Persuading an inattentive and privately informed receiver

Pietro Dall'Ara*

Boston College

August 3, 2024

Abstract

This paper studies the persuasion of a receiver who accesses information only if she exerts costly attention effort. A sender designs an experiment to persuade the receiver to take a specific action. The experiment affects the receiver's attention effort, that is, the probability that she updates her beliefs. As a result, persuasion has two margins: extensive (effort) and intensive (action). The receiver's utility exhibits a supermodularity property in information and effort. By leveraging this property, we prove a general equivalence between experiments and persuasion mechanisms à la Kolotilin et al. (2017). Censoring high states is an optimal strategy for the sender in applications.

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

Contents

*I am grateful for helpful comments from Mehmet Ekmekci, Laurent Mathevet, Utku Ünver, Bumin Yenmez, and numerous seminar audiences. An earlier version of this paper bore the title "The extensive margin of Bayesian persuasion." Contact: pietro.dallara@gmail.com.

1 Introduction

In the "information age," consumers evaluate whether information sources are worth their attention because learning takes effort and time (Simon, 1996; Floridi, 2014). The persuasion literature studies how a sender, such as an advertiser or media outlet, provides information to persuade a receiver to take a specific action (Kamenica, 2019). When attention is costly, the sender faces a dual problem: the receiver can be persuaded only if she pays attention to the information. This paper studies a persuasion model in which the sender's information affects the attention effort of a receiver who privately knows the costs and benefits of information.

The *intensive* margin of persuasion refers to the intensity of the sender's influence on the receiver's action, whereas the *extensive* margin refers to whether the receiver pays attention to the information. The study of the extensive margin is important to understand how consumers allocate attention to product advertisements and news content. This allocation of attention ultimately determines the success of marketing campaigns and the spread of information across heterogeneous audiences.

To study the extensive and intensive margins of persuasion, we introduce the receiver's attention decision into a persuasion game between two players: Sender (he) and Receiver (she). In the first stage of the game, Sender designs a signal, which is a random variable correlated with the unknown state θ. Receiver chooses her attention *effort* e, knowing the signal's distribution but not its realization. Increasing effort is costly but raises the probability of observing the signal's realization. In the last stage of the game, Receiver takes a binary action: 1 or 0. The players' interests conflict because Receiver chooses 1 only if she expects the state θ to exceed her outside option, whereas Sender wants her to choose 1 regardless of the state. The Receiver's outside option and effort cost constitute her privately known *type*. The outside option captures the benefit of information, because it is unlikely that a given piece of information is useful if an extremely beneficial outside option is available.

Sender considers that increasing the correlation between the state and the signal affects both the Receiver's attention effort e (the extensive margin) and her action upon observing the signal (the intensive margin). Specifically, Receiver updates her beliefs with probability e, and does not update with the remaining probability. The choice of effort captures the choice of acquiring information, and the cost of effort may be monetary or cognitive. This model of attention is less general than models of flexible information acquisition (Caplin et al., 2022; Denti, 2022; Pomatto et al., 2023) because Receiver only chooses the probability with which she uniformly observes every signal realization. This parsimonious model allows for including asymmetric information and a general functional form of effort cost.[1]

[1] Typical applications of flexible information acquisition rely on functional-form assumptions and define cost functions over belief distributions, which this model avoids. Specifically, the Receiver's information cost is "experimental" (Denti et al., 2022) because she chooses mixtures of full information and null information about the Sender's signal, represented by effort e.

In the model, the Receiver's utility is supermodular in information and effort (Corollary 1): the return from effort increases in a type-specific informativeness order, which agrees with Blackwell's order whenever possible. This property is a complementarity between information and attention effort. Complementarity is a feature of information acquisition from sources such as news outlets and product advertisements. For instance, voters' willingness to subscribe to a newspaper increases as the newspaper dedicates more space to election news, and TV audiences are more likely to pay attention during commercial breaks if advertisements are informative.[2] This paper analyzes the extent of persuasion in these scenarios.

We establish the equivalence between persuasion mechanisms and signals (Theorem 1). A persuasion mechanism is a menu of signals, one for every Receiver's report of her type. Under a persuasion mechanism, Receiver makes a report and chooses her effort. Specifically, Receiver chooses the probability with which she observes the signal that corresponds to her report. A mechanism is incentive-compatible if Receiver finds it optimal to reveal her type. For every incentive-compatible persuasion mechanism, there is a signal that induces the same action and effort choices of all Receiver's types. The key step in the proof is to construct a signal that "allocates" to each type t the same t-specific informativeness as the incentive-compatible mechanism. This step establishes the equivalence with respect to effort choices. The constructed signal also replicates Receiver's optimal action, by simple convex analysis given the representation of signals as convex functions (Gentzkow and Kamenica, 2016). Thus, the equivalence in Kolotilin et al. (2017) obtains in the particular case of costless effort. As a result, Sender need not offer a fine collection of information structures, and the analysis of the extensive margin can focus on single signals.

We characterize the optimal signal in commonly studied applications, which censors high states (Theorem 3). An upper censorship is a signal that reveals low states and pools high states, as shown in Figure 1. Upper censorships are optimal signals if the Receiver's outside option follows a single-peaked distribution. In the costless-attention case, the result follows directly from the shape of the noise in the Receiver's action given her posterior belief. The noise — perceived by Sender — is exogenous and due to asymmetric information. The general result accounts for the endogenous

[2] There is empirical evidence that informative advertising leads to awareness Honka et al. (2017); Tsai and Honka (2021). Supermodularity implies that Receiver's equilibrium effort increases as her outside option becomes less extreme. Angelucci et al. (2021) empirically confirm that the acquisition of Covid-related information increases with its benefits.

(a) An upper censorship is a signal that reveals the state θ if θ is below a threshold $\overline{\theta}$ and sends a single message, m, if the state is above $\overline{\theta}$.

(b) A bi-upper censorship is a signal that reveals the state θ if θ is below a lower threshold θ_1, sends a message m_1 if the state is between θ_1 and an upper threshold θ_2, and sends a different message m_2 if the state is above θ_2.

Figure 1: An upper censorship (a) and a bi-upper censorship (b), for a state θ with support $[0, 1]$.

randomness that is due to the Receiver's effort. Moreover, the Sender's equilibrium upper censorship provides less information if effort is costless than if effort comes at a small positive cost. If Sender has preferences over the extensive margin, inspired by models of media capture à la Gehlbach and Sonin (2014), "bi-upper censorships" are optimal signals (Figure 1). The additional censorship region allows Sender to separately control the extensive and intensive margins of persuasion.

Related literature Existing work considers persuasion without Receiver's attention effort.[3] The optimality properties of upper censorships are known, and the equivalence between persuasion mechanisms and signals is shown by Kolotilin et al. (2017).[4] We generalize these results to the case of Receiver's costly effort and privately known effort cost.

The persuasion of an inattentive Receiver is studied without private information. In Wei (2021), Receiver's attention cost is posterior-separable. As a result of costly attention and symmetric information, the optimal signal is binary, and, in equilibrium, Receiver pays full attention. In the main model of Bloedel and Segal (2021), Receiver's attention cost is proportional to the entropy reduction in her belief. In a separate model, the authors study the same effort-cost structure as in this paper. The connection with these approches is discussed in Section 2.4. Certain dynamic models of persuasion

[3] *Inter alia*: Rayo and Segal (2010); Kamenica and Gentzkow (2011); Kolotilin (2018); Dworczak and Martini (2019); see also Section 2.4. Note that this paper's model is not nested in the "single-moment persuasion" because effort is a function of the entire posterior-mean distribution — not of a single posterior mean; see Lemma 3 and the Sender's maximand in Lemma B.3.

[4] For upper censorships, see also: Gentzkow and Kamenica (2016); Kleiner et al. (2021); Kolotilin et al. (2022); Arieli et al. (2023); Feng et al. (2024); for persuasion mechanisms, see also: Guo and Shmaya (2019).

include costly Receiver's learning (Liao, 2021; Jain and Whitmeyer, 2022; Au and Whitmeyer, 2023; Che et al., 2023), although the focus of these binary-state models is on intertemporal flow of information.[5]

Other work studies Receiver's information acquisition with different Sender's incentives or Receiver's information sources than in this paper. The "attention-management" literature considers Receiver's attention given a benevolent Sender, who maximizes Receiver's material payoff (without considering her attention cost, Lipnowski et al., 2020, 2022.) The attention model nests mine in a sense made precise in Section 2.4. The literature on persuasion with "outside-information acquisition" studies Receiver's costs of acquiring extra information beyond what Sender provides (Brocas and Carrillo, 2007; Felgenhauer, 2019; Bizzotto et al., 2020; Ravid et al., 2022; Dworczak and Pavan, 2022; Matysková and Montes, 2023). The focus is on how payoffs and information change as outside information becomes cheaper. [6]

Outline Section 2 describes the model and Section 3 analyzes the Receiver's equilibrium attention and action. Section 4 describes the equivalence between persuasion mechanisms and signals, and Section 5 considers upper censorships and applications. Omitted proofs are in Appendix B.

2 Model

2.1 Players, actions, and payoffs

Two players, Sender (he) and Receiver (she), play the following persuasion game. Receiver chooses action $a \in \{0, 1\}$ and effort $e \in [0, 1]$, knowing her type $(c, \lambda) \in [0, 1]^2$. The material payoff of action a, given state $\theta \in [0, 1]$, is $a(\theta - c)$, and the cost of effort e is $\lambda k(e)$, for a continuous function $k\colon [0, 1] \longrightarrow \mathbb{R}$ and given the Receiver's type

[5] Related research includes: Knoepfle's (2020) dynamic model in which Receiver decides whether to pay attention and Sender only values attention; Board and Lu's (2018) search model in which consumers decide whether to pay for a seller's signal or to purchase a product.

[6] If Sender faces a "psychological" audience, the Receiver's belief arises from an optimization problem, which typically occurs after the signal realization (Lipnowski and Mathevet, 2018; Galperti, 2019; Beauchêne et al., 2019; de Clippel and Zhang, 2022; Augias and Barreto, 2024) — and not before, as in this paper.

(c, λ).[7] The Receiver's utility U_R is her material payoff net of effort cost,

$$U_R(\theta, a, e; c, \lambda) := a(\theta - c) - \lambda k(e).$$

For type (c, λ), the *cutoff type* c represents the opportunity cost of taking the risky action, 1, and the *attention type* λ scales the effort cost. Sender chooses a signal — a measurable $\pi\colon [0,1] \longrightarrow \Delta M$, in which ΔM is the set of Borel probability distributions over the rich message space M — about the state and his utility is given by $U_S(a) := a$.[8]

2.2 Information and timing

Information The state θ is distributed according to an atomless distribution $F_0 \in \mathcal{D}$, the *prior belief*, with mean x_0, letting $\mathcal{D}$ be the set of distributions over $[0,1]$ identified by their distribution functions. The Receiver's type is independent of θ and admits a marginal distribution of the attention type λ, $G \in \mathcal{D}$, and a conditional distribution of the cutoff c given λ, $G(\cdot|\lambda) \in \mathcal{D}$.

Timing First, Sender chooses a signal about the state, without knowing either the state or the Receiver's type (c, λ). Second, Receiver chooses effort e, knowing her type (c, λ) and the signal. Then, Nature draws the state θ according to F_0, and the signal realization from $\pi(\theta)$. Afterwards, with probability e, Receiver observes the signal realization, updates her belief about the state using Bayes' rule, and chooses an action given the *posterior* belief; with probability $1 - e$, Receiver does not observe the signal realization and chooses an action given the prior belief. The equilibrium notion is Perfect Bayesian Equilibrium (Appendix A.2).

Notation We endow $\mathcal{I}$ with the pointwise order. The subdifferential and right derivative of $I \in \mathcal{I}$ at $x \in \mathbb{R}_+$ are denoted by, respectively, $\partial I(x)$ and I'. We use $\leq$ for all partial orders and $<$ for the asymmetric part of $\leq$. For posets S and T, the function $g\colon S \times T \longrightarrow \mathbb{R}$ exhibits *increasing differences* if $t \longmapsto g(s', t) - g(s, t)$ is nondecreasing for all $s', s \in S$ with $s < s'$, and exhibits *strictly increasing differences* if $t \longmapsto g(s', t) - g(s, t)$ is increasing for all $s', s \in S$ with $s < s'$.

[7]The results extend to the case of nonlinear cost in λ (see, e.g., Appendix B.2), linearity is maintained to ease interpretation.

[8]For this game, letting $M = [0,1]$ is sufficient (Section A.2); the representation of signals as convex functions used in the rest of the paper is in Section 2.3.

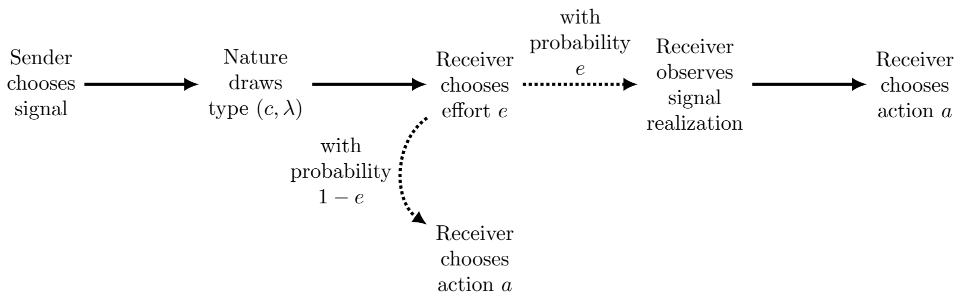


Figure 2: The timing of the game.

