Persuading an Inattentive and Privately Informed Receiver

Pietro Dall'Ara*

Boston College

May 20, 2024

Abstract

I study the persuasion of a receiver who accesses information only if she exerts costly attention effort. The sender designs an experiment to persuade the receiver to take an action. The experiment also influences 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 preferences exhibit a supermodularity property in information and effort. By leveraging this property, I prove a general equivalence result 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.
J.E.L. Codes: D82, D83, D91.

*I am grateful for helpful comments from Giacomo Calzolari, Ryan Chahrour, Vincenzo Denicolò, Mehmet Ekmekci, Jan Knoepfle, Hideo Konishi, Chiara Margaria, Laurent Mathevet, Teddy Mekonnen, Stefano Piasenti, Giacomo Rubbini, Fernando Stragliotto, Utku Ünver, Bumin Yenmez, and participants in the 2023 Spring Meeting of Young Economists, 2023 Econometric Society European Meeting, BC-BU-Brown Theory Workshop, EUI (Micro Group), U. of Bologna, Brown Theory/Experimental Lunch Seminar, BU Reading Group, and Queen Mary Ph.D. Workshop. An earlier version of this paper bore the title "The Extensive Margin of Bayesian Persuasion." Contact: <pietro.dallara@gmail.com>.

Contents

1 Introduction

In the "information age," consumers decide whether an information source deserves attention because acquiring information is costly (Floridi, 2014; Simon, 1996). The persuasion literature studies a sender who supplies information to a receiver to persuade the receiver to take a certain action (Bergemann and Morris, 2019; Kamenica, 2019). When the receiver's attention is costly, the sender faces a dual problem: the receiver can be persuaded only if they pay attention to the sender's information. In this paper, I study the persuasion of a receiver who is privately informed about her costs and benefits of information and in which the sender's information influences the receiver's attention effort.

The *intensive* margin of persuasion captures the intensity of the sender's influence on the receiver's action decision, while the *extensive* margin refers to whether or not the receiver pays attention to the sender's information. The study of the extensive margin is important to understand how consumers allocate attention to product advertisement and news content, which ultimately determines the success of marketing strategies and the spread of information.

To study the extensive and intensive margins of persuasion, I introduce the receiver's attention decision in a canonical persuasion game between two players: Sender (he) and Receiver (she). In the first stage of the game, Sender designs a signal — a random variable correlated with the unknown state θ — called the signal. Knowing the signal's distribution, but not its realization, Receiver chooses her attention *effort* e: high effort is costly and increases the probability of observing the signal 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 and Sender wants Receiver to choose 1 regardless of the state.

Sender takes into account that increasing the correlation between the state and the signal has two effects: on Receiver's attention effort e — the extensive margin —, and on Receiver's action if she observes the signal realization — the intensive margin. Receiver's belief is updated given the signal realization with probability e and it not updated with the remaining probability. The choice of effort captures the choice of acquiring information about the state, and the cost of effort may be monetary or cognitive.

The implied model of attention is less general than in models of flexible information acquisition (Pomatto et al., 2023; Caplin et al., 2022; Denti, 2022; Bloedel and Zhong, 2021), because Receiver only chooses the probability with which she uniformly observes every signal realization. My parsimonious attention model allows to include Receiver's private information and a general functional form of effort cost. In particular, the model exhibits a supermodularity structure in the Receiver's preferences over information and effort (Corollary 1). In contrast, typical applications of flexible information acquisition rely on functional-form assumptions (see the following paragraphs) and define cost functions over posterior-belief distributions, which this model of attention effort avoids (Denti et al., 2022).[1]

I show the equivalence between persuasion mechanisms and signals (Theorem 1). Let's suppose that Sender commits to a persuasion mechanism, a menu of signals, one for every Receiver's report of her type. Under a persuasion mechanism, Receiver reports a type and chooses an effort level. In particular, Receiver chooses the probability with which she observes the signal from the menu that corresponds to her report. A mechanism is incentive-compatible if Receiver finds it optimal to report her type truthfully. For every incentive-compatible persuasion mechanism, I construct a signal that induces the same action and effort decision of all Receiver's types (Theorem 1). The key is to establish a supermodularity property of type-t Receiver's expected utility: the return from effort is increasing in a t-specific informativeness order, which agrees with Blackwell's order whenever possible. I construct a single signal that "allocates" to each Receiver's type t the same t informativeness as an incentive-compatible mechanism. This step establishes equivalence with respect to effort choices. As an implication, the Receiver's action decision is replicated by the signal, which guarantees that the result of Kolotilin et al. (2017) is nested. As an implication Sender does not need to offer a fine collection of information structures,

[1] Receiver effectively chooses mixtures of full information and null information about the Sender's signal — represented by effort e —; thus, the induced cost of information is experimental (Denti et al., 2022).

(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, `Pool`, if the state is above the threshold.

(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 `Pool1` if the state is between the lower threshold θ_1 and an upper threshold θ_2, and sends a different message `Pool2` if the state is above the upper threshold θ_2.

Figure 1: An upper censorship (a) and a bi-upper censorship (b).

and the analysis of attention effort can focus on single signals.

I 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 censors high states, as shown Figure 1. Upper censorships are optimal signals if the Receiver's outside option follows a single-peaked distribution (Theorem 3). I apply my results to the problem of media censorship. If Sender knows Receiver's attention cost and has preferences over the extensive margin, inspired by models of media capture à la Gehlbach and Sonin (2014), bi-upper censorships are optimal signals (see Figure 1.)

Related literature If the receiver's attention is costless, prior work determines the extent of the sender's persuasion.[2] Moreover, optimality properties of upper censorships are known, and the equivalence between persuasion mechanisms and signals is shown by Kolotilin et al. (2017).[3] I generalize these results to the case of receiver's costly attention and privately known attention costs.

To study the extensive margin of persuasion, the model includes Receiver's attention cost and her private information.[4] In particular, Receiver decides about her

[2] *Inter alia*: Kamenica and Gentzkow 2011; Kolotilin 2018; Dworczak and Martini 2019; Rayo and Segal 2010; Brocas and Carrillo 2007; see also Section 2.4.

