The Interval Structure of Optimal Disclosure

Yingni Guo[*] Eran Shmaya[†]

Abstract

A sender persuades a receiver to accept a project by disclosing information about a payoff-relevant quality. The receiver has private information about the quality, referred to as his type. We show that the sender-optimal mechanism takes the form of nested intervals: each type accepts on an interval of qualities and a more optimistic type's interval contains a less optimistic type's interval. This nested-interval structure offers a simple algorithm to solve for the optimal disclosure and connects our problem to the monopoly screening problem. The mechanism is optimal even if the sender conditions the disclosure mechanism on the receiver's reported type.

JEL: D81, D82, D83

Keywords: Information design, sender-receiver games, privately informed receiver, nested intervals

1 Introduction

Sender-receiver games model situations in which an informed party (the sender) tries to influence the action of another party (the receiver). Such games are usually analyzed using cheap-talk equilibria. The more recent information-design literature assumes commitment power on the sender's side and looks for the sender's optimal disclosure mechanism. In this paper, we consider the information design problem in an environment in which the receiver has access to external sources of information.

For instance, an entrepreneur tries to convince an insider investor to fund his project. The insider investor observes the outcome of any experiment done by the entrepreneur to evaluate the project's quality (Diamond (1984), Diamond (1991)). The investor can also generate information by running his own experiments and becoming privately informed (Azarmsa

[*]Department of Economics, Northwestern University, yingni.guo@northwestern.edu.
[†]MEDS, Kellogg School of Management, Northwestern University, e-shmaya@kellogg.northwestern.edu.

and Cong (2018)). In a different example, a lobbyist (e.g., an industry association) tries to sway a policymaker's decision and gets involved in the design of scientific studies that he funds. The policymaker also conducts an internal investigation to gather information about the policy issue and becomes privately informed (Minaudier (2018)). In this paper we study how to design a disclosure mechanism when the receiver has private information about the project or policy issue.

Environment. We consider an environment with two players: the sender and the receiver. The sender promotes a project to the receiver, who decides whether to accept or reject it. The receiver's payoff from accepting increases with the project's quality. His payoff is normalized to zero if he rejects. Therefore, there is a quality threshold such that the receiver benefits from the project if the quality is above the threshold, and loses otherwise. The sender, on the other hand, simply wants the project to be accepted.

The receiver does not know the quality, but has access to an external information source. The information from this source is *the receiver's type*. We assume that the higher the quality, the more likely it is that the receiver has a higher rather than a lower type. Correspondingly, a higher receiver's type is more optimistic that the quality favors a decision to accept than a lower type is. The sender designs a disclosure mechanism to reveal more information about the quality and can commit to this mechanism. The receiver updates his belief based on his private information and the sender's signal. He then takes an action: to accept or to reject.

Main results. To illustrate the structure and the intuition of the optimal mechanism, we consider the case in which the receiver has two possible types: high or low. The sender could design a pooling mechanism under which both types of the receiver always take the same action after any signal. The sender could also design a separating mechanism under which, after some signal, only the high type (who is more optimistic about the quality) accepts. When the receiver's external information is sufficiently informative, the high type is much more optimistic and thus much easier to persuade than the low type, which suggests that a separating mechanism may provide a higher payoff to the sender.

Under any separating mechanism, the high type will accept whenever the low type accepts. In addition, the high type will accept for some other qualities. Our main result states that the optimal separating mechanism takes on a nested-interval structure. This structure has two key properties. First, the set of qualities for which each type accepts is an interval. Second, the high type's interval contains the low type's. These two properties imply that the high type is the only type that accepts when the quality is either quite good or quite bad.

The interval structure translates to an intuitive rule of optimal disclosure. The sender pushes for the project to be accepted by both types for the intermediate qualities. However, when the quality is either quite good or quite bad, the sender pools these qualities and lets the receiver decide based on his own private information.

Pooling the extreme types of the sender in an environment with a privately informed receiver is a phenomenon that has been discussed in recent papers. In the context of advertising, Mayzlin and Shin (2011) argue that ads for high-quality products and low-quality products tend to contain no attribute information about the product. In contrast, ads for medium-quality products contain specific product attributes. The rationale is that an ad with no attributes invites consumers to search, which is likely to uncover positive information for high-quality products. In the market for prescription drugs, Bradford, Turner and Williams (2018) study the rise in off-label use of prescription drugs by physicians. While the FDA's regulations restrict marketing of a drug to only the indications for which it was approved after its efficacy was established, physicians may prescribe it off-label to treat other indications.[1] Thus, a drug company that wishes to sell a drug for indication A can either apply and run efficacy trials for indication A, or apply for indication B and, if successful, support the use of the drug for off-label use for indication A. Due to the FDA restrictions on marketing, the sales of the drug for indication A in the second approach will depend on physicians' knowledge about the drug. Bradford, Turner and Williams (2018) point out that, compared to the on-label use, "off-label use may give a patient either the highest-quality option or a lower-quality option."

The standard explanation for pooling the extreme types of the sender is countersignaling (e.g., Feltovich, Harbaugh and To (2002)). In a countersignalling equilibrium, such pooling is tied to the incentives of different sender's types in the equilibrium; for example, high-quality students choose no education (as do low-quality students) to save the cost of education and to separate themselves from medium-quality students. We give a different explanation for this phenomenon. In our environment, the sender faces a design problem, and pooling the extreme types of the sender is tied to the receiver's incentives.

In our environment with more than two receiver's types, we show that the optimal disclosure mechanism has the following structure: (i) each receiver's type is endowed with an interval of qualities under which this type accepts; (ii) a lower type's acceptance interval is a subset of a higher type's acceptance interval. This interval structure gives us a simple algorithm to find the optimal disclosure: we only need to find the endpoints of the acceptance interval of each receiver's type.

[1] Bradford, Turner and Williams (2018) estimate that the rate of off-label use has risen from 29.9% to 38.3% from 1993 to 2008.

Moreover, the interval structure offers a natural connection between the persuasion problem and the monopoly screening problem analyzed in Mussa and Rosen (1978). We divide the qualities into "positive qualities" and "negative qualities," depending on the sign of the receiver's payoff from accepting. We can interpret the receiver's payoff from accepting on an interval of positive qualities as a buyer's utility from consuming some quantity, while the receiver's payoff from accepting on an interval of negative qualities is the buyer's disutility from paying some price. The receiver's level of optimism (i.e., his type) corresponds to how much the buyer values quantity. Once we use our theorem to impose the interval structure, the sender's problem in our setup has the same structure as the monopoly screening problem.

In the baseline model, we adopt relatively simple payoff functions in order to illustrate the essence of our solution. Later on, we extend the results to more general setups by allowing both sender's and receiver's payoffs to depend on both the quality and the receiver's type. We also show that our solution remains optimal in the environment in which the receiver first reports his type and the sender can disclose different information to different reported types.

Intuition for the interval structure. We again illustrate the intuition with the binary-type case. The acceptance set of each type—the set of qualities for which this type accepts—includes some positive and some negative qualities.

Consider a fixed receiver's type. Suppose that the sender wants to provide a given positive payoff to the receiver from accepting some positive qualities. In order to maximize the probability that the receiver accepts, the sender should recommend the positive qualities that are closest to the threshold. This is because the receiver's payoff from accepting increases in the quality. Similarly, for a given negative payoff to the receiver from accepting some negative qualities, the sender should recommend the negative qualities closest to the threshold. This argument suggests that the set of qualities for which each type accepts is an interval that contains the threshold.

We then fix the positive qualities for which the high type accepts. By excluding some of these positive qualities from the low type, the sender can provide the high type with some surplus over mimicking the low type. For a given surplus to the high type, the sender wants to choose the qualities to exclude from the low type in a way that causes the least damage to the low type. We claim that the sender should exclude the highest qualities from the low type. This is because the low type believes these qualities to be quite unlikely. Similarly, we fix the negative qualities for which the high type accepts. By excluding some of these negative qualities from the low type, the sender can incentivize the low type to follow the acceptance recommendation. However, excluding some negative qualities increases the high type's incentive to mimic the low type. Excluding the lowest qualities creates the smallest increase in the high type's incentive to mimic the low type. This is because the high type believes these qualities to be quite unlikely. These two arguments show that not giving the low type the extreme qualities in the high type's acceptance set is the cheapest way to maintain the incentives for different types.

In order to complete the proof, we need to achieve two properties simultaneously: (i) that each type's acceptance set is an interval, and (ii) that the qualities for which only the high type accepts are the extreme qualities in his acceptance set. The key observation is that these two desirable properties are compatible with each other. Indeed, if an interval strictly contains another interval, then the set of qualities in the former that are not included in the latter are precisely the extreme qualities on both sides.

Related literature. Our paper is related to the literature on information design. Rayo and Segal (2010) and Kamenica and Gentzkow (2011) study optimal persuasion between a sender and a receiver.[2] We study the information design problem in which the receiver has private information about the quality. This is motivated by the observation that the receiver has multiple sources of information. Kamenica and Gentzkow (2011) extend the geometric method to situations in which the receiver has private information. Nonetheless, it is generally difficult to solve for the optimal mechanism using the geometric method. Our approach enables us to explicitly solve for the optimal mechanism.

