Given the equivalence between persuasion mechanisms and signals, we can focus on upper censorships to study the extensive margin of the Sender's persuasion in applications. I apply my results to the problem of media censorship. If Sender knows Receiver's attention cost and has preferences over the extensive margin, inspired by models of media capture à la Gehlbach and Sonin (2014), bi-upper censorships are optimal signals (see Figure 2). I study the effect of changes in Receiver's attention cost on the information provided by the Sender, measured à la Blackwell, through the optimal upper censorship. I do so by isolating the effect of each of the two dimensions of Receiver's private information. Sender provides more information as Receiver's attention cost stochastically increases, if he knows Receiver's outside option (Proposition 2). Moreover, Sender provides more information as the Receiver's attention cost increases, if he knows the attention cost and that cost is sufficiently small (Proposition 3).

Related Literature If the receiver's attention is costless, prior work determines the extent of a sender's intensive margin of persuasion (Kamenica and Gentzkow, 2011; Kolotilin et al., 2017). To study the extensive margin of persuasion, the model of this paper features either the receiver's attention cost, and the receiver's private information. The attention costs lead Receiver to decide whether to become informed, and the private information captures the heterogeneity of attention choice in the audience of a sender. Persuasion of an inattentive receiver has been studied in three models, which do not include Receiver's private information. Wei (2021) studies a receiver who incurs a cost to reduce her uncertainty about the state. Matysková and Montes (2021) study a receiver who acquires costly information about the state from a third party. Differently from these papers, I consider a receiver whose attention cost is not within the rational-inattention paradigm. In the main model of Bloedel and Segal (2021), the receiver bears a cost proportional to the mutual information between the sender's signal and the receiver's signal about the sender's one. In a separate model, the authors study the same cost structure as in my paper.[1] Differently from these models, I include Receiver's private information to study a rich extensive margin of persuasion. The connection with these papers is further discussed in Section 2.

If attention effort is costless, optimality properties of upper-censorship signals are known (Gentzkow and Kamenica, 2016; Kolotilin, 2018; Dworczak and Martini, 2019; Kleiner et al., 2021; Kolotilin et al., 2022; Shishkin, 2019), and the equivalence between persuasion mechanisms and signals is shown by Kolotilin et al. (2017) (see also Guo and Shmaya (2019)). I generalize these results to the case of receiver's costly and privately known attention effort.

The literature on incomplete-information beauty contests studies the supply of Gaussian signals to inattentive receivers.[2] The restriction to Gaussian signals renders many questions about optimal information structures moot. The literature on media capture considers the provision of information to receivers who are privately informed, either about the opportunity cost of supporting an incumbent politician, or about their

[1]Either in Wei (2021) and in the special case of Bloedel and Segal (2021), the analysis assumes that every signal has at most two realizations with positive probability, which is without loss of generality, although for different reasons in the two models. This assumption would imply a loss of generality in my model because Receiver has private information.

[2]Several models characterize the optimal supply of Gaussian signals to inattentive receivers, see Cornand and Heinemann (2008); Chahrour (2014); Myatt and Wallace (2014); Galperti and Trevino (2020); see also Nimark and Pitschner (2019), and references therein, for related models.

attention cost (respectively, Kolotilin et al. (2022) and Gehlbach and Sonin (2014)).[3]

Outline I present the model in the next section. In Section 3.1, I describe the equivalence between persuasion mechanisms and signals. In Section 3.2, I characterize the extensive margin of persuasion. In Section 4.1, I study optimal signals and welfare implications of changes in Receiver's attention cost. In Section 4.2, I discuss implications for the theory of media capture.

2 Model

A Sender (he) and a Receiver (she) play the following persuasion game. Before the *state* $\theta \in \Theta := [0,1]$ is realized, players have a common prior $\mu_0 \in \Delta\Theta$, which admits an absolutely continuous CDF F_0.[4] Receiver's type $t = (\zeta_t, \lambda_t) \in T$, where $T = [0,1]^2$, is distributed independently of the state θ, according to a CDF H. ζ_t is Receiver's threshold type, or outside option, λ_t is Receiver's attention type, or attention cost. Receiver's *material payoff* from taking action $a \in \{0,1\}$ is $u_R(a,\theta,\zeta_t) = a(\theta - \zeta_t)$, when her threshold type is ζ_t and the state is θ. Receiver's effort cost, if her attention *effort* is $e \in [0,1]$, is given by $\lambda_t k(e)$, where k is a continuous function. The Receiver's utility is given by the difference between her material payoff and her effort cost:

$$U_R(a,\theta,e;t) := u_R(a,\theta,\zeta_t) - \lambda_t k(e).$$

Sender always wants Receiver to take action 1, and his utility when Receiver chooses action a is $U_S(a) = a$.

The timing of the game is as follows.

- Sender publicly commits to a signal, which is a measurable function $\sigma : \Theta \to \Delta M$, where M is an exogenous rich space of signal realizations.[5]

- Nature draws Receiver's type t according to H.

- Receiver chooses an effort $e \in [0,1]$, knowing her type t.

[3] See Prat (2015) for a survey of the literature on media censorship.
[4] $\Delta\mathcal{X}$ denotes the set of Borel probability measures over the set $\mathcal{X}$.
[5] It is sufficient that $M = [0,1]$.

- Nature draws the state θ according to μ_0, and a message $m \in M \cup \{\phi\}$. m is drawn from $\sigma(\theta)$ with probability e, and m is equal to ϕ, where $\phi \notin M$, with the remaining probability.

- Receiver observes the message m, and then updates her belief about θ, using Bayes' rule and knowledge of σ and the prior μ_0. Given her *posterior belief*, she chooses an action $a \in \{0, 1\}$.

We analyze Sender-optimal Perfect Bayesian Equilibria of this game, in line with the literature on Bayesian persuasion. We denote by $\overline{F}$ the CDF corresponding to full mass at x_0 which is the prior mean of θ.

Receiver's optimal action and effort Let's describe type-t Receiver's optimal action, given her posterior belief $\mu \in \Delta\Theta$. Letting $t = (\zeta, \lambda)$, the optimal action is 1 if the expected state according to μ, x, exceeds her threshold type ζ, and the optimal action is 0 if the expected state according to μ is such that $x < \zeta$.[6] Thus, Receiver's optimal action depends on belief μ only through its mean $x_\mu := \int_\Theta \theta \, \mathrm{d}\mu(\theta)$. The Receiver's *material payoff at belief* μ is her expected material payoff when her belief is μ:

$$v_t(\mu) := \int_0^1 u_R([x_\mu \geq \zeta], \zeta, \theta) \, \mathrm{d}\mu(\theta),$$

where $[P]$ is the Iverson bracket of the statement P: $[P] = 1$ if the statement P is true, and $[P] = 0$ otherwise. We note that $v_t(\mu)$ depends on the belief μ only through its induced mean x_μ.

Sender's maximization problem After Sender chooses a signal that induces the distribution over posterior beliefs $p \in \Delta\Delta\Theta$, type-t Receiver chooses her effort to maximize her expected utility. In particular, she faces the maximization problem given by

$$\max_{e \in [0,1]} e \int_{\Delta\Theta} v_t(\mu) \, \mathrm{d}p(\mu) + (1-e) v_t(\mu_0) - \lambda_t k(e). \tag{1}$$

[6] We break the Receiver's indifference in favor of Sender. This assumption is without loss of generality given our assumption that H is absolutely continuous. This assumption is necessary for Sender optimality when Sender knows Receiver's threshold type (see, e.g., Gentzkow and Kamenica (2016)).

If k is smooth, the optimal effort is obtained by a simple marginal-cost-marginal-benefit analysis. Type-t Receiver compares the marginal benefit of committing to observing the signal with probability e to the marginal cost of such a commitment. The marginal benefit is the difference between the expected material payoff when Receiver updates her beliefs according to p and the material payoff at the prior belief: $\int_{\Delta\Theta} v_t(\mu)\,\mathrm{d}p(\mu) - v_t(\mu_0)$. We refer to this difference as the marginal benefit of effort at the random posterior p.[7] If k is differentiable, the marginal cost of effort e is given by $\lambda_t \frac{\partial k}{\partial e}(e)$. Since $v_t(\mu)$ depends on belief μ only through its induced mean, the random posterior p influences Receiver's effort decision only through the marginal benefit of effort. In particular, if the signal's informativeness increases in the Blackwell order, the marginal benefit of effort shifts upward for every Receiver's type; while the marginal cost of effort does not change. We denote by $E(p;t)$ the nonempty set of maximizers of the above program, which we study in Section 3.2.

Let's describe the role of the extensive and the intensive margin of persuasion in the Sender's incentives. We use the formalism of random posteriors, as done in the literature on persuasion. Let $\mathcal{R}$ be the set of *feasible random posteriors*: distributions of the Receiver's belief satisfying the martingale condition.[8] We describe the Sender's choice of a feasible random posterior, which is without loss of generality.[9] We define the Sender's payoff at belief μ using Receiver's optimal action as: $V_S(\mu;t) := [x_\mu \geq \zeta_t]$. The Sender's problem is:

$$\sup_{p,e(\cdot)} \int_{\Delta\Theta} \int_T e(t)(V_S(\mu;t) - V_S(\mu_0;t))\,\mathrm{d}H(t)\,\mathrm{d}p(\mu)$$
$$\text{s.t. } p \in \mathcal{R} \text{ and } e(t) \in E(p;t) \text{ for all } t \in T.$$

We decompose the persuasion of a Receiver's type into two terms. The Receiver's optimal action depends only on the mean of the Receiver's belief, which is either

[7]The marginal benefit of effort at a random posterior p is commonly referred to as the value of the information of the signal that induces p.

[8]In particular:

$$\mathcal{R} := \left\{ p \in \Delta\Delta\Theta \ : \ \int_{\Delta\Theta} \mu \mathrm{d}p(\mu) = \mu_0 \right\}.$$

[9]Every signal induces a distribution in $\mathcal{R}$, by the martingale property of Bayesian updating. Moreover, for all $p \in \mathcal{R}$, there exists a signal that induces p as the distribution of the posterior belief; see, e.g., Kamenica and Gentzkow (2011) and Appendix C.2 in Lipnowski and Ravid (2199).

the (random) posterior mean following the information policy p, or the prior mean $\int_\Theta \theta \,\mathrm{d}\mu_0(\theta) =: x_0$. The effort chosen by Receiver is the probability that the mean of the Receiver's belief is the (random) posterior mean following the information policy p. Thus, the Sender's expected payoff depends on the feasible random posterior p in two ways, which can be ascribed to the intensive and the extensive margin of Bayesian persuasion. First, type t acts if the posterior mean is higher than type-t outside option ζ_t. Letting $a^\circ(t) = [x_0 \geq \zeta_t]$ and $a^\star(\mu, t) = [x_\mu \geq \zeta_t]$, and assuming maximum effort, the equilibrium expected action is larger than under an uninformative signal by the following amount:[10]

$$\int_{\Delta\Theta} V_S(\mu; t)\,\mathrm{d}p(\mu) - V_S(\mu_0; t) = \mathbb{E}\{a^\star - a^\circ \mid t,\ e(p, t) = 1\}.$$

Second, each type t has some probability of updating her belief, which is t's effort decision $e(p; t)$. Letting $e(p, t)$ be the effort chosen by type-t Receiver, the expected change in Receiver's action is:

$$\underbrace{e(p, t)}_{\text{extensive margin}} \underbrace{\left(\int_{\Delta\Theta} V_S(\mu; t)\,\mathrm{d}p(\mu) - V_S(\mu_0; t)\right)}_{\text{intensive margin}} = \underbrace{\mathbb{E}\{a^\star - a^\circ \mid t\}}_{\text{persuasion of type } t}.$$

The term $e(p, t)$ captures the extensive margin of persuasion: different posterior distributions may lead to different effort decisions of type t. The second term captures the intensive margin of persuasion: different posterior distributions may lead to different distributions of $a^\star - a^\circ$, given Receiver's type t.

Benchmark cases If infomation is costless, the model is equivalent to persuasion of a privately informed receiver as studied in prior work (e.g., Kolotilin et al. (2017); Kolotilin (2018)). The extensive-margin term in the persuasion decomposition is moot. If information is costly and Receiver's type t is known to Sender, our framework specifies to the model studied by Bloedel and Segal (2021), in which Sender solves

$$\sup_{p,e} \int_{\Delta\Theta} e(V_S(\mu; t) - V_S(\mu_0; t))\,\mathrm{d}p(\mu) \tag{2}$$

$$\text{s.t. } p \in \mathcal{R} \text{ and } e \in E(p; t). \tag{3}$$

[10] The following conditional expectation given t is taken with respect to the random posterior that is distributed according to p.