2.3 Information policies

Without loss, signals can be represented by the distributions they induce about the posterior belief's mean on a Bayesian player who observes the signal realization.[9] Given the presence of Receiver's effort, it pays off to represent signals by the integrals of such distributions, called "information policies" (Lemma 2). We introduce the notation necessary to state this second equivalence, i.e., between posterior-mean distributions and their integrals. Let's define the *information policy of* $F \in \mathcal{D}$ as the function $I_F \colon \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ such that

$$I_F \colon x \longmapsto \int_0^x F(y)\,\mathrm{d}y,$$

the set of feasible distributions $\mathcal{F} := \{F \in \mathcal{D} \mid I_F(1) = I_{F_0}(1), \text{ and } I_F \leq I_{F_0}\}$, and the set of information policies $\mathcal{I} := \{I \colon \mathbb{R}_+ \longrightarrow \mathbb{R}_+ \mid I \text{ is convex and } I_{\overline{F}} \leq I \leq I_{F_0}\}$, in which $\overline{F}$ is the distribution of an atom at the prior mean, so $I_{\overline{F}} \colon x \longmapsto (x - x_0)_+$. Figure 3 illustrates the set $\mathcal{I}$ and Blackwell's order on $\mathcal{I}$. Signals are identified with information policies via to the following result.

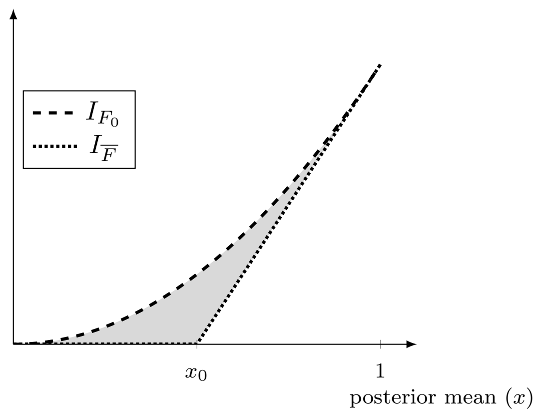
Lemma 1. *The following hold:*

1. *If $F \in \mathcal{F}$, then $I_F \in \mathcal{I}$;*

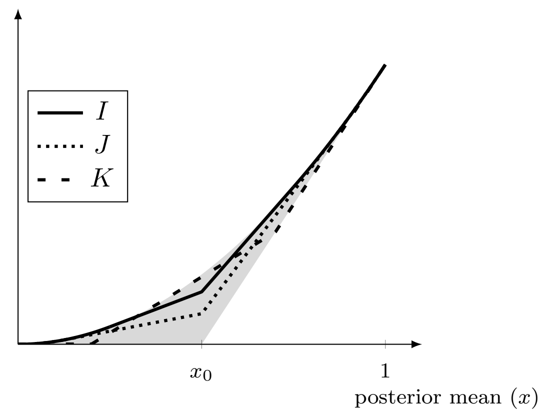
2. *If $I \in \mathcal{I}$, then $I' \in \mathcal{F}$, extending I to take value 0 at every $x < 0$.*

Proof. See Gentzkow and Kamenica (2016) and Kolotilin (2018). **QED**

[9] Signals can be represented by the their posterior-mean distributions in mean-measurable models — as this model — with costless Receiver's attention — unlike this model. Appendix A.1 shows that this equivalence holds for the present model.



(a) The set $\mathcal{I}$ is the set of convex functions that lie between I_{F_0}, corresponding to a fully-informative signal, and $I_{\overline{F}}$, corresponding to an uninformative signal. The prior is a uniform distribution for this figure and the following ones.



(b) Information policy I is more informative than information policy J in the Blackwell sense iff: $I \geq J$. Information policies K and I are not comparable.

Figure 3: Panel (a) illustrates the set of information policies, panel (b) illustrates the Blackwell's order of information policies.

As an implication, Sender chooses $I \in \mathcal{I}$ in the first stage of the game. Thus, the Receiver's posterior mean is drawn from the distribution I' with probability corresponding to her effort, and is equal to x_0 with the remaining probability (Figure 2).

Definition 1. An *equilibrium* is a tuple $\langle I, e(\cdot), \alpha \rangle$, in which $I \in \mathcal{I}$ is the Sender's information policy, $e(c, \lambda, \hat{I}) \in [0, 1]$ is the Receiver's effort given her type (c, λ) and information policy $\hat{I}$, and $\alpha(c, \lambda, x) \in [0, 1]$ is the probability that Receiver chooses action 1 given type (c, λ) and posterior mean x, for a Perfect Bayesian Equilibrium of the game and appropriate measurability requirements (Appendix A.2).

2.4 Discussion and interpretation

Attention effort The term $\lambda k(e)$ in the Receiver's utility represents her attention cost. In particular, let's view e as the *attention effort* exerted by Receiver and look at the effort-choice stage for nondecreasing k. A higher attention effort implies more Receiver's information, in the Blackwell sense, and more costs. The functional form of the effort cost — which includes fixed costs — is general. The model captures a plethora of attention- and non-attention-related phenomena. Examples of costly attention include cognitive difficulties that are psychologically taxing to overcome and limited memory. In contrast, when choosing the probability of being exposed to the media and subscribing to newspapers, the opportunity cost of being attentive is relevant.

Costless-attention benchmark The special case of the model in which effort is costless — i.e., the distribution of λ puts full mass at 0 — is studied by prior work (Kolotilin et al., 2017). There exists an optimal signal that is an upper censorship (Figure 1), for single-peaked distribution of the Receiver's cutoff type (Theorem 3), and signals are equivalent to persuasion mechanisms (Theorem 1).

Symmetric-information benchmark Receiver does not have private information if the type distribution is degenerate. Wei (2021) considers such a model with binary state and two more differences. First, Receiver's attention cost is posterior separable. Second, Receiver's strategy space contains the present one: she chooses signals about the Sender's signal, which include mixtures between no information and full information about the Sender's signal realization. These mixtures constitute the class of signals induced by the choice of effort.

Bloedel and Segal (2021) study a model in which: the state space is a continuum, the cost of attention is proportional to an expected entropy reduction taking into account Receiver's learning about the Sender's signal, and the Receiver's strategy space is fully general.[10] The optimal signal is an upper censorship, although for a different reason than this paper. In particular, Sender perceives Receiver's action as random, given a signal realization, because of her attention strategy; in our model, instead, the randomness arises due to both the Receiver's effort and asymmetric information (as discussed in Section 5.) Bloedel and Segal (2021) also study the symmetric-information benchmark, as one of the alternatives to their model. Due to the binary action and symmetric information, there exists an optimal signal that is a binary signal. There also exists an optimal signal that is an upper censorship (Lemma B.9), and signals are equivalent to persuasion mechanisms (Theorem 1).

[10] This review mentions the results of Bloedel and Segal (2021) in case Sender's utility is U_S and the attention cost is entropic, even if their model encompasses other possibilities.

3 Persuasion

3.1 Receiver's action and effort

This section studies Receiver's equilibrium choices, given type (c, λ).

Given the posterior mean x, Receiver chooses action 1 if $x > c$ and action 0 if $x < c$. Because $\theta \longmapsto U_R(\theta, a, e; c, \lambda)$ is affine, we express the Receiver's expected utility from choosing the action optimally given posterior mean x as

$$U_R(x, e, c, \lambda) := \max_{a \in \{0,1\}} U_R(x, a, e; c, \lambda).$$

To characterize Receiver's effort choice, let's define the *marginal benefit of effort given the information policy* I as the difference between the expected utility from choosing the action optimally with and without the information contained in I: $\int_{[0,1]} U_R(x, e, c, \lambda) - U_R(x_0, e, c, \lambda)\, \mathrm{d}I'(x)$.[11] The *net informativeness of the information policy* I is defined as the difference between I and the uninformative-signal information policy, $I_{\overline{F}}$ (Figure 4a). Using the operator $\Delta \colon I \longmapsto I - I_{\overline{F}}$, the following result shows that the marginal benefit of effort is given by the net informativeness evaluated at c, for all I.

Lemma 2 (Net informativeness)**.** *Given information policy I and Receiver's effort e, the following holds:*

$$\int_{[0,1]} U_R(x, e, c, \lambda) - U_R(x_0, e, c, \lambda)\, \mathrm{d}I'(x) = \Delta I(c).$$

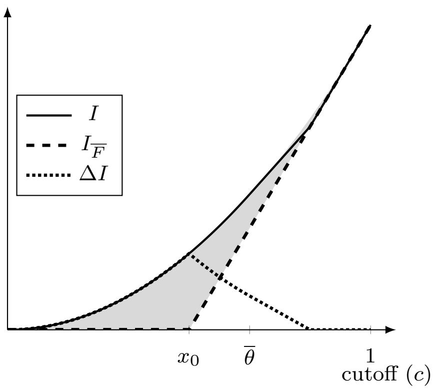
The net informativeness ΔI is single-peaked, with a peak at the prior mean x_0, by construction (Lemma 1). Extreme-cutoff types benefit the least from observing the signal realization because, intuitively, they are the most certain about the optimal action at the prior.

The following result characterizes Receiver's equilibrium choices.

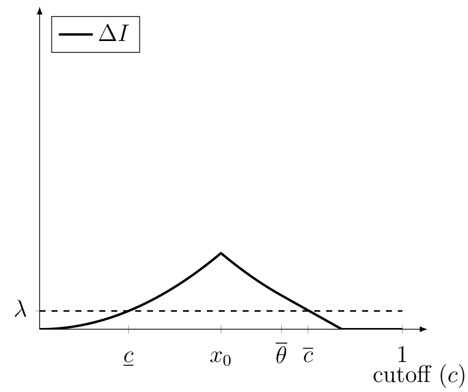
Lemma 3 (Receiver's rationality)**.** *If $\langle I, e(\cdot), \alpha \rangle$ is an equilibrium, then, for every information policy $\hat{I}$:*

1. $1 - \int_{[0,1]} \alpha(c, \lambda, x)\, \mathrm{d}\hat{I}'(x) \in \partial \hat{I}(c)$;

[11] The marginal benefit of effort given I is also referred to as the value of information in the literature.



(a) The net informativeness of I at cutoff c, $\Delta I(c)$, is obtained by subtracting the value of the uninformative-signal information policy at c, $I_{\overline{F}}(c)$, to $I(c)$. The function $c \longmapsto \Delta I(c)$ is single-peaked, with a peak at the prior mean x_0, by construction.



(b) The marginal benefit of effort equals the marginal cost for cutoff types $\underline{c}$ and $\overline{c}$, given linear k and attention type λ (Lemma 2). For cutoffs in $(\underline{c}, \overline{c})$, Receiver chooses effort 1, and for cutoffs in $[0,1] \setminus [\underline{c}, \overline{c}]$ Receiver does not exert effort.

Figure 4: Panel (a) illustrates the construction of the net informativeness of the information policy I, panel (b) illustrates the subset of cutoff types that exert positive effort, given I and linear k, which is an interval. The information policy I is an "upper censorship" in both panels, defined in Section 5.

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

Proof. Part 1. follows from the definition of information policies and the equilibrium properties of α, part 2. follows from Lemma 2 and the equilibrium properties of e. **QED**

The takeaway of Lemma 3 is Part 2., which identifies the net informativeness of I at the Receiver's cutoff as a sufficient statistic for her effort decision. As an implication, the two dimensions of Receiver's type, c and λ, represent her private information about, respectively, her benefit and cost of attention. Part 1. restates the equilibrium conditions that the Receiver's action satisfies.

3.2 Interval structure of the extensive margin

This section studies the Receiver's choice of effort.

The Receiver's *value of information policy $I \in \mathcal{I}$*, given her type (c, λ) and effort

e, is $V_\lambda(e, \Delta I(c)) := e\Delta I(c) - \lambda k(e)$.[12] The value of I exhibits strictly increasing differences in net informativeness and effort, by Lemma 3.

Corollary 1 (Supermodularity)**.** *The Receiver's value of information policy I, $V_\lambda(e, \Delta I(c))$, exhibits strictly increasing differences in e and $\Delta I(c)$.*

As an implication, a more informative Sender's information policy, in the Blackwell sense, makes ex-ante Receiver better off. In particular, I is Blackwell more informative than J iff: $J \leq I$. Hence, if I is more informative than J, I allocates more net informativeness to every type than J. By the increasing-differences property and the envelope theorem (Lemma C.10), Receiver is better off facing I than J.[13] The following result uses the property to characterize the set of cutoffs that exert positive effort, given λ.