Kolotilin (2018) also examines optimal persuasion when the receiver has private information. He provides a linear programming approach and establishes the conditions under which either full or no revelation is optimal. Kolotilin et al. (2017) assume that the receiver privately learns about his threshold for accepting. The receiver's threshold is independent of the quality. Au (2015) studies dynamic disclosure in a similar setting where the receiver's private information is independent of the quality. Our paper differs in that the receiver's type is informative about the quality. Bergemann and Morris (2018) provide the incentive conditions for a decision rule to be implementable by the sender both when he can elicit the receiver's private information and when he cannot (propositions 2 and 3). They show that these two scenarios admit the same set of implementable decision rules with a binary state space (a single receiver and a binary action space), but the equivalence breaks down with more than two states. Azarmsa and Cong (2018) examine the optimal disclosure by an entrepreneur to a privately informed insider investor, with a focus on information hold-up and security design.

[2]Rayo and Segal (2010) assume that the receiver's taste is his private information and examine optimal persuasion when all types observe the same signal by the sender.

In terms of our assumption about the receiver's private information, our paper is closely related to Li and Shi (2017) and to Bergemann, Bonatti and Smolin (2018). Like ours, both papers assume that the buyer has private information about the quality. Li and Shi (2017) consider a seller who sells a product to a buyer whose type is his information about the quality. For each type of buyer, the seller designs an experiment along with a strike price and an advance payment. They show that disclosing different information to different types dominates full disclosure. Bergemann, Bonatti and Smolin (2018) consider a monopolist who sells experiments to a buyer. The monopolist designs a menu of experiments and a tariff function to maximize his profit. Both papers examine the optimal design when the seller can disclose different information to different types. In contrast, we consider both settings in which the sender discloses the same information to all types, and those in which the sender discloses different information to different types. Moreover, we focus on settings in which transfers are not allowed.

Our model admits both the interpretation of a single receiver and that of a continuum of receivers. For this reason, our paper is also related to information design with multiple receivers. Lehrer, Rosenberg and Shmaya (2010), Lehrer, Rosenberg and Shmaya (2013), Bergemann and Morris (2016*a*), Bergemann and Morris (2016*b*), Mathevet, Perego and Taneva (2018), and Taneva (2018) examine the design problem in a general environment. Inostroza and Pavan (2018) study persuasion in global games. Our paper is most closely related to the work in Arieli and Babichenko (2017), which studies optimal persuasion when receivers have different thresholds for acceptance. They assume that receivers' thresholds are common knowledge, so there is no private information.

Our paper is also related to cheap talk games (Crawford and Sobel (1982)) with a privately informed receiver (e.g., Seidmann (1990), Watson (1996), Olszewski (2004), Chen (2009), and Lai (2014)). The setup in Chen (2009) is the closest to ours; she assumes that the receiver has a signal about the sender's type, and shows that nonmonotone equilibria may arise. Our model differs from Chen's in that we focus on the optimal disclosure mechanism. In terms of the receiver's private information, our paper is related to signaling or disclosure games with privately informed receivers (e.g., Feltovich, Harbaugh and To (2002), Angeletos, Hellwig and Pavan (2006), and Quigley and Walther (2018)).

2 An example

Consider an e-commerce platform (Sender) promoting a product to a customer (Receiver), who decides whether or not to buy the product. The customer's payoff from buying depends on the product's quality s, which is uniform on $[0, 1]$. The customer's payoff if he buys is $s - 3/4$. The platform's payoff is one if the customer buys. If not, both players get a zero payoff.

The platform designs a mechanism for disclosing information about s and can commit to this mechanism (e.g., Rayo and Segal (2010) and Guo and Shmaya (2018)). If the customer has no private information, the platform will reveal whether s is above or below $1/2$. The platform extracts the entire surplus, so that the customer's expected payoff from buying is zero.

The customer has some private information about the product's quality, captured by his type. Given s, the customer's type is H with probability s and L with probability $1 - s$. Naturally, the higher the product's quality, the more likely that the customer's type is H. Thus, type H is more optimistic about s and easier to persuade than type L is.

We show that the optimal disclosure takes the form shown in figure 1. The x-axis is the product's quality. The platform signals whether the quality is in the dark-gray or light-gray regions. When the quality is in the dark-gray region (concentrated around 3/4), the product is a solid one. When the quality is in one of the light-gray regions, the product can be either good or bad. These regions are such that, given their private information, both types want to buy if they learn that the quality is in the dark-gray region, but only type H wants to buy if they learn that the quality is in one of the light-gray regions. Note that the platform does not extract the entire surplus, since type H gets a positive payoff from buying when the quality is in the dark-gray region.

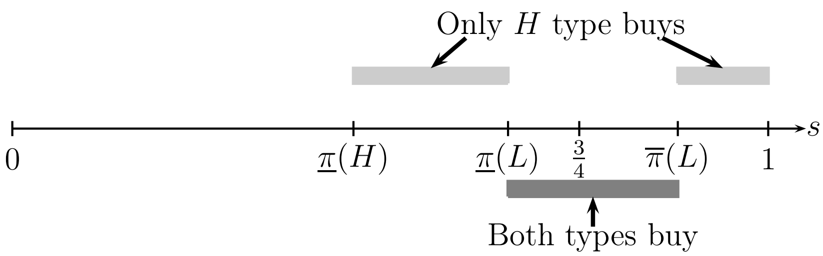

Figure 1: Optimal disclosure for binary types

Thus, the platform should "spend its capital" on intermediate products, and push for the buying decision. It should pool relatively extreme products, for which the customer decides based on his own information. In particular, this implies that the platform lets its best products sell themselves to the customer.

A central concept in our solution is the notion of nested intervals. Both types buy on an interval of qualities: type L buys only in the dark-gray region and type H buys in both the dark-gray and the light-gray regions. Type L's buying interval is a subset of type H's. Compared to type H's interval, type L does not buy when the quality is either quite good or quite bad.

3 Environment and main results

Let $\mathcal{S}$, the set of *Sender's types*, be a bounded interval on the real line equipped with Lebesgue measure μ; let $\mathcal{T}$, the set of *Receiver's types*, be a subset of the real line equipped with a σ-finite measure λ of full support. In all our examples, $\mathcal{T}$ is either a discrete space equipped with the counting measure or an interval equipped with Lebesgue measure. Consider a distribution over $\mathcal{S} \times \mathcal{T}$ with density f with respect to $\mu \times \lambda$. We also make the technical assumptions that the set $\mathcal{T}$ is closed and bounded from below, with the lowest type denoted by $\underline{\mathrm{t}}$, and that the density function $f(s,t)$ is bounded and continuous in t.

Let $u\colon \mathcal{S} \to \mathbf{R}$ be a bounded and nondecreasing function representing *Receiver's payoff* from accepting. Receiver's payoff from rejecting is zero. We assume that Sender's payoff is one if Receiver accepts and is zero otherwise. We adopt this payoff structure in order to simplify the exposition. In section 5.1, we extend our results by allowing Sender's and Receiver's payoffs from accepting to depend on both Sender's and Receiver's types.

What we call "Sender's type" is called "state" in most persuasion papers (e.g., Kamenica and Gentzkow (2011)). In most papers, Receiver has no private information about Sender's type. However, in games with incomplete information, the term "state" refers to all the uncertainty, so it means the pair of types in our setup. We follow the terminology of games with incomplete information.

We make the following assumption about the distribution of Sender's and Receiver's types:

Assumption 1. [i.m.l.r.] The density function f is *increasing in monotone likelihood-ratio order*: For every $t' < t$, the ratio $f(s,t)/f(s,t')$ is weakly increasing in s.

Different types of Receiver have different beliefs about Sender's type. Given assumption 1, the higher t is, the more optimistic Receiver is about the distribution over Sender's types.

Example 1. We extend the platform example from section 2. Sender's type space is $\mathcal{S} = [0,1]$. Receiver's payoff from accepting is $u(s) = s - \zeta$, where ζ is a parameter. Receiver's type space is $\mathcal{T} = \{L, H\}$ with $L < H$. The density $f(s,t)$ is given by

$$f(s,H) = 1/2 + \phi(s - 1/2) \text{ and } f(s,L) = 1/2 - \phi(s - 1/2), \quad \text{for } 0 \leqslant s \leqslant 1.$$

The parameter $\phi \in [0,1]$ captures the accuracy of Receiver's private information. When ϕ is zero, Receiver has no private information. As ϕ increases, Receiver's type becomes more informative about Sender's type. $\square$

Example 1 shows that, as in games with incomplete information, Sender's type captures two aspects of the game. First, it affects Receiver's payoff from accepting. Second, it determines Sender's belief about Receiver's type: when Sender's type is s, Receiver's type is H with probability $1/2 + \phi(s - 1/2)$ and L with probability $1/2 - \phi(s - 1/2)$. Both aspects have implications for designing the optimal mechanism. The first aspect appears in all information design papers. The second aspect appears in setups, like ours, in which Receiver has some private information about Sender's type. Throughout the paper, when we use "type" alone, we mean "Receiver's type."

3.1 Disclosure mechanisms

A *disclosure mechanism* is given by a triple $(\mathcal{X}, \kappa, r)$ where $\mathcal{X}$ is a set of *signals*, κ is a Markov kernel from $\mathcal{S}$ to $\mathcal{X}$,[3] and $r : \mathcal{X} \times \mathcal{T} \to \{0,1\}$ is a *recommendation function.* When Sender's type is s, the mechanism randomizes a signal x according to $\kappa(s, \cdot)$ and recommends that type t accept if and only if $r(x,t) = 1$.