As Bloedel and Segal (2021) observe, we can use a first-order approach when k is sufficiently smooth; moreover, there exists an optimal signal that is a binary signal by a revelation-principle argument. The problem in Equation 2 is similar to that studied in the *attention-management* literature (Lipnowski et al., 2020, 2022a; Wei, 2021). If we assume that the attention-management Sender wants Receiver to take action 1 regardless of the state, the maximization in 2 is a *constrained* version of the attention-management one.[11] In particular, in our model, Receiver effectively chooses an element from a specific set of garblings of the posterior: the mixtures of the Sender's signal and an uninformative signal. In the attention-management literature, Receiver's choice of garbling is unrestricted.

Information policies Receiver chooses her effort to maximize her expected utility (Problem 1), and the marginal benefit of effort depends on the random posterior p only through the distribution of posterior means. Thus, we identify a feasible random posterior with the induced posterior mean distribution, and here we formalize this representation (similarly to, e.g., Gentzkow and Kamenica (2016)). Let $\mathcal{D}$ be the collection of CDF's over $[0, 1]$. A CDF F is feasible if it represents the posterior mean distribution of a feasible random posterior. By Blackwell's theorem, a CDF F is feasible if, and only if: F is a mean preserving contraction of F_0. Let's define the *information policy of a CDF* $F \in \mathcal{D}$ as:

$$\begin{aligned} I_F \colon \mathbb{R}_+ &\to \mathbb{R}_+ \\ x &\mapsto \int_0^x F(y)\,\mathrm{d}y. \end{aligned}$$

The information policy of a feasible F, I_F, is upper bounded pointwise by F_0, due to Blackwell's theorem. I_F is lower bounded pointwise by $\overline{F}$, because the uninformative signal does not change the mean of the receiver's belief. Moreover, I_F is convex because F is nondecreasing. These are the only three constraints on feasible information policies, so we identify a feasible random posterior with its induced information policy (Gentzkow

[11] Sender wants Receiver to take action 1 regardless of the state in Wei (2021), while Sender maximizes Receiver's material payoff in Lipnowski et al. (2020, 2022a). The optimal signal for a Sender who wants action 1 regardless of the state is not characterized in attention management, except in the binary-state case (Wei (2021)).

and Kamenica, 2016). The set of feasible information policies is:

$$\mathcal{I} := \{I \colon \mathbb{R}_+ \to \mathbb{R}_+ \mid I \text{ is convex and } I_{F_0}(x) \geq I(x) \geq I_{\overline{F}}(x) \text{ for all } x \in \mathbb{R}_+\}.$$

We analyze the Sender's problem as a choice of an information policy $I \in \mathcal{I}$. There are two reasons why this choice of formalism pays off. First, I is a measure of the Blackwell informativeness of the corresponding signal. In particular, σ is a more informative signal than τ if, and only if, $I_\sigma(x) \geq I_\tau(x)$, $x \in [0,1]$, where I_S denotes the information policy corresponding to the posterior mean's CDF induced by signal S. Thus, the pointwise ranking of information policies correspond to Blackwell's information order. Second, information policies offer a tractable characterization of optimal effort, as the next Lemma shows.

Preliminary results Receiver chooses effort by comparing her payoff from updating her belief and her payoff from remaining uninformed. Let's develop notation to deal with this comparison. We define the operator Δ as:

$$\Delta \colon I \mapsto I - I_{\overline{F}}.$$

We denote by ΔI the composite function $\Delta(I)$. For the information policy I, ΔI is a measure of the "net" informativeness, where the Blackwell's informativeness of the uniformative signal, given by $I_{\overline{F}}$, is used as a benchmark. We characterize Receiver's marginal benefit of effort in terms of the Sender's information policy I. For an information policy I, we let $I'(x)$ denote the right derivative evaluated at x, which is the value attained by a CDF evaluated at x, and $I'(x^-)$ denote its left derivative.

Lemma 1 (Marginal Benefit of Effort)**.** Receiver's marginal benefit of effort given the information policy I and her type (ζ, λ) is:

$$\int_0^1 u_R([x \geq \zeta], x, \zeta) \,\mathrm{d}I'(x) - \int_0^1 u_R([x \geq \zeta], x, \zeta) \,\mathrm{d}I'_{\overline{F}}(x) = \Delta I(\zeta).$$

Proof. For an information policy $I \in \mathcal{I}$:

$$\begin{aligned}\int_0^1 u_R([x \geq \zeta], x, \zeta) \,\mathrm{d}I'(x) &= \int_\zeta^1 x - \zeta \,\mathrm{d}I'(x) \\ &= x_0 - \zeta + I(\zeta)\end{aligned}$$

The second equality follows from Riemann-Stjeltes integration by parts, using $I'(1) = 1$ and $I(1) = 1 - x_0$. ■

The marginal benefit of effort is increasing in the informativeness of I, measured à la Blackwell. We define type-(c, λ) indirect utility V at the information policy I

$$V(\Delta I(c); \lambda) = \max_{e \in [0,1]} e\Delta I(c) - \lambda k(e) - v_t(\mu_0),$$

and we refer to $V(\Delta I(c); \lambda)$ as the value of information policy I to type (c, λ). The optimal effort of type (c, λ), given information policy I, is an element of $E(\Delta I(c); \lambda)$, where:

$$E(\Delta I(c); \lambda) = \arg\max_{e \in [0,1]} e\Delta I(c) - \lambda k(e).$$

Type-t Receiver's payoff is increasing in the Blackwell information of the Sender's signal, by Blackwell's theorem. Thus, type-t Receiver's value of information policy$I$ is increasing in the informativeness of I. This fact arises as an implication of monotone comparative statics and the envelope theorem, stated in the next Lemma.

Lemma 2. Type-t Receiver's value of information $V(\Delta I(\zeta_t), \lambda_t)$ is a nondecreasing, absolutely continuous and convex function of $\Delta I(\zeta_t)$.

Proof. We observe that $f\colon (e, \Delta I(\zeta_t)) \mapsto e\Delta I(\zeta_t)$ is supermodular. The result follows from the envelope theorem for supermodular optimization (Fact 2 in the Appendix). ■

Unsurprisingly, this result states that Receiver's payoff is increasing in the Blackwell information of the Sender's information policy. However, Blackwell's order is incomplete. We leverage the Lemma to construct a natural type-specific completion of the Blackwell order over information policies. Let's construct a t-specific informativeness order over information policies: $\preceq_t$ over $\mathcal{I}$, such that

$$J \preceq_t I \quad \text{iff} \quad \Delta J(\zeta_t) \leq \Delta I(\zeta_t), \quad \text{for every } J, I \in \mathcal{I}.$$

$\preceq_t$ is a complete order that agrees with Blackwell's order whenever possible. $\preceq_t$ is a local informativeness measure that is a sufficient to caracterize a type-t Receiver's behavior. To prove Lemma 2, we leverage the fact that type-t Receiver's expected utility is supermodular in informativeness and effort, ordering informativeness by $\preceq_t$. In the next section we leverage this observation to prove a strong equivalence between persuasion mechanisms and signals, and in the following section we leverage this observation to study the extensive margin of Bayesian persuasion.

3 Main Results

3.1 Equivalence of Persuasion Mechanisms and Signals

In this section, we consider a more general setup than the previous model. We expand the Sender's strategy space, to include "menues" of signals. We ask whether Sender attains a larger payoff by committing to such a menu, so that each Recever's type self-selects into her preferred signal, than by choosing a single information policy.

A *persuasion mechanism* is a collection of information policies: $(I_r)_{r\in T}$, where $I_r \in \mathcal{I}$ for all reports $r \in T$. We refer to a persuasion mechanism as $I_\bullet$, omitting the reference to reports. A persuasion mechanism $I_\bullet$ is *incentive compatible* (IC) if:

$$V(\Delta I_t(\zeta_t), \lambda_t) \geq V(\Delta I_r(\zeta_t), \lambda_t), \quad \text{for all types } t \in T \text{ and reports } r \in T.$$

We interpret a persuasion mechanism as a rule allocating a signal to every report of Receiver's type. Thus, a persuasion mechanism is IC if it is optimal for Receiver to report her type truthfully. In particular, after her report r, Receiver chooses an effort optimally given the information policy I_r. By a revelation-principle arguments, a Sender who commits to a persuasion mechanisms can, without loss of optimality, commit to an IC persuasion mechanism. We focus on IC persuasion mechanisms in what follows.

The following two definitions characterize a notion of equivalence between an IC persuasion mechanism $I_\bullet$ and a single information policy J. An IC persuasion mechanism $I_\bullet$ and an information policy J induce the same effort and action distributions if the following two conditions hold.

(1)

$$E(\Delta I_t(\zeta_t); \lambda_t) \subseteq E(\Delta J(\zeta_t); \lambda_t), \quad \text{for all } t \in T. \tag{4}$$

(2)

$$I_t'(\zeta_t^-) = J'(\zeta_t^-), \quad \text{for all } t \in T \text{ such that } E(\Delta I(\zeta_t); \lambda_t) \cap (0,1] \neq \emptyset.$$

The following result allows us to study effort and action distributions of persuasion mechanisms via information policies, thus bypassing the screening problem.

Theorem 1. *For every IC persuasion mechanism $I_\bullet$ there exists an information policy J such that: $I_\bullet$ and J induce the same effort and action distributions.*

Proof. Section B in the Appendix. ■

There is a simple intuition for this result, which leverages the local informativeness order $\preceq_t$. $\preceq_t$ determines the choice of type-t Receiver from the collection of information policies of a persuasion mechanism $I_\bullet$, by Lemma 2. Letting c be t's threshold type, we know that type t chooses that information policy $I_\star$ from the mechanism $(I_r)_{r\in T}$ such that $\Delta I_\star(c) \geq \Delta I_r(c)$ for every $r \in T$. It is readily established that $J := \sup_{r\in T} I_r$ is a feasible information policy: J is pointwise bounded by I_{F_0} and $I_{\overline{F}}$ because I_r is an information policy for every $r \in T$, and J is convex because J the pointwise supremum of convex functions. In the proof we show that J replicates effort and action decisions of every type t given the IC mechanism $(I_r)_{r\in T}$.

In the next section we study Sender's optimization by choice of a single information policy. In light of Theorem 1, the following results are relevant to the study of persuasion mechanisms. In particular, a takeaway of Theorem 1 is that single information policies are without loss of generality for welfare analysis.

Remark 1. Let's recall that type-t Receiver's expected utility is supermodular in informativeness, as orderd by $\preceq_t$, and effort. As a confirmation that supermodularity the key intuition for Theorem 1, in Section B of the Appendix, we prove the result assuming supermodular Receiver's interim payoff, as a function of informativeness and effort. This specification nests the original model where Receiver's ex-post utility is given by U_R.

3.2 Characterization of the extensive margin

We characterize effort decisions assuming smoothness conditions on k.

Assumption 1 (Smooth Effort Cost)**.** k is a differentiable convex function on $[0,1]$, and satisfies: $k'(1) > 1 - x_0$.

Under Assumption 1, we denote by $k'(e)$ the derivative of k at effort e.

Lemma 3. Let Assumption 1 hold, $I \in \mathcal{I}$, and e_t be any element of $E(\Delta I(\zeta_t), \lambda_t)$.

$$\Delta I(\zeta_t) \leq \lambda_t k'(e_t),$$

with equality if $e_t > 0$.