[3] For upper censorships, see also: Kolotilin et al. 2022; Kleiner et al. 2021; for persuasion mechanisms, see also: Guo and Shmaya 2019.

[4] The literature on incomplete-information beauty contests studies the supply of Gaussian signals to inattentive receivers (Cornand and Heinemann, 2008; Chahrour, 2014; Myatt and Wallace, 2014;

attention, given that attention is costly, and Receiver's private information captures heterogeneity in the Sender's endogenous audience. Persuasion of an inattentive receiver has been studied without Receiver's private information. In Wei (2021), Receiver's attention cost is proportional to the entropy reduction in her belief about the state. As a result of the attention cost 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 the Receiver's signal about the Sender's signal. In a separate model, the authors study the same cost structure as in my paper. The connection with these papers is further discussed in Section 2.4.

The "attention-management" literature studies Receiver's inattention given a Sender who is benevolent because he maximizes Receiver's material payoff and does not consider her information cost (Lipnowski et al., 2020, 2022). Even if the typical application considers a entropic information cost, the attention model nests mine in a sense made precise in Section 2.4. The literature on persuasion with "parallel information acquisition" studies Receiver's costs of acquiring extra information than from Sender (Matysková and Montes, 2023; Bizzotto et al., 2020; Brocas and Carrillo, 2007). The focus is on how Receiver's utility and Sender's information change as attention is costlier (Section 5).

Outline Section 2 describes and model, Section 3 studies the Receiver's equilibrium attention and action. In Section 4, I describe the equivalence between persuasion mechanisms and signals. In Section 5, I study upper censorships and applications.

Galperti and Trevino, 2020) with different incentives of information recipients and providers than in persuasion models. The literature on media capture considers the provision of information to receivers who are privately informed, either about the cost of supporting a politician, or about their attention cost Gehlbach and Sonin, 2014; Kolotilin et al., 2022; Gitmez and Molavi, 2023; see Section 5 for an application of our model, and Prat (2015) for a literature review.

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 (c, λ). Receiver's utility is her material payoff net of her 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 λ, referred to as the *attention type*, scales the effort cost. Sender chooses a signal — a measurable $\pi\colon [0, 1] \longrightarrow \Delta M$, in which ΔM is the set of probability distributions over the rich message space M — about the state and his payoff given action a is $U_S(a) := a$.[5]

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 cost 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, but not the signal realization. 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

[5] $M = [0, 1]$ suffices for the game (Section A.2), Section 2.3 presents the representation of signals as convex functions used in the rest of the paper.

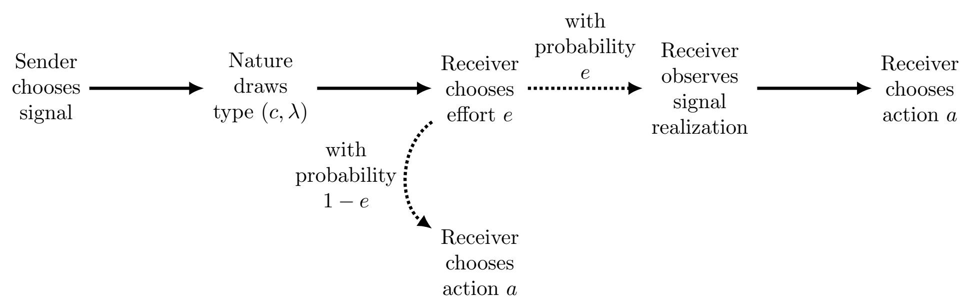


Figure 2: The timing of the game.

Bayes' rule, and chooses an action given the updated, *posterior*, belief; with the remaining probability, 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 product order and the topology induced by the L^1 norm, which metrizes weak convergence (Machina, 1982, Lemma 1). $\partial I(x)$ and I' denote, respectively, the subdifferential and right derivative $I \in \mathcal{I}$ at $x \in \mathbb{R}_+$, and $\circ$ denotes the composition of functions. We use $\leq$ to denote 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 $s < s'$; the function $g\colon S \times T \longrightarrow \mathbb{R}$ exhibits *strictly increasing differences* if $t \longmapsto g(s', t) - g(s, t)$ is increasing for all $s', s \in S$ with $s < s'$. T denotes the support of Receiver's type. Omitted proofs are in Appendix B.

2.3 Signals as 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.[6]

[6] Appendix A.1 verifies this claim. Signals can be represented by the their posterior-mean distributions in persuasion games with ex-post linear Receiver's utility in the state — as in this model — and with costless Receiver's attention — unlike this model. First, every signal induces a distribution function of the posterior belief's mean F that is a mean-preserving contraction of F_0, by Jensen's inequality; conversely, if F is a mean-preserving contraction of F_0, then there exists

Given the presence of Receiver's effort, it pays off to represent signals by the integrals of such distributions, called "information policies" (see Lemma 2).[7] 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

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

the set of feasible distribution functions $\mathcal{F} := \{F \in \mathcal{D} : I_F(1) = I_{F_0}(1), \text{ and } I_F \leq I_{F_0}\}$, and the set of information policies as

$$\mathcal{I} := \{I \colon \mathbb{R}_+ \longrightarrow \mathbb{R}_+ : I \text{ is convex and } I_{\overline{F}} \leq I \leq I_{F_0}\},$$