There are many ways in which Sender can disclose information. To illustrate, we give two examples of signaling structures $(\mathcal{X}, \kappa)$:

- Sender can reveal whether his type is above or below some threshold $\underline{s} \in \mathcal{S}$. The signal space is $\mathcal{X} = \{above, below\}$. For any $s \geqslant \underline{s}$, $\kappa(s, \cdot) = \delta_{above}$. For any $s < \underline{s}$, $\kappa(s, \cdot) = \delta_{below}$.
- Sender can randomize. Let $\mathcal{X} = \{above, below, null\}$ and let $\underline{s} \in \mathcal{S}$ be some threshold. For any $s \geqslant \underline{s}$, $\kappa(s, \cdot) = 1/2\delta_{above} + 1/2\delta_{null}$. This means that, for any $s \geqslant \underline{s}$, Sender says "above" with probability 1/2 and "null" with probability 1/2. For any $s < \underline{s}$, $\kappa(s, \cdot) = 1/2\delta_{below} + 1/2\delta_{null}$. Under this mechanism, Sender sometimes reveals whether his type is above or below $\underline{s}$, and sometimes reveals nothing.

Once Receiver observes the signal x, the recommendation function $r(x, \cdot)$ contains no further information about Sender's type. We could define a disclosure mechanism to be a signaling structure $(\mathcal{X}, \kappa)$. However, we choose to include the recommendation function in our definition of a disclosure mechanism, because it allows us to succinctly define the incentive-compatibility constraints and Sender's problem.

[3]That is, $\kappa : \mathcal{S} \times \mathcal{B}(\mathcal{X}) \to [0,1]$ such that $\kappa(s, \cdot)$ is a probability measure over $\mathcal{X}$ for every $s \in S$, where $\mathcal{B}(\mathcal{X})$ is the sigma-algebra of Borel subsets of $\mathcal{X}$.

3.2 Incentive compatibility

Receiver observes the signal and decides whether to accept or not. Incentive compatibility means that, after observing the signal, each type will follow the mechanism's recommendation. We now proceed to formally define incentive compatibility.

A *strategy* for type t is given by $\sigma : \mathcal{X} \to \{0, 1\}$. Let $\sigma_t^* = r(\cdot, t)$ be the strategy that follows the mechanism's recommendation for type t. We say that the mechanism is *incentive-compatible (IC)* if, for every type t,

$$\sigma_t^* \in \arg\max \int f(s,t)u(s) \left(\int \sigma(x)\ \kappa(s, \mathrm{d}x) \right) \mu(\mathrm{d}s), \tag{1}$$

where the argmax ranges over all strategies σ. The expression inside the argmax is, up to normalization, the expected payoff of type t under σ.

For an IC mechanism, the recommendation function r is almost determined by the signaling structure $(\mathcal{X}, \kappa)$.[4] Type t receives a recommendation to accept if his expected payoff conditional on the signal is positive, and to reject if it is negative.

3.3 Sender's optimal mechanism

Sender's problem is:

$$\text{Maximize} \iint f(s,t) \left(\int r(x,t)\ \kappa(s, \mathrm{d}x) \right)\ \mu(\mathrm{d}s)\lambda(\mathrm{d}t) \tag{2}$$

among all IC mechanisms. We assume a common prior between Sender and Receiver, which is reflected by the fact that the same density function f appears in (1) and (2).

If Receiver has no private information (i.e., if $\mathcal{T}$ is a singleton), then under the optimal mechanism Sender reveals whether or not his type is above some $\underline{s} \in \mathcal{S}$, and recommends accepting only when his type is above $\underline{s}$. The threshold $\underline{s}$ is chosen such that Receiver's expected payoff from accepting is zero.

In this paper we solve for the optimal disclosure when Receiver's type is correlated with Sender's type, so that Receiver has some private information about Sender's type. As we shall see, each type t still accepts on an interval of Sender's types and the interval expands as t increases. However, unlike the case of no private information, the acceptance intervals are typically bounded from above. While the lowest type of Receiver gets a payoff of zero,

[4] We say "almost determined" because of the possible indifference and because the conditional expected payoff is defined up to an event with zero probability. We omit the formal statement of this assertion, in order not to get into the technical intricacies involving conditional distributions.

higher types receive some information rent. We formally define these properties of the optimal mechanism here, and provide the proof in section 6.

We say that the mechanism is a *cutoff mechanism* if (i) the signal space is given by $\mathcal{X} = \mathcal{T} \cup \{\infty\}$, and (ii) $r(x,t) = 1$ if and only if $t \geqslant x$. Thus, a cutoff mechanism announces a cutoff type x and recommends that type x and all higher types accept. If the mechanism announces infinity to be the cutoff type, then it recommends that all types reject. A *deterministic cutoff mechanism* is one in which $\kappa(s,\cdot)$ is Dirac's measure on $z(s)$ for some function $z : \mathcal{S} \to \mathcal{T} \cup \{\infty\}$. Thus, when Sender uses a deterministic cutoff mechanism, he performs no randomization and, when his type is s, he announces a cutoff type $z(s)$.

Under a deterministic cutoff mechanism, the *acceptance set* of type t (i.e., the set of Sender's types for which type t accepts) is $\{s : t \geqslant z(s)\}$. By the definition of a cutoff mechanism, whenever the mechanism recommends that type t accept, it also recommends that any higher type accept. Therefore, for a deterministic cutoff mechanism, type t's acceptance set is a subset of a higher type's acceptance set. Lastly, we say that the mechanism *recommends accepting on intervals* if the acceptance sets of all Receiver's types are intervals.

Figure 2 illustrates two deterministic cutoff mechanisms for our platform example. The y-axis is the signal space $\mathcal{X} = \{L, H\} \cup \{\infty\}$. In both mechanisms, the solid line illustrates the cutoff function $z(s)$. The dashed line illustrates type H's acceptance set, and the dotted line gives type L's acceptance set. The mechanism on the right-hand side recommends accepting on intervals, since both types' acceptance sets are intervals. On the left-hand side, neither type accepts on an interval.

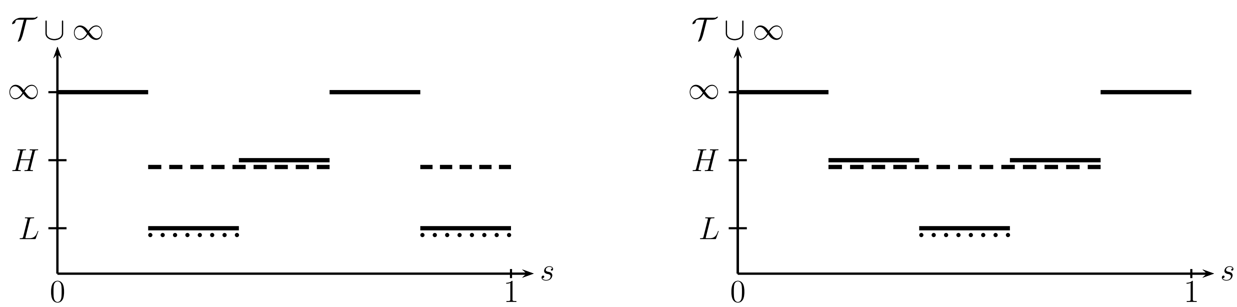

Figure 2: Examples of deterministic cutoff mechanisms

The following theorem is our main result. It states that Sender's optimal mechanism is a deterministic cutoff mechanism that recommends accepting on intervals, and formulates Sender's problem in terms of the endpoints of each type t's acceptance interval.

Theorem 3.1. *Under assumption 1, the optimal IC mechanism is a deterministic cutoff mechanism that recommends accepting on intervals. The acceptance intervals* $[\underline{\pi}(t), \overline{\pi}(t)]$ *are* *the solution to the following optimization problem:*

$$\begin{aligned} &\textit{Maximize} \quad \int W\left(\underline{\pi}(t), \overline{\pi}(t), t\right)\ \lambda(\mathrm{d}t) \\ &\textit{subject to} \quad U\left(\underline{\pi}(t), \overline{\pi}(t), t\right) \geqslant U\left(\underline{\pi}(t'), \overline{\pi}(t'), t\right) \textit{ for all } t' \leqslant t, \\ &\qquad\qquad\quad U\left(\underline{\pi}(\underline{t}), \overline{\pi}(\underline{t}), \underline{t}\right) \geqslant 0, \end{aligned} \tag{3}$$

over all functions $\underline{\pi}, \overline{\pi} : \mathcal{T} \to \mathcal{S}$ *such that* $\underline{\pi}$ *is monotone-decreasing,* $\overline{\pi}$ *is monotone-increasing, and* $\underline{\pi}\left(\underline{t}\right) \leqslant \overline{\pi}\left(\underline{t}\right)$*. The functions* $W, U : \mathbf{R}^3 \to \mathbf{R}$ *are given by*

$$W\left(\underline{q}, \overline{q}, t\right) = \int_{\underline{q}}^{\overline{q}} f(s,t)\ \mu(\mathrm{d}s), \ \textit{and}\ U\left(\underline{q}, \overline{q}, t\right) = \int_{\underline{q}}^{\overline{q}} u(s) f(s,t)\ \mu(\mathrm{d}s).$$

It is a standard argument that under assumption 1 every IC mechanism is essentially a cutoff mechanism. This is because we can replace each signal with the lowest Receiver's type that is still willing to accept given that signal. Due to the i.m.l.r. assumption, all higher types are willing to accept.[5] The contribution of theorem 3.1 is that the optimal mechanism has the additional property that every type accepts on an interval. Moreover, when choosing the endpoints of each type's interval, Sender needs to ensure only that each type's payoff from accepting on his interval is weakly higher than that from accepting on a lower type's interval, and that the lowest type's payoff from accepting is weakly positive.