Proof. Receiver maximizes a concave function over a compact set. A solution exists and, by differentiability of k, we can use standard Lagrangean arguments to show that it has the prescribed form, so long as optimal effort is in $[0,1)$. Let's see that the requirement that $k'(1) > 1 - x_0$ assumes away boundary solutions at 1. Because $I(1) = 1 - x_0$, we have $\Delta I(\zeta) \leq 1 - x_0 < k(1)$, for every $\zeta \in [0,1]$. ■

The marginal-cost-marginal-benefit analysis of Receiver's effort decision is depicted in Figure 3. Net informativeness ΔI defines a continuous function of Receiver's outside option, with a peak at the cutoff type that is equal to the prior mean x_0. The proof of this statement is in the Appendix (Lemma 6). The intuition for single-peakedness comes from the observation that the marginal benefit of effort is $\Delta I(\zeta)$, given I (Lemma 1). Type x_0 finds it valuable to observe any signal about θ, in order to make a more informed choice than if she is left at the prior. Extreme types have, instead, the least to gain from committing to observe a signal: the ex-ante probability that a signal realization modifies t's optimal action is low because only extreme realizations of the state are pivotal for their optimal action. If we intersect the marginal benefit of effort ΔI with $\lambda_t k'(0)$, we observe that, in general, intermediate types will exert a positive effort, and extreme types will not exert any effort. This result is depicted in figure 3, and implies that the set of Receiver types who become informed is defined by the two cutoff types who are just indifferent between exerting positive effort and not exerting any effort. Under Assumption 1, the cutoff types given information policy I and attention cost λ are:

$$\underline{c}^{\lambda}(\Delta I) := \min\{c \in [0,1] \mid \Delta I(c) \geq \lambda k'(0)\},$$
$$\overline{c}^{\lambda}(\Delta I) := \max\{c \in [0,1] \mid \Delta I(c) \geq \lambda k'(0)\},$$

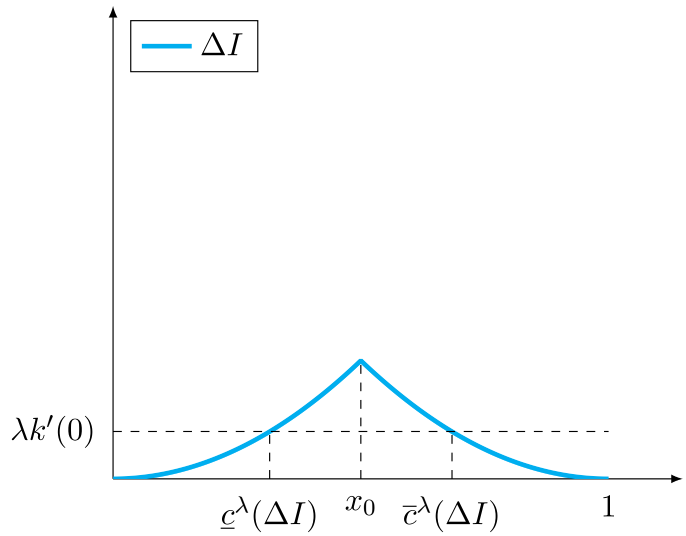

Figure 3: The set of types c who exert positive attention effort is an interval.

with the requirement that if either of the two sets is empty, the relevant cutoff type is x_0. The next observation is that the interval shape of the extensive margin generalizes. By the supermodularity property of the Receiver's value function, type-t's optimal effort is nondecreasing in $\Delta I(\zeta_t)$ without differentiability hypotheses.

Proposition 1. *Let I be an information policy, and $e(c, \lambda) \in E(\Delta I(c), \lambda)$ for every type (c, λ). Then, $e(\cdot, \lambda)$ is single-peaked, with a peak at x_0.*

Proof. We observe that $f : (e, \Delta I(\zeta_t)) \mapsto e\Delta I(\zeta_t)$ is strictly supermodular. Thus, by Lemma 1, we establish that every selection from the optimal effort correspondence is nondecreasing, using monotone comparative statics results (Fact 1 in the Appendix). The result follows from single-peakedness of ΔI, with a peak at x_0, established in Lemma 6 in the Appendix. ■

By our result, every selection from the optimal effort correspondence exhibits cutoff outside-option types, given an attention cost type. In particular, extreme types — whose outside option is above or below the cutoffs — do not exert any effort.

3.3 Sender's Value of an Information Policy

We now express the Sender's problem as a maximization by choice of a feasible information policy, using the previous results on the extensive margin. First, we describe the extensive and intensive margins.

Lemma 4. The Sender's maximization problem is given by:

$$\sup -\int_T e(t)\Delta I'(\zeta_t^-)\,\mathrm{d}H(t) \tag{5}$$

$$\text{s.t. } I \in \mathcal{I} \text{ and } e(t) \in E(\Delta I(\zeta_t), \lambda_t) \text{ for all } t \in T. \tag{6}$$

Proof. Letting G be the marginal CDF of information cost consistent with H. Sender's value of I, given $e(\zeta,\lambda) \in E(\Delta I(\zeta),\lambda)$ for all $(\zeta,\lambda) \in T$, is:

$$\int_{[0,1]}\int_{[0,1]}(1 - I'(\zeta^-) - [\zeta \leq x_0])e(\zeta,\lambda)\,\mathrm{d}H(\zeta|\lambda) + (1 - H(x_0|\lambda))\,\mathrm{d}G(\lambda),$$

because type-t Receiver chooses action 1 when indifferent, by Sender-optimality. The Lemma follows after normalizing Sender's expected payoff from $I_{\overline{F}}$ to 0, and the observation that: $1 - [\zeta_t \leq x_0] = I'_{\overline{F}}(\zeta^-)$. ■

The Sender's value of the information policy I depends on I in two ways: the intensive and the extensive margin of Bayesian Persuasion. First, the probability that threshold type c chooses action 1 is the probability that the posterior mean is higher than her outside option c: $1 - I'(c^-)$. We note that what matters for Sender is not the probability of action 1, but the degree to which the information policy changes the prior action decisions towards action 1. Thus, to $1 - I'(c^-)$ we subtract $[c < x_0]$, and we note that: $1 - I'(c^-) - [c < x_0] = -\Delta I'(\zeta^-)$. Second, a Receiver's type updates her beliefs with probability equals to her effort decision. The next result re-writes the Sender's problem in way to shows that Sender is effectively allocating information to Receiver's types, without smoothness assumptions.

Lemma 5. Let H admit a PDF h that is decomposed as: $h(\zeta,\lambda) = h_{\zeta|\lambda}(\zeta|\lambda)h_\lambda(\lambda)$, and let the conditional PDF $h_{\zeta|\lambda}$ admit a derivative with respect to ζ, $h'_{\zeta|\lambda}$. The Sender's maximization problem is given by:

$$\max_{I\in\mathcal{I}} -\int_{[0,1]}\int_{[0,1]} V(\Delta I(\zeta),\lambda)h'_{\zeta|\lambda}(\zeta|\lambda)h_\lambda(\lambda)\,\mathrm{d}\zeta\,\mathrm{d}\lambda. \tag{7}$$

Proof. See Section C in the Appendix. ■

Under the hypotheses of this Lemma, we define the Sender's value of an information policy I as $v(I)$, which is the maximand in the optimization above. And we say that an information policy I is optimal if it solves the maximization in 7. From the above result, we know that Sender prefers to allocate (Blackwell) informativeness to a Receiver's type (ζ, λ) so long as the measure induced by $h'_{\zeta|\lambda}(\zeta|\lambda)h_\lambda(\lambda)$ is positive, and he prefers to not allocate informativeness to types such that $h'_{\zeta|\lambda}(\zeta|\lambda)h_\lambda(\lambda)$ is negative. In the next section, we make use of this intuition to solve for the optimal signal in applied models.

Remark 2. Shishkin (2019) uses a similar information-allocation intuition, in a model without the extensive margin because Receiver's attention is costless.

4 Applications

4.1 Single-Peaked Distribution of Receiver's Outside Option

In applications, it's common to assume that the distribution of Receiver's outside option is single-peaked (Shishkin (2019); Gitmez and Molavi (2020), and also particular cases considered by Kolotilin (2018); Lipnowski et al. (2022b)).

Assumption 2 (Single-Peakedness of Outside Option Distribution)**.** (1) The attention cost λ is independent of threshold c, and distributed according to the CDF H.

(2) The distribution of the threshold ζ admits an absolutely continuous quasiconcave PDF f, with CDF F.

4.1.1 Optimality of Upper Censorships

Under Assumption 2, the two dimensions of Receiver's type are independently distributed and $h'_{\zeta|\lambda}(\zeta|\lambda)$ is nonpositive before a peak threshold type, and nonnegative after the peak (Lemma 5). Thus, it is optimal to reveal a lot of information through low posterior means, and not much information through high posterior means. There exists a class of signals that achieve this "information allocation," the class of upper-censorship signals. An upper-censorship signal implies full revelation conditional on the state being lower than a cutoff, and full censorship conditional on the state being above the cutoff (Figure 1). Since we work in the space of information policies, we define an upper censorship as an information policy which is induced by an upper-censorship signal.

Definition 1. The $\theta^\star$ *upper censorship* is the unique information policy $I_{\theta^\star} \in \mathcal{I}$ such that:

$$I_{\theta^\star}(x) = \begin{cases} I_{F_0}(x) & , x \in [0, \theta^\star] \\ I_{F_0}(\theta^\star) + (x - \theta^\star)F_0(\theta^\star) & , x \in (\theta^\star, x^\star] \\ I_{\overline{F}}(x) & , x \in (x^\star, \infty), \end{cases}$$

where $x^\star = \int_{\theta^\star}^1 \theta \,\mathrm{d}\frac{F_0(\theta)}{1-F_0(\theta^\star)}$.

The next result shows that upper censorships are optimal information policies under independent cost and threshold types, whenever threshold types are single-peakedly distributed.

Theorem 2. *Let Assumption 2 hold. There exists a $\theta^\star \in \Theta$ such that the $\theta^\star$ upper censorship is an optimal information policy.*

Proof. Section C.1 in the Appendix ■

A reading of this result is as a revelation principle result. In particular, in order to maximize Sender's payoff, Sender can focus on upper censorships under single-peakedness assumptions. A similar result can be proved using single-dipped distributions, where "lower censorships" arise as optimal signals. In light of Theorem 1, we know that the study of upper censorships informs us about properties of persuasion mechanisms. Theorem 1 nests many known results about the optimality of upper censorships in models without Receiver's private information, or effort cost. The next remark discusses uniqueness issues.

4.1.2 Welfare Analysis

Does Receiver benefit from her attention cost? In particular, Receiver's inattentiveness may act as a bargaining power: Sender is forced to increase the informativeness of his signal to induce Receiver to pay attention. This observation holds without Receiver's private information, as we establish in the Appendix (Section C.3). In this section, we assume that k is linear: $k(e) = \kappa e$, $\kappa > 0$. In order to isolate the effect of each of the two dimensions of Receiver's private information, we ask whether Receiver is better off as her attention cost λ increases in two separate cases: (1) when Sender knows Receiver's outside option, and (2) when Sender knows Receiver's attention cost. In the first case, we study an increase of the distribution of the attention cost in a stochastic order.

The next result characterizes the optimal upper censorship with known action threshold.

Proposition 2. *Let the distribution of attention cost admit a log-concave CDF with support $[0, \bar{\kappa}]$, and a continuous PDF, with $\bar{\kappa} > 1 - x_0$, and the outside option be known to Sender. Moreover, let's assume that k is linear: $k(e) = \kappa e$. There exists a solution to the Sender's problem that is a θ upper censorship, where $\theta \in \{0, \theta^\star, \zeta\}$ and $\theta^\star$ solves:*

$$(1 - F_0(\theta))(\zeta - \theta) = (\chi(\Delta I_\theta(\zeta)))^{-1}$$

Proof. Without loss of generality: $\kappa = 1$ and

$$K(\lambda) = \exp\biggl(-\int_\lambda^{\bar{\kappa}} \chi(t)\,\mathrm{d}t\biggr),$$

for some $\chi : (0, \bar{\kappa}) \to \mathbb{R}_+$ such that: $\int_\lambda^{\bar{\kappa}} \chi(t)\,\mathrm{d}t < \infty$ and $\lim_{\lambda \to 0} \int_\lambda^{\bar{\kappa}} \chi(t)\,\mathrm{d}t = \infty$. Our primitive is the nonincreasing reverse hazard rate χ. Without loss of optimality, Sender maximizes his payoff by choice of an upper censorship (see Section C of the Appendix). In particular, Sender's optimization is:

$$\max_{I \in \mathcal{I}^u} (1 - I'(\zeta^-))K(\Delta I(\zeta)),$$

where $\mathcal{I}^u \subseteq \mathcal{I}$ is the collection of upper censorships. Suppose there exists a solution $I \in \mathcal{I}^u$, such that $I = I_{\theta^\star}$, for some $\theta^\star \in (0,1)$. By definition of I, at $y = I(\zeta)$ the next condition holds:

$$I_{F_0}(\theta^\star) + F_0(\theta^\star)(\zeta - \theta^\star) - y = 0.$$

By the implicit function theorem, there exists a differentiable function t:

$$\begin{aligned} t\colon (0,1) &\to (0,1) \\ y &\mapsto \theta^\star, \end{aligned}$$

such that:

$$t'(y) = \begin{cases} \frac{1}{(\zeta - t(y))F_0'(t(y))} & , 0 < \zeta < t(y) \\ \frac{1}{F_0'(t(y))} & , 1 > \zeta \geq t(y). \end{cases}$$

Let the value of I_θ be:

$$\begin{aligned} v \colon (0,1) &\to [0,1] \\ \theta &\mapsto (1 - I_\theta'(\zeta^-))K(\Delta I_\theta(\zeta)). \end{aligned}$$

Since $I'_{\theta^\star}(\zeta^-) = F_0(\theta^\star)$, v is differentiable in θ at $\theta^\star$. Let $\zeta > \theta^\star$. Using the chain rule, the derivative of v with respect to $I(\zeta)$ is nonnegative if:

$$(1 - F_0(\theta^\star))(\zeta - \theta^\star) \geq (\chi(\Delta I_{\theta^\star}(\zeta)))^{-1},$$

and nonpositive if:

$$(1 - F_0(\theta^\star))(\zeta - \theta^\star) \leq (\chi(\Delta I_{\theta^\star}(\zeta)))^{-1}.$$

The function $\theta \mapsto (1 - F_0(\theta))(\zeta - \theta)$ is decreasing on $(0, \zeta)$, and the function $\theta \mapsto \chi(\Delta I_\theta(\zeta))$ is decreasing on $(0, \zeta)$. As a result, finding the optimal $\theta^\star$ upper censorship is a concave program. ■

As a corollary to the above result, let full-information and no-information not be optimal information policies when the reverse hazard rate is χ_1 and when it is χ_2. The optimal upper censorship under χ_1 has a lower censorship point than the optimal upper censorship under χ_2 if $\chi_1(\lambda) \leq \chi_2(\lambda)$ for all λ. Thus, if attention cost stochastically increases (in the reverse hazard rate order), the optimal upper censorship is more Blackwell informative. This result is consistent with the symmetric-information special case of our model, where Sender knows Receiver's type.