Lemma 4 (Interval structure)**.** *Let $\langle \hat{I}, e(\cdot), \alpha\rangle$ be an equilibrium, and define the function $e_\lambda \colon c \longmapsto e(c, \lambda, I)$ for information policy I and attention type λ. The set $e_\lambda^{-1}((0,1])$ is an interval if type (x_0, λ) chooses positive effort, i.e., $e_\lambda(\Delta I(x_0)) > 0$, and it is empty otherwise.*

Proof. Let $\langle \hat{I}, e(\cdot), \alpha\rangle$ be an equilibrium, and let $\lambda \in [0,1]$, $I \in \mathcal{I}$. We start with three preliminary observations. First, $e(c, \lambda, I)$ equals $e^* \circ \Delta I(c)$ for some selection e^* from $\Delta J(c) \longmapsto \arg\max_{e\in[0,1]} V_\lambda(e, \Delta J(c))$, by Lemma 3. Second, every selection from $\Delta J(c) \longmapsto \arg\max_{e\in[0,1]} V_\lambda(e, \Delta J(c))$ is nondecreasing, because V_λ satisfies strictly increasing differences by Corollary 1 via known results (Topkis, 1978, Theorem 6.3). It follows that $e^* \circ \Delta I$ is nondecreasing on $[0, x_0]$ and nonincreasing on $[x_0, 1]$ because ΔI is nondecreasing on $[0, x_0]$ and ΔI is nonincreasing on $[x_0, 1]$.

If $e_\lambda(\Delta I(x_0)) = 0$, then every cutoff c has $e_\lambda(\Delta I(c)) = 0$, by the above observations. Let's suppose that $e_\lambda(\Delta I(x_0)) > 0$. We define $\underline{c} = \sup\{c \in [0, x_0] : e^* \circ \Delta I(c) = 0\}$, if $\{c \in [0, x_0] : e^* \circ \Delta I(c) = 0\} \neq \emptyset$, and $\underline{c} = 0$ otherwise. We define $\overline{c} = \inf\{c \in [x_0, 1] : e^* \circ \Delta I(c) = 0\}$, if $\{c \in [x_0, 1] : e^* \circ \Delta I(c) = 0\} \neq \emptyset$, and $\overline{c} = 1$ otherwise. The claim follows from the next two observations. First, we note that $e^* \circ \Delta I(c) > 0$ only if: $c \in [\underline{c}, \overline{c}]$; second, $c \in (\underline{c}, \overline{c})$ only if $e^* \circ \Delta I(c) > 0$. **QED**

[12] Up to a constant term, $V_\lambda(e, \Delta I(c))$ equals the expected Receiver's payoff: $V_\lambda(e, \Delta I(c)) = \int_{[0,1]} U_R(x, e, c, \lambda)\, \mathrm{d}I'(x) + x_0 - c + I_{\overline{F}}(c)$.

[13] This observation is also an implication of Blackwell's theorem, Corollary 1 is a stronger result (see Section 4.)

For intuition, let's consider a linear effort cost, which captures a price or fixed cost of gathering information, i.e., $k(e) = e$. Receiver compares her marginal cost, λ, and marginal benefit, $\Delta I(c)$, of effort. As shown in Figure 4b, in equilibrium:

$$e(c, \lambda, I) = 1 \implies \Delta I(c) \geq \lambda,$$
$$e(c, \lambda, I) = 0 \implies \Delta I(c) \leq \lambda.$$

Moreover, the net informativeness of I at a cutoff is single-peaked as a function of the cutoff (Figure 3). As an implication, the set of cutoff types that exert positive effort is an interval.[14] The proof of Lemma 4 generalizes the first part of the argument. Specifically, Receiver's effort is nondecreasing in net informativeness at her cutoff type, by the supermodularity of Receiver's preferences through comparative statics à la Topkis (1978).

4 Persuasion mechanisms

This section studies the equivalence between information policies and persuasion mechanisms.

Definition 2. A *persuasion mechanism* $I_\bullet$ is a list of information policies: $I_\bullet = (I_r)_{r \in R}$, with R equal to the support of the Receiver's type. A persuasion mechanism $I_\bullet$ is *incentive-compatible* if

$$\max_{e \in [0,1]} V_\lambda(e, \Delta I_{(c,\lambda)}(c)) \geq \max_{e \in [0,1]} V_\lambda(e, \Delta I_r(c)),$$

for every type (c, λ) and report r.

Our focus on IC mechanisms references to an auxiliary game. First, Sender publicly commits to a mechanism that selects an information policy for every *type report*. Second, Receiver makes a report $r \in R$, knowing her true type (c, λ). The rest of the game unfolds as in Section 2.2: Receiver chooses effort e, then she observes the realization of a signal corresponding to information policy I_r with probability e, and lastly chooses an action. We are interested in equilibria in which Receiver truthfully reports the type, which is without loss via a revelation-principle argument.

[14] Receiver's effort at the boundary is determined through equilibrium selection. The selection is not relevant in equilibrium, for atomless cutoff distributions (Lemma B.3).

We consider a persuasion mechanism $I_\bullet$ to be implementable by an information policy J if: every Receiver's type chooses the same action and effort under truthful reporting given mechanism $I_\bullet$, and in some equilibrium of the subgame that starts with the Sender's choice of information policy J (Section 2.2).

Definition 3. An IC persuasion mechanism $I_\bullet$ is *equivalent to information policy J* if, for every type (c, λ):

1. $\arg\max_{e\in[0,1]} V_\lambda(e, \Delta I_{(c,\lambda)}(c)) \subseteq \arg\max_{e\in[0,1]} V_\lambda(e, \Delta J(c))$,
2. $\partial I_{(c,\lambda)}(c) \subseteq \partial J(c)$ if $(0, 1] \cap \arg\max_{e\in[0,1]} V_\lambda(e, \Delta I_{(c,\lambda)}(c)) \neq \emptyset$.

If attention is costless, Definition 3 is the same as in Kolotilin et al. (2017, p. 1954). The novelty is item 1., which requires type (c, λ) to choose the same effort under $I_\bullet$ as under the signal that implements $I_\bullet$.[15] Item 2. in Definition 3 does not deal with a type who only exerts 0 effort under truthful reporting given $I_\bullet$. The reason is that the equilibrium action given the prior belief does not depend on Sender's information.[16]

We show that every IC persuasion mechanism is equivalent to a signal.

Theorem 1. *Every persuasion mechanism is equivalent to an information policy.*

This result guarantees that the characterization of the extensive margin of persuasion in Section 3 holds in more general environments, including applications in which multiple information structures are available to decision-makers.

We sketch the intuition and proof of Theorem 1, which leverage Corollary 1. The proof verifies that supermodularity is key by establishing the result for more general Receiver's interim payoff functions (Appendix B.2). Let's claim that the IC mechanism $I_\bullet$ is equivalent to its upper envelope J (Figure 5), defined as $J\colon x \longmapsto \sup_{r\in R} I_r(x)$, under positive Receiver's effort. A report r is *active* at x if $I_r(x) \geq I_{r'}(x)$ for all $r' \in R$. Let's also fix Receiver's type (c, λ). First, we observe that an active report at c maximizes Receiver's expected utility. By Lemma 3, report r affects Receiver's utility only through the local net informativeness $\Delta I_r(c)$. By increasing differences, an

[15] Theorem 1 holds under a slightly stronger version of item 1. in Definition 3, as clear in the proof (Section B.2).

[16] Formally, the reason is that the equivalence of the action decision holds as a consequence of item 1. "for this type." Specifically, $\arg\max_{e\in[0,1]} V_\lambda(e, \Delta I_{(c,\lambda)}(c)) = \{0\}$ implies that $0 \in \arg\max_{e\in[0,1]} V_\lambda(e, \Delta J(c))$ by item 1., and prior decisions are the same under the IC $I_\bullet$ and J, possibly via equilibrium selection.

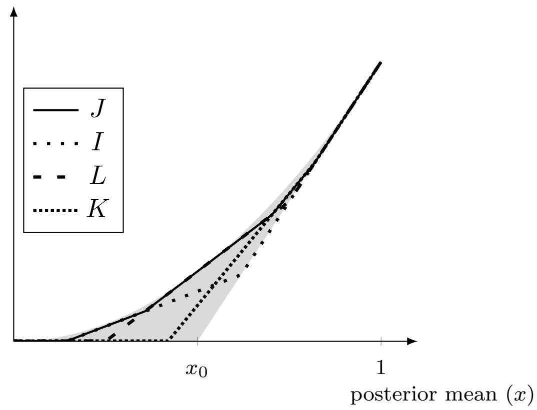


Figure 5: The upper envelope J of the information policies in the persuasion mechanism $I_\bullet = (I, L, K)$ is an information policy. The proof of Theorem 1 shows that the upper envelope of an IC persuasion mechanism $I_\bullet$ implements the same Receiver's action and effort in the game (Section 2) as $I_\bullet$ under truthful reporting.

active report at c makes type (c, λ) weakly better off than any other report (Corollary 1, via the envelope theorem for supermodular programming, Lemma C.10.) Hence, an active report at c maximizes Receiver's expected utility at the reporting stage.

Towards the equivalence with respect to effort, we strengthen the observation: A non-active report makes Receiver strictly worse off than an active report. This conclusion uses both the fact that Corollary 1 establishes strictly increasing differences and type (c, λ)'s positive effort (Lemma C.10). To build on this conclusion, let's order information policies according to the type-specific relation $\leq_c$, defined by $\hat{I} \leq_c \hat{J}$ iff $\Delta\hat{I}(c) \leq \Delta\hat{J}(c)$. The linear order $\leq_c$ is a completion of Blackwell's order and ranks the menu's items according to Receiver's expected utility. By the IC property of the mechanism $I_\bullet$, the policy I_r maximizes $\leq_c$ on $I_\bullet$ only if $\Delta I_r(c) = \Delta I_{(c,\lambda)}$.[17] Hence, $J(c) = I_{(c,\lambda)}(c) \geq I_r(c)$, for every report r. Then, an application of Lemma 3 completes the argument for the equivalence with respect to effort. In particular, the net informativeness, $\Delta J(c)$, is the only component of the information policy $I_{(c,\lambda)}$ that affects the effort decision in the IC mechanism $I_\bullet$.

The equivalence with respect to action decisions follows from simple convex analysis, due to our results. In particular, it holds that $\partial I_r(x) \subseteq \partial J(x)$ if report r is active at x. The proof in Appendix B.2 uses a continuity argument to cover the case of zero effort.

[17]Blackwell's theorem does not suffice for this conclusion, which uses (i) Corollary 1, (ii) the envelope theorem applied to our setup (Appendix, Lemma C.10), and (iii) the completeness property of $\leq_c$.

5 Optimality properties of upper censorships

This section discusses the properties of the following class of information policies.

Definition 4. The $\overline{\theta}$ *upper censorship* is the unique information policy $I_{\overline{\theta}} \in \mathcal{I}$ such that:

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

for $\overline{\theta} \in [0, 1]$.

The case of a single-peaked marginal distribution of the cutoff type is relevant for applications (Romanyuk and Smolin, 2019; Kolotilin et al., 2022; Gitmez and Molavi, 2023; Shishkin, 2024; Augias and Barreto, 2024; Sun et al., 2024).

Assumption 1 (Single-peakedness)**.** The conditional cutoff distribution given attention type λ admits a density $g(\cdot|\lambda)$ such that: (i) $g(\cdot|\lambda)$ is absolutely continuous, and (ii) there exists $p \in [0, 1]$ such that: for all λ, $g(\cdot|\lambda)$ is nondecreasing on $[0, p]$ and nonincreasing on $[p, 1]$.

The class of single-peaked distributions includes the standard uniform and the $[0, 1]$-truncated normal. We say that *strict single-peakedness* holds if: Assumption 1 holds and $g(\cdot|\lambda)$ is increasing on $[0, p]$ and decreasing on $[p, 1]$.

We first establish that an equilibrium exists, and that the Sender's equilibrium expected utility is unique.

Theorem 2. *Under Assumption 1, the Sender's expected utility is the same in every equilibrium and an equilibrium exists.*

In the Appendix (Lemma B.3), we establish that continuity of the cutoff distribution (in Assumption 1) ensures that Sender is indifferent among all Receiver's best responses.[18]

The next result shows that an optimal signal that is an upper censorship exists.

[18] Lipnowski et al. (2024) show that uniqueness obtains in a general model, which does not nest ours. Their Corollary 1 is similar to our observation, even if our proof leverages the convexity of (i) information policies and (ii) Receiver's interim utility $a \longmapsto \max_{e \in [0,1]} V_\lambda(e, a)$ (which, in turn, obtains from the envelope theorem for supermodular optimization, Lemma C.10.)

Theorem 3. *Under Assumption 1, there exists an equilibrium in which the Sender's information policy is an upper censorship.*

Given Theorem 1, Theorem 3 shows that the extensive margin of a complicated optimal persuasion mechanism can be studied via an upper censorship. Moreover, Theorem 3 reduces the Sender's optimization to a uni-dimensional problem.

In the case of costless attention and Sender-optimal equilibria, the argument for Theorem 3 rests on the shape of the exogenous noise in Receiver's action given a posterior belief, from the Sender's perspective. The Sender's expected utility at posterior mean x is $H(x)$, letting H be the distribution of the cutoff type. By single-peakedness, H is "S-shaped." So, Sender is risk-lover conditionally on low posterior means, i.e., $x < p$; and he is risk-averse for high posterior means. In particular, a mean-preserving spreads around a low posterior mean increases Sender's expected utility. Second-order dominance is related to the informativeness of Sender's strategy, because: $F \in \mathcal{F}$ is a mean-preserving spread of $\hat{F} \in \mathcal{F}$ iff F is more Blackwell informative than $\hat{F}$ (i.e., iff $I_{\hat{F}} \leq I_F$, Figure 1b.) Moreover, the upper censorship $I_{\overline{\theta}}$ induces either full information conditionally on the state being lower then the threshold $\overline{\theta}$, or no information except that $\theta > \overline{\theta}$. Thus, intuitively, upper censorships induce posterior-mean distributions that align with the Sender's interests. In the next paragraph, we describe how this intuition is adjusted if effort is endogenous; so, the relevant information policy becomes $x \longmapsto eI(x) + (1-e)I_{\overline{F}}(x)$.