The i.m.l.r. assumption is essential for theorem 3.1. The fact that the optimal mechanism does not randomize follows from the assumption that Sender's type space is nonatomic. We show in the online appendix how to relax this assumption. The assumption that $f(s,t)$ is continuous in t is not essential. Without it, we only need to modify the definition of a cutoff mechanism: in addition to announcing the cutoff type, the mechanism needs to announce whether the cutoff type is supposed to accept or reject.

We now explain Sender's problem in (3) by relating it to the monopoly screening problem, as in Mussa and Rosen (1978). For ease of exposition, we assume that Sender's type space $\mathcal{S}$ is the real line and that $u(0) = 0$. We refer to $s > 0$ as "positive types" and $s < 0$ as "negative types," so that Receiver gains a positive payoff from accepting on positive types and a negative payoff from accepting on negative types.

From theorem 3.1 we know that under the optimal mechanism each type t accepts on an interval of Sender's types. When type t accepts on the interval $\left[\underline{q}, \overline{q}\right]$ with $\underline{q} < 0 < \overline{q}$, his

[5] We do not formalize and prove this assertion because we do not need it. We say "essentially" because of the possible indifference and some intricacies of zero-probability events.

(normalized) payoff is given by:

$$U\left(\underline{q},\overline{q},t\right)=\overline{U}\left(\overline{q},t\right)+\underline{U}\left(\underline{q},t\right).$$

Here $\overline{U}\left(\overline{q},t\right)=\int_0^{\overline{q}} u(s)f(s,t)\ \mu(\mathrm{d}s)$ and $\underline{U}\left(\underline{q},t\right)=\int_{\underline{q}}^0 u(s)f(s,t)\ \mu(\mathrm{d}s)$ are, respectively, the payoff from accepting on $(0,\overline{q}]$ and the payoff from accepting on $\left[\underline{q},0\right)$.

Consider now the following monopoly screening problem. A seller offers a divisible good and is uncertain about the type of the buyer. A type t buyer gets payoff $\overline{U}(\overline{q},t)$ from consuming a good of quantity $\overline{q}$ and gets negative payoff $\underline{U}\left(\underline{q},t\right)$ from paying price $-\underline{q}$. This is a standard screening problem, except that the negative payoff $\underline{U}\left(\underline{q},t\right)$ from paying $-\underline{q}$ is usually assumed to be $\underline{q}$.

A mechanism for the monopoly screening problem is one that offers the buyer of type t the quantity $\overline{\pi}(t)$ and charges the price $-\underline{\pi}(t)$. The constraints of Sender's problem given in (3) are the familiar "downward" IC constraints of the screening problem, namely, that no type t would mimic a lower type t' in the screening problem.

We emphasize that, while the connection between Sender's problem and the monopoly screening problem is intuitive, it relies on two assertions in theorem 3.1. First, the theorem asserts that the optimal disclosure mechanism recommends that each type accept on an interval, which allows us to interpret the endpoints of the intervals as quantity and price. Second, the theorem asserts that the IC constraints of the monopoly screening problem are sufficient for Sender's problem, even though in our environment Receiver has many deviation strategies other than just following the mechanism's recommendation for some lower type: Given a cutoff mechanism, each type t can choose an arbitrary set of announced cutoff types, after which type t accepts.

4 The binary-type case

In this section, we fully characterize the optimal mechanism for the binary-type case. The analysis for any finite-type space is similar. Our goal is threefold. First, we show how theorem 3.1 reduces Sender's problem to a finite-dimensional constrained-optimization problem. Second, we provide the necessary and sufficient condition for pooling to be optimal. Third, we explain why, even in the binary-type case of example 1, a solution to Sender's problem could not be derived by the arguments provided in previous papers.

Let $\mathcal{T}=\{L,H\}$ and $\mathcal{S}=[0,1]$. Receiver's payoff u is strictly monotone-increasing, and $u(\zeta)=0$ for some fixed $\zeta\in[0,1]$. Assumption 1 means that $\frac{f(s,H)}{f(s,L)}$ is monotone-increasing in s. Sender's problem is trivial if type L accepts without further information, so we assume otherwise.

Theorem 3.1 shows that the optimal mechanism takes the form of nested intervals: type L accepts when $s \in [\underline{\pi}(L), \overline{\pi}(L)]$ and type H accepts when $s \in [\underline{\pi}(H), \overline{\pi}(H)]$. The endpoints are the solution to the following problem, in which we have one incentive constraint for each type:

$$\begin{aligned}
\underset{\underline{\pi}(H),\underline{\pi}(L),\overline{\pi}(L),\overline{\pi}(H)}{\text{Maximize}} \quad & \int_{\underline{\pi}(L)}^{\overline{\pi}(L)} f(s,L)\ \mu(\mathrm{d}s) + \int_{\underline{\pi}(H)}^{\overline{\pi}(H)} f(s,H)\ \mu(\mathrm{d}s) \\
\text{subject to} \quad & 0 \leqslant \underline{\pi}(H) \leqslant \underline{\pi}(L) \leqslant \overline{\pi}(L) \leqslant \overline{\pi}(H) \leqslant 1, \\
& \int_{\underline{\pi}(H)}^{\underline{\pi}(L)} u(s)f(s,H)\ \mu(\mathrm{d}s) + \int_{\overline{\pi}(L)}^{\overline{\pi}(H)} u(s)f(s,H)\ \mu(\mathrm{d}s) \geqslant 0, \\
& \int_{\underline{\pi}(L)}^{\overline{\pi}(L)} u(s)f(s,L)\ \mu(\mathrm{d}s) \geqslant 0.
\end{aligned} \tag{4}$$

The structure of the problem allows us to further reduce the number of variables to one. First, the constraint $\overline{\pi}(H) \leqslant 1$ is binding; otherwise, increasing $\overline{\pi}(H)$ would increase Sender's payoff without violating the constraints. Second, type H's constraint is binding; otherwise, increasing $\overline{\pi}(L)$ would increase Sender's payoff without violating the constraints. Lastly, type L's constraint is binding; if not, decreasing $\underline{\pi}(L)$ would benefit Sender while still satisfying the constraints. We are left with one variable and can derive the condition under which Sender pools the two types.

Proposition 4.1. *Pooling is optimal if and only if*

$$\frac{f(1,H)}{f(1,L)} - \frac{f\left(\underline{\pi}^*(L),H\right)}{f\left(\underline{\pi}^*(L),L\right)} < 1 - \frac{u\left(\underline{\pi}^*(L)\right)}{u(1)}, \tag{5}$$

where $\underline{\pi}^*(L)$ is such that type L is indifferent between accepting on $[\underline{\pi}^*(L), 1]$ and rejecting, i.e., $\int_{\underline{\pi}^*(L)}^1 u(s)f(s,L)\ \mu(\mathrm{d}s) = 0$. If (5) holds, then the mechanism recommends that both types accept on $[\underline{\pi}^*(L), 1]$.

Condition (5) states that Sender does not benefit if he marginally shrinks type L's interval in order to expand type H's interval. More explicitly, starting from the pooling interval $[\underline{\pi}^*(L), 1]$, Sender can replace type L's interval with

$$\left[\underline{\pi}^*(L) + \frac{f(1,L)}{f\left(\underline{\pi}^*(L),L\right)}\frac{u(1)}{-u\left(\underline{\pi}^*(L)\right)}\varepsilon, 1-\varepsilon\right] \quad \text{for small } \varepsilon > 0,$$

without violating type L's incentive constraint. Due to the i.m.l.r. assumption, this change allows Sender to lower $\underline{\pi}(H)$, so that type H accepts more often. Condition (5) states that Sender will not benefit from this operation. Therefore, this condition is necessary for pooling to be optimal. We show that this condition is also sufficient. This follows from the fact that the condition is the first-order optimality condition in the constrained-optimization problem (4). Intuitively, the gain from type H when we marginally shrink type L's interval is the greatest when we start from the pooling interval $[\underline{\pi}^*(L), 1]$. If this gain is less than the loss from type L, then pooling is optimal.

Returning to example 1, corollary 4.1 states that separating is strictly optimal when the accuracy of Receiver's signal is sufficiently high.

Corollary 4.1. *In example 1, there exists an increasing function $\Phi(\cdot)$ such that the optimal mechanism is separating if $\phi > \Phi(\zeta)$, and pooling if $\phi < \Phi(\zeta)$.*

Finally, we compare our approach with the concavification (cav-V) approach, following Aumann and Maschler (1995) and Kamenica and Gentzkow (2011, section VI A). Let $V : \Delta(\mathcal{S}) \to \mathbf{R}$ be Sender's payoff when the distribution over $\mathcal{S}$ induced by his signal is γ. The cav-V approach states that Sender's optimal payoff is given by $\text{cav}V(\gamma_0)$, where $\text{cav}V$ is the concave envelope of V and γ_0 is the prior over $\mathcal{S}$.