Is it a robust feature of Bayesian persuasion that information costs increase Receiver's information? Let's consider the case of symmetric information (about Receiver's type, in particular). As pointed out by Wei (2021); Bloedel and Segal (2021) and Matysková and Montes (2021), the answer is no. In particular, under mutual-information cost the informativeness of Sender's signal is nonmonotone in the commonly known information-cost parameter. To see this, consider the following two extremes. If information is costless, Sender uses a partially informative signal, as established in the literature. If information is prohibitevely costly, only poorly informative signals induce Receiver to acquire some information. Our result uncovers a natural way to model larger cost in a stochastic sense to maintain symmetric-information comparative statics, by using the reverse-hazard rate dominance order.

Does Receiver benefit from small, known, attention cost? The next result shows that Receiver benefits from a small, known, cost when she is privately informed about her belief threshold for action.

Proposition 3. *Let Assumption 2 hold, attention cost be known to Sender, and f be strictly single-peaked. For a sufficiently small $\varepsilon > 0$:*

(1) There exists a unique optimal upper censorship when $\lambda = 0$: I°;

(2) There exists a unique optimal upper censorship when $\lambda = \varepsilon$: $I^{\star}$;

(3) $I^{\star}$ is more Blackwell informative than I°, that is: $I^{\circ}(x) \leq I^{\star}(x)$, $x \in \mathbb{R}_+$.

Proof. By Lemma 5, the derivative of the Sender's value of the $\overline{\theta}$ upper censorship with respect to the $\overline{\theta}$ is:

$$\partial F_0/\partial\theta(\overline{\theta}) \int_{\overline{\theta}}^{\overline{c}^{\lambda}} (x - \overline{\theta})\partial h/\partial\zeta(x)\,\mathrm{d}x \leq \partial F_0/\partial\theta(\overline{\theta}) \int_{\overline{\theta}}^{1} (x - \overline{\theta})\partial h/\partial\zeta(x)\,\mathrm{d}x,$$

where the inequality is strict if λ is sufficiently small, because h is decreasing on $(p, 1]$, $p < 1$. The right-hand side of the inequality is the derivative of the Sender's value of the $\overline{\theta}$ upper censorship with respect to the $\overline{\theta}$ when $\lambda = 0$. Because h is increasing on $[0, p)$ and on $(p, 1]$, $p < 1$, and ε is small, both sides of the above inequality are decreasing in $\overline{\theta}$. As a result, there exists a unique optimal upper censorship either when $\lambda = 0$, and when $\lambda = \varepsilon$.[12] ■

In Wei (2021), Receiver is better off with $\varepsilon > 0$ costs than with 0 costs, due to the "bargaining-power" idea described above. In Matysková and Montes (2021), Receiver is worse off with $\varepsilon > 0$ costs than with 0 costs, because Receiver becomes arbitrarily informed at almost 0 cost. So, the welfare analysis of attention cost is dependent on the information-cost model.

[12]Uniqueness with costless information is readily established also using the tools from Kolotilin (2018); Kolotilin et al. (2022).

4.2 Media Censorship

Let's suppose a government wants people to stay home, and has control over the media. If media start to suggest to stay home, for instance by showing how bad a pandemic situation is, individuals may change their behavior only so long as they paid attention to the media. Thus the government must take into account that releasing information has two effects: information impacts behavior if individuals are attentive, a change in the intensive margin of persuasion; information determines attention decisions, a change in the extensive margin of persuasion. We introduce an advertising market á la Gehlbach and Sonin (2014) in the model of Kolotilin et al. (2022), thus providing a bridge between the two approaches to model media censorship.

Assumption 3 (Media Censorship)**.** We assume that:

(1) Sender knows Receiver's attention-cost type λ.

(2) k is linear, and Assumption 2 (Single-Peakedness of Outside Option) holds.

(3) Sender's payoff is given by:

$$U_G(a, e; \cdot) = \psi a + \gamma e.$$

Part (1) is isomorphic to binary effort decision, which is the assumption in Gehlbach and Sonin (2014). Part (3) makes our model's Sender correspond to the government of the media censorship model of Gehlbach and Sonin (2014). Part (2) is made for tractability. ψ captures the mobilizing character of the government. A larger mobilizing character implies that Sender cares more about the support from the population of agents, where an agent is identified by a Receiver's type. γ captures the size of the media market. A larger market implies that Sender cares more about providing information, due to larger advertising revenues. If $\gamma = 0$, we know that an upper censorship is optimal, be previous results. Let's recall that an upper censorship leads to poorly informative large posterior means. Thus, because Sender with $\gamma > 0$ cares more about the extensive margin, it may be suboptimal to provide so little information as an upper censorship does. Let's define a class of information policies that nests upper censorships.

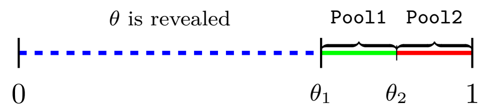

Figure 4: A bi-upper censorship is a signal that reveals low states and separates high from very high states.

Definition 2. A *bi-upper censorship* is an information policy I such that, for $\theta_1, \theta_2 \in \Theta$:

$$I_{\theta^\star}(x) = \begin{cases} I_{F_0}(x) & , x \in [0, \theta_1] \\ I_{F_0}(\theta_1) + (x - \theta_1)F_0(\theta_1) & , x \in (\theta_1, x_1] \\ I_{F_0}(x_1) + (x - x_1)F_0(x_1) & , x \in (x_1, x_2] \\ I_{\underline{F}}(x) & , x \in (x_2, \infty), \end{cases}$$

where $x_1 = \int_{\theta_1}^{\theta_2} \theta \, \mathrm{d}\frac{F_0(\theta)}{F_0(\theta_2) - F_0(\theta_1)}$, $x_2 = \int_{\theta_2}^{1} \theta \, \mathrm{d}\frac{F_0(\theta)}{1 - F_0(\theta_2)}$.

Proposition 4. *Let Assumption 3 hold, and the peak of the PDF of outside option be p, with $p \geq x_0$. There exists an optimal signal that is a bi-upper censorship.*

Proof. The definition of a bi-upper censorship and the proof are in Section C.2. ∎

Figure 4.2 depicts a bi-upper censorship. We can interpret the assumption that the peak of the PDF of ζ is sufficiently large as a sufficiently high disagreement between Sender and the ex-ante Receiver (Shishkin (2019)). The current results nests the media-censorship result in Kolotilin et al. (2022), which shows that an upper censorship is an optimal signal without attention cost.

5 Discussion

Why do people have different information? One possibility is that information providers target some individuals and not others, with the consequence that some receive precise information, and others poor information. This papers investigates this phenomenon. We show that the study of targeting mechanisms can be done using single information policies (Section 3.1). We show that intermediate outside-option types are attentive, while extreme outside-option types exert no attention effort (Section 3.2). This finding points at a possible mechanism for persistence of polarization: extreme positions make Bayesian updating less profitable, and so information providers may not target individuals with extreme opinions. We leverage our results to generalize optimality properties of upper-censorship signals (Section 4.1), to show that Receiver may benefit from her attention cost (Section 4.1.2), and to advance the study of media capture (Section 4.2).

The current model can be extended to study more general cost structures of information, especially in light of an application to central banking. Today, we are amidst a second wave of the communications revolution in central banking: the inclusion of laypeople in the audience of central bank communication (Haldane et al. (2020)).[13] A persuasion mechanism (Section 3.1) captures the strategy used by the European Central Bank and the Bank of England.[14] These institutions use "layered communication:" reports with a simpler, and less precise, part, and a complex, and more precise, part. The idea is that experts and non-experts self-select into the appropriate information. For this application of my model, Theorem 1 has a dual interpretation. On the one hand, it allows to study the extensive margin of layered communication using a fictitious simple, one-layer, communication. On the other hand, it suggests that layered communication has limited advantages. However, my model of attention effort does not capture the richness of real-world information-processing limits. In particular, more Blackwell information implies an increase in the marginal benefit of effort, but has no impact on marginal cost of effort. Thus, by monotone comparative statics, a more informative signal induces every Receiver's type to exert more effort in the model. This observation permits to study the extensive margin of persuasion of a privately informed receiver, at the cost of possibly reducing the richness of the results. A more realistic study of layered communication extends the cost structure of the current paper, so that additional information has an impact on both costs and benefits of information.

[13] Recent evidence asserts that some central banks have successfully reached the general public (Ehrmann and Wabitsch (2022)).

[14] "[...] The Governing Council decided to complement its monetary policy communication with "layered" communication [...] A new visual monetary policy statement was added to explain the ECB's latest decision in a more attractive and simpler format, and in all 24 official EU languages. Using storytelling techniques, relatable visuals and language, the visual monetary policy statement aims to make the ECB's policy decisions more accessible to non-expert audiences across the entire euro area." From ECB Economic Bulletin (November 2021). See Haldane and McMahon (2018), for a discussion of the strategy of the Bank of England.

A Mathematical preliminaries

In this section, we provide background results that we use in the next sections.

Setup $\Delta\mathcal{X}$ is the set of all Borel probability measures on the set $\mathcal{X}$. All mappings are assumed measurable with respect to the relevant Borel σ-algebra. The function $g\colon \mathbb{R}^2 \to \mathbb{R}$ has *increasing differences* if $g(a', e') - g(a', e) \geq g(a, e') - g(a, e)$ whenever $a' \geq a$ and $e' \geq e$; g has *strictly increasing differences* if $g(a', e') - g(a', e) > g(a, e') - g(a, e)$ whenever $a' > a$ and $e' > e$. Let's fix two Euclidean spaces A and B, a *selection from the correspondence* $\beta\colon A \rightrightarrows B$ is a function $f\colon A \to B$ such that $f(x) \in \beta(x)$, $x \in A$. The strong set order on 2^B is defined as: $S \geq_{ss} T$ if $\max\{x, y\} \in S$ and $\min\{x, y\} \in T$, for all $(x, y) \in S \times T$, and $S, T \in 2^B$. We use the following monotonicity notions for a correspondence $\beta\colon A \rightrightarrows B$.

(1) β is *strong set monotonic* if: $a' > a$ implies $\beta(a') \geq_{ss} \beta(a)$, for all $a, a' \in A$.[15]

(2) β is *nondecreasing* if: $a' > a$ implies $b' \geq b$, for all $b \in \beta(a)$, $b' \in \beta(a')$ and $a, a' \in A$.[16]

(3) β is *increasing* if: $a' > a$ implies $b' > b$, for all $b \in \beta(a)$, $b' \in \beta(a')$ and $a, a' \in A$.

We state a result from monotone comparative statics and a version of the envelope theorem.

Fact 1 (Monotone Comparative Statics, MCS)**.** Let $g\colon [0,1]^2 \to \mathbb{R}$ be continuous and exhibit increasing differences. Let $\beta\colon [0,1] \rightrightarrows [0,1]$ such that $\beta(a) = \arg\max_{e\in[0,1]} g(a, e)$, $a \in [0,1]$.

(1) β is nonempty-valued, compact-valued and strong set monotonic.

(2) $\sup \beta$ is a nondecreasing selection of β.

(3) If g exhibits strictly increasing differences, then β is nondecreasing, or, equivalently, every selection from β is nondecreasing.

Proof. For a proof, see Vives (1990). ■

[15] It is common to call "increasing", or "increasing in the strong set order", a correspondence β that is *strong set monotonic* here.

[16] It is common to call "strongly increasing", or "monotone", a correspondence β that is *nondecreasing* here.