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 $I_F \in \mathcal{I}$;*

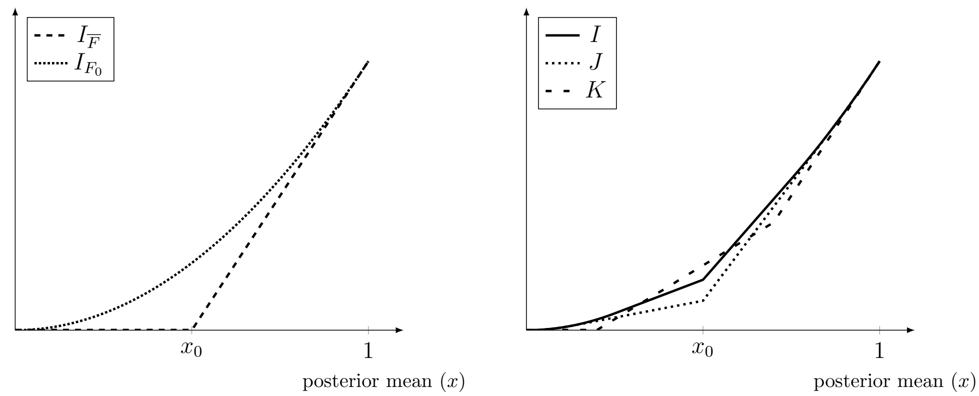
2. *If $I \in \mathcal{I}$, then $I' \in \mathcal{F}$, once I is extended to take value 0 at every $x < 0$.*

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

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 function I' with probability corresponding to her effort, and is equal to x_0 with the remaining probability.

a signal that induces F as the distribution of the posterior mean (Gentzkow and Kamenica, 2016; Kolotilin, 2018, *inter alia.*) Appendix A.1 shows that the equivalence holds for this model too.

[7] Information policies are also called "left side integrals" (An, 1995; Bagnoli and Bergstrom, 2005), "spread functions" (Romanyuk and Smolin, 2019), and "integral CDFs" (Shishkin, 2023). A belief distribution that has barycenter equal to a given prior belief is also called an information policy (Lipnowski and Mathevet, 2018; Romanyuk and Smolin, 2019; Lipnowski et al., 2020; Lipnowski and Ravid, 2020). Ravid et al. (2022) use the symbol "I_F" for the function $I_{F_0} - I_F$ in our notation.



(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.

(b) Information policy I is more informative than information policy J in the Blackwell's 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.

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 measurabilty requirements (Appendix A.2).

2.4 Discussion and Interpretation

Attention effort The term $\lambda k(e)$ in the Receiver's utility can be interpreted as 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 costly to overcome and limited memory. In contrast, when choosing the probability of being exposed to the media and evaluating the monetary costs of purchasing newspapers, the opportunity cost of being attentive to communication is relevant.

Costless-Attention benchmark The special case of the model in which effort is costless for Receiver — i.e., the distribution of λ puts full mass at 0 — is studied by prior work (Kolotilin et al., 2017, e.g.). This model captures the persuasion of a privately informed receiver who is fully attentive to the Sender's communication. There exists an optimal signal that is an upper censorship, see 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 If Receiver's type distribution puts full mass at some (c°, λ°), Receiver does not have private information. Wei (2021) considers such a model with binary state and two more differences. First, Receiver's attention cost is proportional to the expected entropy reduction in her belief. Second, Receiver's strategy space contains the present one: Receiver chooses signals about the Sender's signal, which include mixtures between the Sender's signal and no information. The latter class of signals is the one induced by the sole choice of effort.[8] The optimal signal is a binary signal in Wei (2021).

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 that takes into account that Receiver learns about the Sender's signal, and the Receiver's strategy space is fully general.[9] The optimal signal is an upper censorship, although for a different phenomenon than this paper, if Sender's utility function is U_S. In particular, Sender perceives Receiver's action as random given a signal realization because of the entropic attention cost in Bloedel and Segal (2021), while the perceived noise is due to both the Receiver's attention effort and asymmetric information (about her type) in this paper. Bloedel and Segal (2021) also study the symmetric-information case,

[8] Receiver's strategy space is fully general and her attention cost is entropy-based in Lipnowski et al. (2020, 2022), even though "pure-persuasion motives" are absent because the Sender's utility equals the Receiver's "material utility".

[9] See Bloedel and Segal (2021) and Lipnowski et al. (2022) for the differences between the two entropy-based costs.

as one of the variations of their main specification. Due to both the binary action and the symmetric information, there exists an optimal signal that is a binary signal, as discussed in Section 5.[10] In this case, there exists an optimal signal that is an upper censorship (Section B.11, and signals are equivalent to persuasion mechanisms (Theorem 1).

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

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

In order to study the Receiver's effort choice, we 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 observing the Sender's signal: $\int_{[0,1]} U_R(x, e, c, \lambda) - U_R(x_0, e, c, \lambda) \, \mathrm{d}I'(x)$.[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 4). Using the operator $\Delta \colon I \longmapsto I - I_{\overline{F}}$, the following result shows that the marginal benefit of effort given cutoff type c is given by the net informativeness evaluated at c.

Lemma 2 (Net Informativeness)**.** *Given information policy I and Receiver's effort*

[10] The joint restriction of binary action and symmetric information leads to considering binary signals without loss for the current model of attention cost (see, however, Section 5); with more general Receiver's strategies and a different attention cost, Bloedel and Segal (2021) show that the optimal signal need not be binary.

[11] The marginal benefit of effort given an information policy is often referred to as the value of information.

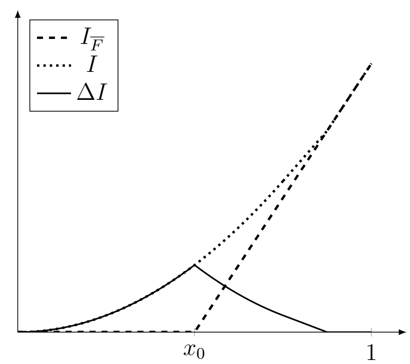


Figure 4: The net informativeness of I at a cutoff c, $\Delta I(c)$, is obtained by subtracting the value of the uninformative-signal information policy at c, $I_{\overline{F}}(c)$, to $I(c)$. By construction (Lemma 1) net informativeness of I at a cutoff c is single-peaked as a function of c, with a peak at the prior mean x_0.

e, the following holds:

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

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:

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

By absolute continuity of I:

$$V = 1 - c - I(1) + I(c). \tag{1}$$

Because $I(1) = 1 - x_0$, repeating the steps for $I_{\overline{F}}$ yields

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

QED

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. $1 - \int_{[0,1]} \alpha(c,\lambda,x)\,\mathrm{d}\hat{I}'(x) \in \partial\hat{I}(c)$*;*
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 the derivation in the proof of Lemma 2, Equation (1), and the equilibrium properties of e. **QED**

The takeaway of Lemma 3 in 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 information. Part 1. restates the equilibrium conditions that the Receiver's action satisfies.[12]

3.2 Interval Structure of the Extensive Margin

This section studies Receiver's effort.