We claim that Receiver's effort is affected by the signal's informativeness in a way that aligns with Sender's interests. In particular, let's suppose that Sender increases the net informativeness of posterior mean x: $\Delta I(x)$. This change induces cutoff-type x to pay extra attention, via the envelope theorem for supermodular optimization (lemmata 3 and C.10.) If cutoff-type x increases her effort, she gathers more information, because $x \longmapsto eI(x) + (1-e)I_{\overline{F}}(x)$ increases in the Blackwell's order as e increases. Thus, by increasing the net informativeness of I — the Sender's information policy — at x, Sender spreads out the Receiver's posterior-mean distribution around x. This argument, however, is "local." Specifically, net informativeness $\Delta I(x)$ increases only via a new information policy that satisfies the convexity constraint in $\mathcal{I}$. The proof uses the described argument to construct $I_{\overline{\theta}} \in \mathcal{I}$ that improves upon I, for arbitrary I.

Sender optimally provides more information as Receiver's attention cost increases, for small attention costs. We say that $I \in \mathcal{I}$ is an *optimal information policy* if there exists an equilibrium in which Sender chooses I.

Proposition 1. *Let strict single-peakedness hold, F_0 admit a density, k be linear, and the attention type put full mass at λ. Let I_{θ_ε} be an optimal upper censorship if $\lambda = \varepsilon$, and I_η an optimal upper censorship if $\lambda = 0$, with $\eta \in (0,1)$. Then: $\theta_\varepsilon > \eta$ for all sufficiently small $\varepsilon > 0$.*

The same qualitative result holds in Wei (2021, Proposition 7), as well as an in Bizzotto et al. (2020, Figure 1), Brocas and Carrillo (2007, p. 944), and in the two-state-two-action case of Matysková and Montes (2023). Let's describe the intuition in the symmetric-information benchmark, for $c > x_0$. In order to persuade Receiver to take action 1, Sender solves the maximization of Receiver's action, given the constraint that she exerts effort 1. Let's observe that the "participation constraint" binds (Lemma B.4). Let's suppose this were not the case. Sender increases the probability of a posterior mean x with $x \geq c$ as much as possible. Specifically, he induces the mean $x = c$ with the highest probability that satisfies Bayes' rule (Kamenica and Gentzkow, 2011). Hence, Receiver faces two possibilities: she is indifferent between the actions, with some probability ($x = c$); she finds it optimal to choose the riskless action, 0, with the remaining probability ($x < c$). So, information brings no value, and the constraint binds. Thus, Sender provides more useful information if $\lambda > 0$ than if $\lambda = 0$.[19] Proposition 1 shows that the insight generalizes, for small λ. In general, a change in the censorship state θ_ε affects the extensive margin given Receiver's private information. However, only the extensive margin's upper bound ($\bar{c}$ in Figure 4b) is affected by small changes in θ_ε around η, because a nontrivial upper censorship is optimal if $\lambda = 0$.[20] This argument leads to Proposition 1.

In applications to media capture, Sender cares directly about Receiver's material payoff (Kolotilin et al., 2022) and attention (Gehlbach and Sonin, 2014). In the former case, Sender is a government that weighs social utility. In the latter case, Sender is a dictator and owns a state's media, so he collects advertisement revenues. The next result shows that an extension of the class of upper censorships contains a Sender-optimal signal for general Sender's utility.[21] A *bi-upper censorship* is an

[19] Lemma B.4 establishes that Sender provides more Blackwell information to Receiver if $\lambda > 0$ than if $\lambda = 0$ under symmetric information.

[20] Figure 4a shows the net informativeness of an upper censorship, which is 0 for cutoff types higher than the censorship state $\overline{\theta}$.

[21] Equilibrium existence is not established for this extension. The difficulty lies in establishing

information policy I such that:

$$I(x) = \begin{cases} I_{F_0}(x) & , x \in [0, \theta_1] \\ I_{F_0}(\theta_1) + F_0(\theta_1)(x - \theta_1) & , x \in (\theta_1, x_1] \\ I_{\overline{F}}(x_2) - m(x_2 - x) & , x \in (x_1, x_2], \end{cases}$$

for $m = \frac{I_{\overline{F}}(x_2) - [I_{F_0}(\theta_1) + F_0(\theta_1)(x_1 - \theta_1)]}{x_2 - x_1}$ and $0 \leq \theta_1 \leq x_1 \leq x_2 \leq 1$ (Figure 1).

Proposition 2. *Let Assumption 1 hold, with peak $p \geq x_0$, k be linear, the attention type put full mass at λ, and Sender's utility be given by $U_G(\theta, a, e, c, \lambda) := a + \gamma e + \rho U_R(\theta, a, e, c, \lambda)$, for $\gamma \geq 0$ and $\rho \geq 0$. For every equilibrium $\langle I, e(\cdot), \alpha \rangle$, there exists a bi-upper censorship with a weakly greater Sender's expected utility given $e(\cdot)$ and α than I.*

The case of $\gamma = 0$ is studied by Kolotilin et al. (2022), who find that upper censorships are optimal signals, and by Gitmez and Molavi (2023). Sender's preferences given $\rho = 0$ are introduced by Gehlbach and Sonin (2014), who assume binary state and Sender's signal. The requirement that the cutoff's peak satisfies $p \geq x_0$ represents sufficient ex-ante disagreement between Sender and Receiver, as in Shishkin (2024) and for symmetric cutoff densities. The proof constructs a bi-upper censorship that replicates the same extensive margin as an arbitrary $I \in \mathcal{I}$, and that improves upon I in terms of expected Receiver's action and utility. Thanks to the same intuition as for Theorem 3, we construct an upper censorship $I_{\overline{\theta}}$ that improves upon I for $\gamma = 0$, because the argument for $\rho = 0$ is preserved if $\rho > 0$ (as in Kolotilin et al., 2022.) The role of the additional censorship region is to modify $I_{\overline{\theta}}$ in a way that replicates the extensive margin of I. Specifically, the second threshold state is constructed to increase the marginal benefit of effort of certain types in case $I_{\overline{\theta}}$ induces fewer cutoff types than I to exert effort, for $\gamma > 0$.

continuity of the extensive margin, i.e., continuity of $F \longmapsto \underline{c}(\Delta I_F)$ and $F \longmapsto \overline{c}(\Delta I_F)$ when $\mathcal{F}$ is endowed with the L^1-norm topology (which makes $\mathcal{F}$ compact, see Kleiner et al., 2021.)

Appendices

A Equilibrium

A.1 Preliminaries

We claim that the Sender's signal impacts the decisions and payoffs of both Sender and Receiver only through the distribution of the posterior mean that it induces on a Bayesian agent who always observes the signal realization.

Type-t Receiver's optimal action, given posterior belief $\mu \in \mathcal{D}$ and $t = (c, \lambda)$, depends on the belief μ only through its mean $\overline{x}_\mu := \int_{[0,1]} \theta \, \mathrm{d}\mu(\theta)$. The Receiver's *material payoff at belief* μ is her expected material payoff given belief μ:

$$v_t(\mu) := \begin{cases} \int_{[0,1]} (\theta - c) \, \mathrm{d}\mu(\theta), & \text{if } \overline{x}_\mu \geq c, \\ 0, & \text{if } \overline{x}_\mu < c. \end{cases}$$

We note that $v_t(\mu)$ depends on the belief μ only through x_μ. If the Sender's signal induces the Bayes-plausible distribution over posterior beliefs p (Kamenica and Gentzkow, 2011), type-t Receiver chooses effort to maximize her expected utility

$$e \int_{\mathcal{D}} v_t(\mu) \, \mathrm{d}p(\mu) + (1 - e) v_t(F_0) - \lambda k(e). \tag{1}$$

Thus, Receiver's action, effort and her payoff depend on the Sender's signal only via the distribution of the posterior mean (i.e., the distribution of x_μ implied by p.) The claim follows from the Sender's payoff function, which depends on the signal only via the Receiver's choice of action (we note that the same conclusion holds if Sender's utility is U_G as defined in Proposition 2.)

A.2 Equilibrium definition

We define a Perfect Bayesian Equilibrium in which Sender directly chooses an experiment $F \in \mathcal{F}$. From Section A.1, this approach is without loss. From Lemma 1, the equilibrium notion is essentially the same as in the text (Section 2.2). Let T denote the support of Receiver's type. Given $F \in \mathcal{F}$ and an effort $e \in [0, 1]$, we define $e \odot F = eF + (1-e)\overline{F}$, and note that $e \odot F \in \mathcal{F}$. An equilibrium is a tuple $\langle F, e(\cdot), \alpha \rangle$, in which $F \in \mathcal{F}$, $e(\cdot, \hat{F}) \colon T \longrightarrow [0, 1]$ is measurable for all $\hat{F} \in \mathcal{F}$, $\alpha(\cdot, x) \colon T \longrightarrow [0, 1]$ is measurable for all $x \in [0, 1]$, and $\alpha(c, \lambda, \cdot) \colon [0, 1] \longrightarrow [0, 1]$ is measurable for all $(c, \lambda) \in T$, such that:

1. α satisfies a Opt:

$$\alpha(c, \lambda, x) > 0 \text{ only if } 1 \in \arg\max_{a \in \{0,1\}} U_R(x, a, e, c, \lambda)$$

for all $x \in [0, 1]$, $(c, \lambda) \in T$;

2. $e(\cdot)$ satisfies e Opt:

$$e(c, \lambda, \hat{F}) \in \arg\max_{e \in [0,1]} \int_{[0,1]} \max_{a \in \{0,1\}} U_R(x, a, e, c, \lambda) \, \mathrm{d}(e(c, \lambda, F) \odot F)(x)$$

for all $(c, \lambda) \in T$, $\hat{F} \in \mathcal{F}$;

3. F is rational for Sender, given $(\alpha, e(\cdot))$, that is: F maximizes

$$F \longmapsto \int_{[0,1]} \int_{[0,1]} \int_{[0,1]} \alpha(x, c, \lambda) \, \mathrm{d}(e(c, \lambda, F) \odot F)(x) \, \mathrm{d}G(c|\lambda) \, \mathrm{d}G(\lambda)$$

on $\mathcal{F}$.

In the Appendix, we use e to denote both a typical level of effort in $[0, 1]$ and the typical function $e(\cdot)$ in the equilibrium definition, for notational convenience. Note that: (i.) U_R is bounded so the maximization in a Opt is well-defined; (ii.) $e \longmapsto U_R(x, a, e, c, \lambda)$ is continuous for all x, a, c, λ, so the set of maximizers in e Opt is nonempty; (iii.) lemmata B.1 and B.3 establish that the maximization in (3.) is well-defined, given items 1. and 2.

B Proofs

We endow $\mathcal{F}$ with the L^1 norm, which metrizes weak convergence (Machina, 1982, Lemma 1).

B.1 Proof of Lemma 2

Proof. Let's define Receiver's equilibrium expected material payoff given $I \in \mathcal{I}$ and $e \in [0, 1]$, $V := \int_{[0,1]} U_R(x, e, c, \lambda) \, \mathrm{d}I'(x) + \lambda k(e)$. By definition of U_R, letting $\alpha(c, x)$ be any distribution over $\{0, 1\}$ such that $\alpha(c, x)\big(\arg\max_{a \in \{0,1\}} a(x - c)\big) = 1$ for every posterior mean x, we have:

$$\begin{aligned} V &= \int_{[c,1]} x - c \, \mathrm{d}I'(x) - \big(1 - \alpha(c, c)(\{1\})\big)\big(I'(c) - I'(c^-)\big)(c - c), \\ &= \int_{[c,1]} x - c \, \mathrm{d}I'(x). \end{aligned}$$

By Riemann–Stieltjes integration by parts (Machina, 1982, Lemma 2),

$$V = (1-c) - \int_{[c,1]} I'(x)\,\mathrm{d}x.$$

By absolute continuity of I,

$$V = x_0 - c + I(c). \tag{2}$$

Hence, we have

$$\int_{[0,1]} U(x,e,c,\lambda)\,\mathrm{d}I'(x) - \int_{[0,1]} U(x,e,c,\lambda)\,\mathrm{d}\overline{F}(x) = \Delta I(c).$$

QED

B.2 Proof of Theorem 1

Theorem 1 is implied by the result proved in this section as Proposition B.1.

For this section, we fix a function $f\colon [0,1]\times[0,1] \longrightarrow \mathbb{R}^2$ that satisfies strictly increasing differences, and such that: $f(\cdot,a)$ is continuous for all $a\in[0,1]$, $f(e,\cdot)$ is nondecreasing for all $e\in[0,1]$, the derivative with respect to the variable a, $\frac{\partial f}{\partial a}(e,\cdot)$, exists, is nonnegative and bounded for all $e\in[0,1]$, and $f(e,\cdot)$ is increasing for all $e\in(0,1]$. We also maintain the definitions of the main text except that the following definitions replace the corresponding ones given in the main text.