Fact 2 (Envelope Theorem, ET)**.** Let $g\colon \mathbb{R}^2 \to \mathbb{R}$ be continuous and exhibit increasing differences. Let the derivative with respect to the variable a, $\partial g/\partial a(\cdot, e)$, exist, be bounded and nonnegative for all $e \in [0,1]$. Let's define $\beta\colon [0,1] \rightrightarrows [0,1]$ such that $\beta(a) = \arg\max_{e\in[0,1]} g(a,e)$, and $V(a) = \max_{e\in[0,1]} g(a,e)$, $t \in [0,1]$. The following hold:

(1) V is absolutely continuous, nondecreasing and convex on $[0,1]$.

Proof. V is absolutely continuous and nondecreasing. The assumptions of the envelope theorem, see Theorem 2 in Milgrom and Segal (2002), are satisfied. In particular, V is absolutely continuous because $\partial g/\partial a(\cdot, e)$ exists and is bounded on $[0,1]$, so that $g(\cdot, e)$ is absolutely continuous for all $e \in [0,1]$. Because $\partial g/\partial a$ is nonnegative, V is nondecreasing, which follows from the integral representation of V.

V is convex. Let $\sup \beta =: e^\star$. $e^\star$ is a well-defined nondecreasing selection from β, by Fact 1. $\partial g/\partial a(a,e)$ is nondecreasing in a, by assumption, and is nondecreasing in e by increasing differences (Fact 1). $e^\star$ is nondecreasing (Fact 1), so $a \mapsto \partial g/\partial a(a, e^\star(a))$ is nondecreasing. The mapping $a \mapsto \partial g/\partial a(a, e^\star(a))$ is a selection from the subgradient of a convex function U, defined up to a constant term (Theorem 24.8, Rockafellar (1970), noting that the mapping is uni-dimensional). By the integral representation of V (Milgrom and Segal, 2002):

$$V(a_1) = V(0) + \int_0^{a_1} \partial g/\partial a(\tilde{a}, e^\star(\tilde{a}))\,\mathrm{d}\tilde{a},\ \ a_1 \in [0,1].$$

We have $U(a) = V(a)$ for all $a \in [0,1]$, because the convex function U admits an integral representation that is the same as that of V on $[0,1]$. ■

We state a result from convex analysis.

Fact 3 (Subdifferential of Convex Functions)**.** Let $A \subseteq \mathbb{R}$, $f\colon A \to \mathbb{R}$ be convex, and $\varphi\colon \mathbb{R} \to \mathbb{R}$ be a nondecreasing convex function on the range of f.

(1) The composition $\varphi \circ f$ is convex.

(2) If y is in the domain of $\varphi \circ f$, then:

$$\{\alpha u : (\alpha, u) \in \partial\varphi(f(y)) \times \partial f(y)\} = \partial(\varphi \circ f)(y).$$

Proof. For a proof of (1) and (2), see, respectively, Proposition 8.21 and Corollary 16.72 in Bauschke et al. (2011).[17] ■

Feasible CDF's Let $\mathcal{D}$ be the collection of CDF's over $[0, 1]$. Let $F_0 \in \mathcal{D}$ be absolutely continuous with $\int_0^1 \theta \,\mathrm{d}F_0(\theta) =: x_0$. F_0 is the CDF of the state, θ. Let $\mathcal{S}$ be the collection of all signals about the state. A CDF F is feasible if there exists a signal s about the state θ that induces F as the marginal CDF of $\mathbb{E}\{\theta|s\}$. In order to describe the set of feasible CDF's, let's define the information policy of a CDF $F \in \mathcal{D}$ as

$$\begin{aligned} I_F \colon \mathbb{R}_+ &\to \mathbb{R}_+ \\ x &\mapsto \int_0^x F(y) \,\mathrm{d}y. \end{aligned}$$

By Blackwell's theorem, a CDF F is feasible if, and only if: F is a mean preserving contraction of F_0. We define the set of *feasible CDF's* as

$$\mathcal{F} := \{F \in \mathcal{D} : I_F(1) = I_{F_0}(1), \text{ and } I_F(x) \leq I_{F_0}(x) \text{ for all } x \in \mathbb{R}_+\}.$$

Feasible information policies In order to use information policies to represent feasible CDF's, we use the following notation. Let $(\cdot)_+ := \max\{\cdot, 0\}$, id be the identity function id: $x \mapsto x$ and $\overline{F} := (\mathrm{id} - \theta_0)_+$, we note that $\overline{F} \in \mathcal{F}$. Let's define the set of feasible *information policies*:

$$\mathcal{I} := \{I \colon \mathbb{R}_+ \to \mathbb{R}_+ \ : \ I \text{ convex, and } I_{F_0}(x) \geq I(x) \geq I_{\overline{F}}(x) \text{ for all } x \in \mathbb{R}_+\}.$$

In what follows, we refer to information policies, dropping the feasibility qualifier. For a function I: we denote the right derivative and the left derivative by, respectively, $\partial_+ I$ and $\partial_- I$ (or $I'(x^-)$), and the subdifferential of I by ∂I. For two convex functions, the subdifferential of $I + J$ at x is the (Minkowski) addition of $\partial I(x)$ and $\partial J(x)$, we denote it by $\partial(I + J)(x)$. Let $I \in \mathcal{I}$, then: (i) the right-derivative of I at a point x exists for all $x \in [0, 1]$, and (ii) the left-derivative of I at $x \in [0, 1]$ exists for all $x \in [0, 1]$, once I is extended to take value 0 at $x < 0$. We define I on nonnegative reals to simplify

[17] For a different presentation of the chain rule for subdifferential calculus, see Proposition 1 and 2 in Lemaire (1985) (see also Remark 2, Case 3, part (ii)), which is less general even though it suffices for the current setup.

the following exposition, so when we refer to $\partial_- I(x)$ we implicitely assume that I is extended to take value 0 on $(-\infty, 0)$. The next result allows us to represent feasible CDF's with information policies.

Fact 4. The following hold:

(1) If $F \in \mathcal{F}$, then $I_F \in \mathcal{I}$;

(2) If $I \in \mathcal{I}$, then $I' \in \mathcal{F}$, once I is extended to take value 0 at $x < 0$.

Proof. For a proof, see Gentzkow and Kamenica (2016). ∎

Information allocations We define the operator Δ as:

$$\Delta \colon I \mapsto I - I_{\overline{F}}.$$

And we denote by ΔI the composite function $\Delta(I)$. It holds that $0 \le \Delta I(x) \le 1 - x_0$, $x \in \mathbb{R}$, by definition of I. The set of *information allocations* is:

$$\begin{aligned}\mathcal{A} := \{A \colon \mathbb{R}_+ \to \mathbb{R}_+ : A|_{[0,x_0]} \text{ is convex}, \ A|_{[x_0,1]} \text{ is convex}, \\ A \text{ is continuous at } x_0, \ A(x) \le I_{F_0}(x) - I_{\overline{F}}(x) \text{ for all } x \in \mathbb{R}_+, \\ \text{some } m \in [0,1),\, m' \in [m,1] \text{ exist such that: } \partial_- A(x_0) = m, \\ \partial_+ A(x_0) = m' - 1\}.\end{aligned}$$

The next lemma shows that $\Delta \colon \mathcal{I} \to \mathcal{A}$, and information allocations represent information policies.

Lemma 6. The following hold:

(1) If $A \in \mathcal{A}$, then: $A + I_{\overline{F}} \in \mathcal{I}$.

(2) If $I \in \mathcal{I}$, then $\Delta I \in \mathcal{A}$.

Proof. The only non trivial step is to show convexity of $A + I_{\overline{F}}$. We note that m, in the definition of A, is a subdifferential of $(A + I_{\overline{F}})(x_0)$. Therefore, the function $A + I_{\overline{F}}$ is convex on $[0, x_0]$, convex on $[x_0, \infty)$, and subdifferentiable at x_0. Global convexity follows. ∎

B Proofs for Section 3.1

In this section, we prove Theorem 1 in a more general setup than that of Section 3.1. We define $g\colon [0,1]\times[0,1]\to\mathbb{R}$ to be a continuous function, with bounded and nonnegative derivative in the first argument g_1. We also assume that $g_1(a,e)>0$ for all $e\in(0,1]$. Type-t Receiver's utility given the information policy I and effort e is:

$$g(\Delta I(\zeta_t), e) - \lambda_t k(e).$$

The model of Section 3.1 corresponds to the particular case in which $g(a,e)=ae$.

A *persuasion mechanism* is a menu of information policies $(I_r)_{r\in R}$, with $I_r\in\mathcal{I}$ for all types $r\in R$, and $R=T$. A persuasion mechanism $(I_r)_{r\in R}$ is *incentive compatible* (IC) if:

$$t \in \arg\max_{r\in R}\left\{\max_{e\in[0,1]} g(\Delta I_r(\zeta_t), e) - \lambda_t k(e)\right\}, \quad \text{for all types } t\in T.$$

An IC persuasion mechanism $(I_r)_{r\in R}$ and an information policy I induce the same effort distribution if:

$$\arg\max_{e\in[0,1]} g(\Delta I_t(\zeta_t), e) - \lambda_t k(e) \subseteq \arg\max_{e\in[0,1]} g(\Delta I(\zeta_t), e) - \lambda_t k(e) \quad \text{for all } t\in T. \tag{8}$$

Let's define

$$E(\Delta I_r(\zeta_t),\lambda_t) := \arg\max_{e\in[0,1]}[g(\Delta I_r(\zeta_t), e) - \lambda_t k(e)].$$

An IC persuasion mechanism $(I_r)_{r\in R}$ and an information policy I induce the same action distribution if:

$$\partial_- I_t(\zeta_t) = \partial_- I(\zeta_t), \quad \text{for all } t\in T \text{ such that: } E(\Delta I_t(\zeta_t),\lambda_t)\cap(0,1]\neq\emptyset.$$

Theorem 1 is a corollary of the next result.

Theorem 3. *For every IC persuasion mechanism $(I_r)_{r\in R}$ there exists an information policy J such that $(I_r)_{r\in R}$ and J induce the same effort and action distributions.*

Proof. Let's fix an IC persuasion mechanism $(I_r)_{r\in R}$. First, we define an information policy J, and then we show that it induces the same effort and action distributions as $(I_r)_{r\in R}$.

Definition of information policy J. Let's fix an IC persuasion mechanism $(I_r)_{r\in R}$. Let's define the function $I\colon [0,1]\to\mathbb{R}_+$ as follows:

$$I(c) := \sup_{r\in R} I_r(c),\ \ c\in[0,1] \tag{9}$$

$I(c)$ is well defined because $0\le I_r(c)\le I_{F_0}(c)\le 1-x_0$, $c\in[0,1]$. I is the pointwise supremum of a family of convex functions, so I is convex. It holds that $I_{\overline{F}}(c)\le I(c)\le I_{F_0}(c)$, $c\in[0,1]$, because $I_r\in\mathcal{I}, r\in R$. We extend I on $(1,\infty)$, so that the resulting extended function $J\colon\mathbb{R}_+\to\mathbb{R}_+$ is an information policy, by defining $J(c)=I_{F_0}(c)$, $c\in(1,\infty)$, and $J(c)=I(c)$, $c\in[0,1]$. We have that $J\in\mathcal{I}$.

Equivalence of effort distribution. Let's define

$$\begin{aligned} V(\Delta I_r(\zeta_t),\lambda_t) &:= \max_{e\in[0,1]}[g(\Delta I_r(\zeta_t),e)-\lambda_t k(e)] \\ \text{and } E(\Delta I_r(\zeta_t),\lambda_t) &:= \arg\max_{e\in[0,1]}[g(\Delta I_r(\zeta_t),e)-\lambda_t k(e)], \end{aligned}$$

in which $V(\cdot,\lambda_t)$ is well-defined and $E(\cdot,\lambda_t)$ is nonempty-valued.

There are two cases.

(1) $E(\Delta I_t(\zeta_t),\lambda_t)\cap(0,1]\neq\emptyset$.

(2) $E(\Delta I_t(\zeta_t),\lambda_t)=\{0\}$.