[12] Part 1. is unchanged in models of costless attention effort, except if the focus is on Sender-optimal equilibria (Gentzkow and Kamenica, 2016, p. 600).

We define the Receiver's value of an information policy at the effort-choice stage, in light of Lemma 3, which exhibits strictly increasing differences in net informativeness and effort.

Definition 2. The Receiver's *value of information policy* $I \in \mathcal{I}$, given her type (c, λ) and effort e, is $V_\lambda(e, \Delta I(c)) := e\Delta I(c) - \lambda k(e)$.[13]

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

This result implies that a more informative Sender's information policy, in the Blackwell sense, makes ex-ante Receiver better off, via an application of the envelope theorem. In particular, I is Blackwell more informative than J iff: $J \leq I$; hence, I allocates more net informativeness to every type than J, for $J \leq I$. This observation is known from Blackwell's theorem. The supermodularity property in Corollary 1 is a stronger result.[14] In this section, we apply it to characaterize the set of cutoff types that exert positive effort.

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-cost type λ. The set $e_\lambda^{-1}((0,1])$ is an interval.*

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 two 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 $[x_0, 1]$.

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 $\{c \in [x_0, 1] :$

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

[14] Indeed, a weaker corollary of Lemma 3 is that $(e, I) \longmapsto V_\lambda(e, \Delta I(c))$, exhibits strictly increasing differences in $e \in [0,1]$ and $I \in \mathcal{I}$.

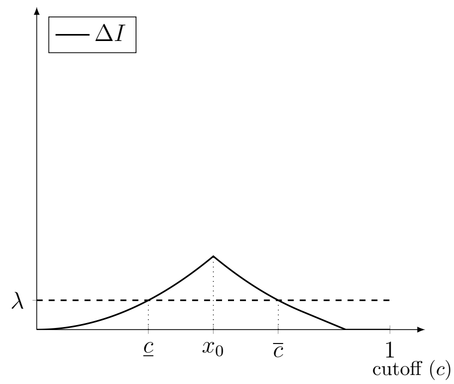


Figure 5: Given linear k and attention cost λ, the marginal benefit of effort equals the marginal cost for cutoff types $\underline{c}$ and $\overline{c}$ (Lemma 2, see also Figure 4). For cutoffs in $(\underline{c}, \overline{c})$, Receiver chooses effort 1, and for cutoffs in $[0,1] \setminus [\underline{c}, \overline{c}]$ Receiver does not exert effort. The subset of cutoff types that observe the signal realization with positive probability, is an interval.

$e^* \circ \Delta I(c) = 0\} \neq \emptyset$, 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**

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 5, 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 5). As an implication, the set of Receiver's cutoff types that exert positive effort is an interval.[15] The proof of Lemma 4 generalizes the first part of the above argument. By the supermodularity property of Receiver's preferences,

[15]Receiver's effort at the boundary is determined through equilibrium selection. The selection is not relevant for atomless type distributions (Lemma B.5).

comparative statics à la Topkis (1978) tells us that Receiver's effort is nondecreasing in net informativeness at her cutoff type.

4 Persuasion Mechanisms

This section studies the equivalence between single signals and menus of signals, called "persuasion mechanisms."

A persuasion mechanism is a menu of information policies, each corresponding to a specific report of Receiver.

Definition 3. A *persuasion mechanism* $I_\bullet$ is a list of information policies: $I_\bullet = (I_r)_{r\in R}$, with $R = T$. 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 the following 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 first chooses effort e, then observes a signal corresponding to information policy I_r with probability e, and lastly chooses an action. We are interested in equilibria in which Receiver reports truthfully, which is without loss via a revelation-principle argument.

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 4. 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 4 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 $I_\bullet$.[16] Item 2. in Definition 4 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.[17]

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 (Section 3) holds in more general environments. Theorem 1 nests the equivalence established by Kolotilin et al. (2017) in the costless-attention case.

We sketch the intuition and proof of Theorem 1 that leverage the supermodularity property of the Receiver's interim payoff (Corollary 1). The proof verifies that supermodularity is the key property by proving 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 the upper-envelope of $I_\bullet$ — the function $x \longmapsto \sup_{r \in R} I_r(x)$ —, which we call J (Figure 6), under positive Receiver's effort. A report r is *active* at x if $I_r(x) \geq I_{r'}(x)$ for all $r' \in R$ and fix Receiver's type (c, λ). Let's observe that an active report at c maximizes the Receiver's expected utility. By Lemma 3, a report r impacts Receiver's utility only

[16] Theorem 1 holds under a slightly stronger version of item 1. in Definition 4, as clear in the proof (Section B.2).

[17] Formally, the reason is that the equivalence of the action decision holds as a consequence of item 1. "for this type". In particular, $\arg\max_{e \in [0,1]} V_\lambda(e, \Delta I_{(c,\lambda)}(c)) = \{0\}$ 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; i.e., it is necessary and sufficient for Receiver's action decision to be rational that $\Pr\{a = 1\} \in 1 - \partial I_{\overline{F}}(c)$.

