

# Conveying value via categories

PAULA ONUCHIC

Nuffield College, University of Oxford

DEBRAJ RAY

Department of Economics, NYU and Department of Economics, University of Warwick

A sender sells an object of unknown quality to a receiver who pays his expected value for it. Sender and receiver might hold different priors over quality. The sender commits to a monotone categorization of quality. We characterize the sender's optimal monotone categorization, the optimality of full pooling or full separation, and make precise a sense in which pooling is dominant relative to separation. As an application, we study the design of a grading scheme by an educational institution that seeks to signal student qualities and simultaneously incentivize students to learn. We show how these incentive constraints are embedded as a distortion of the school's prior over student qualities, generating a monotone categorization problem with distinct sender and receiver priors.

**KEYWORDS.** Monotonic categorization, heterogeneous priors, information design.

**JEL CLASSIFICATION.** D82, D83.

## 1. INTRODUCTION

A sender is about to come into possession of an object of unknown quality. Prior to knowing that quality, she commits to a *categorization*. That is, she partitions the set of qualities into subsets or categories—some possibly singletons—and verifiably commits to reveal the category in which the quality belongs. The categories must be monotone. For instance, she can place qualities between  $a_1$  and  $a_2$  into one category. She cannot, however, lump qualities below  $a_1$  with those above  $a_2$ , where  $a_2 > a_1$ . Monotonicity is a natural restriction in many settings; see Section 2.5.


A receiver buys the object and pays the sender his expected value conditional on the sender's category announcement. The sender seeks to maximize expected payment.

The sender and receiver use distinct distributions to evaluate the expectation of quality. That could mean that they hold different priors, and so disagree about the underlying distribution of qualities. However, there are other situations with common priors that map to a reduced form with different priors. For instance, the sender might

---

Paula Onuchic: [p.onuchic@nyu.edu](mailto:p.onuchic@nyu.edu)

Debraj Ray: [debraj.ray@nyu.edu](mailto:debraj.ray@nyu.edu)

We thank Heski Bar-Isaac, Gregorio Curello, Laura Doval, Ian Jewitt, Navin Kartik, Elliot Lipnowski, Erik Madsen, Laurent Mathivet, Luis Rayo, Ludvig Sinander, Ennio Stacchetti, and Daniel Quigley for comments. Ray acknowledges funding under NSF Grant SES-1851758. Author names are in random order, and this article should be cited as Onuchic  Ray.

be an intermediary for individuals who differ in their optimism or pessimism about the value of the object they own, and she might be more responsive to, say, optimistic owners who are also more generous with their fees. Further, distinct priors may be a stand-in for state-dependent sender payoffs over and above receiver payments or incentive constraints that effectively distort the measure that the sender employs to maximize expected value. Temporarily postponing this discussion (see Sections 2.4 and 4), note that a difference in priors, either primitive or induced, is what makes our categorization problem nontrivial.

A categorization will typically have *pooling intervals* in which all qualities are in the same category and *separating intervals* in which all qualities are revealed. We build an auxiliary function  $x \mapsto H(x)$ , where  $H(x)$  is the probability, from the sender's perspective, that the quality is below the  $x$ -quantile in the receiver's prior. Theorem 1 shows that an optimal categorization can be built by pooling quality intervals where  $H$  differs from its lower convex envelope and by separating in all intervals where these two objects coincide. Section 2.3 compares our approach with the ironing procedure of Myerson (1981) and the procedure in Rayo (2013).

Theorem 1 can be applied to study full pooling as well as local pooling on intervals. The former is optimal when the receiver's prior dominates the sender's prior under first stochastic dominance either throughout (Proposition 1) or on intervals (Proposition 2).<sup>1</sup> However, these findings are not paralleled for separation. Full separation is optimal if and only if the sender's prior dominates the receiver's in the likelihood ratio order (Proposition 3). Moreover, this dominance relationship over an interval continues to be necessary for separation on that interval, but is not sufficient (Proposition 4). This asymmetry across pooling and separation has implications for the relative prevalence of pooling. In Section 3.3, we describe a precise sense in which pooling is more widespread than separation.

Specifically, suppose that pairs of priors are drawn from some universe of allowable priors, which could be equally attached to sender and receiver. Say that a quality is *potentially pooled* if it is pooled (with other qualities) in at least one of the two problems, and is *comprehensively pooled* if it is pooled in both problems. Proposition 5 shows that every quality is potentially pooled, and that the set of problems for which a nondegenerate interval of qualities is comprehensively pooled is open and dense in the space of all prior pairs (under the uniform topology). Pooling is the rule rather than the exception.

That said, it is possible to examine how the extent of pooling and separation varies with the sender's and receiver's prior beliefs. In Section 3.4, we vary the implied optimism in the sender's prior. Proposition 6 shows that a higher prior for the sender (in the likelihood-ratio order) expands optimal separating regions. In Section 3.5, we show that a special case of "nonlinear sender payoff" problems can be rewritten as a version of our benchmark problem, with an implied distortion in the sender's prior. In that special class of problems, an increased convexity of the sender's payoff function is analogous to heightened optimism of the sender's prior in the benchmark problem.

<sup>1</sup>We say a distribution dominates another *on an interval* if the former distribution, conditional on the interval, dominates the latter distribution, conditional on that same interval.

Categorization has several applications. Financial rating agencies classify assets according to riskiness, certifying companies underwrite eco-friendly labels, bond issues are rated by agencies, the Department of Health provides restaurants with sanitary inspection grades, and schools grade students according to their academic achievements. In Section 4, we study an application where a school chooses a grading system both to signal student's underlying abilities and to incentivize students to exert effort to learn. Beyond its intrinsic interest, this application illustrates how the school's problem of incentivizing learning can be nested by our framework. Proposition 7 shows that incentive constraints can be incorporated into our benchmark model, with an appropriately induced sender prior, which is different from the receiver prior even though the primitive priors could be the same.

### 1.1 *Related literature*

We contribute to the literature on “information design,” stemming from [Rayo and Segal \(2010\)](#) and [Kamenica and Gentzkow \(2011\)](#). Within that literature, our work relates mainly to papers studying monotone information design, such as [Mensch \(2021\)](#) and [Ivanov \(2021\)](#), and information design with heterogeneous priors, such as [Alonso and Camara \(2016\)](#). Our paper combines these two starting points and provides a sharp characterization of optimal categorization in this combined environment. In Section 3.4, we compare optimal monotone categorization to non-monotone benchmarks.

Like us, [Dworczak and Martini \(2019\)](#) study optimal signaling structures in environments where the receivers' actions depend on their posterior mean. Much of the analysis in [Kolotilin \(2018\)](#) also focuses on the “posterior mean case.”<sup>2</sup> Both papers study problems where the sender's payoff is affine in the state and potentially nonlinear in the receiver's chosen action. In contrast, the sender's payoff in our environment is assumed to be affine in the receiver's action (which equals the receiver's posterior mean), but potentially nonlinear in the state. See, for example, the discussion in Section 2.3, which clarifies how the heterogeneous priors between the sender and receiver can be expressed as the sender having a payoff function that is nonlinear in the state. Moreover, in our exercise, the sender must employ a monotone categorization, which is not a requirement in [Kolotilin \(2018\)](#) and [Dworczak and Martini \(2019\)](#). [Dworczak and Martini \(2019\)](#) provide conditions for the optimality of monotone signals in their environment. [Kolotilin \(2018\)](#) also provides conditions for a special type of monotone structure, which he calls interval revelation, to be optimal. [Kolotilin and Li \(2021\)](#) provide some characterization results for monotone persuasion, again in a setup where the sender's payoff is state independent.

The solution method described in Theorem 1 is related to the ironing technique originally used in [Myerson \(1981\)](#), and recently extended by [Kleiner, Moldovanu, and Strack \(2021\)](#) to apply to optimization problems subject to majorization constraints. This connection is discussed at length in Section 2.3.

---

<sup>2</sup>Some of [Kolotilin's \(2018\)](#) results apply to a broader class of persuasion problems. However, [Kolotilin \(2018\)](#) remarks that his Assumption 2 (page 614) implies that the “sender's payoff depends only on the expected state” and, therefore, the results that rely on such assumption refer to the “posterior mean case.”

Rayo (2013) studies optimal monotonic signals in a problem where a monopolistic seller designs a menu of good qualities to be offered to buyers who care about the “status” implied by their chosen good. In part, our analysis can be viewed as bridging Rayo’s specific problem to a more abstract setting where sender and receiver have distinct priors. This allows us to interpret the sender’s incentives to pool or separate states within certain intervals as due to greater local pessimism or greater local optimism than the receiver, respectively. Some of Rayo’s (2013) results parallel ours; for example, his characterization of the optimality of full separation is equivalent to ours in Proposition 3. Kartik, Lee, and Suen (2021) find related characterization of full separation in an environment with a different restriction on the set of experiments available to the sender.

## 2. MONOTONE CATEGORIZATION

A sender is about to come into possession of an object, the quality  $a$  of which is currently unknown to her. She has the opportunity to choose a *categorization*—a partition of the space of potential qualities. She commits to naming the element of the partition to which  $a$  belongs. A receiver buys the object and obtains value equal to its quality. He believes that quality is distributed according to a continuous cdf  $R$ , strictly increasing on  $[\underline{a}, \bar{a}]$  with  $R(\underline{a}) = 0$  and  $R(\bar{a}) = 1$ . He stands ready to pay his expected value for the object, where expectations are computed using  $R$  and the category revealed by the sender.

It is assumed that the chosen categorization must be *monotone*, with each element a (possibly degenerate) interval. For instance, the sender can create two categories for qualities between  $\underline{a}$  and  $a_1$  or between  $a_1$  and  $\bar{a}$ . She cannot, however, lump together qualities below  $a_1$  or above  $a_2 > a_1$  into one category without including all qualities in between.

The sender’s objective is to maximize her expected revenue from the sale of the object, by committing to a categorization before quality is revealed to her. That expectation is calculated using a signed measure  $S$  (which is potentially not a cdf). The maximization problem is nontrivial only if  $S \neq R$ ; that is, when sender and receiver have distinct priors. If  $S = R$ , then the expected sale revenue is independent of the sender’s chosen characterization. (In that case, Bayesian plausibility implies that the expected revenue is equal to the object’s expected value according to the common prior  $S = R$ .)