First, we consider case (1). By ET, we have:

$$V(a,\lambda_t)-V(\Delta I_t(\zeta_t),\lambda_t)=\int_{\Delta I_t(\zeta_t)}^{a} g_1(\widetilde{a},e(\widetilde{a}))\,\mathrm{d}\widetilde{a},$$

for a selection e of $E(\cdot,\lambda_t)$. Because g exhibits strictly increasing differences, MCS implies that $e(\widetilde{a})\ge e(\Delta I_t(\zeta_t))$ if $\widetilde{a}\ge\Delta I_t(\zeta_t)$. By the assumption that $g_1(\widetilde{a},\cdot)>0$ on $(0,1]$ for all $\widetilde{a}$

$$V(a,\lambda_t)-V(\Delta I_t(\zeta_t),\lambda_t)>0,\ \text{for all } a>\Delta I_t(\zeta_t).$$

Thus, in case (1) IC implies that

$$\sup_{r\in R} \Delta I_r(\zeta_t) = \Delta I_t(\zeta_t).$$

Let's consider case (2), and, towards a contradiction, let's assume $0 \notin E(\Delta J(\zeta_t), \lambda_t)$. By Berge's Theorem, $E(\cdot, \lambda_t)$ is upper hemi continuous. Therefore, there exists $\overline{a} \in (\Delta I_t(\zeta_t), \Delta J(\zeta_t))$ and $f > 0$ such that $f \in E(\overline{a}, \lambda_t)$. By the assumption that $g_1(\widetilde{a}, \cdot) > 0$ on $(0, 1]$ for all $\widetilde{a}$

$$V(\Delta J(\zeta_t), \lambda_t) - V(\overline{a}, \lambda_t) > 0.$$

The above inequality and ET imply that

$$V(\Delta J(\zeta_t), \lambda_t) - V(\Delta I_t(\zeta_t), \lambda_t) > 0.$$

Thus, in case (2) we also have that IC implies

$$\sup_{r\in R} \Delta I_r(\zeta_t) = \Delta I_t(\zeta_t).$$

The above equality contradicts $E(\Delta J(\zeta_t), \lambda_t) \neq E(\Delta I_t(\zeta_t), \lambda_t)$.

Therefore, J induces the same effort distribution as $(I_r)_{r\in R}$.

Equivalence of action distribution. Let s be such that $E(\Delta I_s(\zeta_s), \lambda_s) \cap (0, 1] \neq \emptyset$. We have $I_s(\zeta_s) = J(\zeta_s)$, be the previous claims. Let's suppose that $\partial_- I_s(\zeta_s) \neq \partial_- J(\zeta_s)$ for some type $t \in T$. Because I_s and J are information policies, they have the same extension on $(-\infty, 0)$, so that $\zeta_s > 0$. By convexity of information policies on $\mathbb{R}_+$ and $\zeta_s > 0$, we have that $\partial_- I_s(\zeta_s)$ and $\partial_- J(\zeta_s)$ are well-defined. Thus, $\partial_- I_s(\zeta_s)$ is a subgradient of I_s at ζ_s, and $\partial_- I_s(\zeta_s)$ is not subgradient of J at ζ_s, in particular, there exists $x \in \mathbb{R}$ such that

$$I_s(x) \geq I_s(\zeta_s) + \partial_- I_s(\zeta_s)(x - \zeta_s) > J(x),$$

which implies $I_s(x) > J(x)$. The last inequality contradicts the definition of J. ■

C Optimal signal characterization

First, we prove three lemmata.

Proof of Lemma 7 By Lemma 4, we define the Sender's problem as follows. We say that the *value of the Sender's problem* is:

$$\begin{aligned}&\sup_{I,e} \int_{[0,1]} \int_{[0,1]} (1 - I'(\zeta^-) - [\zeta \leq x_0]) e(\Delta I(\zeta), \lambda) \, \mathrm{d}H(\zeta|\lambda) \, \mathrm{d}G(\lambda) \\ &\text{s.t. } I \in \mathcal{I} \text{ and } e(\Delta I(\zeta), \lambda) \in E(\Delta I(\zeta), \lambda) \text{ for all } (\zeta, \lambda) \in [0,1]^2.\end{aligned}$$

We define $\widetilde{V}_S : \mathcal{I} \to \mathbb{R}$ as the maximand in the above optimization, which, in general, depends on the selection e. Lemma 7 is a corollary to the next result.

Lemma 7. Let's assume that H admits a conditional PDF of ζ given λ that is absolutely continuous as a function of ζ. Then, the value of the Sender's problem is:

$$\sup_{I \in \mathcal{I}} \int_{[0,1]} \int_{[0,1]} V(\Delta I(\zeta'), \lambda) \frac{\partial}{\partial \zeta} h_{\zeta|\lambda}(\zeta'|\lambda) \, \mathrm{d}\zeta' \, \mathrm{d}G(\lambda) + C,$$

where C is a constant term that does not depend on I and e.

Proof. We note that, by the absolute continuity hypothesis on the conditional PDF of ζ given λ, we have that $\frac{\partial}{\partial \zeta} h_{\zeta|\lambda}(\cdot|\lambda) =: h'_{\zeta|\lambda}(\cdot|\lambda)$ exists at every ζ, except on a countable subset of $[0,1]$, for all $\lambda \in [0,1]$. We define

$$V_S(I; \lambda) := -\int_{[0,1]} e(\Delta I(\zeta), \lambda) \Delta I'(\zeta^-) h_{\zeta|\lambda}(\zeta|\lambda) \, \mathrm{d}\zeta,$$

in order to express $\widetilde{V}_S$ as follows:

$$\begin{aligned}\widetilde{V}_S(I) &= \int_{[0,1]} \left\{ -\int_{[0,1]} e(\Delta I(\zeta), \lambda) \Delta I'(\zeta^-) h_{\zeta|\lambda}(\zeta|\lambda) \, \mathrm{d}\zeta \right\} \mathrm{d}G(\lambda). \\ &= \int_{[0,1]} V_S(I; \lambda) \, \mathrm{d}G(\lambda).\end{aligned}$$

We now apply the Envelope Theorem (Fact 2), and we define $A = \Delta I$. In particular, we let $\partial V(A(\zeta), \lambda)$ be the subdifferential of $V(\cdot, \lambda)$ computed at $A(\zeta)$, and we use $V_1(A(\zeta), \lambda) := e(\Delta I(\zeta), \lambda) \in \partial V(A(\zeta), \lambda)$. Given these results, we have the following equality.

$$V_S(I;\lambda) = -\int_{[0,x_0]} V_1(A(\zeta),\lambda)A'(\zeta^-)h_{\zeta|\lambda}(\zeta|\lambda)\,\mathrm{d}\zeta - \int_{(x_0,1]} V_1(A(\zeta),\lambda)A'(\zeta^-)h_{\zeta|\lambda}(\zeta|\lambda)\,\mathrm{d}\zeta.$$

By Fact 3, part (2), for a selection d from the subdifferential of $V(\cdot,\lambda)\circ A(\cdot)$, we have:

$$V_S(I;\lambda) = -\int_{[0,x_0]} d(\zeta)h_{\zeta|\lambda}(\zeta|\lambda)\,\mathrm{d}\zeta - \int_{(x_0,1]} d(\zeta)h_{\zeta|\lambda}(\zeta|\lambda)\,\mathrm{d}\zeta.$$

In particular, the equality holds because $V(\cdot,\lambda)$ is nondecreasing and convex, and the definition of A as an information allocation implies that A is convex on $[0,x_0]$ and on $(x_0,1]$, so that we apply Fact 3, part (2), on $[0,x_0]$ and on $(x_0,1]$ separately. By continuity of the PDF, we have:

$$V_S(I;\lambda) = -\int_{[0,1]} d(\zeta)h_{\zeta|\lambda}(\zeta|\lambda)\,\mathrm{d}\zeta.$$

By Fact 3, part (1), the composition $V(\cdot,\lambda)\circ A(\cdot)$ is a convex function. Thus, $V\circ A$ is the integral of any selection from the subdifferential of $V(\cdot,\lambda)\circ A(\cdot)$ (see, e.g., Corollary 24.2.1 in Rockafellar (1970)). Then, the next equality follows from integration by parts. In particular, the next equality holds because: (i) the composition $V(\cdot,\lambda)\circ A(\cdot)$ is a convex function, and (ii) $h_{\zeta|\lambda}(\cdot|\lambda)$ is absolutely continuous by hypothesis.

$$\begin{aligned} V_S(I;\lambda) = {} & -V(A(1),\lambda)h_{\zeta|\lambda}(1|\lambda) + V(A(0),\lambda)h_{\zeta|\lambda}(0|\lambda) + \\ & + \int_{[0,1]} V(A(\zeta),\lambda)h'_{\zeta|\lambda}(\zeta|\lambda)\,\mathrm{d}\zeta. \end{aligned}$$

The claim follows, because $A(1) = 0 = A(0)$. ■

The next lemma shows a property of upper censorships. A similar lemma is used in Lipnowski et al. (2021).

Lemma 8. Let $I \in \mathcal{I}$. There exists an information policy $I^u \in \mathcal{I}$ such that:

(CONS) $I^u(\zeta) = I(\zeta)$.

(IMPR) $I^{u\prime}(\zeta_-) \leq I'(\zeta_-)$ and:

$$\begin{aligned} I^u(x) - I(x) &\geq 0, x \in [0, \zeta] \\ I^u(x) - I(x) &\leq 0, x \in [\zeta, \infty]. \end{aligned}$$

.

(CENS) For some $\theta_u \in [0, \zeta]$, I^u is the θ_u upper censorship.

Proof. If $I = I_{\underline{F}}$, the proof is complete, letting I^u be the 0 upper censorship. Let $I \neq I_{\underline{F}}$. Let's define $M := \{m \in [0, I'(\zeta_-)] \mid I(\zeta) + m(x - \zeta) \leq I_{F_0}(x)$ for all $x \in [0, \zeta]\}$, and $m := \min M$. We construct an information policy starting from the line $x \mapsto I(\zeta) + m(x - \zeta)$, by means of the next three claims.

m is well-defined. (i) M is nonempty, because $0 \leq I'(\zeta_-) \leq 1$ (which follows from $I \in \mathcal{I}$), $I'(\zeta_-) \in \partial I(\zeta_-)$ and $I(x) \leq I_{F_0}(x)$ for all x; (ii) M is closed, becase the mapping $m \mapsto I(\zeta) + m(x - \zeta)$ is a continuous function of m for given x, ζ, I; (iii) M is bounded because if $m \in M$ then $0 \leq m \leq I'(\zeta_-)$, and $I'(\zeta_-) \leq 1$ because $I \in \mathcal{I}$.

There exists a unique $\theta_u \in [0, \zeta]$ such that $I_{F_0}(\theta_u) = I(\zeta) + m(\theta_u - x)$. Suppose there does not exist such a θ_u; then there exists $\varepsilon > 0$ such that: $I(\zeta) + (m - \varepsilon)(x - \zeta) < I_{F_0}(x)$ for all $x \in [0, \zeta]$ (if ε is sufficiently small) and $m - \varepsilon \in M$ (because $I \neq I_{\underline{F}}$ and $\zeta > \theta_0$, if ε is sufficiently small). We reach a contradiction with the definition of m. Uniqueness follows from convexity of I_{F_0}.

$m \in \partial I_{F_0}(\theta_u)$ and $I(\zeta) + m(x - \zeta) = I_{F_0}(\theta_u) + (x - \theta_u)F_0(\theta_u)$ for all x. First, we argue that $m \in \partial I_{F_0}(\theta_u)$. By convexity of I_{F_0} and definition of θ_u, $x \mapsto I(\zeta) + m(x - \zeta)$ is tangent to I_{F_0} at θ_u. Thus, m is a subdifferential of I_{F_0} at θ_u. Now, we argue that $I(\zeta) + m(x - \zeta) = I_{F_0}(\theta_u) + (x - \theta_u)F_0(\theta_u)$ for all x. $m = F_0(\theta_u)$ because I_{F_0} is continuously differentiable. The equality follows from $x \mapsto I(\zeta) + m(x - \zeta)$ being equal to I_{F_0} at $x = \theta_u$.

Let's find an information policy that satisfies (CENS). We define the following function.

$$
\begin{aligned}
I^u \colon \mathbb{R}_+ &\to \mathbb{R}_+ \\
x &\mapsto \begin{cases} I_{F_0}(x) & ,x \in [0,\theta_u] \\ I(\zeta) + m(x-\zeta) & ,x \in (\theta_u, \zeta] \\ \max\{I(\zeta) + m(x-\zeta), I_{\underline{F}}(x)\} & ,x \in (\zeta, \infty). \end{cases}
\end{aligned}
$$

To show (CENS) it suffices to show that: (i) for some $x_u \in [0,1]$

$$
I^u(x) = \begin{cases} I_{F_0}(x) & ,x \in [0,\theta_u] \\ I_{F_0}(\theta_u) + (x-\theta_u)F_0(\theta_u) & ,x \in (\theta_u, x_u] \\ I_{\underline{F}}(x) & ,x \in (x_u, \infty), \end{cases}
$$

and (ii) $I^u \in \mathcal{I}$. We claim that (i) holds by means of the next three claims.