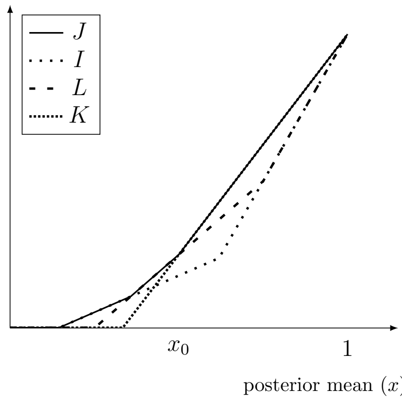


Figure 6: The upper envelope J of the information policies in the IC persuasion mechanism $I_\bullet = (I, L, K)$ is an information policy. In the proof of Theorem 1, we show that J implements the same Receiver's action and effort in the game (Section 2) as $I_\bullet$ under truthful revelation.

through the local net informativeness $\Delta I_r(c)$. By supermodularity, an active report at c makes type (c, λ) weakly better off than any other report (Corollary 1, via monotone comparative statics à la Topkis, 1978). Hence, an active report at c maximizes Receiver's utility at the reporting stage.

Towards the equivalence in terms of Receiver's 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; for a formal statement, see Lemma B.2. 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 (Figure 3) 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 $\Delta I_r(c) = \Delta I_{(c,\lambda)}$.[18] 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 — item 1. of Definition 4. In particular, the localized net

[18]Blackwell's theorem alone does not suffice to reach this conclusion, which uses (i) Corollary 1, (ii) the envelope theorem applied to our setup (Appendix, Lemma B.2), and (iii) the completeness property of $\leq_c$.

informativeness $\Delta I_{(c,\lambda)}(c) = \Delta J(c)$ is the only feature 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. In particular, it holds that $\partial I_r(x) \subseteq \partial J(x)$ if report r is active. The proof in Appendix B.2 uses a continuity argument to cover the case of zero effort.

5 Optimality Properties of Upper Censorships

This section discusses the properties of a class of signals, "upper censorships" (Figure 1).

Definition 5. The θ *upper censorship* is the unique information policy $I_\theta \in \mathcal{I}$ such that:

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

for $\theta \in [0, 1]$.

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

Assumption 1 (Single-Peakedness)**.** The Receiver's type satisfies single-peakedness, that is: the conditional cutoff distribution given attention cost 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 $[p, 1]$; p is called the cutoff's *peak*.

The class of single-peaked distributions includes the standard uniform and the $[0, 1]$-truncated normal.

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

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

For costless attention, the uniqueness result rests on the continuity of the cutoff distribution, which ensures that Sender is indifferent among all Receiver's best responses. The complete proof exploits the convexity of information policies and the envelope theorem applied to the Receiver's supermodular utility (Corollary 1) to generalize the argument (Lemma B.5).[19]

Our next results shows that an optimal signal that is an upper censorship exists.

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 dimensionality of the Sender's optimization to a uni-dimensional problem.

In case of costless attention and Sender-optimal equilibria, the argument for Theorem 3 is as follows. The Sender's expected utility at posterior mean x is $H(x)$, letting H be the distribution of Receiver's cutoff c. By single-peakedness, H is S-shaped, with a saddle point at the peak. So, Sender is risk-lover, conditional on low posterior means, i.e., $x < p$, and he is risk-averse for high posterior means. Mean-Preserving spreads around low posterior means increase Sender's expected utility. Second-Order dominance is related to the informativeness of Sender's strategy: $F \in \mathcal{F}$ is a mean-preserving spread of $\hat{F} \in \mathcal{F}$ iff F is more Blackwell informative than $\hat{F}$, i.e., if $I_{\hat{F}} \leq I_F$ (Figure 3). Moreover, the upper censorship I_u induces either full information conditionally on the state being lower then the threshold u, or no information except that $\theta > u$. Intuitively, upper censorships imply a posterior-mean distribution that aligns with Sender's interests. In the next paragraph, we describe

[19] Lipnowski et al. (2024) show that uniqueness of Sender's equilibrium expected utility obtains in a general model that does not nest ours given the presence of attention effort; their Corollary 1 is similar to our observation.

how this intuition changes if effort is endogenous, namely, if the relevant information policy is $x \longmapsto eI(x) + (1-e)I_{\overline{F}}(x)$.

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 to the Sender's signal, via the envelope theorem applied to the Receiver's supermodular utility (Lemma B.2 applied to Lemma 3.) 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 also spreads out the Receiver's posterior-mean distribution given that she endogenously chooses effort, around x. As an implication, the intuition in the above paragraph goes through. This argument is "local". In particular, $\Delta I(x)$ increases only if the resulting information policy satisfies the convexity constraints in Lemma 1. The proof uses the above argument to construct an information policy $I_u \in \mathcal{I}$ that improves upon I, for arbitrary I.

Sender optimally provides more information as Receiver's attention cost increases, for small attention costs.

Proposition 1. *Let Assumption 1 hold, F_0 admit a density, k be linear, and the attention cost put full mass at λ. Let's denote by I_{θ_ε} an optimal upper censorship if $\lambda = \varepsilon$, and by I_η an optimal upper censorship if $\lambda = 0$. Then: $\theta_\varepsilon \geq \eta$ for all sufficiently small ε.*

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 case, for $c > x_0$. In order to persuade Receiver to take action 1, Sender's posterior distribution 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.6). Let's suppose this were not the case. Then, Sender increases the probability of a Receiver's posterior mean x with $x \geq c$ as much as possible, by inducing a *single* posterior mean $x = c$, with the highest probability that satisfies Bayes' rule (Kamenica and Gentzkow, 2011; Gentzkow and Kamenica, 2016). As a result, Receiver is facing two scenarios: she is indifferent between the two actions with some probability ($x = c$); she finds it optimal to choose the riskless action, 0, with the remaining probability ($x < c$). Hence, information brings no value. As an implication, the constraint binds and Sender provides more Blackwell information to Receiver, if $\lambda > 0$, than if $\lambda = 0$. Proposition 1 shows that the insight generalizes, for small λ. In general, as Sender changes the censorship state θ_ε, he affects the extensive margin given Receiver's private information. However, only the extensive margin's upper bound is impacted by small changes in θ_ε around η, because a nontrivial censorship is optimal if $\lambda = 0$.[20] This mechanism leads to Proposition 1.

In certain applications, Sender cares directly about Receiver's material payoff (Kolotilin et al., 2022; Lipnowski et al., 2020) and about Receiver's attention decisions (Gehlbach and Sonin, 2014). In the former case, Sender is a benevolent planner who cares about the quality of Receiver's decision-making. In the latter case, Sender owns a state's media, so he collects the advertisement's revenues. The next result shows that an extension of the class of upper censorships contains the Sender's optimal signal.