The *value of an information policy* $I\in\mathcal{I}$ is $V_\lambda(e,\Delta I(c)) := f(e,\Delta I(c)) - K(e,\lambda)$, Receiver's cost of effort $e\in[0,1]$ is $K(e,\lambda)$ for a continuous function $K(\cdot,\lambda)$, we use the shorthand $t=(c_t,\lambda_t)$, and we define the set of optimal efforts

$$E_{\lambda_t}(\Delta I(\zeta_t)) := \arg\max_{e\in[0,1]} V_{\lambda_t}(e,\Delta I(\zeta_t)),$$

and $V_{\lambda_t}(\Delta I(\zeta_t)) := \max_{e\in[0,1]} V_{\lambda_t}(e,\Delta I(\zeta_t))$, for $I\in\mathcal{I}$. A persuasion mechanism $I_\bullet$ is *incentive compatible* (IC) if:

$$t\in\arg\max_{r\in R} V_{\lambda_t}(\Delta I_r(\zeta_t)), \text{for all types } t\in T.$$

Definition 5. An IC persuasion mechanism $I_\bullet$ is *equivalent to an experiment* if there exists information policy I such that, for all $t \in T$:

$$1.E_{\lambda_t}(\Delta I_t(\zeta_t)) \subseteq E_{\lambda_t}(\Delta I(\zeta_t)), \tag{3}$$

$$2.\partial I_t(\zeta_t) \subseteq \partial I(\zeta_t) \quad \text{if } (0,1] \cap E_{\lambda_t}(\Delta I_t(\zeta_t)) \neq \emptyset. \tag{4}$$

Proposition B.1. *Every IC persuasion mechanism is equivalent to an experiment.*

Proof. Let's fix an IC persuasion mechanism $I_\bullet$. The proof has three steps: (1) we define an information policy J, (2) we show that J induces the same effort and (3) action distributions as $I_\bullet$.

(1) Definition of information policy J Let's define the function $I\colon [0,1] \longrightarrow [0,1]$ as follows:

$$I(c) := \sup_{r \in R} I_r(c), \ c \in [0,1]. \tag{5}$$

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

(2) Effort distribution There are two cases.

1. $E_{\lambda_t}(\Delta I_t(\zeta_t)) \cap (0,1] \neq \emptyset$.
2. $E_{\lambda_t}(\Delta I_t(\zeta_t)) = \{0\}$.

First, we consider case (1.). By the envelope theorem (Lemma C.10), we have:

$$V_{\lambda_t}(a) - V_{\lambda_t}(\Delta I_t(\zeta_t)) = \int_{\Delta I_t(\zeta_t)}^{a} \frac{\partial f}{\partial e}(\tilde{a}, e(\tilde{a}))\, \mathrm{d}\tilde{a},$$

for a selection e of E_{λ_t}. Because f exhibits strictly increasing differences, $e(\tilde{a}) \geq e(\Delta I_t(\zeta_t))$ if $\tilde{a} \geq \Delta I_t(\zeta_t)$. By the assumption that $\frac{\partial f}{\partial e}(\tilde{a}, \cdot) > 0$ on $(0,1]$ for all $\tilde{a}$

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

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 $0 \notin E_{\lambda_t}(\Delta J(\zeta_t))$. By Berge's Maximum Theorem (Aliprantis and Border, 2006, Theorem 17.31), E_{λ_t} is upper hemi-continuous and has compact values. Hence, by the sequential characterization of upper hemi-continuity of compact-valued correspondences (Aliprantis and Border, 2006, Theorem 17.16), there exists $\overline{a} \in (\Delta I_t(\zeta_t), \Delta J(\zeta_t))$ and $f > 0$ such that $f \in E_{\lambda_t}(\overline{a})$ (else, define $a_n := \frac{1}{n}\Delta I_t(\zeta_t) + \left(1 - \frac{1}{n}\right)\Delta J(\zeta_t),\ n \in \mathbb{N}$, to get: $a_n \longrightarrow \Delta J(\zeta_t)$ as $n \longrightarrow \infty$, $E_{\lambda_t}(a_n) = \{0\},\ n \in \mathbb{N}$, and $0 \notin E_{\lambda_t}(\Delta J(\zeta_t))$, which contradicts upper hemi-continuity of E_{λ_t}.) By the assumption that $\frac{\partial f}{\partial e}(\tilde{a}, \cdot) > 0$ on $(0, 1]$ for all $\tilde{a}$

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

The above inequality and the envelope theorem imply that

$$V_{\lambda_t}(\Delta J(\zeta_t)) - V_{\lambda_t}(\Delta I_t(\zeta_t)) > 0.$$

Hence, IC does not hold, which is a contradiction. Thus, $0 \in E_{\lambda_t}(\Delta J(\zeta_t))$.

(3) Action distribution Let's suppose that $d \in \partial I_s(\zeta_s)$ and $d \notin \partial J(\zeta_s)$ for some type $s \in T$. Because I_s and J are information policies, they have the same extension on $(-\infty, 0)$ and $\zeta_s > 0$. We have that d is a subgradient of I_s at ζ_s, and d is not subgradient of J at ζ_s; from the fact that $J(\zeta_s) = I_s(\zeta_s)$ — established above —, there exists $x \in \mathbb{R}$ such that

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

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

B.3 Proof of Theorem 2

We establish existence and payoff uniqueness of equilibria under the assumption that the conditional density $c \longmapsto g(c|\lambda)$ is absolutely continuous for all λ, which is maintained in this section.

Definition 6. The experiment $\hat{F} \in \mathcal{F}$ is an *equilibrium experiment* if there exists an equilibrium $\langle F, e, \alpha \rangle$ with $\hat{F}(x) = F(x)$ for all $x \in \mathbb{R}$. The Receiver's *value from the experiment*

$F \in \mathcal{F}$ is: $V_\lambda(\Delta I_F(c)) := \max_{e\in[0,1]} V_\lambda(e, \Delta I_F(c))$. We say that there are *multiple Sender's payoffs* if: there exist equilibria $\sigma = \langle F, e, \alpha\rangle$ and $\tilde\sigma = \langle \tilde F, \tilde e, \tilde\alpha\rangle$ such that: $\hat W(\sigma) \neq \hat W(\tilde\sigma)$, for

$$\hat W(\sigma) := \int_{[0,1]}\int_{[0,1]}\int_{[0,1]} \alpha(x, c, \lambda)\,\mathrm{d}(e(c, \lambda)\odot F)(x)\,\mathrm{d}G(c|\lambda)\,\mathrm{d}G(\lambda).$$

We define the function

$$W\colon F \longmapsto \int_{[0,1]}\int_{[0,1]} V_\lambda(\Delta I_F(c))\frac{\partial g}{\partial c}(c|\lambda)\,\mathrm{d}c\,\mathrm{d}G(\lambda).$$

and

$$W_\lambda\colon F \longmapsto \int_{[0,1]} V_\lambda(\Delta I_F(c))\frac{\partial g}{\partial c}(c|\lambda)\,\mathrm{d}c.$$

We say that $F \in \mathcal{F}$ is W *maximal* if F maximizes W on $\mathcal{F}$.

Lemma B.1. *W is continuous on $\mathcal{F}$.*

Proof. Let's fix λ, $F \in \mathcal{F}$, and $\varepsilon > 0$, and define $p_\lambda := \int_{[0,1]}\left|\frac{\partial g}{\partial c}(c|\lambda)\right|\mathrm{d}c$. Let $\delta := \frac{\varepsilon}{p_\lambda}$ if $p_\lambda > 0$, and let δ be an arbitrary positive number otherwise. Let $H \in \mathcal{F}$ be such that $\int_{[0,1]}|H(x) - F(x)|\,\mathrm{d}x < \delta$. The proof consists of three steps.

We first establish the claim that: $|V_\lambda(\Delta I_H(c)) - V_\lambda(\Delta I_F(c))| < \delta$. By definition of V_λ and the envelope theorem (Lemma C.10), there exists a selection e from $c \longmapsto \arg\max_{e\in[0,1]} e\Delta I_F(c) - \lambda k(e)$ such that:

$$|V_\lambda(\Delta I_H(c)) - V_\lambda(\Delta I_F(c))| = \int_{[\min\{\Delta I_H(c),\Delta I_F(c)\},\max\{\Delta I_H(c),\Delta I_F(c)\}]} e(a)\,\mathrm{d}a.$$

The codomain of e is $[0, 1]$, so, by the above equality:

$$|V_\lambda(\Delta I_H(c)) - V_\lambda(\Delta I_F(c))| \le |\Delta I_H(c) - \Delta I_F(c)|.$$

We have the following chain of inequalities,

$$\begin{aligned}|V_\lambda(\Delta I_H(c)) - V_\lambda(\Delta I_F(c))| &\le \left|\int_{[0,c]} H(x) - F(x)\,\mathrm{d}x\right|\\ &\le \int_{[0,c]}|H(x) - F(x)|\,\mathrm{d}x\\ &\le \delta,\end{aligned}$$

which establishes the claim. Next, we establish the continuity of the function W_λ on $\mathcal{F}$. We have the following chain of inequalities:

$$\begin{aligned}|W_\lambda(H) - W_\lambda(F)| &\leq \int_{[0,1]} |V_\lambda(\Delta I_H(c)) - V_\lambda(\Delta I_F(c))| \left|\frac{\partial g}{\partial c}(c|\lambda)\right| \mathrm{d}c \\ &\leq \delta p_\lambda \\ &\leq \varepsilon.\end{aligned}$$

Thus, W_λ is continuous on $\mathcal{F}$. The result follows from the following chain of inequalities:

$$\begin{aligned}|W(H) - W(F)| &\leq \int_{[0,1]} |W_\lambda(H) - W_\lambda(F)| \, \mathrm{d}G(\lambda) \\ &\leq \varepsilon.\end{aligned}$$

QED

Lemma B.2. *There exists a measurable selection from* $(c, \lambda, x) \longmapsto \max_{a \in \{0,1\}} U_R(x, a, e; c, \lambda)$*, for all* $e \in [0, 1]$*, and there exists a measurable selection from* $(c, \lambda) \longmapsto \arg\max_{e \in [0,1]} e\Delta I_F(c) - \lambda k(e)$*, for all* $F \in \mathcal{F}$*.*

Proof. The nontrivial part is to show (2). Receiver is maximizing a real-valued function that is continuous in c, λ, and the choice variable e. Thus, the Measurable Maximum Theorem holds (Aliprantis and Border, 2006, Theorem 18.19). **QED**

The next result establishes that the Sender's payoff from any information policy is the same for every equilibrium, which is a slightly stronger version of the uniqueness condition in Definition 6. The comparison holds because Definition 6 compares Sender's expected utility from the *equilibrium information policy*, across equilibria, while the proof compares Sender's expected utility from an arbitrary, fixed, information policy, across equilibria.

Lemma B.3. $F \in \mathcal{F}$ *is an equilibrium experiment if, and only if:* F *is* W *maximal. Moreover, there are not multiple Sender's payoffs.*

Proof. We first show that: F is W maximal if, and only if: F is rational for Sender, given (α, e), α satisfies a Opt, and e satisfies e Opt. It suffices to that the mapping $D_\lambda(\cdot, \alpha, e)$ such that

$$D_\lambda(\cdot, \alpha, e) \colon F \longmapsto \int_{[0,1]} \int_{[0,1]} \alpha(x, c, \lambda) \, \mathrm{d}(e(c, \lambda, F) \odot F)(x) \, \mathrm{d}G(c|\lambda) - W_\lambda(F)$$

is constant (in F,) for all λ. As a preliminary step, we note that $e(c, \lambda, F) = e^*_\lambda(\Delta I_F(c))$, for all $c \in [0, 1]$ and a selection e^* from $\Delta I_F(c) \longmapsto \arg\max_{e \in [0,1]} e\Delta I_F(c) - \lambda k(e)$, by e Opt, given F.