Some $x_u \in [\zeta, 1]$ exists such that:

$$
I(\zeta) + m(x-\zeta) \geq I_{\underline{F}}(x) \quad ,x \in [0, x_u] \tag{10}
$$

$$
I(\zeta) + m(x-\zeta) \leq I_{\underline{F}}(x) \quad ,x \in [x_u, 1]. \tag{11}
$$

Let's note that: (a) $I(\zeta) \geq I_{\underline{F}}(\zeta)$; (b) by $m \in \partial I_{F_0}(\theta_u)$ and $I_{F_0}(1) = I_{\underline{F}}(1)$, we have that $I_{\underline{F}}(1) \geq I(\zeta) + m(1-\zeta)$, and (c) the two functions, $x \mapsto I(\zeta) + m(x-\zeta)$ and $I_{\underline{F}}$, are affine with slopes, respectively, m and 1, such that: $m \leq 1$. We proceed to verify that (ii) holds, i.e. $I^u \in \mathcal{I}$. (ii) follows from the next two claims.

$I_{\underline{F}}(x) \leq I^u(x) \leq I_{F_0}(x)$ for all $x \in \mathbb{R}_+$ and I^u locally convex at all $x \notin \{\theta_u, x_u\}$. If $x \in [0, \theta_u)$, I^u is locally convex and $I_{\underline{F}}(x) \leq I^u(x) \leq I_{F_0}(x)$. If $x \in (\theta_u, \zeta)$, I^u is affine, $I_{\underline{F}}(x) \leq I(x) \leq I^u(x)$ by construction of I^u and definition of I, and $I^u(x) \leq I_{F_0}(x)$ by $m \in \partial I_{F_0}(x)$. If $x \in [\zeta, \infty)$, I is locally convex (because it is the maximum of affine functions), $I_{\underline{F}}(x) \leq I^u(x)$ by construction of I^u, $I^u(x) \leq I_{F_0}(x)$ because: (i) $m \in \partial I_{F_0}(\zeta)$ and (ii) $I_{\underline{F}}(x) \leq I_{F_0}(x)$. To verify global convexity, it suffices to verify the next claim.

I^u is subdifferentiable at $x \in \{\theta_u, x_u\}$. First, we argue that m is a subdifferential of I^u at θ_u. This follows from the fact that the slope of I^u at θ_u is a subdifferential of I_{F_0} at θ_u, and $I^u(\theta_u) = I_{F_0}(\theta_u)$. On $[0, \theta_u]$, $I^u = I_{F_0}$, and on $[\theta_u, \infty]$ I^u is above the line

$x \mapsto I(\zeta) + m(x - \zeta)$. Thus, $m \in \partial I^u(\theta^u)$. Second, the fact that m is a subdifferential of I^u at ζ follows from the claim in (10).

We have shown that (CENS) holds. (IMPR) and (CONS) hold by construction. ■

C.1 Proof of Proposition 2

Proof. If an optimal information policy exists, the Proposition follows from Lemma 5 and Lemma 8. Existence is readily established using the tools from Kleiner et al. (2021), observing that the maximand is a continuous functional (Lemma 5 and the envelope theorem in Fact 2). ■

C.2 Proof of Proposition ??

We prove several lemmata.

Lemma 9. Let $I \in \mathcal{I}$ such that $p \geq \overline{c}(\Delta I)$, there exists another information policy $I^\star$ such that:

(FEAS) $I^\star$ is feasible: $I^\star \in \mathcal{I}$,

(EM) $I^\star$ produces the same extensive margin as I: $\overline{c}(\Delta I^\star) = \overline{c}(\Delta I)$, $\underline{c}(\Delta I^\star) = \underline{c}(\Delta I)$.

(IMPR)

$$\Delta I^\star(x) \geq 0, \text{ for all } x \in [\underline{c}(\Delta I), \overline{c}(\Delta I)]$$

.

(CENS) There exist $x_\ell, \theta_\ell, \theta_m, x_m$ such that $0 \leq x_\ell \leq \theta_\ell \leq \theta_m \leq x_m \leq 1$, and:

$$I^\star(x) = \begin{cases} I_{\underline{F}}(x) & , x \in [0, x_\ell] \\ I_{F_0}(\theta_\ell) + F_0(\theta_\ell)(x - \theta_\ell) & , x \in (x_\ell, \theta_\ell] \\ I_{F_0}(x) & , x \in (\theta_\ell, \theta_m] \\ I_{F_0}(\theta_m) + F_0(\theta_m)(x - \theta_m) & , x \in (\theta_m, x_m] \\ \underline{I}_{\underline{F}}(x) & , x \in (x_m, \infty]. \end{cases}$$

Proof. We use the following notation: $\overline{c}(I - \underline{I}) =: \overline{c}$, $\underline{c}(I - \underline{I}) =: \underline{c}$. In the first step, we prove the lemma in the case where there is a feasible information policy that is a straight line between the points $\underline{p} := (\underline{c}, I(\underline{c}))$ and $\overline{p} := (\overline{c}, I(\overline{c}))$. In the second step, we prove the lemma in the case where there is not a feasible information policy that is a straight line between the points $\underline{p}$ and $\overline{p}$.

First Step. Let's define the line line i such that $x \mapsto I(\underline{c}) + \lambda^*(x - \underline{c})$, with slope $\lambda^* := \frac{I(\overline{c})-I(\underline{c})}{\overline{c}-\underline{c}}$. We claim that $i^\star(x) := \max\{i(x), \underline{I}_0(x)\}$ satisfies all properties. It is FEAS by hypothesis. It is EXT because $i(\underline{c}) = I(\underline{c})$ and $i(\overline{c}) = I(\overline{c})$. It is IMPR because I is convex and $i^\star$ is EXT. It is CENS with $\theta_\ell = \theta_m = x_m$, because: (i) EXT of $i^\star$ and convexity of I imply that $i^\star$ is affine in $[\underline{c}, \overline{c}]$, (ii) $\lambda^* \in [0,1]$ and EXT imply, with $I \in \mathcal{I}_0$ that there are intersections $\widetilde{x}_1, \widetilde{x}_2$, with $\widetilde{x}_1 \leq \underline{c} \leq \overline{c} \leq \widetilde{x}_2$, where: $i^\star(x) = \underline{I}(x)$ if $x \in [0, \widetilde{x}_1] \cup [\widetilde{x}_2, 1]$.

Second Step. In this case, $i^\star$ is not FEAS. Since $i^\star$ satisfies FEAS at x if $x \leq \underline{c}$ and if $x \geq \overline{c}$, there is a point $x^* \in (\underline{c}, \overline{c})$ such that $i(x^*) > I_{F_0}(x^*)$.

$$L := \{\lambda \in [I'(\underline{c}), 1] \mid I(\underline{c}) + \lambda(x - \underline{c}) \leq I_{F_0}(x) \text{ for all } x \in [\underline{c}, \infty)\},$$
$$M := \{\lambda \in [0, I'(\overline{c})] \mid I(\overline{c}) + \lambda(x - \overline{c}) \leq I_{F_0}(x) \text{ for all } x \in [0, \overline{c}]\}.$$

$\ell := \max L$, $m := \min M$. We define two lines:

$$y_\ell \text{ is: } x \mapsto I(\underline{c}) + \ell(x - \underline{c})$$
$$y_m \text{ is: } x \mapsto I(\overline{c}) + m(x - \overline{c}).$$

We prove a lemmata.

Lemma 10. ℓ, m are well-defined.

Proof. L is nonempty because $I'(\underline{c}) \in L$, which follows from: (i) $I_{F_0}(x) \geq I(x)$ for all x and (ii) $I'(\underline{c}) \in \partial I(\underline{c})$. M is nonempty because $I'(\overline{c}) \in M$, which follows from: (i) $I_{F_0}(x) \geq I(x)$ for all x and (ii) $I'(\overline{c}) \in \partial I(\overline{c})$. L, M are closed because I_{F_0} is continuous. L, M are bounded. ■

Lemma 11. that there exists a unique pair of numbers $(\theta_\ell, \theta_m) \in [\underline{c}, 1] \times [0, \overline{c}]$ such that:

$$y_\ell(\theta_\ell) = I_{F_0}(\theta_\ell)$$
$$y_m(\theta_m) = I_{F_0}(\theta_m)$$

Proof. Suppose there does not exists such a ℓ. There exists a sufficiently small $\varepsilon > 0$ such that: (i) $\ell + \varepsilon \in L$ and (ii) $I(\underline{c}) + (\ell + \varepsilon)(x - \underline{c}) < I_{F_0}(x)$ for all $x \in [\underline{c}, \infty)$; we note that $\ell = 1$ contradicts $\ell \in L$ because $I'_{F_0}(x) < 1$ if $x < 1$. Uniquenss of ℓ follows from convexity of I_{F_0}.

Suppose there does not exists such an m. There exists a sufficiently small $\varepsilon > 0$ such that: (i) $\ell - \varepsilon \in M$ and (ii) $I(\bar{c}) + (m - \varepsilon)(x - \bar{c}) < I_{F_0}(x)$ for all $x \in [0, \bar{c})$; we note that $m = 0$ contradicts $I \neq I_{\underline{F}}$. Uniquenss of m follows from convexity of I_{F_0}. ■

Lemma 12. *$\theta_\ell \leq \theta_m$.*

Proof. Let's prove that it suffices to show that: $\ell \leq m$. Suppose $\ell \leq m$, then: since $\ell \in \partial I_{F_0}(\theta_\ell)$ and $m \in \partial I_{F_0}(\theta_m)$, and I_{F_0} is strictly convex, we have: $\theta_\ell \leq \theta_m$.

First, we show that $\ell \leq \lambda^*$. Suppose that: $\ell > \lambda^*$. Then: $I(x) + \ell(x - \underline{c}) > I(\underline{c}) + \lambda^*(x - \underline{c})$ for all $x > \underline{c}$. Therefore, since $\ell > 0$:

$$I_{F_0}(x^*) \geq I(\underline{c}) + \lambda^*(x^* - \underline{c}).$$

We reached a contradiction with the definition of x^*, so: $\ell \leq \lambda^*$.

Let's prove that $m \geq \lambda^*$. Suppose $m < \lambda^*$. Then: $I(x) + m(x - \overline{c}) > I(\overline{c}) + \lambda^*(x - \overline{c})$ for all $x < \overline{c}$. Therefore, since $m > 0$:

$$I_{F_0}(x^*) \geq I(\underline{c}) + \lambda^*(x^* - \underline{c}).$$

We reached a contradiction with the definition of x^*, so: $m \geq \lambda^*$. Therefore, we have $m \geq \lambda^* \geq \ell$, which implies $\theta_m \geq \theta_\ell$. ■

We define a candidate $I^\star$ and we verify that it has the desired properties.

$$I^\star(x) := \begin{cases} \max\{I_{\underline{F}}(x), I(\underline{c}) + \ell(x - \underline{c})\} & , x \in [0, \theta_\ell] \\ I_{F_0}(x) & , x \in [\theta_\ell, \theta_m] \\ \max\{I_{\underline{F}}(x), I(\overline{c}) + m(x - \overline{c})\} & , x \in [\theta_m, \infty] \end{cases}$$

Let's first verify that $I^\star$ is well-defined. We know that $\ell \in \partial I_{F_0}(\theta_\ell)$ and $m \in \partial I_{F_0}(\theta_m)$. Since $I(\underline{c}) + \ell(0 - \underline{c}) < I_{F_0}(0)$ and $I(\underline{c}) \geq I_{F_0}(\underline{c})$, $\max\{I_{F_0}(x), I(\underline{c}) + \ell(x - \underline{c})\} = I_{F_0}(x)$ if $x < x_0$; and $\max\{I_{F_0}(x), I(\underline{c}) + \ell(x - \underline{c})\} = I(\underline{c}) + \ell(x - \underline{c})$ if $x > x_0$; for some $x_0 \in [0, \theta_\ell]$. In a similar way, we can show that there exists a $x_2 \in [\theta_m, 1]$ such that:

$\max\{I_{F_0}(x), I(\bar{c}) + m(x-\bar{c})\} = I_{F_0}(x)$ if $x > x_2$, and $\max\{I_{F_0}(x), I(\bar{c}) + m(x-\bar{c})\} = I(\bar{c}) + m(x-\bar{c})$ if $x < x_2$.

(CENS) follows from the definition of $I^\star$ and its well-definedness, using $y_0 = x_0$, $y_1 = x_1$, $y_2 = x_3$, $y_3 = x_4$, and $\alpha_1 = \lambda_1$ and $\alpha_2 = \lambda_3$.

(IMPR) IMPR on $[\underline{c}, x_1]$ and $[x_3, \bar{c}]$ follows from convexity of I, and on $[x_1, x_3]$ follows from FEAS of I in that region.

(EM) follows from $I^\star(\underline{c}) = I(\underline{c}) + \lambda_1(x - \underline{c})$, and $I^\star(\bar{c}) = I(\bar{c}) + \lambda_3(x - \bar{c})$.