A *bi-upper censorship* is an information policy I such that:

$$I(x) = \begin{cases} I_{F_0}(x) & , x \in [0, \theta_1] \\ I_{F_0}(\theta_1) + F_0(\theta_1)(x - \theta_1) & , x \in (\theta_1, x_1] \\ I_{\overline{F}}(1) + m(x - 1) & , x \in (x_1, 1], \end{cases}$$

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

Proposition 2. *Let Assumption 1 hold, with peak $p \geq x_0$, k be linear, and the attention cost put full mass at λ. There exists an equilibrium in which the Sender's information policy is a bi-upper censorship if Sender's utility function is given by*

[20] To see this observation, Figure 4 depits the net informativeness of an upper censorship, which is 0 for types higher than the censorship state. The observation then follows from Receiver's equilibrium behavior (Lemma 3).

$U_G(\theta, a, e, c, \lambda) := a + \gamma e + \rho U_R(\theta, a, e, c, \lambda)$, *for* $\gamma \geq 0,\ \rho \geq 0$.

The case of $\gamma = 0$ is studied by Kolotilin et al. (2022), who find that upper censorships are optimal signals. 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 peak is $p \geq x_0$ may represent sufficient ex-ante disagreement between Sender and Receiver, as in Shishkin (2023) and for symmetric cutoff densities. In the proof, we construct a bi-upper censorship that replicates the same extensive margin as an arbitrary information policy I, and that improves upon I in terms of Receiver's action.

6 Conclusion

This paper introduces the receiver's choice of attention effort into a canonical model of persuasion with private information. Higher effort increases the probability of observing the messages of the sender, and the sender's audience is endogenous. The receiver's preferences exhibit a supermodularity property that allows us to establish a general equivalence result between single signals and persuasion mechanisms. The equivalence takes into account the receiver's action and effort choices. The sender's optimization problem is solved by censoring high states, a strategy relevant in applications such as media capture, where the sender values receiver's effort directly.

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. In this section, for notational convenience, we break Receiver's indifferences in the Sender's favor. This is without loss of generality in terms of the Receiver's payoff and the Sender's payoff given that the marginal distribution of the cutoff type c is atomless, by adapting the argument for Lemma B.5.

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

$$v_t(\mu) := \int_{[0,1]} [x_\mu \geq c](\theta - c) \, \mathrm{d}\mu(\theta),$$

where $[P]$ is the Iverson bracket of the statement P: $[P] = 1$ if the statement P is true, and $[P] = 0$ otherwise. We note that $v_t(\mu)$ depends on the belief μ only through x_μ. If the Sender's signal induces the distribution over posterior beliefs $p \in \Delta(\mathcal{D})$, letting $\Delta(\mathcal{D})$ be the set of Bayes-plausible posterior distributions (Kamenica and Gentzkow, 2011), type-t Receiver chooses effort to maximize her expected utility

$$e \int_{\mathcal{D}} v_t(\mu) \, \mathrm{d}p(\mu) + (1 - e) v_t(\mu_0) - \lambda k(e). \tag{2}$$

Thus, Receiver's action and 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 (and effort in the extensions of Section 5.)

A.2 Equilibrium Definition

We define a Perfect Bayesian Equilibrium in which the 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). 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:

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

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

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

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.

B Proofs

B.1 Auxiliary Results

Fact B.1 (Subdifferential of Convex Functions)**.** *Let $S \subseteq \mathbb{R}$, $f : S \longrightarrow \mathbb{R}$ be convex, and $\varphi : \mathbb{R} \longrightarrow \mathbb{R}$ 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 set of *information allocations* is:

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

Lemma B.1. *The following hold:*

1. *If* $A \in \mathcal{A}$*, then:* $A + I_{\overline{F}} \in \mathcal{I}$*.*
2. *If* $I \in \mathcal{I}$*, then* $\Delta I \in \mathcal{A}$*.*

Proof. The only nontrivial step is to show convexity of $A + I_{\overline{F}}$. We note that m, in the definition of A, is a subgradient of $A + I_{\overline{F}}$ at x_0. **QED**

Remark B.1. We note that any element of $\mathcal{F}$, $\mathcal{I}$ and $\mathcal{A}$ can be identified with its restriction on $[0,1]$, which without loss has codomain $[0,1]$. By the Lemmata 1 and B.1, there exists a bijection between any two of $\mathcal{F}$, $\mathcal{I}$, and $\mathcal{A}$ — e.g., take $F \longmapsto I_F$ with inverse $I \longmapsto I'$. Moreover, if we endow $\mathcal{F}$ with the Blackwell order and use the product order for $\mathcal{A}$ and $\mathcal{I}$, then the bijection is an order isomorphism.

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

Lemma B.2 (Envelope Theorem and Comparative Statics)**.** *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

$$V(a) = V(0) + \int_0^a \frac{\partial f}{\partial a}(e^\star(\tilde{a}),\tilde{a})\,\mathrm{d}\tilde{a}.$$

Since $\frac{\partial f}{\partial a}$ is nonnegative, V is nondecreasing. 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):

$$V(a') - V(a) \geq \int_a^{a'} \frac{\partial f}{\partial a}(e', \tilde{a}) \, \mathrm{d}\tilde{a}$$
$$\geq \int_a^{a'} \frac{\partial f}{\partial a}(e', a) \, \mathrm{d}\tilde{a},$$

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**

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 $V_\lambda(e, \Delta I(c)) := f(e, \Delta I(c)) - \lambda k(e)$, we use the shorthand $t = (c_t, \lambda_t)$, and we define the set of optimal efforts

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

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

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

Definition 6. 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.
$$\partial I_t(\zeta_t) \subseteq \partial I(\zeta_t) \quad \text{if } (0,1] \cap E_{\lambda_t}(\Delta I_t(\zeta_t)) \neq \emptyset.$$

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:

$$I(c) := \sup_{r \in R} I_r(c), \; c \in [0,1]. \tag{4}$$

$I(c)$ 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 B.2), 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 $a_n := \frac{1}{n}\Delta I_t(\zeta_t) + \big(1 - \frac{1}{n}\big)\Delta J(\zeta_t),\ n \in \mathbb{N}$, 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; since $J(\zeta_s) = I_s(\zeta_s)$ — as 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 7. (1) $\hat{F} \in \mathcal{F}$ is an *equilibrium experiment* if there exists an equilibrium $\langle F, e, \alpha \rangle$ with $\hat{F}(x) = F(x)$ for all $x \in \mathbb{R}$.