An interpretation of our model is that the sender and receiver hold distinct priors about the object’s value, and we, therefore, informally think of  $S$  as a cdf. However, our analysis does not rely on such a presumption, and some of our applications and interpretations make use of the fact that  $S$  is a signed measure. We assume that  $S$  has bounded variation on the same support  $[\underline{a}, \bar{a}]$  as  $R$ , with  $S(\underline{a})$  and  $S(\bar{a})$  both finite, and that it only has upward jumps.<sup>3</sup> Measure  $S$  would indeed be a cdf if it were nondecreasing with the usual endpoint conditions.

<sup>3</sup>The assumed absence of downward jumps avoids notational complexity, but is dispensable.

2.1 Sender’s categorization problem

For any countable—possibly empty—collection  $\mathcal{P}$  of disjoint intervals, each of the form  $[p, p') \subset [\underline{a}, \bar{a}]$ , define

$$A(x) = \begin{cases} \mathbb{E}_R[y|y \in [p, p')] & \text{if } x \in [p, p') \in \mathcal{P}, \\ x & \text{otherwise.} \end{cases} \tag{1}$$

Let  $\mathcal{A}_R$  be the collection of all such functions. This set is nonempty; e.g.,  $A(x) = x$  for all  $x$  lies in it. The sender picks  $A \in \mathcal{A}_R$  to maximize her expected return under the integrating function  $S$  or, more precisely, under the signed measure that  $S$  represents.

Observe that any  $A \in \mathcal{A}_R$  has (at most) countably many disjoint intervals on which it is constant; these are the pooling intervals. Adjacent pooling intervals have distinct constant images. Qualities that are not in pooling intervals are in separating regions. By convention, all pooling intervals are closed on the left and open on the right, so  $A$  is right-continuous. Accordingly, we always separate  $\bar{a}$  (so  $A(\bar{a}) = \bar{a}$ ). These choices are without loss of generality because (a)  $R$  has no mass points and (b)  $S$  has only upward jumps by assumption, so the sender would prefer to close any discontinuity in  $A$  on the right rather than the left.

For this last reason, and to write expected value formally as an integral, we adopt the convention that  $S$  is left-continuous. The sender solves<sup>4</sup>

$$\text{maximize } \int_{\underline{a}}^{\bar{a}} A(x) dS(x), \tag{2}$$

noting that this Stieltjes integral is just the expected value under  $S$  when  $S$  is a cdf. It is well defined because the integrand  $A$  is right-continuous, and because  $S$  has bounded variation and is taken to be left-continuous. We could have adopted a right-continuous representation for  $S$ , except that the integral in (2) would have to be rewritten to accommodate the left-hand limits of  $S$ , which is notationally cumbersome.

2.2 Optimal categorization

Define a *weighting function*  $\Psi$  associated with  $A \in \mathcal{A}_R$  by

$$\Psi(x, A) = \begin{cases} S(p) + [R(x) - R(p)] \left[ \frac{S(p') - S(p)}{R(p') - R(p)} \right] & \text{if } x \text{ is in a pooling interval } [p, p'), \\ S(x) & \text{otherwise.} \end{cases} \tag{3}$$

With  $R$  strictly increasing,  $\Psi$  is well defined. If  $S$  is a cdf, so is  $\Psi$ . If  $S$  has bounded variation, so does  $\Psi$ . Next, for any left-continuous function  $H : [0, 1] \rightarrow \mathbb{R}_+$ , define its

---

<sup>4</sup>If  $S$  has all the properties of a cdf except that  $S(\bar{a}) < 1$ , this can be interpreted as  $S$  having a mass point at  $\bar{a}$ . Under that interpretation, the objective in (1) represents the sender’s expected value over the interval  $[\underline{a}, \bar{a})$ . However, note that for reasons (a) and (b) described above, it is without loss to let  $A(\bar{a}) = \bar{a}$  and to write the sender’s objective as the expected value over  $[\underline{a}, \bar{a}]$ .

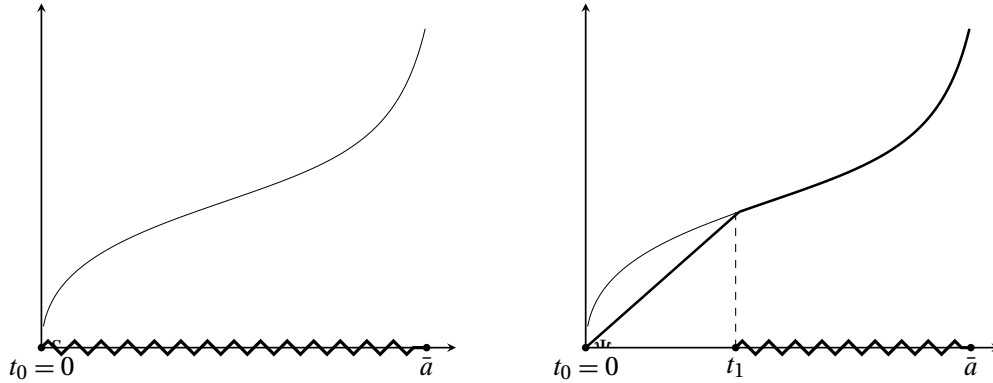


FIGURE 1. Function  $R$  (not shown) is uniform on  $[0, \bar{a}]$  and  $S$  is as pictured. Jagged lines indicate separating regions. In the first panel, there is full separation and  $\Psi = S$ . In the second panel, the sender pools on  $[t_0, t_1]$  and separates elsewhere, and  $\Psi$  is shown as a thicker line.

lower convex envelope by

$$\check{H}(x) \equiv \min\{y \mid (x, y) \in \text{Co}(\text{Graph}(H))\}.$$

This chalks out uniquely the largest convex function we can place below  $H$ . In what follows, we study the particular function  $H = S \circ R^{-1}$ .

LEMMA 1. (i) The value to the sender,  $\int_{\underline{a}}^{\bar{a}} A(x) dS(x)$ , equals  $\int_{\underline{a}}^{\bar{a}} x d\Psi(x, A)$ .  
 (ii) Furthermore, for any  $A \in \mathcal{A}_R$  and  $z \in [0, 1]$ ,  $\Psi(R^{-1}(z), A) \geq \check{H}(z)$ .

Part (i) states that the sender’s value under any categorization  $A$  is found by integrating  $x$  over  $[\underline{a}, \bar{a}]$  under the weighting function  $\Psi(\cdot, A)$ , which in separating regions “follows” the sender’s prior  $S$  and in pooling regions follows the receiver’s prior  $R$ . See Figure 1. Moreover, integration by parts reveals that

$$\begin{aligned} \int_{\underline{a}}^{\bar{a}} x d\Psi(x, A) &= [1 - \Psi(\bar{a}, A)]\bar{a} - [1 - \Psi(\underline{a}, A)]\underline{a} + \int_{\underline{a}}^{\bar{a}} (1 - \Psi(x, A)) dx \\ &= [1 - S(\bar{a})]\bar{a} - [1 - S(\underline{a})]\underline{a} + \int_{\underline{a}}^{\bar{a}} (1 - \Psi(x, A)) dx, \end{aligned} \tag{4}$$

where we use  $\Psi(\underline{a}, A) = S(\underline{a})$  and  $\Psi(\bar{a}, A) = S(\bar{a})$ . Therefore, we will have found our optimal categorization if we can find a suitable pointwise lower bound to every  $\Psi$ . That motivates part (ii), which connects the weighting function  $\Psi$  to the lower convex envelope  $\check{H}$  of  $H = S \circ R^{-1}$ . It is easy to see that  $\check{H}$  has zones where it coincides with locally convex segments of  $H$  (not necessarily all of them), and other intervals where it is a straight line connecting two points of the form  $(z, H(z))$  and  $(z', H(z'))$ . We can thus fashion a categorization  $A^*$  by pooling the intervals where  $\check{H}$  is a straight line and by separating everywhere else. The  $\Psi$  induced by  $A^*$  will coincide with  $\check{H}$ .

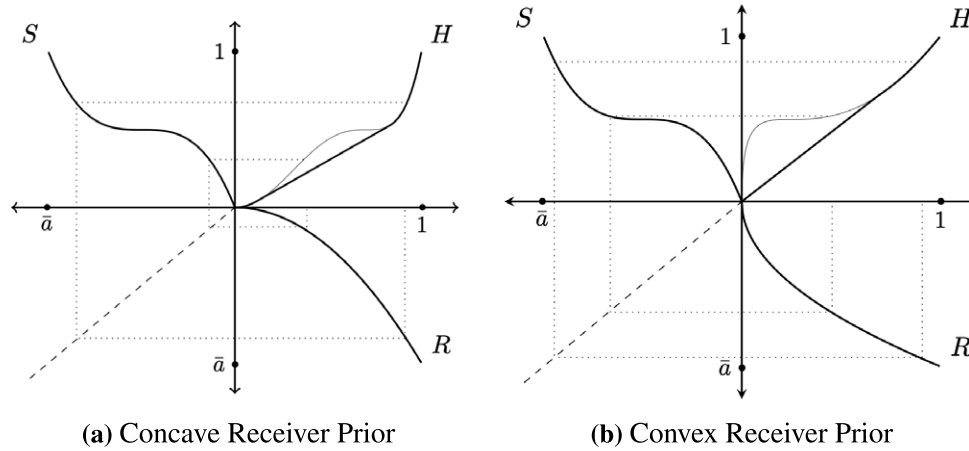


FIGURE 2. The construction of  $H = S \circ R^{-1}$  and its lower convex envelope. In panel (a),  $R$  is concave. In panel (b),  $R$  is convex. In both cases,  $S$  has a reverse-logistic shape. Function  $H$  is plotted as a thin line along with its lower convex envelope, plotted as a thick line.

LEMMA 2. *There exists  $A^* \in \mathcal{A}_R$  such that  $\Psi(R^{-1}(z), A^*) = \check{H}(z)$  for all  $z \in [0, 1]$ .*

Together, Lemmas 1(ii) and 2 imply that there is a categorization  $A^*$  such that for every categorization  $A \in \mathcal{A}_R$  and  $z \in [0, 1]$ ,  $\Psi(R^{-1}(z), A) \geq \Psi(R^{-1}(z), A^*)$ . In other words,

$$\Psi(x, A) \geq \Psi(x, A^*) \quad \text{for every } x \in [a, \bar{a}] \text{ and } A \in \mathcal{A}_R. \tag{5}$$

Consequently, (4) and (5) imply that sender value is weakly higher under  $A^*$  than under any  $A \in \mathcal{A}_R$ , which yields the following theorem.