First, let's express Sender's expected utility in equilibrium as follows,[22]

$$\hat{W}(F) := \int_{[0,1]}\int_{[0,1]} e^*_\lambda(\Delta I_F(c))(\alpha(x,c,\lambda) - \alpha(x_0,c,\lambda))\,\mathrm{d}F(x)\,\mathrm{d}G(c|\lambda) + \int_{[0,1]} \alpha(x_0,c,\lambda)\,\mathrm{d}G(c|\lambda).$$

Thus, by Lemma 3, there exists a selection d_I^1 from the subdifferential of ΔI_F on $[0,x_0]$ and a selection d_I^2 from the subdifferential of ΔI_F on $(x_0,1]$ such that:

$$-(\hat{W}(F) - \hat{W}(\overline{F})) = \int_{[0,x_0]} e^*_\lambda(\Delta I_F(c))d_I^1(c)\,\mathrm{d}G(c|\lambda) + \int_{(x_0,1]} e^*_\lambda(\Delta I_F(c))d_I^2(c)\,\mathrm{d}G(c|\lambda)$$

By the envelope theorem (Lemma C.10), e^*_λ is a selection from the subdifferential of the convex and nondecreasing function V_λ. By $\Delta I_F \in \mathcal{A}$, ΔI_F is: (i) convex on $[0,x_0]$, and (ii) convex on $(x_0,1]$. Hence: by the rules of subdifferential calculus (Fact C.1), there exists a selection d from the subdifferential of $V_\lambda \circ \Delta I_F$ such that: $d(c) = e^*_\lambda(\Delta I_F(c))d_I^1(c)$, for all $c \in [0,x_0]$, and $d(c) = e^*_\lambda(\Delta I_F(c))d_I^2(c)$, for all $c \in (x_0,1]$. Hence:

$$\begin{aligned} -(\hat{W}(F) - \hat{W}(\overline{F})) &= \int_{[0,x_0]} d(c)\,\mathrm{d}G(c|\lambda) + \int_{(x_0,1]} d(c)\,\mathrm{d}G(c|\lambda) \\ &= \int_{[0,x_0]} d(c)\,\mathrm{d}G(c|\lambda) + \int_{[x_0,1]} d(c)\,\mathrm{d}G(c|\lambda), \end{aligned}$$

in which the second equality uses absolute continuity of $G(\cdot|\lambda)$. By Fact C.1, the composition $V_\lambda \circ \Delta I_F$ is a convex function on $[0,x_0]$, so $V_\lambda \circ \Delta I_F$ is the integral of any selection from the its subdifferential on $[0,x_0]$ (Rockafellar, 1970, Corollary 24.2.1). Similarly, $V_\lambda \circ \Delta I_F$ is a convex function on $[x_0,1]$. By absolute continuity of $g(\cdot|\lambda)$, we integrate by parts to obtain:

$$-(\hat{W}(F) - \hat{W}(\overline{F})) = V_\lambda \circ \Delta I_F(1)g(1|\lambda) - V_\lambda \circ \Delta I_F(0)g(0|\lambda) - \int_{[0,1]} V_\lambda \circ \Delta I_F(c)\frac{\partial g}{\partial c}(c|\lambda)\,\mathrm{d}c.$$

[22]The symbol $\hat{W}$ is used for a slightly different function in Definition 6 because, for notational convenience, the current proof establishes a slightly stronger uniqueness statement than Definition 6 for an additional reason than the aforementioned one. Namely, the proof looks at the conditional expected Sender's utility given λ.

The fact that $\Delta I_F(1) = \Delta I_F(0) = 0$ implies

$$-(\hat{W}(F) - \hat{W}(\overline{F})) = (g(1|\lambda) - g(0|\lambda))V_\lambda(0) - \int_{[0,1]} V_\lambda \circ \Delta I_F(c) \frac{\partial g}{\partial c}(c|\lambda)\,\mathrm{d}c.$$

Hence:

$$\hat{W}(F) = W(F) + \hat{W}(\overline{F}) - (g(1|\lambda) - g(0|\lambda))V_\lambda(0).$$

So:

$$D_\lambda(F, \alpha, e) = \int_{[0,1]} \alpha(x_0, c, \lambda)\,\mathrm{d}G(c|\lambda) - (g(1|\lambda) - g(0|\lambda))V_\lambda(0)$$

Hence, $D_\lambda(\cdot, \alpha, e)$ is constant on $\mathcal{F}$. Hence, F is W maximal if, and only if: F is rational for Sender, given (α, e), α satisfies a Opt, and e satisfies e Opt.

From the above equivalence, it follows that: if $\langle \hat{F}, e, \alpha \rangle$ is an equilibrium, then $\hat{F}$ is W maximal. For the other direction, let F be W maximal. By Lemma B.2, there exist e and α that satisfy the equilibrium measurability conditions, a Opt, and e Opt, given F. Because F is W maximal, F is rational for Sender, given (α, e), by the above equivalence. Thus, $\langle F, e, \alpha \rangle$ is an equilibrium.

As an implication, there are not multiple Sender's payoffs. **QED**

Proposition B.2. *An equilibrium exists.*

Proof. First, we observe that the set $\mathcal{F}$ is compact in the topology induced by the L^1 norm (Kleiner et al., 2021, Proposition 1). The result follows from Weierstrass' Theorem and Lemma B.3 via upper semi continuity of the Sender's maximand in the definition of W maximality (Lemma B.1). **QED**

Proof of Theorem 2

Proof. Theorem 2 is implied by Lemma B.3 and Proposition B.2, given that Assumption 1 contains the continuity requirements assumed in this section. **QED**

B.4 Proof of Theorem 3

Theorem 3 is a consequence of Lemma B.3 and the following property of upper censorship, a version of the following result appears in the working paper Lipnowski et al., 2021, Appendix

A.5; similar results appear in Kolotilin et al. (2017, Theorem 2) and Romanyuk and Smolin (2019, Theorem 2).

Lemma B.4. *Let $I \in \mathcal{I}$ and $\zeta \in [0,1]$. There exists $\theta \in [0,\zeta]$ such that:*

(1.) $I_\theta(\zeta) = I(\zeta)$;

(2.) $I'_\theta(\zeta^-) \leq I'(\zeta^-)$, and

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

.

Proof. Let $\zeta \in [0,1]$. Let $M := \{m \in [0, I'(\zeta^-)] : I(\zeta) + m(x-\zeta) \leq I_{F_0}(x) \text{ for all } x \in [0,\zeta]\}$, and $m := \min M$. We construct an information policy starting from the line $x \longmapsto I(\zeta) + m(x-\zeta)$, via the next three claims.

(1) 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 $m \longmapsto I(\zeta) + m(x-\zeta)$ is a continuous function on $[0, I'(\zeta^-)]$; (iii) M is bounded because $I'(\zeta^-) \leq 1$, from $I \in \mathcal{I}$.

(2) There exists $\theta \in [0,\zeta]$ such that $I_{F_0}(\theta) = I(\zeta) + m(\theta-\zeta)$. If $m = 0$, then $0 = I_{F_0}(0) \geq I(\zeta) \geq 0$. Hence, taking $\theta = 0$ verifies our claim. Let $m > 0$, and suppose there does not exist $\theta \in [0,\zeta]$ such that $I_{F_0}(\theta) = I(\zeta) + m(\theta - \zeta)$. There exists $\overline{\varepsilon} > 0$ such that: $I(\zeta) + (m-\varepsilon)(x-\zeta) < I_{F_0}(x)$ for all $x \in [0,\zeta]$ and $0 < \varepsilon \leq \overline{\varepsilon}$. Moreover, for a sufficiently small $\varepsilon > 0$, we have $m - \varepsilon \in M$. Thus, we have a contradiction with the definition of m.

(3) $m \in \partial I_{F_0}(\theta)$ and $I(\zeta)+m(x-\zeta) = I_{F_0}(\theta)+(x-\theta)F_0(\theta)$ for all x. First, we argue that $m \in \partial I_{F_0}(\theta)$. By convexity of I_{F_0} and definition of θ, $x \longmapsto I(\zeta)+m(x-\zeta)$ is tangent to I_{F_0} at θ. Thus, m is a subgradient of I_{F_0} at θ. Now, we argue that $I(\zeta)+m(x-\zeta) = I_{F_0}(\theta)+(x-\theta)F_0(\theta)$ for all x. $m = F_0(\theta)$ because I_{F_0} is differentiable (by the fact that $F_0(x^-) = F_0(x), x \in \mathbb{R}$.) The equality follows because $x \longmapsto I(\zeta) + m(x-\zeta)$ is equal to I_{F_0} at $x = \theta$.

We define the following function.

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

Now, we claim that $I^u = I_\theta$. It suffices to show that: (i) for some $x_u \in [0,1]$

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

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

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

$$I(\zeta) + m(x-\zeta) \geq I_{\overline{F}}(x) \quad , x \in [0, x_u] \tag{6}$$

$$I(\zeta) + m(x-\zeta) \leq I_{\overline{F}}(x) \quad , x \in (x_u, 1]. \tag{7}$$

Let's note that: (a) $I(\zeta) \geq I_{\overline{F}}(\zeta)$; (b) by $m \in \partial I_{F_0}(\theta)$ and $I_{F_0}(1) = I_{\overline{F}}(1)$, we have that $I_{\overline{F}}(1) \geq I(\zeta) + m(1-\zeta)$, and (c) the two functions, $x \longmapsto I(\zeta) + m(x-\zeta)$ and $I_{\overline{F}}$, are affine with slopes, respectively, m and 1, such that: $m \leq 1$.

We proceed to verify that (ii) holds, i.e. $I^u \in \mathcal{I}$, via the next two claims.

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

(2) I^u *is subdifferentiable at* $x \in \{\theta, x_u\}$. First, we argue that m is a subgradient of I^u at θ. This follows from the fact that the slope of I^u at θ is a subgradient of I_{F_0} at θ, and $I^u(\theta) = I_{F_0}(\theta)$. On $[0,\theta]$, $I^u = I_{F_0}$, and on $[\theta_u, \infty)$ I^u is above the line $x \longmapsto I(\zeta) + m(x-\zeta)$. Thus, $m \in \partial I^u(\theta)$. Second, the fact that m is a subgradient of I^u at x_u follows from the claim in (6).

We have established that $I^u = I_\theta$. (1.) and (2.) hold by construction. **QED**

Proof of Theorem 3

Proof. By Lemma B.3, the optimal experiment maximizes

$$\begin{aligned} W\colon F \longmapsto \int_{[0,1]} \int_{[0,p]} & V_\lambda(\Delta I_{\hat{F}}(c)) \frac{\partial g}{\partial c}(c|\lambda)\,\mathrm{d}c \\ & + \int_{[p,1]} V_\lambda(\Delta I_{\hat{F}}(c)) \frac{\partial g}{\partial c}(c|\lambda)\,\mathrm{d}c\,\mathrm{d}G(\lambda). \end{aligned}$$

Suppose two experiments $F, H \in \mathcal{F}$ such that $I_F(x) \geq I_H(x)$ for all $x \in [0, p]$ and $I_F(x) \leq I_H(x)$ for all $x \in [p, 1]$. Because (i) V_λ is nondecreasing, (ii) $\frac{\partial g}{\partial c}(\cdot|\lambda)$ is nonnegative on $[0, p]$ and nonpositive on $[p, 1]$, it follows that $W(F) \geq W(H)$.

The result follows from Lemma B.4. **QED**

B.5 Proof of Proposition 1

The proof has three steps. First, we establish a single-crossing property of the derivative of Sender's payoff from I_θ with respect to the threshold state θ, in a series of three claims. Second, we establish a strictly increasing property of Sender's payoff from I_θ with respect to the threshold state θ and the attention-cost type, given certain conditions, in two claims. The third step verifies that the optimality properties and the hypotheses in the statement of the Proposition imply the aforementioned conditions. The rest of the proof completes the argument.

Proof. Let's fix an equilibrium $\langle F, e(), \alpha \rangle$. $e(c, \lambda, I)$ equals $e^*_\lambda \circ \Delta I(c)$ for some selection e^*_λ from $\Delta J(c) \longmapsto \arg\max_{e \in [0,1]} V_\lambda(e, \Delta J(c))$, by Lemma 3. We define $\underline{c}_\lambda(\Delta I) = \sup\{c \in [0, x_0] : e^*_\lambda \circ \Delta I(c) = 0\}$, if $\{c \in [0, x_0] : e^*_\lambda \circ \Delta I(c) = 0\} \neq \emptyset$, and $\underline{c}_\lambda(\Delta I) = 0$ otherwise. We define $\bar{c}_\lambda(\Delta I) = \inf\{c \in [x_0, 1] : e^*_\lambda \circ \Delta I(c) = 0\}$, if $\{c \in [x_0, 1] : e^*_\lambda \circ \Delta I(c) = 0\} \neq \emptyset$, and $\bar{c}_\lambda(\Delta I) = 1$ otherwise.

(1.) Let strict single-peakedness hold. We claim that the function $(\theta, \zeta) \longmapsto \int_{[\theta,\zeta]} (c - \theta) \frac{\partial}{\partial c} g(c|\lambda)\, \mathrm{d}c$ crosses zero at most once and from above, that is:

$$\int_{[\theta,\zeta]} (c - \theta) \frac{\partial}{\partial c} g(c|\lambda)\, \mathrm{d}c \leq 0 \implies \int_{[\theta',\zeta']} (c - \theta') \frac{\partial}{\partial c} g(c|\lambda)\, \mathrm{d}c < 0,$$

for all $\theta \leq \theta'$ and $\zeta \leq \zeta'$, with $\theta' < \zeta'$, $\theta < \zeta$. If $p \leq \theta'$, the result holds. If $\int_{[\theta,\zeta]} (c - \theta) \frac{\partial}{\partial c} g(c|\lambda)\, \mathrm{d}c \leq 0$, then $p < \zeta$. It holds that:

$$\begin{aligned}\int_{[\theta,\zeta]} (c - \theta) \frac{\partial}{\partial c} g(c|\lambda)\, \mathrm{d}c = \int_{[\theta,\theta')} (c - \theta) \frac{\partial}{\partial c} g(c|\lambda)\, \mathrm{d}c + \int_{[\theta',p)} (c - \theta) \frac{\partial}{\partial c} g(c|\lambda)\, \mathrm{d}c \\ + \int_{[p,\zeta]} (c - \theta) \frac{\partial}{\partial c} g(c|\lambda)\, \mathrm{d}c.\end{aligned}$$

Let $\int_{[\theta,\zeta]}(c-\theta)\frac{\partial}{\partial c}g(c|\lambda)\,\mathrm{d}c \leq 0$. Then:

$$\int_{[\theta,\theta')}(c-\theta)\frac{\partial}{\partial c}g(c|\lambda)\,\mathrm{d}c + \int_{[\theta',p)}(c-\theta)\frac{\partial}{\partial c}g(c|\lambda)\,\mathrm{d}c \leq \\ -\int_{[p,\zeta]}(c-\theta)\frac{\partial}{\partial c}g(c|\lambda)\,\mathrm{d}c,$$

which implies, by $\theta' < p$:

$$\int_{[\theta',p)}(c-\theta)\frac{\partial}{\partial c}g(c|\lambda)\,\mathrm{d}c < -\int_{[p,\zeta]}(c-\theta)\frac{\partial}{\partial c}g(c|\lambda)\,\mathrm{d}c.$$

From the above inequality and $p < \zeta$, we have:

$$\int_{[\theta',p)}(c-\theta)\frac{\partial}{\partial c}g(c|\lambda)\,\mathrm{d}c + \int_{[p,\zeta]}(c-\theta)\frac{\partial}{\partial c}g(c|\lambda)\,\mathrm{d}c \\ + \int_{(\zeta,\zeta']}(c-\theta)\frac{\partial}{\partial c}g(c|\lambda)\,\mathrm{d}c < 0,$$

so the claim follows.