Calculating the concave envelope over the (infinite-dimensional) space of distributions is typically difficult, and most papers that derive an explicit solution to Sender's problem require V to be a function of the posterior expectation $\int s\,\gamma(\mathrm{d}s)$ alone (see Gentzkow and Kamenica (2016), Dworczak and Martini (2018), and Kolotilin et al. (2017)). Our example 1 is beyond the scope of these papers. Even though Receiver's optimal action in this example depends only on his posterior expectation of Sender's type given his own type and Sender's signal, Sender's payoff $V(\gamma)$ from a posterior distribution γ is not a function of the posterior expectation $\int s\,\gamma(\mathrm{d}s)$ alone. Indeed, assume for convenience that $\phi = 1$, so $f(s, H) = s$, and $f(s, L) = 1 - s$. Then, Sender's payoff from the distribution γ induced by his signal is given by:

$$V(\gamma) = \begin{cases} 1, & \text{if } \int (s-\zeta)(1-s)\,\gamma(\mathrm{d}s) \geqslant 0; \\ 0, & \text{if } \int (s-\zeta)s\,\gamma(\mathrm{d}s) < 0; \\ \int s\,\gamma(\mathrm{d}s), & \text{otherwise,} \end{cases}$$

where the three regions that define V correspond, respectively, to the distributions under which both types accept, the distributions under which neither type accepts, and the distributions under which only type H accepts.

5 Discussion

In this section we explain how our results extend to cases in which Sender's and Receiver's payoffs depend on both s and t, and there is no common prior. Our goal is not the extension per se, but to use the more general setup in order to compare our results to those in Kolotilin (2018) and Kolotilin et al. (2017). For simplicity, in this section we use examples in which Sender's type space is atomic. In the online appendix, we explain how the definition of the interval structure and our results extend to an atomic type space for Sender.

5.1 More general setups

We make the following modifications to our model. We use $u(s,t)$ and $v(s,t)$ to denote Receiver's and Sender's payoffs, respectively, from accepting. We allow heterogeneous beliefs. We continue to denote Receiver's belief by $f(s,t)$, but now denote Sender's belief by $g(s,t)$. We assume that $v(s,t), g(s,t) > 0$ for every s,t.

Extending the definitions of incentive compatibility and Sender's problem to this environment is straightforward. Our proof of theorem 3.1 can be modified to show that the assertion in the theorem holds under the following assumptions 2 and 3, which replace the monotonicity of Receiver's payoff function and the i.m.l.r. assumption (assumption 1), respectively.

Assumption 2. For every $t \in \mathcal{T}$, the ratio $\frac{f(s,t)u(s,t)}{g(s,t)v(s,t)}$ increases in s.

Assumption 2 connects the beliefs and payoffs of Receiver and Sender. Without it, the optimal mechanism does not recommend accepting on intervals even when $\mathcal{T}$ is a singleton (i.e., even when Receiver has no private information). Example 2 in the online appendix shows that theorem 3.1 may not hold when assumption 2 fails.

Assumption 3. There exists some $s_0 \in \mathcal{S}$ such that $u(s,t) \geqslant 0$ for $s \geqslant s_0$, and $u(s,t) \leqslant 0$ for $s \leqslant s_0$. Moreover, for every $t', t \in \mathcal{T}$ such that $t' < t$, the ratio $\frac{f(s,t)u(s,t)}{f(s,t')u(s,t')}$ weakly increases in s.

Assumption 3 captures the idea that Receiver's types are ranked. It implies the more general ranking assumption in Kolotilin (2018). For the sake of comparison with Kolotilin (2018) and later with Kolotilin et al. (2017), we repeat this assumption:

Assumption 4. For every $t', t \in \mathcal{T}$ such that $t' < t$, and every probability measure Q over $\mathcal{S}$, if $\int f(s,t')u(s,t')\ Q(\mathrm{d}s) \geqslant (>)0$, then $\int f(s,t)u(s,t)\ Q(\mathrm{d}s) \geqslant (>)0$.

Except for some minor technical detail, this is assumption 1 in Kolotilin (2018). It says that given any Sender's signal, if type t' is willing to accept, then a higher type t is also willing to accept. Assumption 4 is satisfied, for example, when the density $f(s,t)$ satisfies the i.m.l.r. assumption and Receiver has the additive payoff function given by $u(s,t) = s - t$. In Kolotilin et al. (2017), the optimal IC mechanism does not always recommend accepting on intervals. This shows that in theorem 3.1, assumption 3 cannot be replaced by assumption 4.

5.2 Private incentive compatibility

Up to now we have assumed that the information disclosed to all Receiver's types is the same. We now consider a different environment in which Receiver first reports his type and Sender can disclose different information to different types. Our definition of a disclosure mechanism is amenable to this environment: each Receiver's type can report any type $t' \in \mathcal{T}$; instead of observing the signal x, he observes the recommendation $r(x,t')$ for the reported type.

This environment restricts the set of possible deviation strategies and gives rise to a weaker notion of incentive compatibility, which we call "private incentive compatibility." Formally, a mechanism is *privately incentive-compatible* (or *privately IC)* if (1) holds for every type t where the argmax ranges over all strategies σ of the form $\sigma(x) = \bar{\sigma}(r(x,t'))$ for some type $t' \in \mathcal{T}$ and some $\bar{\sigma} : \{0,1\} \to \{0,1\}$.

Note that we use the same definition of a disclosure mechanism for the environment in which Sender discloses the same information to all types, and the environment in which Sender discloses different information to different types. We make the distinction between the two environments at the level of incentive compatibility. This approach makes straightforward the logical implication between incentive compatibility and private incentive compatibility.

It is easy to see that every mechanism that is IC is also privately IC. The following example shows that the converse is not true.

Example 2. Let $\mathcal{S} = \{-1000, 1, 10\}$ and $\mathcal{T} = \{L, H\}$. Receiver's payoff from accepting is given by $u(s,t) = s$. The density $f(s,t) = g(s,t)$ is given by:

	-1000	1	10
H	$5/22$	$5/22$	$1/22$
L	$20/82$	$20/82$	$1/82$

Consider a mechanism which recommends that H accept if s is 10 and reject otherwise, and that L accept if s is 1 and reject otherwise. This mechanism is privately IC, but is not IC because if Sender announces the recommendation to both types, then Receiver will want to accept whenever it is recommended that some type should accept. □

The mechanism in example 2 is, of course, not optimal. The following corollary shows that Sender does no better under privately IC mechanisms than under IC mechanisms.

Corollary 5.1. *Under assumptions 2 and 3, no privately IC mechanism gives a higher payoff to Sender than the optimal IC mechanism.*

It is interesting to compare this result with a result of Kolotilin et al. (2017). In their setup, Receiver's and Sender's types are independent and Receiver's payoff is given by $u(s,t) = s - t$. They show that, given the independence and the additive-payoff structure, any payoffs that are implementable by a privately IC mechanism are implementable by an IC one. This is not true in our setup: not every privately IC mechanism in our setup can be duplicated by an IC one. Indeed, under the privately IC mechanism in example 2, type H accepts with the interim probability $1/11$ and type L accepts with the interim probability $20/41$, but there exists no IC mechanism with these interim acceptance probabilities.

To understand the connection between corollary 5.1 and the equivalence result of Kolotilin et al. (2017), it is useful to start with the more general ranking assumption 4. With this assumption we have the following proposition:

Proposition 5.1. *Under assumption 4, every privately IC cutoff mechanism is IC.*

In both our setup and that of Kolotilin et al. (2017), Receiver's types are ranked in the sense of assumption 4. In the type-independence and additive-payoff environment of Kolotilin et al. (2017), every privately IC mechanism can be transformed into an equivalent cutoff mechanism. Therefore, from proposition 5.1 every privately IC mechanism can be transformed to an equivalent IC mechanism. In our setup, the main argument of our proof (lemma 6.2) shows how to create a cutoff mechanism from a privately IC mechanism. But this cutoff mechanism weakly increases the probability with which each Receiver's type accepts, so it is not necessarily equivalent to that privately IC mechanism.

Example 1 in the online appendix shows that the result that the optimal privately IC mechanism is also IC does not extend to an environment in which $f(s,t)$ satisfies the i.m.l.r. assumption and Receiver's payoff is additive. This example satisfies assumptions 2 and 4, but not assumption 3.

6 Proofs

6.1 Proof of theorem 3.1 and corollary 5.1

We assume without loss of generality that $u(s) \geqslant 0$ for $s \geqslant 0$, and $u(s) \leqslant 0$ for $s \leqslant 0$. For a mechanism $(\mathcal{X}, \kappa, r)$, if type t follows the recommendation, then *type t's acceptance probability in s* is given by

$$\rho(s,t) = \int r(x,t)\ \kappa(s, \mathrm{d}x). \tag{6}$$

Let $\chi(t,t')$ be the payoff for type t if he follows the mechanism's recommendation for t':

$$\chi(t,t') = \iint f(s,t)u(s)\rho(s,t')\,\mu(\mathrm{d}s)\,\lambda(\mathrm{d}t).$$

The mechanism is *downward incentive-compatible (or downward IC)* if and only if (i) every t prefers to follow the recommendation for his type than to follow the recommendation for any lower type; and (ii) the lowest type $\underline{t}$ prefers to follow the recommendation for his type over always rejecting. Formally:

$$\chi(t,t) \geqslant \chi(t,t'), \text{ for every } t' \leqslant t \in \mathcal{T}, \text{ and} \tag{7}$$

$$\chi(\underline{t},\underline{t}) \geqslant 0. \tag{8}$$

Since downward incentive compatibility depends on the mechanism only through the acceptance probabilities given by the function ρ, we sometimes refer to ρ as a mechanism and say that ρ is downward IC if (7) and (8) hold.