THEOREM 1. *A solution to the sender’s problem exists. A categorization  $A^*$  is a solution if and only if  $\Psi(R^{-1}(z), A^*) = \check{H}(z)$  for all  $z \in [0, 1]$ , where  $H = S \circ R^{-1}$ .*

Figure 2 displays the optimal categorization. Function  $S$  is taken to be a cdf. It is shown horizontally flipped and has a “reverse-logistic” shape. In the illustration in panel (a),  $R$  is concave (shown vertically flipped) and in panel (b),  $R$  is convex (again vertically flipped). In each panel, the function  $H = S \circ R^{-1}$  is derived, and displayed in the north-east quadrant. Its lower convex envelope is also displayed. To see informally how this categorization behaves, note that if  $R$  is concave (panel (a)), this concavity compresses the concave segment of the derived  $H$  and elongates the convex segment. The resulting pool is, therefore, small. This is reasonable: the concavity of  $R$  implies receiver pessimism about quality, so it is better that the sender separates qualities to a greater degree (relative to uniform  $R$ ). In fact, panel (a) shows two distinct zones of separation. In panel (b),  $R$  is convex: the receiver is relatively optimistic. That accentuates the concave segment of  $H$  and induces greater pooling—another intuitive observation.

2.3 Relation between Theorem 1, the ironing procedure, and Rayo (2013)

An “ironing” procedure, as introduced in Myerson (1981), can often solve problems of the form

$$\max_{A \in \mathcal{A}} \int_a^{\bar{a}} \pi(x) A(x) dR(x), \tag{6}$$

where  $R$  is some distribution with support  $[a, \bar{a}]$ ,  $\pi$  is some potentially non-monotonic function (it is the virtual surplus function in Myerson), and  $\mathcal{A}$  is a given subset of all non-decreasing functions from  $[a, \bar{a}]$  into some given interval  $[\underline{b}, \bar{b}]$ . For expositional ease, we assume  $R$  is uniform on  $[0, 1]$ , but this is not necessary.

Suppose first that  $\mathcal{A}$  is the space of all nondecreasing functions. The procedure defines a new version of  $\pi$ , which “irons out” its non-monotonicities. Let  $\Pi(x) \equiv \int_0^x \pi(z) dz$  and  $\check{\pi}(x) \equiv d\check{\Pi}(x)/dx$ , where  $\check{\Pi}$  is the lower convex envelope of  $\Pi$ . Because  $\check{\Pi}$  is convex, the ironed version  $\check{\pi}$  of  $\pi$  is nondecreasing. Moreover,  $\check{\pi}$  coincides with  $\pi$  on intervals where both are strictly increasing, and (generically) differs from  $\pi$  on intervals where  $\check{\pi}$  is constant. Replacing  $\pi$  by  $\check{\pi}$  in (6), we obtain the auxiliary expression

$$\int_a^{\bar{a}} \check{\pi}(x) A(x) dR(x). \tag{7}$$

Next, choose  $A \in \mathcal{A}$  so as to maximize the objective in (7). One solution is the step function  $A^*$  with  $A^*(x) = \underline{b}$  when  $\check{\pi}(x) \leq 0$  and  $A^*(x) = \bar{b}$  when  $\check{\pi}(x) > 0$ . Importantly, this solution is such that  $A$  is constant wherever  $\check{\pi}$  differs from  $\pi$ . and we can, consequently, show that such  $A^*$  must also be a solution to the original maximization problem.<sup>5</sup>

Previous literature has shown that the ironing method introduced above can be applied to a broader set of problems in which  $\mathcal{A}$  is further restricted. For example, additional constraints arise in Myerson’s allocation problems with multiple potential buyers, or from the production costs as in Mussa and Rosen (1978), or to account for a variety of capacity constraints. In all such cases, the ironing solution is such that  $A$  is strictly increasing on intervals where  $\check{\pi}$  is strictly increasing and constant on intervals where  $\check{\pi}$  is constant.

In our categorization problem, assuming further that  $S$  is absolutely continuous with respect to  $R$ , (2) can be converted to (6), with  $\pi(x) \equiv \frac{dS}{dR}(x)$ . However, our sender must choose  $A$  from  $\mathcal{A}_R$ , the space of all nondecreasing functions that additionally satisfy the Bayesian constraint in (1).<sup>6</sup> Rayo (2013) studies the maximization problem in (6), subject to Bayesian constraints analogous to (1), with the additional assumption that  $\pi$  is a

<sup>5</sup>Formally, we have

$$\int_a^{\bar{a}} \pi(x) A(x) dR(x) \leq \int_a^{\bar{a}} \check{\pi}(x) A(x) dR(x) \leq \int_a^{\bar{a}} \check{\pi}(x) A^*(x) dR(x) = \int_a^{\bar{a}} \pi(x) A^*(x) dR(x),$$

where the last equality follows from the fact that  $A^*$  is constant on each interval  $I$  where  $\pi$  differs from  $\check{\pi}$  and  $\int_I \pi(x) dR(x) = \int_I \check{\pi}(x) dR(x)$  by the construction of  $\check{\pi}$ .

<sup>6</sup>Such Bayesian constraints are similar in spirit to capacity constraints, as they can be thought of as the sender allocating a limited amount of “beliefs” about the object’s quality across categories.



smooth function. The solution methods proposed both by Rayo (2013) and by us yield pooling and separating regions that coincide with the ironing solution just described. Therefore, one interpretation of our Theorem 1 and the methods used in Rayo (2013) is that the ironing procedure applies to design problems subject to Bayesian constraints as in (1).<sup>7</sup>

Now we remark on some differences. As already noted, our objective function can be rewritten in the form (6) if  $S$  is absolutely continuous with respect to  $R$ . However, the proof of Theorem 1 applies more broadly even when the sender's objective *cannot* be written in that form—so that the ironing procedure, or Rayo's (2013) method, cannot be applied directly. Specifically, our approach circumvents the step of rewriting the objective in (6)—to obtain (7)—by relying on the special structure of the Bayesian constraints in (1).

Instead, part (i) of Lemma 1 achieves a different rewriting of the sender's objective. It shows that  $\int_a^{\bar{a}} A(x) dS(x)$  can be equivalently written as  $\int_a^{\bar{a}} x d\Psi(x, A)$ , which is the expected value of the object under the weighting function  $\Psi$  (so named because  $S$ , and, therefore,  $\Psi$ , may not be a proper cdf). This weighting function is a composite of  $S$  and  $R$ : in the separating intervals of  $A$ , it equals the sender's prior  $S$ ; in the pooling intervals of  $A$ , it is an affine function of  $R$ . See the precise definition of  $\Psi$  in (3). The sender's problem is then to pick an  $A$  that yields the “best” such weighting function.

Part (ii) of Lemma 1 shows that for every  $A$ , the implied weighting function  $\Psi$  is weakly dominated by the lower convex envelope of  $H \equiv S \circ R^{-1}$  in the sense of first-order stochastic dominance. Additionally, Lemma 2 shows that there exists some  $A \in \mathcal{A}_R$  whose induced weighting function is *equal* to the lower convex envelope of  $H$ . That  $A$  must, therefore, be a solution to the sender's problem.

Thus, both our procedure and the ironing method invoke convex envelopes, but in different ways. In the latter, the convex envelope is used to define the auxiliary problem in (7), the solution to which is then the solution to the original problem. In contrast, in our method the convex envelope of  $H = S \circ R^{-1}$  is itself the *solution* to the problem; it equals the weighting function induced by the sender's optimal categorization. Moreover, the first-order stochastic dominance argument described above and the optimality of the overall procedure do not rely on  $S$  being a cdf that is absolutely continuous with respect to  $R$ .

#### 2.4 Remarks on non-common priors

The literal interpretation of our model is that sender and receiver hold distinct priors: they agree to disagree. But there are other, “common prior,” models that might lead to the sender acting as if she holds a prior distinct from the receiver's.<sup>8</sup> As a first example,

<sup>7</sup>In a recent paper, written concurrently to ours, Kleiner, Moldovanu, and Strack (2021) demonstrate that ironing can be derived as a special case of a general solution method that applies to optimization problems subject to majorization constraints, such as Bayes-plausibility restrictions.

<sup>8</sup>This literal interpretation is also taken by Alonso and Camara (2016), who extend the “concavification” argument of Kamenica and Gentzkow (2011) to a context with heterogeneous priors. Further, see Van den Steen (2004, 2009, 2010, 2011), Che and Kartik (2009), Galperti (2019), de Clippel and Zhang (2020), and (Kartik, Lee, and Suen (2017, 2021), who also consider environments with heterogeneous priors.

the sender might be interested in a robust solution such as one that generates the highest return to her under the most pessimistic receiver prior in some exogenously given class. Additionally, in Section 3.6, we describe how the profit maximization problem of a retailing intermediary, studied in [Rayo and Segal \(2010\)](#), can be interpreted as that of a sender with state-dependent preferences (or, equivalently, a distorted prior). Finally, in Section 4, we provide a model with incentive constraints arising from moral hazard and show that it maps into our setting with non-common priors.

### 2.5 *Remarks on monotonicity*

Our sender is restricted to choosing monotonic categorizations. In many situations, monotonicity is warranted, either coming from external constraints or from internal constraints within a larger problem. As an example of the former, a non-monotone categorization by a credit rating agency of the riskiness of debt issuers could invite a lawsuit from a relatively safe issuer. (See [Goldstein and Leitner \(2018\)](#) for a discussion.) As an instance of the latter, monotonicity could emerge as the outcome of a broader design problem in which incentive constraints need to be respected and natural single-crossing conditions hold over the type space of agents. Our application in Section 4 is a case in point (though our goal in that section is broader), and so is the design problem in [Rayo \(2013\)](#).

Compared to unrestricted persuasion problems, this additional imposition of monotonicity calls for a new solution method, because some concavifications of sender values in receiver posteriors, achieved through belief-splitting, are no longer available when categories are required to be monotone. In Section 3.6, we compare our results on monotonic categorization to categorization benchmarks without the monotonicity assumption, such as in [Alonso and Camara \(2016\)](#) and [Rayo and Segal \(2010\)](#).

There are problems that can also be mapped into the framework of monotone persuasion with state-dependent sender preferences. For instance, [Kolotilin and Zapechelnnyuk \(2019\)](#) establish an equivalence between balanced delegation problems and monotone persuasion problems. Our method finds the optimal signal in their linear persuasion environment when the sender's value is linear in the receiver's action. [Board \(2009\)](#) studies the optimal design of groups accounting for peer effects. His case of average quality peer effects is also closely connected to a monotone persuasion problem. In [Rayo \(2013\)](#), buyers value both the underlying quality of a status good and the average type of agents who purchase it. In that setting, the monopolist seller's problem is also a monotone categorization problem.