(2.) Let strict single-peakedness hold. $\int_{[\theta,\overline{c}]}(c-\theta)\frac{\partial}{\partial c}g(c|\lambda)\,\mathrm{d}c$ is increasing in λ if $p \leq \overline{c}$, for $\overline{c} := \overline{c}_\lambda(\Delta I_\theta)$ and $\overline{c} \in (x_0, 1)$. The claim holds because $\lambda \longmapsto \overline{c}_\lambda(\Delta I_\theta)$ is decreasing under our hypotheses.

(3.) Let Assumption 1 hold. We claim that $\overline{c}_\lambda(\Delta I_\theta) > \theta$, $\underline{p} \leq \overline{c}_\lambda(\Delta I_\theta)$, and $\overline{c}_\lambda(\Delta I_\theta) \in (x_0, 1)$, if: I'_θ maximizes W on $\mathcal{F}$ and $F_0, \overline{F}$ do not maximize W on $\mathcal{F}$. If $\overline{c}_\lambda(\Delta I_\theta) \leq \theta$, then F_0 maximizes W on $\mathcal{F}$. If $\overline{c}_\lambda(\Delta I_\theta) < \underline{p}$, then F_0 maximizes W on $\mathcal{F}$. The rest of the claim follows from similar arguments.

Let $x_{\overline{\theta}} := \frac{\int_{\overline{\theta}}^1 \theta\,\mathrm{d}F_0(\theta)}{1-F_0(\overline{\theta})}$, for threshold state $\overline{\theta} \in [0,1]$. By lemmata B.3, we compute the derivative of the Sender's expected utility, given information policy $I_{\overline{\theta}}$, with respect to $\overline{\theta}$, which is:

$$\frac{\partial}{\partial\overline{\theta}}W(I'_{\overline{\theta}}) = \begin{cases} \frac{\partial F_0}{\partial\theta}(\overline{\theta})\int_{[\max\{\overline{\theta},\underline{c}_\lambda(\Delta I_{\overline{\theta}})\},\overline{c}_\lambda(\Delta I_{\overline{\theta}})]}(x-\overline{\theta})\frac{\partial g}{\partial c}(x|\lambda)\,\mathrm{d}x, & \text{if } \theta < \overline{c}_\lambda(\Delta I_\theta) \\ 0, & \text{if } \theta > \overline{c}_\lambda(\Delta I_\theta). \end{cases}$$

As claimed above, under our hypotheses, $\theta_\varepsilon < \overline{c}_\lambda(\Delta I_{\theta_\varepsilon})$. Moreover, by strict single-peakedness, there exists a unique optimal upper censorship I_η if $\lambda = 0$, with $\eta \in (0,1)$ (Kolotilin et al., 2022, Lemma 7). Let's complete the proof.

First, claim 1. implies that $\overline{\theta} \longmapsto W(I'_{\overline{\theta}})$ crosses zero only once and from above: at θ_ε. By claims 2. and 3., $\theta_\varepsilon > \eta$, for $\varepsilon > 0$. **QED**

B.6 Proof of Proposition 2

The proof of Proposition 2 has two steps. The first and main step has the same structure as that of Theorem 3. In particular, Lemma B.5 generalizes the construction of Lemma B.4 to construct: an information policy $I^\star$ that preserves the extensive margin and improves upon an arbitrary information policy I, for large p. $I^\star$ induces two censorship regions, separated by a full-revelation region. The second step of the proof: (1) adds a second censorship region at the top to include the general case of $p > x_0$, and (2) verifies that eliminating the bottom censorship region improves upon Sender's payoff.

Let's fix an equilibrium $\langle F, e(\cdot), \alpha \rangle$. $e(c, \lambda, I)$ equals $e^*_\lambda \circ \Delta I(c)$ for some selection e^*_λ from $\Delta J(c) \longmapsto \arg\max_{e \in [0,1]} V_\lambda(e, \Delta J(c))$, by Lemma 3. We define $\underline{c}_\lambda(\Delta I) = \sup\{c \in [0, x_0] : e^*_\lambda \circ \Delta I(c) = 0\}$, if $\{c \in [0, x_0] : e^*_\lambda \circ \Delta I(c) = 0\} \neq \emptyset$, and $\underline{c}_\lambda(\Delta I) = 0$ otherwise. We define $\overline{c}_\lambda(\Delta I) = \inf\{c \in [x_0, 1] : e^*_\lambda \circ \Delta I(c) = 0\}$, if $\{c \in [x_0, 1] : e^*_\lambda \circ \Delta I(c) = 0\} \neq \emptyset$, and $\overline{c}_\lambda(\Delta I) = 1$ otherwise. For the rest of this section, we omit reference to λ.

Lemma B.5. *Let $I \in \mathcal{I}$ and define $c^* := \overline{c}(\Delta I)$. There exists another information policy $I^\star$ that satisfies the following properties:*

1. *(FEAS) $I^\star$ is feasible, i.e., $I^\star \in \mathcal{I}$,*
2. *(EM) $I^\star$ produces the same extensive margin as I, i.e., $\overline{c}(\Delta I^\star) = c^*$ and $\underline{c}(\Delta I^\star) = \underline{c}(\Delta I)$.*
3. *(IMPR)*

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

.

4. *(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:*

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

Proof. We use the notation: $\bar{c}(I - I_{\overline{F}}) =: \bar{c}$, $\underline{c}(I - I_{\overline{F}}) =: \underline{c}$. In the first step, we prove the lemma for the case in which there is a feasible information policy that is a straight line between the points $\underline{p} := (\underline{c}, I(\underline{c}))$ and $\overline{p} := (\bar{c}, I(\bar{c}))$. In the second step we analyze the other case.

First Step. Let's define the line i such that $x \longmapsto I(\underline{c}) + \lambda^*(x - \underline{c})$, with slope $\lambda^* := \frac{I(\bar{c}) - I(\underline{c})}{\bar{c} - \underline{c}}$. We claim that $i^\star(x) := \max\{i(x), I_{\overline{F}}(x)\}$ satisfies all properties. $i^\star$ is FEAS by hypothesis. $i^\star$ is EXT because $i(\underline{c}) = I(\underline{c})$ and $i(\bar{c}) = I(\bar{c})$. $i^\star$ is IMPR because I is convex and $i^\star$ is EXT. $i^\star$ 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 on $[\underline{c}, \bar{c}]$, (ii) $\lambda^* \in [0, 1]$ and EXT imply, with $I \in \mathcal{I}$, that there are intersection points $\widetilde{x}_1, \widetilde{x}_2$, with $\widetilde{x}_1 \le \underline{c} \le \bar{c} \le \widetilde{x}_2$, such that: $i^\star(x) = I(x)$ if $x \in [0, \widetilde{x}_1] \cup [\widetilde{x}_2, 1]$.

Second Step. In this case, $i^\star$ is not FEAS. Because $i^\star$ satisfies FEAS at x if $x \le \underline{c}$ and if $x \ge \bar{c}$, there exists a point $x^* \in (\underline{c}, \bar{c})$ such that: $i(x^*) > I_{F_0}(x^*)$. Let's define:

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

$\ell := \max L$, $m := \min M$, and the lines

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

As part of the rest of the proof, we establish some lemmata.

Lemma B.6. *ℓ, m are well-defined.*

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

Lemma B.7. *that there exists a unique pair of numbers $(\theta_\ell, \theta_m) \in [\underline{c}, 1] \times [0, \bar{c}]$ 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 θ_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 $\theta_\ell = 1$ contradicts $\ell \in L$ because $I'_{F_0}(x) < 1$ if $x < 1$. Uniquenss of θ_ℓ follows from convexity of I_{F_0}.

Suppose there does not exists such θ_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 $\theta_m = 0$ contradicts $I \neq I_{\overline{F}}$. Uniquenss of θ_m follows from convexity of I_{F_0}. **QED**

Lemma B.8. $\theta_\ell \leq \theta_m$.

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

Next, 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, because $\ell > 0$, we get:

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

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

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

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

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

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

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

Let's first verify that $I^\star$ is well-defined. We know that $\ell \in \partial I_{F_0}(\theta_\ell)$ and $m \in \partial I_{F_0}(\theta_m)$. Because $I(\underline{c}) + \ell(0 - \underline{c}) < I_{F_0}(0)$ 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 $x < x_2$.

1. CENS follows from the definition of $I^\star$ and the conclusion of the above paragraph.

2. IMPR on $[\underline{c}, \theta_\ell]$ and $[\theta_m, \bar{c}]$ follows from convexity of I, and on $[\theta_\ell, \theta_m]$ follows from FEAS of I in that region.

3. EM follows by construction of $I^\star$.

4. FEAS is established in a similar way as in the last step of the proof of Lemma B.4.

QED

Proof of Proposition 2

Proof. Let's define information policy J by: letting J equal $I^\star$, constructed as in Lemma B.5 by replacing c^* with p, for $x \in [0, x_m^\circ]$, defining the point x_m° in which $I^\star$ intercepts the line $j\colon x \longmapsto I(\overline{c}) + I'(\overline{c})(x - \overline{c})$; and letting J equal $x \longmapsto \max\{I_{\overline{F}}(x), j(x)\}$ on $[x_m^\circ, \infty)$.

It suffices to show that: if the resulting information policy J induces a censorship region at the bottom, then there is an improvement over J that is a bi-upper censorship. Suppose that $I^\star$ is affine on $[x_\ell, \theta_\ell]$ and $I^\star$ equals $I_{\overline{F}}$ on $[0, x_\ell]$, for $0 < x_\ell < \theta_\ell$ (for notation, see Lemma B.5.) By construction, $I^\star(\theta_\ell) = I_{F_0}(\theta_\ell)$. Let's define information policy K by

$$K(x) = \begin{cases} I_{F_0}(x) & , 0 \le x \le \theta_\ell, \\ J(x) & , x \ge \theta_\ell. \end{cases}$$

$K \ge J$, so: K induces a weakly higher Receiver's ex-ante payoff (by Blackwell's theorem) and weakly decreases $\underline{c}_\lambda$, with respect to J. Hence, by $\rho \ge 0$ and $\gamma \ge 0$, it is left to verify that the expected Receiver's action is weakly higher under K than under J. Because $p \ge x_0$, the argument of Theorem 3 suffices. Specifically, by Lemma B.3, we have that:

$$\begin{aligned} W(K') - W(J') &= \int_{[0,\theta_\ell]} \big(V_\lambda(\Delta K(c)) - V_\lambda(\Delta J(c))\big) \frac{\partial g}{\partial c}(c|\lambda)\, \mathrm{d}c \\ &\ge 0, \end{aligned}$$

in which the inequality follows from the definition of $I^\star$, which includes $p \ge \theta_\ell$. Hence K is a bi-upper censorship that improves upon I, for arbitrary I, in terms of U_G. **QED**

B.7 Symmetric-information benchmark

For this section, Sender knows both $c = \zeta$ and $\lambda = \kappa$, k is linear, and F_0 admits a density. The Sender's *problem* is:

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

because an experiment F is an equilibrium experiment iff I_F solves the above problem, due to a generalization of the argument of Gentzkow and Kamenica (2016). If $\zeta > 1$, any information policy is optimal. If $\zeta \le x_0$, $I_{\overline{F}}$ is optimal. Let $1 \ge \zeta \ge \theta_0$.

Lemma B.9. *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:*

$$\Delta I_\theta \leq \kappa,$$

with equality if $\theta > 0$.