Let $\nu(t)$ be the normalized probability that type t accepts:

$$\nu(t) = \int f(s,t)\rho(s,t)\,\mu(\mathrm{d}s).$$

We say that a mechanism *weakly dominates* another mechanism if for every t the former has a weakly higher $\nu(t)$. We say that a mechanism *dominates* another mechanism if (i) the former weakly dominates the latter, and (ii) for some t the former has a strictly higher $\nu(t)$.

We say that a cutoff function $z : \mathcal{S} \to \mathcal{T} \cup \{\infty\}$ is *U-shaped* if there exists some $\hat{s}$ such that z is monotone-decreasing for $s \leqslant \hat{s}$, and monotone-increasing for $s \geqslant \hat{s}$. A deterministic cutoff mechanism has a U-shaped cutoff function if and only if it recommends accepting on intervals. We also call such a mechanism a U-shaped cutoff mechanism.

Theorem 3.1 is the immediate consequence of the following theorem:

Theorem 6.1. *The optimal downward IC mechanism is a deterministic cutoff mechanism that recommends accepting on intervals. The endpoints of the acceptance intervals are the solution to Problem* (3)*. The cutoff function* $z(s)$ *is increasing on the set* $\{s : s \geqslant 0\}$ *and decreasing on the set* $\{s : s \leqslant 0\}$*. This optimal mechanism is IC.*

The proof of theorem 6.1 will use lemmas 6.2 and 6.3. Lemma 6.2 establishes the fact that when one searches for the optimal downward IC mechanism, it is sufficient to search among U-shaped cutoff mechanisms that are downward IC. Lemma 6.3 asserts that, for any cutoff mechanism that is downward IC, if we publicly declare the cutoff type to be x, then all types which are supposed to accept (i.e., types t such that $t \geqslant x$) will still accept. (It is possible that lower types will also accept.)

We prove Theorem 6.1 in three steps. First, Problem (3) admits an optimal solution $(\underline{\pi}^*, \overline{\pi}^*)$, and this solution can also be chosen to be pointwise maximal. Second, Sender's payoff under any downward-IC mechanism is bounded by the optimal value of Problem (3). Third, there exists an IC, deterministic U-shaped cutoff mechanism whose acceptance intervals are $[\underline{\pi}^*(t), \overline{\pi}^*(t)]$.

For the case in which the type space $\mathcal{T}$ is finite, the first step of the proof is immediate since Problem (3) is finite-dimensional. For the case in which $\mathcal{T}$ is a continuous-type space, we need to be more careful in establishing the existence of an optimal solution and also in showing that it is not dominated, because of a possible problem with sets of types of measure zero. Aside from these nuisances, the core of the proof is in the second and third steps. These steps use lemmas 6.2 and 6.3.

Proof of theorem 6.1. **Step 1.** Consider the set Π of all pairs $\underline{\pi}, \overline{\pi} : \mathcal{T} \to \mathcal{S}$ such that $\underline{\pi}$ is monotone-decreasing, $\overline{\pi}$ is monotone-increasing, and $\underline{\pi}(t) \leqslant \overline{\pi}(t)$ for every t. This set is sequentially compact by Helly's selection theorem. The set of pairs which satisfy the downward-IC constraints in Problem (3) is closed and the objective function is continuous. Therefore an optimal solution exists. Pick such a solution $(\underline{\pi}^\circ, \overline{\pi}^\circ)$. Let Π° be the set of all feasible solutions $(\underline{\pi}, \overline{\pi})$ such that $\underline{\pi}(t) \leqslant \underline{\pi}^\circ(t)$ and $\overline{\pi}^\circ(t) \leqslant \overline{\pi}(t)$ for every t. Clearly, all elements of Π° are optimal solutions. Note that there exists no element $(\underline{\pi}, \overline{\pi}) \in \Pi^\circ$ such that $\underline{\pi}(t) < \underline{\pi}^\circ(t)$ on some continuity point t of $\underline{\pi}^\circ$ or $\overline{\pi}(t) > \overline{\pi}^\circ(t)$ on some continuity point t of $\overline{\pi}^\circ$, since this would imply that the inequality holds on a set of types with positive measure, and therefore would contradict the optimality of $(\underline{\pi}^\circ, \overline{\pi}^\circ)$. Since the set of discontinuities of $(\underline{\pi}^\circ, \overline{\pi}^\circ)$ is countable, we can apply Helly's selection theorem again to find an element $(\underline{\pi}^*, \overline{\pi}^*)$ in Π° that is pointwise maximal, in the sense that there exists no other feasible solution $(\underline{\pi}, \overline{\pi})$ such that $\underline{\pi}(t) \leqslant \underline{\pi}^*(t)$ and $\overline{\pi}^*(t) \leqslant \overline{\pi}(t)$ for every t, with some inequalities strict.

Step 2. For any cutoff mechanism that recommends accepting on intervals $[\underline{\pi}(t), \overline{\pi}(t)]$, Sender's payoff under this mechanism is the objective function in Problem (3), and the mechanism is downward IC if and only if $(\underline{\pi}, \overline{\pi})$ is a feasible solution of Problem (3). Therefore, it follows from Lemma 6.2 that Sender's payoff under every downward-IC mechanism is bounded by the solution to Problem (3).

Step 3. Let $\rho(s,t) = 1_{[\underline{\pi}^*(t), \overline{\pi}^*(t)]}(s)$. Then $\rho(s,t)$ is a downward-IC mechanism. By Lemma 6.2 and 6.3 it is weakly dominated by an IC, deterministic cutoff mechanism, and by Lemma 6.2 again, this IC mechanism is itself weakly dominated by a deterministic cutoff mechanism that recommends acceptance on an interval. Since $(\underline{\pi}^*, \overline{\pi}^*)$ was chosen to be pointwise maximal, all these weak dominations must be equalities, and therefore there exists a cutoff IC mechanism that recommends acceptance on $[\underline{\pi}^*(t), \overline{\pi}^*(t)]$. □

Lemma 6.2. *For every downward IC mechanism, there exists a deterministic cutoff mechanism such that it recommends accepting on intervals, is downward IC, and weakly dominates the original mechanism. The cutoff function $z(s)$ is increasing on the set $\{s : s \geqslant 0\}$ and decreasing on the set $\{s : s \leqslant 0\}$.*

The proof of lemma 6.2 has two steps. We begin with an arbitrary downward IC mechanism. In the first step we concentrate the acceptance probabilities of each type t to an interval $\left[\underline{p}(t), \overline{p}(t)\right]$ around 0, in such a way that the positive (or negative) payoff that each type gets from positive (or negative) Sender's types is the same as in the original mechanism. This step preserves the downward IC conditions and weakly increases the acceptance probability of each type.

In the second step we make the mechanism a cutoff mechanism by essentially letting each type accept at the union of his interval and all intervals of lower types. This creates the new acceptance intervals $[\underline{\pi}(t), \overline{\pi}(t)]$, which are nested in the sense that $\underline{\pi}$ is monotone-decreasing and $\overline{\pi}$ is monotone-increasing. If $\mathcal{T}$ is finite, then we can now define a cutoff function $z(s) = \min\{t \colon s \in [\underline{\pi}(t), \overline{\pi}(t)]\}$ that induces these acceptance intervals, i.e., such that

$$z(s) \leqslant t \longleftrightarrow \underline{\pi}(t) \leqslant s \leqslant \overline{\pi}(t). \tag{9}$$

The case of a continuous-type space has an additional complication, because in order to obtain a cutoff function z such that (9) holds, we need an additional semi-continuity property of $\underline{\pi}(t), \overline{\pi}(t)$.[6] This additional complication is due to our insistence that the cutoff type will accept. A more general definition of a cutoff mechanism, which allows the mechanism to

[6]For example, if $[\underline{\pi}(t), \overline{\pi}(t)] = \begin{cases} [-2, 2], & \text{if } t > 1 \\ [-1, 1] & \text{if } t \leqslant 1 \end{cases}$, then no such z exists.

recommend that the cutoff type either accept or reject, would have spared us some technical difficulties in the proof, but we chose to make the definitions simpler.

Proof of lemma 6.2. For every type t, let $\underline{p}(t)$ and $\overline{p}(t)$ be such that $\underline{p}(t) \leqslant 0 \leqslant \overline{p}(t)$, and

$$\begin{aligned} \int_0^\infty f(s,t)u(s)\rho(s,t)\,\mu(\mathrm{d}s) &= \int_0^{\overline{p}(t)} f(s,t)u(s)\,\mu(\mathrm{d}s), \text{ and} \\ \int_{-\infty}^0 f(s,t)u(s)\rho(s,t)\,\mu(\mathrm{d}s) &= \int_{\underline{p}(t)}^0 f(s,t)u(s)\,\mu(\mathrm{d}s). \end{aligned} \tag{10}$$

From (10) and the i.m.l.r. assumption, it follows that, for $t' < t$:

$$\begin{aligned} \int_0^\infty f(s,t)u(s)\rho(s,t')\,\mu(\mathrm{d}s) &\geqslant \int_0^{\overline{p}(t')} f(s,t)u(s)\,\mu(\mathrm{d}s), \text{ and} \\ \int_{-\infty}^0 f(s,t)u(s)\rho(s,t')\,\mu(\mathrm{d}s) &\geqslant \int_{\underline{p}(t')}^0 f(s,t)u(s)\,\mu(\mathrm{d}s). \end{aligned} \tag{11}$$

From (10), (11), and (7), it follows that:

$$\int f(s,t)u(s)\left(\mathbf{1}_{[\underline{p}(t),\overline{p}(t)]} - \mathbf{1}_{[\underline{p}(t'),\overline{p}(t')]}\right)\mu(\mathrm{d}s) \geqslant 0 \tag{12}$$

for every $t' < t$.