## 3. POOLING AND SEPARATION UNDER OPTIMAL CATEGORIZATION

Theorem 1 has implications for pooling and separation, both globally and on subintervals. Propositions 1–4 relate separating regions of the optimal categorizations to the sender's local optimism relative to the receiver and its pooling regions to the sender's relative local pessimism. The relevant notions of local optimism and pessimism are discussed below. It is of interest that they are not symmetric.

### 3.1 Full pooling and pooling intervals

Proposition 1 states that if the receiver is more optimistic than the sender in the sense of first-order stochastic dominance (or  $\succsim_1$ ), then the sender should not reveal any information that might make the receiver more pessimistic, and so full pooling on  $[\underline{a}, \bar{a}]$  is optimal.

**PROPOSITION 1.** *Full pooling on  $[\underline{a}, \bar{a}]$  is optimal if and only if  $S(x) - S(\underline{a}) \geq [1 - S(\underline{a})]R(x)$  for all  $x \in [\underline{a}, \bar{a}]$ . If  $S$  is a cdf with  $S(\underline{a}) = 0$ , then this condition is equivalent to  $R \succsim_1 S$ .*

This intuitive idea extends to local regions over which the receiver is more optimistic than the sender. Say that  $R \succsim_1 S$  on an interval  $I$  if  $R(\cdot|I) \succsim_1 S(\cdot|I)$ .

**PROPOSITION 2.** *Let  $S$  be a cdf. If  $R \succsim_1 S$  on  $[a, b]$ , then there exists an optimal categorization where  $[a, b]$  belongs to a pooling interval. Conversely,  $R \succsim_1 S$  on any pooling interval under an optimal categorization.*

If  $x < \bar{a}$  is a discontinuity point of  $S$ , then there exists some  $x' > x$  such that  $R \succ_1 S$  on  $[x, x']$ . Therefore, a consequence of Proposition 2 is that, for all mass points in  $S$ , there exists an optimal categorization such that the mass point belongs to a pooling category.

### 3.2 Full separation and separating intervals

The same ideas do not carry over in symmetric fashion to zones of separation. Unlike the case of full pooling, first-order stochastic dominance is not the relevant criterion. When  $S$  is a cdf, full separation is optimal if and only if  $S$  dominates  $R$  in the *likelihood ratio order*, denoted by  $\succsim_\ell$ .<sup>9</sup> Put another way, the converse of full pooling is not separation. It is true that if  $S \succsim_1 R$ , she would gain from splitting  $[\underline{a}, \bar{a}]$  into two or more pools, but for *full* separation to be optimal, she must want to split *every* pool.

**PROPOSITION 3.** *Full separation is optimal if and only if  $H = S \circ R^{-1}$  is convex on  $[0, 1]$ . If  $S$  is a cdf, then this condition is equivalent to  $S \succsim_\ell R$ .*

Unlike Proposition 2, this assertion admits only a partial extension to subintervals:  $S \succsim_\ell R$  over  $[a, b]$  is necessary but not sufficient for  $[a, b]$  to be nested in a separating interval.

**PROPOSITION 4.** *Let  $S$  be a cdf. If  $[a, b]$  is a subset of some separating interval of qualities in any optimal categorization, then  $S \succsim_\ell R$  on  $[a, b]$ .*

<sup>9</sup>We have  $S \succsim_\ell R$  if  $\mathbb{P}_S[x \in X] \mathbb{P}_R[x \in Y] \leq \mathbb{P}_S[x \in Y] \mathbb{P}_R[x \in X]$  for all measurable sets  $X$  and  $Y$  with  $X \leq Y$ , where  $X \leq Y$  means that  $x \in X$  and  $y \in Y$  imply  $x \leq y$ . Shaked and Shanthikumar (2007) show that this is equivalent to  $S \circ R^{-1}$  being convex (see equation (1.C.4)). In Theorem 1.C.5, Shaked and Shanthikumar (2007) also show that  $S \succsim_\ell R$  is equivalent to  $S \succsim_1 R$  over every subinterval of  $[\underline{a}, \bar{a}]$ .

### 3.3 Pooling versus separation

From our characterization, it should be obvious that zones of pooling and separation will typically alternate, and, in fact, must alternate when  $S$  and  $R$  are smooth.<sup>10</sup> That lends an air of symmetry to pooling versus separation. At the same time, the distinct conditions for pooling and separation in the earlier propositions hint at an asymmetry between the two outcomes. We will now argue that there is a well defined sense in which pooling, or the deliberate concealment of information, is more prevalent than its transparent disclosure (or separation).

Fix a pair of priors  $P = (P_1, P_2)$  that admits continuous, strictly positive densities  $(p_1, p_2)$  on  $[\underline{a}, \bar{a}]$  and let  $h(x) \equiv p_1(x)/p_2(x)$ . We say that  $h$  is *regular* if it is either weakly monotone or has an isolated turn. Specifically, a (possibly degenerate) interval  $[c, d]$ , with  $\underline{a} < c \leq d < \bar{a}$ , is an isolated turn of  $h$  if there is  $\epsilon > 0$  such that (a)  $h$  is constant on  $[c, d]$ , (b)  $h$  is strictly monotone on  $[c - \epsilon, c]$  and  $[d, d + \epsilon]$ , and (c)  $h$  is strictly increasing on one of these intervals and strictly decreasing on the other. Regularity is an extremely mild property that rules out some pathological functions.<sup>11</sup>

Consider  $\mathcal{P}$ , the space of all pairs of strictly positive and continuous densities  $p = (p_1, p_2)$  on  $[\underline{a}, \bar{a}]$ , with  $h(x) \equiv p_1(x)/p_2(x)$  regular. Endow  $\mathcal{P}$  with the product topology of uniform convergence on densities.<sup>12</sup> Each such  $p$  admits two sender–receiver problems, one with  $(s, r) = (p_1, p_2)$  and one with  $(s, r) = (p_2, p_1)$ , where  $s$  and  $r$  are the densities of the sender’s and receiver’s priors, respectively. A quality  $x \in [\underline{a}, \bar{a}]$  is *potentially pooled under  $p$*  if it is pooled (with some subset of other qualities) in some optimal categorization in at least one of these two problems, and it is *comprehensively pooled under  $p$*  if it is pooled in *any* optimal categorization in *both* the problems.

**PROPOSITION 5.** *Every  $x \in (\underline{a}, \bar{a})$  is potentially pooled under every  $p \in \mathcal{P}$ . The set of prior pairs  $p$  for which some nondegenerate interval of qualities is comprehensively pooled under  $p$  is open and dense in  $\mathcal{P}$ .*

The first part of Proposition 5 states that for every quality  $x$  and every prior pair, there is at least one assignment of that pair to sender and receiver for which an optimal solution pools  $x$ . The second part of the proposition makes it clear that the same cannot be said of separation, at least outside a sparse closed set with empty interior. Topologically, the set of prior pairs that comprehensively pools some intervals of qualities is open and dense in the space of all prior pairs.

Figure 3(a) illustrates these points. Panel (a) shows how there can be full pooling in one problem and partial pooling in the other (with priors flipped). Panel (b) shows that even if some pooling and some separation occurs in both problems, there is still more pooling than separation, in the sense that some interval of qualities must be pooled

<sup>10</sup>Rayo (2013) also finds that alternating pooling and separating intervals is optimal.

<sup>11</sup>A turn would fail to be isolated if it is the limit of a sequence of turns accumulating arbitrarily close to it. This would still not eliminate regularity as long as at least one of the other turns is isolated. However, there are pathological examples that are neither monotone nor have an isolated turn *anywhere*, such as the Weierstrass fractal function, and these are eliminated via regularity.

<sup>12</sup>We thank an anonymous referee for suggesting this line of exposition.

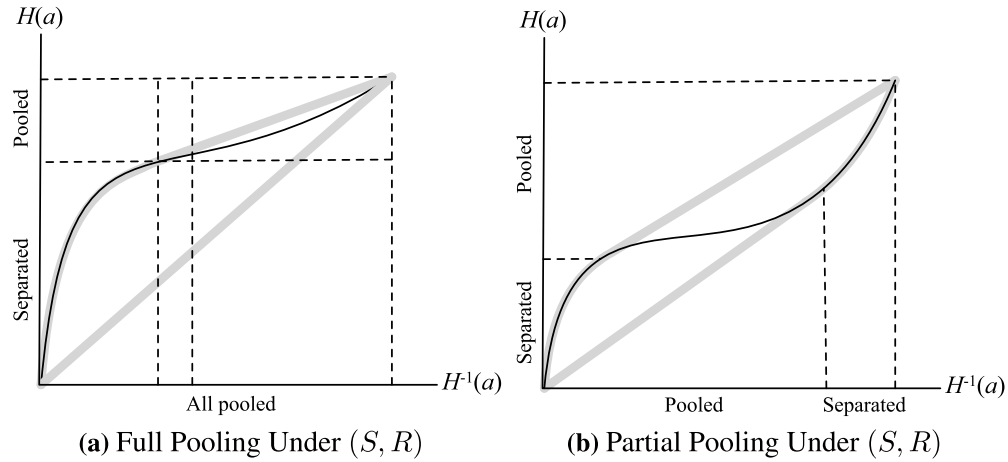


FIGURE 3.  $H$  is shown as the thin line; its envelopes as thick grey lines. In panel A, full pooling occurs under  $(S, R)$  and partial pooling under  $(R, S)$ . In panel B, partial pooling occurs under both  $(S, R)$  and  $(R, S)$ . In both panels, pooling is more widespread on average than separation.

in both problems, whereas the same is not true of separation. To read these figures, note that optimal categorization under the  $(S, R)$  problem is given by the lower convex envelope of  $H = S \circ R^{-1}$  and, therefore, by exactly the same logic, by the upper concave envelope of  $H$  under the  $(R, S)$  problem.

### 3.4 Comparative statics on priors

In this section, we apply the characterization in Theorem 1 to examine the implications of changing optimism in sender and receiver priors. Continue to assume that  $S$  is a cdf. In what follows,  $\text{Sep}(A)$  denotes the union of all separating intervals under a categorization  $A$ .