(2) The Receiver's *value from the experiment* $F \in \mathcal{F}$ is: $V_\lambda(\Delta I_F(c)) := \max_{e \in [0,1]} V_\lambda(e, \Delta I_F(c))$.

(3) 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

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

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.3. *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 B.2), there exists a selection e from $c \longmapsto \arg\max_{e \in [0,1]} e\Delta I_F(c) - \lambda k(e)$ such that:

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

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

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

We have the following chain of inequalities,

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

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:

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

QED

Lemma B.4. *(1) There exists a measurable selection from* $(c, \lambda, x) \longmapsto \max_{a \in \{0,1\}} U_R(x, a, e; c, \lambda)$ *for all* $e \in [0, 1]$*; (2) 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 7. The comparison holds because Definition 7 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.5 (Uniqueness of Sender's Payoff)**.** $F \in \mathcal{F}$ *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 Sende'rs expected utility in equilibrium, $\hat{W}(F) := \int_{[0,1]} \int_{[0,1]} \alpha(x, c, \lambda)\, \mathrm{d}(e^*_\lambda(\Delta I_F(c)) F)(x)\, \mathrm{d}G(c|\lambda)$, as follows,[21]

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

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

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

By the envelope theorem (Lemma B.2), 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 B.1), there exists a selection d from the subdifferential of $V_\lambda \circ \Delta I_F$ such that: $d(c) = e^*_\lambda(\Delta I_F(c)) d^1_I(c)$, for all $c \in [0, x_0]$, and $d(c) = e^*_\lambda(\Delta I_F(c)) d^2_I(c)$, 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 B.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).

[21] The symbol $\hat{W}$ is used for a slightly different function in Definition 7 because, for notational convenience, the current proof establishes a slightly stronger uniqueness statement than Definition 7 for an additional reason than the aforementioned one. Namely, the proof looks at the conditional expected Sender's utility given λ.

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:

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

Since $\Delta I_F(1) = \Delta I_F(0) = 0$, we have:

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

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.4, there exist e and α that satisfy the equilibrium measurability conditions, a Opt, and e Opt, given F. Since 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 $\mathcal{F}$, when we identify functions that are equal almost everywhere, 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.5 via upper semi continuity of the Sender's maximand in the definition of W maximality (Lemma B.3). **QED**

Proof of Theorem 2

Proof. Theorem 2 is implied by Lemma B.5 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.5 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.6. *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:*

$$I_\theta(x) - I(x) \geq 0, x \in [0,\zeta]$$
$$I_\theta(x) - I(x) \leq 0, x \in [\zeta,\infty).$$

.

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

(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:

$$
I(\zeta) + m(x - \zeta) \geq I_{\overline{F}}(x) \quad , x \in [0, x_u] \tag{5}
$$

$$
I(\zeta) + m(x - \zeta) \leq I_{\overline{F}}(x) \quad , x \in (x_u, 1]. \tag{6}
$$

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 (5).

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

Proof of Theorem 3

Proof. By Lemma B.5, the optimal experiment maximizes W defined as:

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

Suppose two experiments $F, H \in \mathcal{F}$ have information policies given by $I = I_F$, $J = I_H$ such that: $I(x) \geq J(x)$ for all $x \in [0,p]$ and $I(x) \leq J(x)$ 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 nonnpositive on $[p,1]$, it follows that $I_F \geq I_H$, it follows that $W(F) \geq W(H)$.

The result follows from Lemma B.6. **QED**

B.5 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, the following lemmata generalize the construction of Lemma B.6 to construct: an information policy that preserves the extensive margin and improves upon an arbitrary information policy. This information policy has three censorship regions, of which one is at the bottom, and its construction is complete after the lemmata. This bottom censorship region is payoff-irrelevant for Sender, and this last observation completes the 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 $\underline{c}_\lambda = 0$ 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 $\overline{c}_\lambda = 1$ otherwise. For the rest of this section, we omit reference to λ.

Lemma B.7. *Let $I \in \mathcal{I}$ such that $p \geq \overline{c}(\Delta I)$, there exists another information policy $I^\star$ such that:*

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

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

(IMPR)

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

.

(CENS) There exist $x_\ell, \theta_\ell, \theta_m, x_m$ *such that* $0 \le x_\ell \le \theta_\ell \le \theta_m \le x_m \le 1$, *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 following notation: $\overline{c}(I - I_{\overline{F}}) =: \overline{c}$, $\underline{c}(I - I_{\overline{F}}) =: \underline{c}$. In the first step, we prove the lemma in the case where there is a feasible information policy that is a straight line between the points $\underline{p} := (\underline{c}, I(\underline{c}))$ and $\overline{p} := (\overline{c}, I(\overline{c}))$. In the second step, we prove the lemma in the case where there is not a feasible information policy that is a straight line between the points $\underline{p}$ and $\overline{p}$.