In addition, for every type t the monotonicity of $u(s)$ in s and the fact that $0 \leqslant \rho(s,t) \leqslant 1$ together imply that:

$$\int_{-\infty}^\infty f(s,t)\rho(s,t)\,\mu(\mathrm{d}s) \leqslant \int_{\underline{p}(t)}^{\overline{p}(t)} f(s,t)\,\mu(\mathrm{d}s) \tag{13}$$

by the Neyman-Pearson lemma.

Thus, the mechanism with acceptance intervals $[\underline{p}(t),\overline{p}(t)]$ is downward IC and weakly dominates the original mechanism, as expressed in (12) and (13), respectively. Note, however, that this is not yet a cutoff mechanism.

We now introduce nested acceptance intervals $[\underline{\pi}(t),\overline{\pi}(t)]$ which are downward IC and will give rise to a cutoff mechanism. Let $\overline{\pi}(t) = \inf_{\varepsilon>0}\sup\{\overline{p}(t')\colon t' < t+\varepsilon\}$ be the least right-continuous, monotone majorant of $\overline{p}$. Similarly, let $\underline{\pi}(t) = \sup_{\varepsilon>0}\inf\{\underline{p}(t')\colon t' < t+\varepsilon\}$. Let z be given by (9). Then the cutoff mechanism given by z with acceptance intervals $[\underline{\pi}(t),\overline{\pi}(t)]$ weakly dominates the mechanism with acceptance intervals $[\underline{p}(t),\overline{p}(t)]$. We claim that it is also downward IC.

Let $t' < t$. We first need to show that for the mechanism $(\underline{\pi}, \overline{\pi})$, type t prefers to accept on his interval $[\underline{\pi}(t), \overline{\pi}(t)]$ over accepting on type t''s interval $[\underline{\pi}(t'), \overline{\pi}(t')]$, i.e., that

$$\int f(s,t)u(s) \left(\mathbf{1}_{[\underline{\pi}(t),\overline{\pi}(t)]} - \mathbf{1}_{[\underline{\pi}(t'),\overline{\pi}(t')]}\right) \mu(\mathrm{d}s) \geqslant 0. \tag{14}$$

Let t_k, t'_k be such that

$$\lim_{k\to\infty} t_k = t_\infty \leqslant t \text{ and } \lim_{k\to\infty} \underline{p}(t_k) = \underline{\pi}(t), \text{ and}$$
$$\lim_{k\to\infty} t'_k = t'_\infty \leqslant t' \text{ and } \lim_{k\to\infty} \overline{p}(t'_k) = \overline{\pi}(t').$$

If $t_\infty \leqslant t'$, then $\underline{\pi}(t') = \underline{\pi}(t)$, implying that the integrand in (14) is nonnegative. Therefore, we can assume that $t' < t_\infty$ and, therefore, that $t'_k < t_k$ for every k.

From the definition of $\overline{\pi}$, the continuity assumption on f, and the fact that $\lim_{k\to\infty} t_k = t_\infty \leqslant t$, it follows that $\limsup_{k\to\infty} \overline{p}(t_k) \leqslant \overline{\pi}(t)$ and $\lim_{k\to\infty} f(s,t_k) = f(s,t_\infty)$. In addition, we know that $\lim_{k\to\infty} \overline{p}(t'_k) = \overline{\pi}(t')$. From these properties and Fatou's Lemma, it follows that:

$$\limsup_{k\to\infty} \int_{\overline{p}(t'_k)}^{\overline{p}(t_k)} f(s,t_k)u(s)\, \mu(\mathrm{d}s) \leqslant \int_{\overline{\pi}(t')}^{\overline{\pi}(t)} f(s,t_\infty)u(s)\, \mu(\mathrm{d}s).$$

A similar argument shows that

$$\limsup_{k\to\infty} \int_{\underline{p}(t_k)}^{\underline{p}(t'_k)} f(s,t_k)u(s)\, \mu(\mathrm{d}s) \leqslant \int_{\underline{\pi}(t)}^{\underline{\pi}(t')} f(s,t_\infty)u(s)\, \mu(\mathrm{d}s).$$

Therefore, it follows that:

$$\begin{aligned}
&\int f(s,t_\infty)u(s) \left(\mathbf{1}_{[\underline{\pi}(t),\overline{\pi}(t)]} - \mathbf{1}_{[\underline{\pi}(t'),\overline{\pi}(t')]}\right) \mu(\mathrm{d}s) \\
&= \int_{\underline{\pi}(t)}^{\underline{\pi}(t')} f(s,t_\infty)u(s)\, \mu(\mathrm{d}s) + \int_{\overline{\pi}(t')}^{\overline{\pi}(t)} f(s,t_\infty)u(s)\, \mu(\mathrm{d}s) \\
&\geqslant \limsup_{k\to\infty} \left(\int_{\overline{p}(t'_k)}^{\overline{p}(t_k)} f(s,t_k)u(s)\, \mu(\mathrm{d}s) + \int_{\underline{p}(t_k)}^{\underline{p}(t'_k)} f(s,t_k)u(s)\, \mu(\mathrm{d}s) \right) = \\
&\limsup_{k\to\infty} \int f(s,t_k)u(s) \left(\mathbf{1}_{[\underline{p}(t_k),\overline{p}(t_k)]} - \mathbf{1}_{[\underline{p}(t'_k),\overline{p}(t'_k)]}\right) \mu(\mathrm{d}s) \geqslant 0,
\end{aligned}$$

where the last inequality follows from (12). Given the i.m.l.r. assumption, since $t \geqslant t_\infty$, $\int f(s,t)u(s) \left(\mathbf{1}_{[\underline{\pi}(t),\overline{\pi}(t)]} - \mathbf{1}_{\underline{\pi}(t'),\overline{\pi}(t')]}\right) \mu(\mathrm{d}s)$ must be positive as well. This proves (14).

Finally, we need to show that the lowest type gets a payoff of at least zero from obeying under the mechanism $(\underline{\pi}, \overline{\pi})$. In the case of a discrete-type space, this follows from the corresponding property of the original mechanism since in this case $\overline{\pi}(\underline{t}) = \overline{p}(\underline{t})$ and $\underline{\pi}(\underline{t}) = \underline{p}(\underline{t})$. In the general case we need to appeal to an argument that is similar to the one with which we proved the downward IC conditions and that uses converging sequences of types. We omit this argument here. □

Lemma 6.3. *Assume assumption 4. For every cutoff mechanism κ that is downward IC, the IC mechanism induced by this mechanism has the property that type t accepts when $t \geqslant x$, where x is the announced cutoff type.*

Proof of lemma 6.3. Assumption 4 implies that, for every cutoff mechanism, if $t'' \leqslant t' \leqslant t$ are types such that type t' prefers following the recommendation for his type to following the recommendation for t'', then type t prefers following the recommendation for t' to following the recommendation for t'':

$$\chi(t', t') \geqslant \chi(t', t'') \rightarrow \chi(t, t') \geqslant \chi(t, t''), \text{ for every } t'' \leqslant t' \leqslant t. \tag{15}$$

Fix a type t. We need to show that in the induced IC mechanism, type t accepts on the event $\{x \in B\}$ for every Borel subset $B \subseteq [\underline{t}, t]$, where x is the public signal announced by the mechanism. That is, we need to show that

$$\int f(s,t)u(s)\kappa(s,B)\,\mu(\mathrm{d}s) \geqslant 0.$$

It is sufficient to prove the assertion for sets B of the form $B = \{x : t'' < x \leqslant t'\}$ for some $\underline{t} \leqslant t'' < t' \leqslant t$ and for the set $B = \{\underline{t}\}$, since these sets generate the Borel sets. Indeed, for $B = \{x : t'' < x \leqslant t'\}$, the following holds:

$$\int f(s,t)u(s)\kappa(s,B)\,\mu(\mathrm{d}s) = \chi(t,t') - \chi(t,t'') \geqslant 0,$$

where the inequality follows from downward incentive-compatibility for the case $t = t'$, and is extended to the case $t \geqslant t'$ by (15). The case $B = \{\underline{t}\}$ follows by a similar argument from (8). □

Theorems 3.1 and 6.1 hold under assumptions 2 and 3. The proof is similar and hence omitted.

Proof of corollary 5.1. Downward incentive-compatibility is a weaker notion than private incentive-compatibility in the sense that every mechanism that is privately IC is also downward IC. This corollary follows directly from theorem 6.1. Because the optimal downward IC mechanism is IC, it is also the optimal privately IC mechanism. □

6.2 Proof of proposition 5.1

Assumption 4 implies that, for every cutoff mechanism, if $t \leqslant t'' < t'$ are types such that type t prefers following the recommendation for t' to following the recommendation for t'', then type t'' prefers following the recommendation for t' to following the recommendation for t'':

$$\chi(t,t') > \chi(t,t'') \to \chi(t'',t') > \chi(t'',t''), \text{ for every } t \leqslant t'' < t'. \tag{16}$$

Based on lemma 6.3, we have shown that in the induced IC mechanism type t accepts when $t \geqslant x$. We need to also show that type t rejects on the event $\{x \in B\}$ for every Borel subset $B \subseteq (t, \infty]$, where x is the public signal announced by the mechanism. That is, we need to show that

$$\int f(s,t)u(s,t)\kappa(s,B)\,\mu(\mathrm{d}s) \leqslant 0.$$

It is sufficient to prove the assertion for sets B of the form $B = \{x : t'' < x \leqslant t'\}$ for some $t \leqslant t'' < t'$ since these sets generate the Borel sets. Indeed, for $B = \{x : t'' < x \leqslant t'\}$, the following holds:

$$\int f(s,t)u(s,t)\kappa(s,B)\,\mu(\mathrm{d}s) = \chi(t,t') - \chi(t,t'') \leqslant 0,$$

where the inequality follows from private incentive-compatibility for the case $t = t''$, (type t'' will not mimic type t'), and is extended to the case $t'' \geqslant t$ by (16).