**PROPOSITION 6.** Fix a receiver prior  $R$  and consider two sender priors  $S$  and  $\hat{S}$ , with  $\hat{S} \succ_{\ell} S$ . There exists an optimal categorization  $\hat{A}$  for the sender with prior  $\hat{S}$  such that, for any optimal categorization  $A$  for the sender with prior  $S$ ,

$$\text{int}(\text{Sep}(A)) \subseteq \text{int}(\text{Sep}(\hat{A})).$$

That is, increasing the sender’s prior in the likelihood-ratio order leads to an expansion of the separating regions and, therefore, an increase in the Blackwell informativeness of the optimal categorization (from the receiver’s perspective).<sup>13</sup> To understand this, recall that  $\hat{S} \succ_{\ell} S$  implies that  $\hat{S} = \varphi \circ S$  for some increasing and convex function  $\varphi$ . A consequence is that the “separating regions” where  $\varphi \circ S \circ R^{-1}$  coincides with its lower

<sup>13</sup>Proposition 6 shows that the interior of optimal separating regions under  $S$  is contained in the interior of optimal separating regions under  $\hat{S}$ . This is enough to rank these optimal categorizations in terms of Blackwell informativeness from the receiver’s perspective, because  $R$  has no mass points.

convex envelope is larger than the separating regions where  $S \circ R^{-1}$  coincides with its convex envelope.

Curello and Sinander (2022) perform a related exercise. They study a persuasion environment in which sender and receiver have the same prior, but the sender’s payoff is a potentially nonlinear function of the receiver’s posterior mean. Their main results provide conditions under which increasing the convexity of the sender’s payoff function increases the informativeness of the optimal signal (in the Blackwell sense). Our Proposition 6 analogously shows that when sender payoff is linear but priors are asymmetric, an increase in the convexity of the sender’s prior increases the informativeness of the optimal signal.

Suppose instead that we fix the sender’s prior  $S$  and vary the optimism of the receiver from  $R$  to  $\hat{R}$ , with  $R \succ_{\ell} \hat{R}$ . Interestingly, the converse of Proposition 6 does not hold: there exist  $S, R$ , and  $\hat{R}$ , such that separating regions do not “expand” from problem  $(S, R)$  to problem  $(S, \hat{R})$ . Formally, there is some optimal categorization  $A$  under  $(S, R)$  such that for any optimal categorization  $\hat{A}$  under  $(S, \hat{R})$ ,  $\text{int}(\text{Sep}(A)) \not\subseteq \text{int}(\text{Sep}(\hat{A}))$ , and so Theorem 1 does not imply any clear ranking between the informativeness of optimal categorizations under  $R$  versus  $\hat{R}$ .<sup>14</sup>

### 3.5 Remarks on nonlinear sender payoffs

Throughout the paper, we maintain that the sender’s payoff is linear in the posterior mean induced on the receiver by the observed category. However, we note that there is a special class of “nonlinear sender payoff” problems that can be rewritten as in the linear benchmark problem. Suppose that sender and receiver share a common prior ( $S = R$ ). Denoting the state by  $x$  and the receiver’s posterior mean by  $a$ , let the sender’s payoff function be  $U(x, a) = \lambda_1(x)a + \lambda_2 a^2$  for  $\lambda_1 : [\underline{a}, \bar{a}] \rightarrow \mathbb{R}$  and  $\lambda_2 \in \mathbb{R}$ . In this case, the sender picks  $A \in \mathcal{A}_R$  to maximize

$$\int_{\underline{a}}^{\bar{a}} [\lambda_1(x)A(x) + \lambda_2 A(x)^2] dS(x).$$

Using  $S = R$  and the definition of  $A(x)$ , it is easy to see that

$$\begin{aligned} \int_{\underline{a}}^{\bar{a}} [\lambda_1(x)A(x) + \lambda_2 A(x)^2] dS(x) &= \int_{\underline{a}}^{\bar{a}} [\lambda_1(x) + \lambda_2 A(x)] A(x) dS(x) \\ &= \int_{\underline{a}}^{\bar{a}} [\lambda_1(x) + \lambda_2 A(x)] A(x) dR(x) \\ &= \int_{\underline{a}}^{\bar{a}} [\lambda_1(x) + \lambda_2 x] A(x) dR(x) = \int_{\underline{a}}^{\bar{a}} A(x) d\hat{S}(x), \end{aligned}$$

<sup>14</sup>The relation  $R \succ_{\ell} \hat{R}$  is equivalent to there being an increasing and convex function  $\varphi$  such that  $\hat{R}^{-1} = R^{-1} \circ \varphi$ . Despite the convexity added by  $\varphi$ , it is not always true that the set where  $H = S \circ R^{-1}$  equals its lower convex envelope is a subset of the set where  $\hat{H} = S \circ R^{-1} \circ \varphi$  equals its lower convex envelope.

where  $d\hat{S}(x) = [\lambda_1(x) + \lambda_2x] dR(x)$ . Note that in this special case, the nonlinearity of  $U$  can also be viewed as a change of the sender's prior. We can use this rewriting to perform analogous exercise to [Curello and Sinander's \(2022, Theorems 1 and 2\)](#): an increase in  $\lambda_2$ , which increases the convexity of  $U$ , maps into increased convexity of  $\hat{S}$ . With an argument parallel to that in [Proposition 6](#), we can show that this heightened convexity thus implies an increase in the informativeness of optimal categorizations.<sup>15</sup>

### 3.6 Non-monotonic categorization

[Alonso and Camara \(2016\)](#) study a persuasion problem with heterogeneous priors, without restricting the sender to monotone signals. [Rayo and Segal's \(2010\)](#) seminal paper can also be regarded as a problem of persuasion with heterogeneous priors, but no monotonicity constraint.<sup>16</sup>

Specifically, [Rayo and Segal \(2010\)](#) consider an information intermediary who is paid a fee whenever the sale of a prospect takes place. The intermediary receives different unit fees for different prospects. Let  $x$  be the value of a prospect to a potential buyer and let  $\pi(x)$  be the unit fee paid to the intermediary if a sale takes place. For most of the paper, [Rayo and Segal \(2010\)](#) assume that the probability that a sale takes place is equal to the buyer's expectation of the object's value (given any information they observe). Further, suppose no two prospects have the same value, so each  $x$  is associated with a single fee  $\pi(x)$ . Let  $R$  be the prior about the prospect's value, commonly held by the intermediary and the potential buyer.<sup>17</sup> Then the intermediary's profit when she picks a categorization  $A \in \mathcal{A}_R$  is

$$\int_{\underline{a}}^{\bar{a}} \pi(x) A(x) dR(x) = \int_{\underline{a}}^{\bar{a}} A(x) dS(x),$$

where we set  $dS(x) = \pi(x) dR(x)$ . In that case,  $d[S \circ R^{-1}](x) = dS(x)/dR(x) = \pi(x)$ , so that regions where  $\pi$  is increasing are equivalent to regions where  $S \circ R^{-1}$  is convex, and regions where  $\pi$  is decreasing are equivalent to regions where  $S \circ R^{-1}$  is concave.

[Rayo and Segal \(2010\)](#) show that two prospects  $(x, \pi)$  and  $(x', \pi')$  can only be optimally pooled if they are *not ordered*:  $(x - x')(\pi - \pi') < 0$ . Our results show that this is not true of optimal *monotonic* categorizations. Indeed, an optimal pooling interval may contain a region where  $\pi$  is increasing ( $S \circ R^{-1}$  is convex) and, therefore, contains

<sup>15</sup>[Proposition 6](#) cannot be applied directly because it uses a different notion of increased convexity than the one implied by an increase in  $\lambda_2$ : increasing  $\lambda_2$  is equivalent to adding a convex function to the original prior of the sender. Despite [Proposition 6](#) not applying directly, we can show that an analogous result holds under this alternative notion of increased convexity.

<sup>16</sup>Our [Propositions 1–4](#) are more immediately comparable to results in [Rayo and Segal's \(2010\)](#). However, if the sender's payoff is assumed to be linear in the receiver's action, then [Alonso and Camara's \(2016\)](#) model is equivalent to [Rayo and Segal's \(2010\)](#). In that case, our discussion in this section is also a relevant comparison to [Alonso and Camara's \(2016\)](#) benchmark.

<sup>17</sup>[Rayo and Segal \(2010\)](#) assume that there are only finitely many prospects, so that  $R$  has finite support. This is also assumed in [Alonso and Camara \(2016\)](#). In our exposition, we take  $R$  to be a continuous distribution.

ordered prospects.<sup>18</sup> Further, they show that all prospects that are optimally pooled together must have payoffs  $(x, \pi)$  that lie on a straight line with nonpositive slope. Consequently, if  $\pi(x)$  does not have an interval where it is linear and downward sloping, then every optimal signal realization pools together at most two qualities.<sup>19</sup> Optimal monotone categories, on the other hand, may pool together intervals containing more than two qualities.

Finally, without the restriction to monotonic signals, full separation is optimal when all prospects are ordered:  $(x - x')(\pi(x) - \pi(x')) > 0$  for every  $x$  and  $x'$ . This condition, equivalent to  $S \circ R^{-1}$  being globally convex, is also the condition for full separation when monotonicity is imposed (Proposition 3). Conversely, Rayo and Segal's (2010) condition for full pooling is that all prospects lie on a straight line with negative slope. That is,  $\pi(x)$  must be *linear and decreasing*. In contrast, the condition for full pooling under the monotonicity restriction is much weaker; see Proposition 1.

The discussion above contrasts properties of optimal categorizations with and without the restriction to monotonicity. More recently, Jewitt and Quigley (2022) study a class of persuasion games in which the sender has rank-dependent preferences Yaari (1987). Their problem can also be interpreted as a persuasion problem with heterogeneous priors across sender and receiver (without the restriction to monotonic signals). They show that in the class of rank-dependent sender preferences, the optimal signal can always be taken to be monotone (so that the monotonicity restriction does not bind).

#### 4. MORAL HAZARD AND EDUCATIONAL GRADES

In this section, we study an application where a school chooses a grading system both to signal student's underlying abilities and to incentivize students to exert effort to learn. Beyond its intrinsic interest, this application illustrates how incentive constraints arising from moral hazard can generate all the key features of our benchmark model, namely a distortion in the sender's state-dependent preferences and the restriction to monotone signals.<sup>20</sup>

*The Setting.* A school—the sender—designs a grading system to maximize tuition revenues. This is part of a larger problem in which it might choose an admissions cutoff.