(FEAS) First, $I^\star$ is always above $\underline{I}_0$. Second $I^\star$ is always below $\overline{I}_0$, which follows from $\lambda_\ell \in \partial\overline{I}_0(x_\ell)$ for all $\ell \in \{1,3\}$. The maximum of affine functions is convex, and $\overline{I}_0$ is convex. Global convexity then follows if $I^\star$ is subdifferentiable at x_1 and x_3. We now claim that $\lambda_\ell \in \partial I^\star(x_\ell)$ for all $\ell \in \{1,3\}$. This claim follows from $\lambda_\ell \in \partial\overline{I}_0(x_\ell)$ for all $\ell \in \{1,3\}$, and the fact that $\overline{I}_0(x_1) = I(\underline{c}) + \lambda_1(x_1 - \underline{c})$ and $\overline{I}_0(x_3) = I(\bar{c}) + \lambda_3(x_3 - \bar{c})$ (together with convexity of $I^\star$ in $[0, x_1]$ and $[x_3, 1]$). We established that the subdifferential of $I^\star$ at x_1 and x_3 nonempty, which finalizes the proof that $I^\star$ is globally convex.

■

Proof of Proposition ??

Proof. By the previous lemmata, to prove Proposition **??** we only need to prove the following claim.

Let $I \in \mathcal{I}$ such that $p < \bar{c}(\Delta I)$, there exists another information policy I° such that:

(FEAS) I° is feasible: $I^\circ \in \mathcal{I}$,

(EM) I° produces the same extensive margin as I: $\bar{c}(\Delta I^\circ) = \bar{c}(\Delta I)$, $\underline{c}(\Delta I^\circ) = \underline{c}(\Delta I)$.

(IMPR)

$$\Delta I^\circ(x) \geq 0, \text{ for all } x \in [\underline{c}(\Delta I), \bar{c}(\Delta I)]$$

.

(CENS) There exist $x_\ell, \theta_\ell, \theta_m, x_m, \theta_u, x_u$ such that $0 \le x_\ell \le \theta_\ell \le \theta_m \le x_m^\circ \le x_u \le 1$, and:

$$I^\circ(x) = \begin{cases} I_{\underline{F}}(x) & , x \in [0, x_\ell] \\ I_{F_0}(\theta_\ell) + F_0(\theta_\ell)(x - \theta_\ell) & , x \in (x_\ell, \theta_\ell] \\ I_{F_0}(x) & , x \in (\theta_\ell, \theta_m] \\ I_{F_0}(\theta_m) + F_0(\theta_m)(x - \theta_m) & , x \in (\theta_m, x_m^\circ] \\ I_{F_0}(\theta_u) + F_0(\theta_u)(x - \theta_u) & , x \in (x_m^\circ, x_u] \\ I_{\underline{F}}(x) & , x \in (x_u, \infty). \end{cases}$$

The claim follows from taking $I^\star$ from the previous lemmata until the point x_m° where $I^\star$ intercepts the line j: $x \mapsto I(\bar{c}) + I'(\bar{c})(x - \bar{c})$, and $\max\{I_{\underline{F}}, j\}$ after x_m°. ■

C.3 Known ζ and κ

We assume, in this section only, that Sender knows both ζ and κ. If $\zeta > 1$, any information policy is optimal. If $\zeta \le \theta_0$, $I_{\underline{F}}$ is optimal. Let $1 \ge \zeta \ge \theta_0$.

The Sender's problem is:

$$\max_{I \in \mathcal{I}} (1 - I'(\zeta_-))[\Delta I(\zeta) \ge \kappa].$$

Lemma 13. There exists a solution to the Sender's problem $I \in \mathcal{I}$ such that: for $\theta \in [0, \zeta]$, I is the θ upper censorship and:

$$\Delta I_\theta \le \kappa,$$

with equality if $\theta > 0$.

Proof. Let $\mathcal{I}^u := \{I \in \mathcal{I} \ : \ I = I_\theta, \text{ for some } \theta \in [0,1] \text{ such that } \theta \le \zeta\}$. Suppose the solution is not I_{F_0}. The Sender's problem is, without loss of optimality by lemma 8:

$$\max_{I \in \mathcal{I}^u} (1 - I'(\zeta_-))[\Delta I(\zeta) \ge \kappa].$$

Suppose there exists a solution $I \in \mathcal{I}^u$, such that $I = I_{\theta^\star}$, for some $\theta^\star \in (0,1)$. We distinguish three cases.

(1) If $\Delta I(\zeta) < \kappa$, then $I_{\underline{F}}$ achieves the same Sender payoff. (2) If $\Delta I(\zeta) = \kappa$, the lemma holds. (3) Let's suppose $\Delta I(\zeta) > \kappa$. By definition of I, at $y = I(\zeta)$ the next condition holds:

$$I_{F_0}(\theta^\star) + F_0(\theta^\star)(\zeta - \theta^\star) - y = 0.$$

By the implicit function theorem, there exists a differentiable function t:

$$\begin{aligned} t \colon (0,1) &\to (0,1) \\ y &\mapsto \theta^\star, \end{aligned}$$

such that:

$$t'(y) = \begin{cases} \frac{1}{(\zeta - t(y))F_0'(t(y))} & , 0 < \zeta < t(y) \\ \frac{1}{F_0'(t(y))} & , 1 > \zeta \geq t(y). \end{cases}$$

Let the value of I_θ be:

$$\begin{aligned} v \colon (0,1) &\to [0,1] \\ \theta &\mapsto (1 - I_\theta'(\zeta_-)) \end{aligned}$$

Because $I_{\theta^\star}'(\zeta_-) = F_0(\theta^\star)$, v is differentiable in θ at $\theta^\star$. Using the chain rule, the derivative of v with respect to $I(\zeta)$ is:

$$-F_0'(t(I(\zeta)))\frac{1}{(\zeta - t(I(\zeta)))F_0'(t(I(\zeta)))},$$

whenever $\zeta > t(I(\zeta))$, and -1 otherwise. It follows that we can consider without loss solutions $I \in \mathcal{I}^u$ that satisfy: $\Delta I_\theta(\zeta) = \kappa$ and $I = I_\theta$, or $\Delta I(\zeta) < \kappa$. ■

References

Bauschke, Heinz H, Patrick L Combettes, and others 2011. *Convex analysis and monotone operator theory in Hilbert spaces*. 408: Springer.

Bergemann, Dirk, and Stephen Morris 2019."Information Design: A Unified Perspective."

J. Econ. Lit.. 57 (1): 44–95, https://ideas.repec.org/a/aea/jeclit/v57y2019i1p44-95.html.

Bloedel, Alexander W., and Ilya Segal 2021."Persuading a Rationally Inattentive Receiver." working Paper, Stanford University.

Chahrour, Ryan 2014."Public communication and information acquisition." *Am. Econ. J. Macro.*. 6 (3): 73–101.

Cornand, Camille, and Frank Heinemann 2008."Optimal Degree of Public Information Dissemination." *Econ. J.*. 118 (528): 718–742, 10.1111/j.1468-0297.2008.02139.x, _eprint: https://academic.oup.com/ej/article-pdf/118/528/718/26155047/ej0718.pdf.

Dworczak, Piotr, and Giorgio Martini 2019."The Simple Economics of Optimal Persuasion." *J. Polit. Econ.*. 127 (5): 1993–2048, 10.1086/701813.

Ehrmann, Michael, and Alena Wabitsch 2022."Central bank communication with non-experts–A road to nowhere?" *J. Monetary. Econ.*. 127 69–85, Publisher: Elsevier.

Floridi, Luciano 2014. *The Fourth Revolution: How the Infosphere is Reshaping Human Reality.* : OUP Oxford, Google-Books-ID: 65eAAwAAQBAJ.

Galperti, Simone, and Isabel Trevino 2020."Coordination motives and competition for attention in information markets." *J. Econ. Theory.* 188, 105039, 10.1016/j.jet.2020.105039.

Gehlbach, Scott, and Konstantin Sonin 2014."Government control of the media." *Journal of public Economics.* 118 163–171.

Gentzkow, Matthew, and Emir Kamenica 2016."A Rothschild-Stiglitz Approach to Bayesian Persuasion." *Am. Econ. Rev.*. 106 (5): 597–601, https://ideas.repec.org/a/aea/aecrev/v106y2016i5p597-601.html.

Gitmez, Arda, and Pooya Molavi 2020."Media Capture: A Bayesian Persuasion Approach." working Paper.

Guo, Yingni, and Eran Shmaya 2019."The interval structure of optimal disclosure." *Econometrica.* 87 (2): 653–675.

Haldane, Andrew, Alistair Macaulay, and Michael McMahon 2020.“The 3 E’s of Central Bank Communication with the Public.” *SSRN Electronic Journal*. 10.2139/ssrn.3520658.

Haldane, Andrew, and Michael McMahon 2018.“Central Bank Communications and the General Public.” *AEA Papers and Proceedings*. 108 578–583, 10.1257/pandp.20181082.

Kamenica, Emir 2019.“Bayesian Persuasion and Information Design.” *Annual Review of Economics*. 11 (1): 249–272, 10.1146/annurev-economics.

Kamenica, Emir, and Matthew Gentzkow 2011.“Bayesian Persuasion.” *Am. Econ. Rev.*. 101 (6): 2590–2615, https://ideas.repec.org/a/aea/aecrev/v101y2011i6p2590-2615.html.

Kleiner, Andreas, Benny Moldovanu, and Philipp Strack 2021.“Extreme Points and Majorization: Economic Applications.” *Econometrica*. 89 (4): 1557–1593, 10.3982/ECTA18312.

Kolotilin, Anton 2018.“Optimal information disclosure: a linear programming approach.” *Theoretical Economics*. 13 (2): , https://ideas.repec.org/a/the/publsh/1805.html.

Kolotilin, Anton, Timofiy Mylovanov, and Andriy Zapechelnyuk 2022.“Censorship as optimal persuasion.” *Theoretical Economics*. 17 (2): 561–585.

Kolotilin, Anton, Tymofiy Mylovanov, Andriy Zapechelnyuk, and Ming Li 2017.“Persuasion of a Privately Informed Receiver.” *Econometrica*. 85 (6): 1949–1964, 10.3982/ECTA13251.

Lemaire, B. 1985.“Application of a Subdifferential of a Convex Composite Functional to Optimal Control in Variational Inequalities.” In *Nondifferentiable Optimization: Motivations and Applications* edited by Vladimir F. Demyanov, and Diethard Pallaschke, Lecture Notes in Economics and Mathematical Systems 103–117, Berlin, Heidelberg: Springer, 10.1007/978-3-662-12603-5_10.

Lipnowski, Elliot, Laurent Mathevet, and Dong Wei 2020.“Attention management.” *American Economic Review: Insights*. 2 (1): 17–32.

——— 2022a.“Optimal attention management: A tractable framework.” *Game Econ. Behav.*. 133 170–180.

Lipnowski, Elliot, and Doron Ravid, "Predicting Choice from Information Costs."

Lipnowski, Elliot, Doron Ravid, and Denis Shishkin 2021."Persuasion via Weak Institutions." July.

——— 2022b."Persuasion via Weak Institutions." *J. Polit. Econ.*. 130 (10): 2705–2730, 10.1086/720462.

Matysková, Ludmila, and Alfonso Montes 2021."Bayesian Persuasion With Costly Information Acquisition."

Milgrom, Paul, and Ilya Segal 2002."Envelope theorems for arbitrary choice sets." *Econometrica*. 70 (2): 583–601, Publisher: Wiley Online Library.

Myatt, David P, and Chris Wallace 2014."Central bank communication design in a Lucas-Phelps economy." *J. Monetary. Econ.*. 63 64–79.

Nimark, Kristoffer P., and Stefan Pitschner 2019."News media and delegated information choice." *J. Econ. Theory*. 181 160–196, 10.1016/j.jet.2019.02.001.

Prat, Andrea 2015."Chapter 16 - Media Capture and Media Power." In *Handbook of Media Economics* edited by Simon P. Anderson, Joel Waldfogel, and David Strömberg 1 of Handbook of Media Economics: North-Holland, 669–686, 10.1016/B978-0-444-63685-0.00016-4.

Rockafellar, R Tyrrell 1970."Convex Analysis." Publisher: Citeseer.

Shishkin, Denis 2019."Evidence Acquisition and Voluntary Disclosure." December.

Simon, Herbert A 1996."Designing organizations for an information-rich world." *International Library of Critical Writings in Economics*. 70 187–202.

Vives, Xavier 1990."Nash equilibrium with strategic complementarities." *J. Math. Econ.*. 19 (3): 305–321, Publisher: Elsevier.

Wei, Dong 2021."Persuasion under costly learning." *J. Math. Econ.*. 94, 102451.