6.3 Proof of proposition 4.1

We first note that, under our assumption that type L rejects without additional information, both IC constraints in (4) bind. Indeed, for type H this holds trivially in the pooling case; in the separating case, if the constraint is not binding, then slightly increasing $\overline{\pi}(L)$ would increase Sender's payoff without violating either type's IC constraint. For type L, we know from theorem 3.1 that $u(\underline{\pi}(L)) \leqslant 0$. If the IC constraint for type L is not binding, then making $\underline{\pi}(L)$ slightly smaller will increase Sender's payoff without violating either type's IC constraint.

Since both IC constraints are binding and since $\overline{\pi}(H) = 1$, the variables in Sender's problem (4) are determined by a single variable. If $\overline{\pi}(L) = y$ for some $\zeta \leqslant y \leqslant 1$, then (i)

$\underline{\pi}(L) = \ell_L(y)$ where $\ell_L : [\zeta, 1] \to [0, \zeta]$ is given by

$$\int_{\ell_L(y)}^{y} u(s) f(s, L)\ \mu(\mathrm{d}s) = 0;$$

and (ii) $\underline{\pi}(H) = \ell_H(y)$ where $\ell_H : [\zeta, 1] \to [0, \zeta]$ is given by

$$\int_{\ell_H(y)}^{\ell_L(y)} + \int_{y}^{1} u(s) f(s, H)\ \mu(\mathrm{d}s) = 0.$$

By the implicit function theorem, ℓ_L and ℓ_H are differentiable and their respective derivatives are given by:

$$\ell_L'(y) = \frac{u(y)}{u(\ell_L(y))} \frac{f(y, L)}{f(\ell_L(y), L)}, \quad \ell_H'(y) = \frac{u(y)}{-u(\ell_H(y))} \frac{f(y, H) f(\ell_L(y), L) - f(y, L) f(\ell_L(y), H)}{f(\ell_H(y), H) f(\ell_L(y), L)}.$$

It is easy to see that ℓ_L is monotone-decreasing and that the i.m.l.r. assumption implies that ℓ_H is monotone-increasing. In terms of the variable y, Sender's payoff is given by

$$R(y) = \int_{\ell_L(y)}^{y} f(s, L)\ \mu(\mathrm{d}s) + \int_{\ell_H(y)}^{1} f(s, H)\ \mu(\mathrm{d}s).$$

Thus, Sender's payoff is differentiable. After we substitute $\ell_L'(y)$ and $\ell_H'(y)$ into $R'(y)$, it follows that $R'(y)$ is positive if and only if:

$$\frac{f(y, H)}{f(y, L)} - \frac{f(\ell_L(y), H)}{f(\ell_L(y), L)} < -u(\ell_H(y)) \left(\frac{1}{u(y)} - \frac{1}{u(\ell_L(y))} \right).$$

Since $\ell_L'(y) \leqslant 0 \leqslant \ell_H'(y)$, the left-hand side increases in y due to the i.m.l.r. assumption and the right-hand side decreases in y since $u(s)$ increases in s. Therefore, $R'(y)$ is positive if and only if y is small enough. Therefore, pooling is optimal if and only if Sender's payoff achieves maximum at $y = 1$, which is equivalent to $R'(1) \geqslant 0$. That is, the inequality above holds when y equals 1.

6.4 Proof of Corollary 4.1

Proof of Corollary 4.1. Substituting $f(s, H), f(s, L)$, and $u(s)$ into condition (5) in proposition 4.1, we find that this condition holds if and only if

$$\zeta \geqslant \frac{3\phi^3 + 13\phi^2 - (\phi - 1)^2 \sqrt{9\phi^2 - 6\phi + 33} + 21\phi - 5}{8\phi(3\phi + 1)},$$

and the right-hand side is monotone-increasing in ϕ, as desired. □

References

Angeletos, George-Marios, Christian Hellwig, and Alessandro Pavan. 2006. "Signaling in a Global Game: Coordination and Policy Traps." *Journal of Political Economy*, 114(3): 452–484.

Arieli, Itai, and Yakov Babichenko. 2017. "Private Bayesian Persuasion." *Working paper.*

Aumann, Robert J., and Michael Maschler. 1995. *Repeated Games with Incomplete Information.* The MIT Press.

Au, Pak Hung. 2015. "Dynamic information disclosure." *The RAND Journal of Economics*, 46(4): 791–823.

Azarmsa, Ehsan, and Lin Cong. 2018. "Persuasion in Relationship Finance." *Working paper.*

Bergemann, Dirk, Alessandro Bonatti, and Alex Smolin. 2018. "The Design and Price of Information." *American Economic Review*, 108(1): 1–48.

Bergemann, Dirk, and Stephen Morris. 2016*a*. "Bayes Correlated Equilibrium and the Comparison of Information Structures in Games." *Theoretical Economics*, 11(2): 487–522.

Bergemann, Dirk, and Stephen Morris. 2016*b*. "Information Design, Bayesian Persuasion, and Bayes Correlated Equilibrium." *American Economic Review: Papers and Proceedings*, 106(5): 586–591.

Bergemann, Dirk, and Stephen Morris. 2018. "Information Design: A Unified Perspective." *Journal of Economic Literature, forthcoming.*

Bradford, David W., John Turner, and Jonathan W. Williams. 2018. "Off-Label Use of Pharmaceuticals: A Detection Controlled Estimation Approach." *Working paper.*

Chen, Ying. 2009. "Communication with Two-sided Asymmetric Information." *Working paper.*

Crawford, Vincent P., and Joel Sobel. 1982. "Strategic Information Transmission." *Econometrica*, 50(6): 1431–1451.

Diamond, Douglas W. 1984. "Financial Intermediation and Delegated Monitoring." *The Review of Economic Studies*, 51(3): 393–414.

Diamond, Douglas W. 1991. "Monitoring and Reputation: The Choice between Bank Loans and Directly Placed Debt." *Journal of Political Economy*, 99(4): 689–721.

Dworczak, Piotr, and Giorgio Martini. 2018. "The Simple Economics of Optimal Persuasion." *Working paper.*

Feltovich, Nick, Richmond Harbaugh, and Ted To. 2002. "Too cool for school? Signalling and countersignalling." *RAND Journal of Economics*, 33(4): 630–649.

Gentzkow, Matthew, and Emir Kamenica. 2016. "A Rothschild-Stiglitz Approach to Bayesian Persuasion." *American Economic Review*, 5(106): 597–601.

Guo, Yingni, and Eran Shmaya. 2018. "Costly Miscalibration." *Working paper.*

Inostroza, Nicolas, and Alessandro Pavan. 2018. "Persuasion in Global Games with Application to Stress Testing." *Working paper.*

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

Kolotilin, Anton. 2018. "Optimal Information Disclosure: A Linear Programming Approach." *Theoretical Economics*, 13(2): 607–635.

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

Lai, Ernest K. 2014. "Expert Advice for Amateurs." *Journal of Economic Behavior & Organization*, 103: 1–16.

Lehrer, Ehud, Dinah Rosenberg, and Eran Shmaya. 2010. "Signaling and Mediation in Games with Common Interests." *Games and Economic Behavior*, 68(2): 670–682.

Lehrer, Ehud, Dinah Rosenberg, and Eran Shmaya. 2013. "Garbling of Signals and Outcome Equivalence." *Games and Economic Behavior*, 81: 179–191.

Li, Hao, and Xianwen Shi. 2017. "Discriminatory Information Disclosure." *American Economic Review*, 107(11): 3363–85.

Mathevet, Laurent, Jacopo Perego, and Ina Taneva. 2018. "Information Design: The Belief Hierarchy Approach." *Working paper.*

Mayzlin, Dina, and Jiwoong Shin. 2011. “Uninformative Advertising as an Invitation to Search.” *Marketing Science*, 30(4): 666–685.

Minaudier, Clement. 2018. “The value of confidential policy information: persuasion, transparency, and influence.” *Working paper.*

Mussa, Michael, and Sherwin Rosen. 1978. “Monopoly and Product Quality.” *Journal of Economic Theory*, 18(2): 301–317.

Olszewski, Wojciech. 2004. “Informal Communication.” *Journal of Economic Theory*, 117(2): 180–200.

Quigley, Daniel, and Ansgar Walther. 2018. “Inside and outside information.” *Working paper.*

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

Seidmann, Daniel. 1990. “Effective Cheap Talk with Conflicting Interests.” *Journal of Economic Theory*, 50(2): 445–458.

Taneva, Ina. 2018. “Information Design.” *Working paper.*

Watson, Joel. 1996. “Information Transmission When the Informed Party Is Confused.” *Games and Economic Behavior*, 12(1): 143–161.