<sup>18</sup>Revisit Figure 3(a) and further suppose that  $R$  is uniform over  $[\underline{a}, \bar{a}] = [0, 1]$ . The optimal monotonic categorization is full pooling, even though  $S \circ R^{-1}$  is convex over some interval or, equivalently,  $\pi(\cdot)$  is increasing over that same interval. In their Lemma 3, Rayo and Segal (2010) show that a non-monotonic signal structure can improve over full pooling in that case. In our example, there exists some small interval  $[a, b] \subset [0, 1/2)$ , depicted by the two vertical dotted lines, such that the average slope of  $H$  over that interval is less than 1, which is the average slope over the full interval  $[0, 1]$ . Equivalently, this means that the average fee  $\pi$  over  $[a, b]$  is smaller than the average fee over  $[0, 1]$ . Additionally, because the common prior is uniform and  $a < b \leq 1/2$ , every value in the interval is below the ex ante average value. Therefore, a signal  $\sigma_1$  that pools  $[0, a) \cup [b, 1)$  has larger average value *and* larger average fee than a signal  $\sigma_2$  that pools  $[a, b)$ . Consequently, if the sender is not constrained by monotonicity, the signal structure that consists of the two pools  $\sigma_1$  and  $\sigma_2$  is an improvement over full pooling. (For more details, see Rayo and Segal (2010), Lemmas 1 and 3.)

<sup>19</sup>Kolotilin, Corrao, and Wolitzky (2022) make the similar point that optimal signals are “pairwise”, in the sense that each induced posterior distribution has at most binary support.

<sup>20</sup>In this application, the link between ability, effort, and output is deterministic. It would be of interest to extend these arguments to stochastic output, after conditioning on ability and effort.



The choice of that cutoff can be incorporated with no change to the discussion that follows. We emphasize the “second stage” where that cutoff—and the student body—has already been chosen. Normalize admitted students to a unit measure, with abilities  $a$  distributed according to  $R$  on  $[\underline{a}, \bar{a}]$ . Upon entering, students fully learn  $a$ , but before that they have only some prior, represented as a distribution over  $[\underline{a}, \bar{a}]$ . For instance, they may be close to a degenerate distribution (full self-knowledge), or have the population belief  $R$ , or have some third belief. Among these, we assume that there is some continuous prior  $R_0$ , with  $R_0(\underline{a}) = 0$ , which is the lowest according to first-order stochastic dominance among all possible post-entry priors.

A market (our receiver) with shared prior  $R$  infers student abilities from a learning level or grade  $\ell$  that the school chooses to make observable. Let  $\lambda > 0$  be the value of learning, relative to inherent ability, so that the market pays

$$\mathbb{E}(a|\ell) + \lambda\ell \tag{8}$$

to a student with learning  $\ell$ , where the conditional expectation is determined by the prior  $R$  as well as equilibrium strategies. Notice that  $\ell$  has both intrinsic and signaling value.

*Incentive-Compatible Learning.* The school chooses a compact set of certified learning levels  $L$ . Learning nothing is always an option, so  $0 \in L$ . A student of ability  $a$  chooses  $\ell \in L$  by exerting effort at cost  $c(a)\ell$ , where  $c'(a) < 0$ . We assume the following condition:

[C] We have  $c(\bar{a}) > \lambda$ , so that rewards to learning alone do not motivate any student.

Every nonzero  $\ell \in L$  will presumably be occupied by some ability type. A student could choose  $\ell \notin L$ , but the market observes only  $\ell' = \max\{\ell'' \in L | \ell'' \leq \ell\}$ , so there is no point in doing so. In the spirit of direct mechanisms, suppose that the school assigns learning  $\ell(a) \in L$  to each ability type  $a$ .<sup>21</sup> The value to an obeying student of ability  $a$  is

$$\mathbb{E}_R(a'|\ell(a)) + \lambda\ell(a) - c(a)\ell(a),$$

where  $\mathbb{E}_R(a'|\ell(a))$  is the expectation of ability  $a'$  given that  $\ell(a)$  is observed and given that all students follow  $\ell$ . A standard single-crossing argument yields the following lemma.

LEMMA 3. *If  $\ell(a)$  and  $\ell(a')$  are optimal for  $a$  and  $a'$ , with  $a > a'$ , then  $\ell(a) \geq \ell(a')$ .*

So incentive-compatibility restricts the school to a monotone categorization of abilities. Learning  $\ell$  could have separating intervals on which it is strictly increasing and pooling intervals on which it is constant. These obviously correspond to a particular categorization  $A_\ell$ . Incentive-compatibility additionally implies that for every  $a, a' \in [\underline{a}, \bar{a}]$ ,

$$A_\ell(a) + \lambda\ell(a) - c(a)\ell(a) \geq \max\{A_\ell(a') + \lambda\ell(a') - c(a)\ell(a'), \underline{a}\}. \tag{9}$$

<sup>21</sup>The school maps each ability to a deterministic certified learning level. It can be shown that restricting to deterministic grading is without loss in our environment. We thank Ian Jewitt for pointing this out.

The second constraint on the right-hand side of (9) is needed in case  $\ell(\underline{a}) > 0$ . Then the zero-learning choice is off-path, and suitable beliefs will be needed to guarantee incentive-compatibility. We presume that in such cases, the observation of 0 is associated with the belief that the student has the lowest ability  $\underline{a}$ . This implements the best equilibrium from the perspective of a tuition-maximizing school. The following lemma tightly links incentive-compatible learning functions  $\ell$  to their corresponding categorizations  $A_\ell$ .

LEMMA 4. (i) *The following statements are equivalent.*

(a) *A learning function  $\ell(a)$  is incentive-compatible.*

(b) *Learning  $\ell$  is nondecreasing with  $\ell(\underline{a}) \in [0, \frac{A_\ell(\underline{a})-\underline{a}}{c(\underline{a})-\lambda}]$  (where  $A_\ell$  is the corresponding categorization), differentiable on any separating interval with*

$$\ell'(a) = \frac{1}{c(a) - \lambda}, \tag{10}$$

and at any threshold  $t$  dividing two adjacent intervals, using  $\uparrow$  for left-hand limit,

$$\ell(t) = \ell^\uparrow(t) + \frac{A_\ell(t) - A_\ell^\uparrow(t)}{c(t) - \lambda}. \tag{11}$$

(c) *For every  $A \in \mathcal{A}_R$  and  $\underline{\ell} \in [0, \frac{A(\underline{a})-\underline{a}}{c(\underline{a})-\lambda}]$ , there is a unique function  $\ell$  with  $\ell(\underline{a}) = \underline{\ell}$ , and satisfying (10) and (11). That describes all incentive-compatible  $\ell$  such that  $A_\ell = A$ .*

*Tuition and School Payoffs.* The school sets a single tuition level. Type  $R_0$  students have the lowest willingness to pay, so the school must maximize their expected payoff before fees.<sup>22</sup> That is, the school chooses incentive-compatible  $\ell$  to maximize

$$\int_{\underline{a}}^{\bar{a}} [A_\ell(a) + \{\lambda - \sigma c(a)\} \ell(a)] dR_0(a), \tag{12}$$

where  $\sigma \in [0, 1]$  is the extent to which parents internalize effort costs at the ex ante stage. The extent of this internalization will determine not just the level of the tuition (which is a relatively minor consideration, at least for the analysis), but also the school’s “attitude” toward the intrinsic value of learning; more on this immediately below. We now link the school problem to our more abstract setting, thereby permitting a full solution of it.

*Solution to the School Problem.* Consider two cases. If  $\lambda > \sigma \int_{\underline{a}}^{\bar{a}} c(a) dR_0(a)$ , then  $R_0$  parents value, on average, intrinsic learning relative to cost. It is obvious that no matter what categorization  $A$  the school seeks to implement, its associated learning function must have the highest possible starting point. That is, recalling Lemma 4, initial learning  $\underline{\ell} = \ell(\underline{a})$  must be set equal to the upper bound  $\frac{A(\underline{a})-\underline{a}}{c(\underline{a})-\lambda}$  for any choice of  $A \in \mathcal{A}_R$ .

Otherwise,  $\lambda \leq \sigma \int_{\underline{a}}^{\bar{a}} c(a) dR_0(a)$ . Now learning is not intrinsically valued by  $R_0$  parents, so the school optimally sets  $\ell(\underline{a}) = 0$ . That motivates the definition, for any

<sup>22</sup>Because the student body is fixed, the school maximizes its profit by serving the lowest belief students.

$A \in \mathcal{A}_R$ ,

$$\ell^*(A) = \begin{cases} \frac{A(\underline{a}) - \underline{a}}{c(\underline{a}) - \lambda} & \text{if } \lambda > \sigma \int_{\underline{a}}^{\bar{a}} c(a) dR_0(a), \\ 0 & \text{if } \lambda \leq \sigma \int_{\underline{a}}^{\bar{a}} c(a) dR_0(a). \end{cases} \tag{13}$$

PROPOSITION 7. For every  $A \in \mathcal{A}_R$  and  $\underline{\ell} = \ell^*(A)$  as defined in (13), pick unique  $\ell$  as described in Lemma 4. Then school payoff is given by

$$\int_{\underline{a}}^{\bar{a}} [A(a) + \{\lambda - \sigma c(a)\} \ell(a)] dR_0(a) = \int_{\underline{a}}^{\bar{a}} A(a) dS(a) + K, \tag{14}$$

where  $K$  is a constant and

$$S(a) = R_0(a) + \int_a^{\bar{a}} \frac{\sigma c(x) - \lambda}{c(a) - \lambda} dR_0(x) \quad \text{for all } a \in (\underline{a}, \bar{a}], \text{ with} \tag{15}$$

$$S(\underline{a}) = \min \left\{ 0, \int_{\underline{a}}^{\bar{a}} \frac{\sigma c(x) - \lambda}{c(a) - \lambda} dR_0(x) \right\}.$$

The function  $S$  is left-continuous and has bounded variation with at most one discontinuity. Also,  $S(\underline{a})$  and  $S(\bar{a})$  are finite. These conditions, along with the fact that  $A$  is right-continuous with at most countably many discontinuities, guarantee that all the assumptions of the baseline model are satisfied.

The school problem is solved by finding a solution  $A^*$  to the optimal categorization problem with  $R$  as the receiver's distribution and  $S$ , defined in (15), as the sender's distribution. The optimal learning function is the unique  $\ell$  associated with  $A^*$  with  $\ell(\underline{a}) = \ell^*(A^*)$ .

Proposition 7 fully removes  $\ell$ , as well as its attendant moral-hazard implications, from the analysis and converts this model into our simpler categorization model. In so doing, it reveals three reasons for the induced prior  $S$  to be different from  $R$ . First, the sender may, in effect, be delegated to work on behalf of someone with a distinct prior. Here, this is the student or parent with lowest belief  $R_0$ .<sup>23</sup> The school is pushed to cater to their interests in order to maximize tuition revenue. The second stems from ancillary constraints that might be involved in revealing quality; here, these have to do with learning. Third, the actions taken to signal quality (school performance) may have direct payoff effects. All three enter (15). These considerations can additionally cause  $S$  to depart from a cdf.