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

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

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

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

$$\begin{aligned} y_\ell \text{ is: } & x \longmapsto I(\underline{c}) + \ell(x - \underline{c}) \\ y_m \text{ is: } & x \longmapsto I(\overline{c}) + m(x - \overline{c}). \end{aligned}$$

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

Lemma B.8. *ℓ, m are well-defined.*

Proof. L is nonempty because $I'(\underline{c}) \in L$, which follows from: (i) $I_{F_0}(x) \geq I(x)$ for all x and (ii) $I'(\underline{c}) \in \partial I(\underline{c})$. M is nonempty because $I'(\bar{c}) \in M$, which follows from: (i) $I_{F_0}(x) \geq I(x)$ 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.9. *that there exists a unique pair of numbers $(\theta_\ell, \theta_m) \in [\underline{c}, 1] \times [0, \bar{c}]$ such that:*

$$\begin{aligned} y_\ell(\theta_\ell) &= I_{F_0}(\theta_\ell) \\ y_m(\theta_m) &= I_{F_0}(\theta_m) \end{aligned}$$

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

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

Lemma B.10. $\theta_\ell \leq \theta_m$.

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

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

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

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

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

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

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

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

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

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

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

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

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

(FEAS) First, $I^\star$ is always above $I_{\overline{F}}$. Second $I^\star$ is always below I_{F_0}, which follows from $\lambda_\ell \in \partial I_{F_0}(x_\ell)$ for all $\ell \in \{1,3\}$. The maximum of affine functions is convex, and I_{F_0} is convex. Global convexity then follows if $I^\star$ is subdifferentiable at x_1 and x_3. We now claim that $\lambda_\ell \in \partial I^\star(x_\ell)$ for all $\ell \in \{1,3\}$. This claim follows from $\lambda_\ell \in \partial I_{F_0}(x_\ell)$ for all $\ell \in \{1,3\}$, and the fact that $I_{F_0}(x_1) = I(\underline{c}) + \lambda_1(x_1 - \underline{c})$

and $I_{F_0}(x_3) = I(\bar{c}) + \lambda_3(x_3 - \bar{c})$ (together with convexity of $I^\star$ in $[0, x_1]$ and $[x_3, 1]$). We established that the subdifferential of $I^\star$ at x_1 and x_3 nonempty, which finalizes the proof that $I^\star$ is globally convex.

QED

Proof of Proposition 2

Proof. By the previous lemmata, the reulst follows from taking $I^\star$ from the previous lemma until the point x_m° where $I^\star$ intercepts the line j: $x \longmapsto I(\bar{c}) + I'(\bar{c})(x - \bar{c})$, and $\max\{I_{\overline{F}}, j\}$ after x_m°. **QED**

B.6 Proof of Proposition 1

Proof. By Lemma B.5, the derivative of the Sender's expected utility, given information policy $I_{\overline{\theta}}$, with respect to $\overline{\theta}$ satisfies:

$$\frac{\partial F_0}{\partial \theta}(\overline{\theta}) \int_{\overline{\theta}}^{\bar{c}_\lambda(\Delta I_\theta)} (x - \overline{\theta}) \frac{\partial g}{\partial c}(x|\lambda)\, \mathrm{d}x \geq \frac{\partial F_0}{\partial \theta}(\overline{\theta}) \int_{\overline{\theta}}^{1} (x - \overline{\theta}) \frac{\partial g}{\partial c}(x|\lambda)\, \mathrm{d}x,$$

in which the right-hand side of the inequality is the derivative of the Sender's value of the $\overline{\theta}$ upper censorship with respect to the $\overline{\theta}$ when $\lambda = 0$. **QED**

B.7 Symmetric Information

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:

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

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 $\zeta \leq x_0$, $I_{\overline{F}}$ is optimal. Let $1 \geq \zeta \geq \theta_0$.

Lemma B.11. *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.6:

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

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:

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

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

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

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

References

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

An, Mark Yuying (1995), "Log-concave probability distributions: Theory and statistical testing." Working Paper.

Bagnoli, Mark and Ted Bergstrom (2005), "Log-concave probability and its applications." *Economic Theory*, 26, 445–469.

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

Bergemann, Dirk and Stephen Morris (2019), "Information Design: A Unified Perspective." *Journal of Economic Literature*, 57, 44–95.

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.

Bloedel, Alexander W. and Weijie Zhong (2021), "The cost of optimally acquired information." Working Paper.

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

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

Chahrour, Ryan (2014), "Public communication and information acquisition." *American Economic Journal: Macroeconomics*, 6, 73–101.

Cornand, Camille and Frank Heinemann (2008), "Optimal Degree of Public Information Dissemination." *The Economic Journal*, 118, 718–742.

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

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

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

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

Galperti, Simone and Isabel Trevino (2020), "Coordination motives and competition for attention in information markets." *Journal of Economic Theory*, 188, 105039.

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

Gentzkow, Matthew and Emir Kamenica (2016), "A Rothschild-Stiglitz Approach to Bayesian Persuasion." *American Economic Review*, 106, 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, 653–675.

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, 2590–2615.

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

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

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

Kolotilin, Anton, Tymofiy Mylovanov, Andriy Zapechelnyuk, and Ming Li (2017), "Persuasion of a Privately Informed Receiver." *Econometrica*, 85, 1949–1964.

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

Lipnowski, Elliot, Laurent Mathevet, and Dong Wei (2020), "Attention management." *American Economic Review: Insights*, 2, 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 and Doron Ravid (2020), "Cheap talk with transparent motives." *Econometrica*, 88, 1631–1660.

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, 277–324.

Matysková, Ludmila and Alfonso Montes (2023), "Bayesian persuasion with costly information acquisition." *Journal of Economic Theory*, 211, 105678, URL https://www.sciencedirect.com/science/article/pii/S0022053123000741.

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

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

Myatt, David P. and Chris Wallace (2014), "Central bank communication design in a Lucas-Phelps economy." *Journal of Monetary Economics*, 63, 64–79.

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

Prat, Andrea (2015), "Chapter 16 - Media Capture and Media Power." In *Handbook of Media Economics* (Simon P. Anderson, Joel Waldfogel, and David Strömberg, eds.), volume 1 of *Handbook of Media Economics*, 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, 346–387.

Rayo, Luis and Ilya Segal (2010), "Optimal Information Disclosure." *Journal of Political Economy*, 118, 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, 250–76.

Shishkin, Denis (2023), "Evidence Acquisition and Voluntary Disclosure." Working Paper.

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

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

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