Proof. Let $\mathcal{I}^u := \{I \in \mathcal{I} : I = I_\theta, \text{ for } \theta \in [0, \zeta]\}$. Suppose the solution is not I_{F_0}. The Sender's problem is, without loss of optimality by Lemma B.4:

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

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 $I_{\overline{F}}$ 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:

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

such that:

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

Let the value of I_θ be:

$$\begin{aligned} v \colon (0, 1) &\longrightarrow [0, 1] \\ \theta &\longmapsto \big(1 - I'_\theta(\zeta^-)\big) \end{aligned}$$

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:

$$-\frac{\partial F_0}{\partial \theta}(t(I(\zeta)))\frac{1}{(\zeta - t(I(\zeta)))\frac{\partial F_0}{\partial \theta}(t(I(\zeta)))},$$

if $\zeta > t(I(\zeta))$, 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 $\Delta I(\zeta) < \kappa$. **QED**

C Auxiliary Results

Fact C.1 (Subdifferential of convex functions)**.** *Let $S \subseteq \mathbb{R}$, $f\colon S \longrightarrow \mathbb{R}$ be convex, and $\varphi\colon \mathbb{R} \longrightarrow \mathbb{R}$ be a nondecreasing convex function on the range of f.*

1. *The function $\varphi \circ f$ is convex on S.*
2. *For all $y \in S$, letting $t = f(y)$, we have:*

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

Proof. See Bauschke and Combettes (2011, Proposition 8.21 and Corollary 16.72). **QED**

The following lemma states known facts from the envelope theorem and monotone comparative statics.

Lemma C.10 (Envelope theorem)**.** *Let $f\colon [0,1] \times [0,1] \longrightarrow \mathbb{R}^2$ exhibit increasing differences, and be such that: $f(\cdot, a)$ is continuous for all $a \in [0,1]$, $f(e, \cdot)$ is nondecreasing for all $e \in [0,1]$, the derivative with respect to the variable a, $\frac{\partial f}{\partial a}(e, \cdot)$, exists and is bounded for all $e \in [0,1]$. The following hold:*

1. $\arg\max_{e\in[0,1]} f(e,a) \neq \emptyset$ *for all $a \in [0,1]$:*
2. $a \longmapsto \max_{e\in[0,1]} f(e,a)$ *is nondecreasing and absolutely continuous.*
3. *If $a \longmapsto \frac{\partial f}{\partial a}(e,a)$ is nondecreasing for all $e \in [0,1]$, then $a \longmapsto \max_{e\in[0,1]} f(e,a)$ is convex.*
4. *If f exhibits strictly increasing differences, $a \longmapsto \frac{\partial f}{\partial a}(e,a)$ is nondecreasing, $f(e,\cdot)$ is increasing for all $e \in (0,1]$, $\arg\max_{e\in[0,1]} f(e,a) \cap (0,1] \neq \emptyset$, and $1 \geq a' > a \geq 0$, then:*

$$\max_{e\in[0,1]} f(e,a') > \max_{e\in[0,1]} f(e,a).$$

Proof. By upper semi-continuity of f, $\arg\max_{e\in[0,1]} f(e,a) \neq \emptyset$, so 1. holds. Then, by the increasing-differences property of f, there exists a nondecreasing selection $e^\star\colon a \longmapsto \arg\max_{e\in[0,1]} f(e,a)$ on $[0,1]$ (Milgrom and Shannon, 1994). By our hypotheses, we apply the envelope theorem (Milgrom and Segal, 2002), letting $V(a) := \max_{e\in[0,1]} f(e,a)$, to establish that V is absolutely continuous and

$$V(a) = V(0) + \int_{[0,a]} \frac{\partial f}{\partial a}(e^\star(\tilde{a}), \tilde{a})\,\mathrm{d}\tilde{a}.$$

V is nondecreasing because $\frac{\partial f}{\partial a} \geq 0$. Hence, 2. holds.

Let's establish that V is convex if $a \longmapsto \frac{\partial f}{\partial a}(e,a)$ is nondecreasing. By the increasing-differences property of f: (i) $e \longmapsto \frac{\partial f}{\partial a}(e,a)$ is nondecreasing, and (ii) there exists a nondecreasing $e^\star\colon a \longmapsto \arg\max_{e\in[0,1]} f(e,a)$. As a result, $a \longmapsto \frac{\partial f}{\partial a}(e^\star(a),a)$ is nondecreasing. Thus, V is convex (Theorem 24.8 in Rockafellar, 1970, noting that $a \longmapsto \frac{\partial f}{\partial a}(e^\star(a),a)$ is uni-dimensional.) Hence, 3. holds.

Let $a' > a$, for $a', a \in [0,1]$, and $e' \in \arg\max_{e\in[0,1]} f(e,a) \cap (0,1]$. Then: $V(a') - V(a) = \int_{[a,a']} \frac{\partial f}{\partial a}(e^*(\tilde{a}),\tilde{a})\,\mathrm{d}\tilde{a}$ for every selection e^* of $\arg\max_{e\in[0,1]} f(e,a) \cap (0,1]$. We have the following chain of inequalities under the additional hypotheses stated in part 4.:

$$\begin{aligned} V(a') - V(a) &\geq \int_{[a,a']} \frac{\partial f}{\partial a}(e', \tilde{a})\,\mathrm{d}\tilde{a} \\ &\geq \int_{[a,a']} \frac{\partial f}{\partial a}(e', a)\,\mathrm{d}\tilde{a}, \end{aligned}$$

in which the first inequality follows from the strict increasing-differences property of f and the definition of e', the second inequality holds because $a \longmapsto \frac{\partial f}{\partial a}(e,a)$ is nondecreasing (for the first inequality, in particular, we note that: (i) every selection e^* of $\arg\max_{e\in[0,1]} f(e,a) \cap (0,1]$ is nondecreasing, (ii) there exists a selection e^* of $\arg\max_{e\in[0,1]} f(e,a) \cap (0,1]$ such that $e^*(a) = e'$.) Because $\int_{[a,a']} \frac{\partial f}{\partial a}(e',a)\,\mathrm{d}\tilde{a} = (a'-a)\frac{\partial f}{\partial a}(e',a)$, 4. holds. **QED**

References

Aliprantis, Charalambos D. and Kim C. Border (2006), *Infinite Dimensional Analysis: A Hitchhiker's Guide*, third edition. Springer Berlin, Heidelberg.

Angelucci, Charles, Amit K. Khandelwal, Andrea Prat, and Ashley Swanson (2021), "Knowledge acquisition in a high-stakes environment: Evidence from the Covid-19 pandemic." Working Paper.

Arieli, Itai, Yakov Babichenko, Rann Smorodinsky, and Takuro Yamashita (2023), "Optimal persuasion via bi-pooling." *Theoretical Economics*, 18(1), 15–36.

Au, Pak Hung and Mark Whitmeyer (2023), "Attraction versus persuasion: Information provision in search markets." *Journal of Political Economy*, 131(1), 202–245.

Augias, Victor and Daniel M. A. Barreto (2024), "Persuading a wishful thinker." Working Paper.

Bauschke, Heinz H. and Patrick L. Combettes (2011), *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, second edition. Springer Cham.

Beauchêne, Dorian, Jian Li, and Ming Li (2019), "Ambiguous persuasion." *Journal of Economic Theory*, 179, 312–365.

Bizzotto, Jacopo, Jesper Rüdiger, and Adrien Vigier (2020), "Testing, disclosure and approval." *Journal of Economic Theory*, 187, 105002.

Bloedel, Alexander W. and Ilya Segal (2021), "Persuading a Rationally Inattentive Agent." Working Paper.

Board, Simon and Jay Lu (2018), "Competitive information disclosure in search markets." *Journal of Political Economy*, 126(5), 1965–2010.

Brocas, Isabelle and Juan D. Carrillo (2007), "Influence through ignorance." *The RAND Journal of Economics*, 38(4), 931–947.

Caplin, Andrew, Mark Dean, and John Leahy (2022), "Rationally inattentive behavior: Characterizing and generalizing shannon entropy." *Journal of Political Economy*, 130(6), 1676–1715.

Che, Yeon-Koo, Kyungmin Kim, and Konrad Mierendorff (2023), "Keeping the listener engaged: A dynamic model of bayesian persuasion." *Journal of Political Economy*, 131(7), 1797–1844.

de Clippel, Geoffroy and Xu Zhang (2022), "Non-bayesian persuasion." *Journal of Political Economy*, 130(10), 2594–2642.

Denti, Tommaso (2022), "Posterior separable cost of information." *American Economic Review*, 112(10), 3215–59.

Denti, Tommaso, Massimo Marinacci, and Aldo Rustichini (2022), "Experimental cost of information." *American Economic Review*, 112(9), 3106–23.

Dworczak, Piotr and Giorgio Martini (2019), "The Simple Economics of Optimal Persuasion." *Journal of Political Economy*, 127(5), 1993–2048.

Dworczak, Piotr and Alessandro Pavan (2022), "Preparing for the worst but hoping for the best: Robust (bayesian) persuasion." *Econometrica*, 90(5), 2017–2051.

Felgenhauer, Mike (2019), "Endogenous persuasion with costly verification." *The Scandinavian Journal of Economics*, 121(3), 1054–1087.

Feng, Yiding, Chien-Ju Ho, and Wei Tang (2024), "Rationality-robust information design: Bayesian persuasion under quantal response." In *Proceedings of the 2024 ACM-SIAM Symposium on Discrete Algorithms, SODA 2024, Alexandria, VA, USA, January 7-10, 2024* (David P. Woodruff, ed.), 501–546, SIAM.

Floridi, Luciano (2014), *The Fourth Revolution: How the Infosphere is Reshaping Human Reality*. Oxford University Press, Oxford.

Galperti, Simone (2019), "Persuasion: The art of changing worldviews." *American Economic Review*, 109(3), 996–1031.

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." *American Economic Review*, 106(5), 597–601.

Gitmez, A. Arda and Pooya Molavi (2023), "Informational autocrats, diverse societies."

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

Honka, Elisabeth, Ali Hortaçsu, and Maria Ana Vitorino (2017), "Advertising, consumer awareness, and choice: evidence from the u.s. banking industry." *The RAND Journal of Economics*, 48(3), 611–646.

Jain, Vasudha and Mark Whitmeyer (2022), "Competitive disclosure of information to a rationally inattentive agent." Working Paper.

Kamenica, Emir (2019), "Bayesian Persuasion and Information Design." *Annual Review of Economics*, 11, 249–272.

Kamenica, Emir and Matthew Gentzkow (2011), "Bayesian Persuasion." *American Economic Review*, 101(6), 2590–2615.

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

Knoepfle, Jan (2020), "Dynamic competition for attention." Working Paper.

Kolotilin, Anton (2018), "Optimal information disclosure: a linear programming approach." *Theoretical Economics*, 13(2), 607–635.

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.

Liao, Xiaoye (2021), "Bayesian persuasion with optimal learning." *Journal of Mathematical Economics*, 97, 102534.

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

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

Lipnowski, Elliot, Laurent Mathevet, and Dong Wei (2022), "Optimal attention management: A tractable framework." *Games and Economic Behavior*, 133, 170–180.

Lipnowski, Elliot, Doron Ravid, and Denis Shishkin (2021), "Persuasion via weak institutions." Working Paper. Electronic copy available at: https://ssrn.com/abstract=3168103, July 20, 2021.

Lipnowski, Elliot, Doron Ravid, and Denis Shishkin (2024), "Perfect bayesian persuasion."

Machina, Mark J. (1982), "Expected utility analysis without the independence axiom." *Econometrica*, 50(2), 277–324.

Matysková, Ludmila and Alfonso Montes (2023), "Bayesian Persuasion With Costly Information Acquisition." *Journal of Economic Theory*, 211, 105678.

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

Milgrom, Paul and Chris Shannon (1994), "Monotone Comparative Statics." *Econometrica*, 62(1), 157–180.

Pomatto, Luciano, Philipp Strack, and Omer Tamuz (2023), "The cost of information: The case of constant marginal costs." *American Economic Review*, 113(5), 1360–93.

Prat, Andrea (2015), "Media Capture and Media Power." In *Handbook of Media Economics* (Simon P. Anderson, Joel Waldfogel, and David Strömberg, eds.), volume 1, 669–686, North-Holland.

Ravid, Doron, Anne-Katrin Roesler, and Balázs Szentes (2022), "Learning before trading: On the inefficiency of ignoring free information." *Journal of Political Economy*, 130(2), 346–387.

Rayo, Luis and Ilya Segal (2010), "Optimal Information Disclosure." *Journal of Political Economy*, 118(5), 949–987.

Rockafellar, R. Tyrell (1970), *Convex Analysis*. Princeton University Press, Princeton.

Romanyuk, Gleb and Alex Smolin (2019), "Cream skimming and information design in matching markets." *American Economic Journal: Microeconomics*, 11(2), 250–76.

Shishkin, Denis (2024), "Evidence Acquisition and Voluntary Disclosure." Working Paper.

Simon, Herbert A. (1996), "Working paper no. 96–2 – knowledge and the time to attend to it." In *Carnegie Bosch Institute for Applied Studies in International Management*.

Sun, Junze, Arthur Schram, and Randolph Sloof (2024), "Publicly persuading voters: The single-crossing property and the optimality of interval revelation." Working Paper.

Topkis, Donald M. (1978), "Minimizing a submodular function on a lattice." *Operations Research*, 26(2), 305–321.

Tsai, Yi-Lin and Elisabeth Honka (2021), "Informational and noninformational advertising content." *Marketing Science*, 40(6), 1030–1058.

Wei, Dong (2021), "Persuasion under costly learning." *Journal of Mathematical Economics*, 94, 102451.