At a somewhat more technical level, even though  $R_0$  is continuous, the induced  $S$  may be discontinuous. Remember that parental priors are continuous, with  $R_0(\underline{a}) = 0$ . In the first of the two cases where learning is intrinsically valued,  $S(\underline{a})$  as defined in (15)

<sup>23</sup>In an affirmative action setting,  $R_0$  can also represent the subdistribution of “diversity students” in whose outcomes the school may be particularly interested. In that case, the school may inherently care about  $R_0$ , rather than through the tuition channel we consider.

is negative—it is not a cdf any more—but the entire induced function  $S$  is easily seen to be continuous. If, on the other hand, learning is not intrinsically valued,  $S(\underline{a}) > 0$ , thereby effectively generating a mass point under  $S$  at  $\underline{a}$ . Our methods apply in either event, but the latter case must generate an initial pool of zero learning.

*Pooling.* An intriguing question that deserves a more detailed exploration is whether grade pooling is a pervasive property of an educational system. Proposition 8 below, a straightforward implication of Theorem 1 and Proposition 1, formalizes the following claims: a school will want to pool all ability types when (i) the market places a low relative value on learning, as in Lizzeri (1999), where learning has no value, (ii) if students fully internalize their cost of learning ex ante, and (iii) if the lowest belief students are certain ex ante that they are the lowest ability type.

PROPOSITION 8. *A sufficient condition for full pooling to be optimal is*

$$\int_a^{\bar{a}} [\sigma c(x) - \lambda] dR_0(x) \geq 0 \quad \text{for every } a \in [\underline{a}, \bar{a}]. \quad (16)$$

Furthermore, the following statements are true:

(i) Equation (16) is satisfied when  $\lambda = 0$ . Moreover, if there is  $\lambda > 0$  such that it is satisfied, then it is also satisfied for any  $\lambda' < \lambda$ .

(ii) Equation (16) is satisfied when  $\sigma = 1$ . Moreover, if there is  $\sigma > 0$  such that it is satisfied, then it is also satisfied for any  $\sigma' > \sigma$ .

(iii) Equation (16) is satisfied when  $R_0$  is degenerate at  $a = \underline{a}$ . Moreover, if there is  $R_0$  such that it is satisfied, then it is also satisfied for any  $R'_0$  such that  $R_0$  first-order stochastically improves over  $R'_0$ .

This result also highlights more broadly the school's incentives to create grading pools. By pooling, the school boosts the grade of lower-type students, who are then not distinguishable from higher-type students in the same pool. In contrast, separation is necessary to induce learning, because the value of separating oneself from lower types (or pooling oneself with higher types) is a student's only incentive to incur in learning costs. These considerations are distinct from those in previous literature, such as Ostrovsky and Schwarz (2010), where a school's incentives to create grading pools depend on the distribution of job types to which students will be matched.<sup>24</sup>

*Schooling with Lower Censorship.* We now make functional form assumptions that enable a complete characterization of the optimal grading policy. Let  $R$  be uniform on  $[0, 1]$  and let  $R_0 = a^\gamma$  for  $\gamma \in [0, 1]$ . If  $\gamma = 1$ , then  $R_0 = R$ : students have no ex ante information about their own abilities. For lower  $\gamma$ ,  $R_0$  is first-order stochastically dominated

<sup>24</sup>Boleslavsky and Kim (2020) also study an environment in which signaling motivates costly effort. In our setting, agents exert effort after drawing their ability and the incentive constraint implies that signaling structures must be monotone, whereas in Boleslavsky and Kim (2020), costly effort improves an agent's distribution of types. Other close parallels are Rodina and Farragout (2016) and Saeedi and Shourideh (2020), who consider a principal who wants to improve an agent's investment in productivity when the only instrument at hand is an information disclosure policy. Their environment is different from ours both in how the agent's effort decision is set up and in the grading schemes that are permitted.

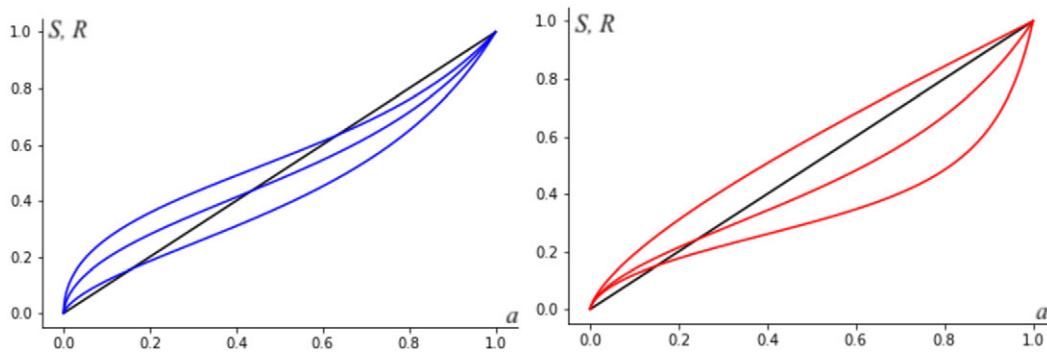


FIGURE 4. Schooling with lower censorship. In both panels,  $R$  is the uniform distribution – plotted as the 45° line. On the left,  $S$  is shown for three values of  $\gamma$ :  $S$  increases in  $\tilde{z}_1$  as  $\gamma$  increases. On the right,  $S$  is shown for three values of  $\lambda$ : again,  $S$  increases in  $\tilde{z}_1$  as  $\lambda$  increases.

by  $R$ , so that the lowest types are pessimistic relative to the population average. The lower is  $\gamma$ , the further is the belief of the most pessimistic type from that of the average agent. Set  $c(a) = 1/a$  and  $\lambda < 1$ . Finally, set  $\sigma = 0$ , which means that the cost of effort is not internalized at all by the parents when choosing to join the school.

With these functional forms, use (15) to map the school's distorted prior  $S$ :

$$S(a) = \frac{a^\gamma - \lambda a}{1 - \lambda a}. \quad (17)$$

In this special case,  $S$  is a cdf (see Figure 4). Since  $S$  has a concave–convex shape, Theorem 1 immediately implies that lower censorship, whereby all abilities below a threshold are pooled together and all abilities thereafter are fully revealed, is the optimal categorization.

**COROLLARY 1** (to Theorem 1 and Proposition 7). *There is  $\tilde{a} \in (0, 1]$  such that the solution to the school problem is to pool students with  $a \in [0, \tilde{a})$  and fully reveal the ability of all students with  $a \geq \tilde{a}$ .*

We can easily compute  $\tilde{a}$ , and perform comparative statics with respect to  $\lambda$  and  $\gamma$ . When the lowest belief type is more optimistic, i.e.,  $\gamma$  is higher,  $\tilde{a}$  is lower and there is more separation. The connection with the market value for learning is not monotonic. Initially, higher value for learning induces more separation, but for high values of  $\lambda$ , increasing  $\lambda$  leads to more pooling.

If, otherwise, the distorted prior were convex–concave, upper censorship would be optimal. Other papers in the literature find that upper or lower censorship provides optimal signals for the sender. Kolotilin, Mylovannov, and Zapechelnyuk (2019) show conditions for optimality of lower and upper censorship, respectively, that the sender's payoff as a function of the receiver's posterior mean be concave–convex and convex–concave. In both those papers, they consider environments where the sender's payoff does not directly depend on the state and the sender's payoff is nonlinear in the receiver's posterior

mean. In our model, the sender's payoff is state dependent, but linear in the receiver's posterior mean. Perhaps surprisingly, these environments are distinct and cannot be mapped to each other.

## 5. CONCLUSION

In this paper, we study a monotone categorization problem. A sender offers an object of unknown quality to a receiver, who pays his expected value for it. That expected value is taken conditional on the receiver's information, which is affected in turn by the sender's choice of a monotone quality categorization. That is, the sender commits to revealing the object's quality up to an information partition, where each element of the partition is an (possibly degenerate) interval. This exercise is nontrivial when sender and receiver hold different priors over quality. We characterize the sender's optimal monotone categorization, obtain several corollaries, such as the characterization of optimality of complete pooling or separation, and make precise a sense in which pooling is dominant relative to separation.

The assumption that sender and receiver hold distinct priors may be literally interpreted, but we also emphasize situations in which distinct priors emerge as reduced forms of a more primitive setting with additional incentive constraints. As an example, we study the design of a grading scheme by an educational institution that seeks to signal student qualities and simultaneously incentivize students to learn. We show how these incentive constraints are embedded as a distortion of the school's prior over student qualities, generating a monotone categorization problem with distinct sender and receiver priors, even if the two agents have the same priors in the original problem.

The categorization problem has several applications. Financial rating agencies classify assets according to riskiness, certifying companies underwrite eco-friendly labels, bond issues are rated by agencies, the Department of Health provides restaurants with sanitary inspection grades, and schools grade students according to their academic achievements. In all of these settings, it is natural to presume that sender and receiver could have differing opinions on the underlying distribution of the relevant state. Alternatively, as in our example, one or more agents could face incentive constraints, resulting in effectively distinct priors. Our framework is broad enough to incorporate such situations.

A limitation of the analysis is that our methods apply without qualification only when the sender's payoff can be written as an affine function of the receiver's posterior expectation. While some limited progress can be made in special nonlinear settings, a general analysis of the nonlinear case is currently beyond the scope of the methods developed in this paper. We put on record here our opinion that progress on this front would represent a significant step forward in our understanding of monotone categorization problems.

6. PROOFS

PROOF OF LEMMA 1. Part (i). Let Pool denote the collection of pooling intervals with generic element  $[p, p']$ . Because  $A \in \mathcal{A}_R$  and  $R$  is strictly increasing, we have

$$\begin{aligned} \int_{\underline{a}}^{\bar{a}} A(a) dS(a) &= \int_{[\underline{a}, \bar{a}] \setminus \text{Pool}} adS(a) + \sum_{\text{Pool}} \mathbb{E}_R(a|a \in [p, p']) [S(p') - S(p)] \\ &= \int_{[\underline{a}, \bar{a}] \setminus \text{Pool}} adS(a) + \sum_{\text{Pool}} \int_p^{p'} adR(a) \left[ \frac{S(p') - S(p)}{R(p') - R(p)} \right] \\ &= \int_{[\underline{a}, \bar{a}] \setminus \text{Pool}} a d\Psi(a, A) + \sum_{\text{Pool}} \int_p^{p'} a d\Psi(a, A) = \int_{\underline{a}}^{\bar{a}} a d\Psi(a, A), \end{aligned}$$

where, in the penultimate step,  $d\Psi$  is well defined as  $\Psi$  has bounded variation, and the last equality follows from the continuity of the integrand.

For any categorization  $A$  and  $z \in [0, 1]$ , consider the quantile weighting function

$$\Phi(z, A) \equiv \Psi(R^{-1}(z), A).$$

Define a *quantile pooling interval* of  $A$  as any interval  $[w, w']$  such that  $[R^{-1}(w), R^{-1}(w')]$  is a pooling interval of  $A$ . Then

$$\Phi(z, A) = \begin{cases} H(w) + (z - w) \left[ \frac{H(w') - H(w)}{w' - w} \right] & \text{if } z \text{ is in some quantile pooling } [w, w'], \\ H(z) & \text{otherwise,} \end{cases} \tag{18}$$

where we recall that  $H(x) = S(R^{-1}(x))$ . In particular, the quantile weighting function associated with  $A \in \mathcal{A}_R$  equals  $H$  in quantile separating regions, and is a straight line connecting  $(w, H(w))$  and  $(w', H(w'))$  in quantile pooling regions of the form  $[w, w']$ . This means that  $\text{Graph}(\Phi(\cdot, A)) \subset \text{Co}(\text{Graph}(H))$ , which immediately implies  $\Phi(z, A) \geq \check{H}(z)$ .  $\square$

PROOF OF PROPOSITION 1. When  $A$  is the categorization that pools every quality, then  $\Phi(z, A) = H(0) + z(H(1) - H(0))$ . By Theorem 1, full pooling is then a solution to the sender’s problem if and only if  $\check{H}(z) = H(0) + z(H(1) - H(0))$  for all  $z \in (0, 1]$ . Now notice that this condition is equivalent to

$$\frac{H(z) - H(0)}{z} \geq H(1) - H(0)$$

for all  $z \in (0, 1]$ . Using the definition of  $H$ , this condition can be rewritten as  $S(x) - S(\underline{a}) \geq (1 - S(\underline{a}))R(x)$  for all  $x \in [\underline{a}, \bar{a}]$ .  $\square$

PROOF OF PROPOSITION 2. If  $R \succsim_1 S$  on the interval  $[a, b]$ , then for all  $x \in [a, b]$ ,

$$\frac{S(x) - S(a)}{R(x) - R(a)} \geq \frac{S(b) - S(a)}{R(b) - R(a)}.$$

Equivalently, for every  $z \in [w, w']$ , where  $w = R(a)$  and  $w' = R(b)$ ,

$$H(z) \geq H(w) + (z - w)[H(w') - H(w)].$$

Because the straight line connecting  $(w, H(w))$  and  $(w', H(w'))$  lies in  $\text{Co}(\text{Graph}(H))$ , it must be that for  $z \in [w, w']$ ,  $\check{H}(z) \leq H(w) + (z - w)[H(w') - H(w)] \leq H(z)$ . However, then there are quantiles  $z'$  and  $z''$  with  $z \leq w < w' \leq z'$  such that for  $z \in [w, w']$ ,  $H(z)$  belongs to the straight line connecting  $(z', H(z'))$  and  $(z'', H(z''))$ . So there exists an optimal categorization that pools the interval of quantiles  $[z', z'']$ , which contains the interval of quantiles  $[w, w']$ . Equivalently, such categorization pools the interval of qualities  $[R^{-1}(z'), R^{-1}(z'')]$ , which contains the interval  $[a, b]$ .

Now let us prove the second statement. Suppose  $[a, b]$  belong to a pooling interval  $[a', b']$  with  $a' \leq a$  and  $b' \geq b$ . Also suppose  $R$  does not  $\succsim_1 S$  on  $[a', b']$ . Then there exists  $z \in (R(a'), R(b'))$  such that

$$H(z) < H(w) + (z - w)[H(w') - H(w)],$$

where  $w = R(a')$  and  $w' = R(b')$ . However, that means that  $H(w) + (z - w)[H(w') - H(w)] \neq \check{H}(z)$ , and so  $[a', b']$  cannot be a pooling interval in the optimal categorization.  $\square$

**PROOF OF PROPOSITION 3.** The first statement is immediate given Theorem 1. As for the second,  $H$  is convex in  $[0, 1]$  if and only if for every  $w, x, z \in [0, 1]$  with  $w < x \leq z$ ,

$$\frac{H(x) - H(w)}{x - w} \leq \frac{H(z) - H(w)}{z - w}.$$

Letting  $a = R^{-1}(w)$ ,  $b = R^{-1}(z)$ , and  $y = R^{-1}(x)$ , this condition is equivalent to, for all  $a, b, y \in [a, \bar{a}]$  with  $a < y \leq b$ ,

$$\frac{S(y) - S(a)}{R(y) - R(a)} \leq \frac{S(b) - S(a)}{R(b) - R(a)}.$$

If  $S$  is a strictly increasing cdf, then this condition is equivalent to  $S(\cdot|(a, b))$  first-order stochastically dominating  $R(\cdot|(a, b))$  for every  $a, b \in [a, \bar{a}]$ , which is in turn equivalent to  $S$  dominating  $R$  in the likelihood-ratio order.  $\square$

**PROOF OF PROPOSITION 4.** Let  $w = R(a)$  and  $w' = R(b)$ . If  $[a, b]$  is a subset of some separating interval of qualities of categorization  $A$ , then for all  $x \in [w, w']$ ,  $\Phi(x, A) = H(x)$ , and if  $H(x) \neq \check{H}(x)$  for some  $x \in [w, w']$ , then by Theorem 1,  $A$  is not an optimal categorization.

Now notice that if  $H = \check{H}$  on this interval, then  $H$  is convex on it, and so  $S$  dominates  $R$  in the likelihood-ratio order over this interval.  $\square$

The proof of Proposition 5 repeatedly uses the lemma below, which is an immediate consequence of Theorem 1 and, thus, is stated without proof.



LEMMA 5. Take  $x \in (a, \bar{a})$  and let  $z = R(x)$ . Suppose there exist quantiles  $z_1 < z < z_2$  and  $\lambda \in (0, 1)$ , with  $z = \lambda z_1 + (1 - \lambda)z_2$ , such that  $\lambda H(z_1) + (1 - \lambda)H(z_2) \leq H(z)$ . Then there exists an optimal categorization such that  $x$  belongs to a pooling category. Moreover, if the inequality is strict, then  $x$  belongs to a pooling category in any optimal categorization.

PROOF OF PROPOSITION 5. Part (i). Take  $p \in \mathcal{P}$  and let  $P = (P_1, P_2)$  be the associated pair of prior cdfs. Let quality  $x \in (a, \bar{a})$  belong to some separating interval in some optimal categorization of problem  $(S, R) = (P_1, P_2)$ , and let  $z = R(x) = P_2(x)$ . By Lemma 5, it must be that for all  $(z_1, z_2)$  with  $0 \leq z_1 < z < z_2 \leq 1$ ,

$$\lambda H(z_1) + (1 - \lambda)H(z_2) \geq H(z) \quad \text{with } z = \lambda z_1 + (1 - \lambda)z_2, \tag{19}$$

where  $H \equiv P_1 \circ P_2^{-1}$ . Fix any such pair  $(z_1, z_2)$ , and let  $y_1 = H(z_1)$ ,  $y_2 = H(z_2)$ , and  $y = H(z)$ . Then  $0 \leq y_1 < y < y_2 \leq 1$ , and we can rewrite (19) as

$$\lambda y_1 + (1 - \lambda)y_2 \geq y \quad \text{and} \quad \lambda H^{-1}(y_1) + (1 - \lambda)H^{-1}(y_2) = H^{-1}(y), \tag{20}$$

where  $H^{-1} = P_2 \circ P_1^{-1}$ , and so corresponds to the  $H$  function in the “mirrored” problem  $(S, R) = (P_2, P_1)$ . Reduce  $\lambda$  to  $\lambda'$  if needed, so that the inequality in (20) holds with equality. Then, because  $H^{-1}$  is increasing, we must have

$$\lambda' H^{-1}(y_1) + (1 - \lambda')H^{-1}(y_2) \leq H^{-1}(y), \quad \text{where } y = \lambda' y_1 + (1 - \lambda')y_2. \tag{21}$$

Using (21) and Lemma 5 again, we have that quantiles  $(y_1, y_2)$  of distribution  $P_1$  are pooled in some optimal categorization of the problem  $(S, R) = (P_2, P_1)$ . Noting that quantiles  $(y_1, y_2)$  of  $P_1$  correspond to quantiles  $(z_1, z_2)$  of  $P_2$ —and thus contain  $x$ —we conclude that  $x$  belongs to a pooling interval in some optimal categorization of problem  $(S, R) = (P_2, P_1)$ .

Part (ii).

Step 1. Fix  $p = (p_1, p_2) \in \mathcal{P}$ , with  $h(x) = p_1(x)/p_2(x)$ , let  $P = (P_1, P_2)$  be its associated pair of cdfs, and let  $H = P_1 \circ P_2^{-1}$ . In this step, we prove the following claim.

CLAIM 1. If  $h$  is not monotone, there exists a nondegenerate interval of qualities that is comprehensively pooled under  $p$ .

PROOF OF CLAIM 1. Because  $h$  is regular and non-monotone, it has an isolated turn, say at some interval of qualities  $[x^*, x^{**}]$ . By the definition of an isolated turn, there is  $x' < x^*$  and  $x'' > x^{**}$  such that  $h$  is constant on  $[x^*, x^{**}]$ , and (without loss) strictly decreasing on  $[x', x^*]$  and strictly increasing on  $[x^{**}, x'']$ . Equivalently, letting  $\{z', z^*, z^{**}, z''\} = \{P_2(x'), P_2(x^*), P_2(x^{**}), P_2(x'')\}$ ,  $H$  is strictly concave on  $[z', z^*]$ , linear on  $[z^*, z^{**}]$ , and strictly convex on  $[z^*, z'']$ . (We use notation  $z$  to indicate quantiles of distribution  $P_2$ .)

Because  $H$  is strictly concave on  $[z', z^*]$ , Lemma 5 implies that qualities in quantiles  $[z', z^*)$  of  $P_2$  are pooled in any optimal categorization of problem  $(S, R) = (P_1, P_2)$ . Further, in any such optimal categorization, quantiles  $[z', z^*)$  must be contained in a strictly larger pooling interval, which strictly extends not just beyond  $z^*$ , but up to some quantile  $z_2 \in (z^{**}, z'']$ . The reason is that for any quantile  $z$  between  $z^*$  and such  $z_2$ , the strict

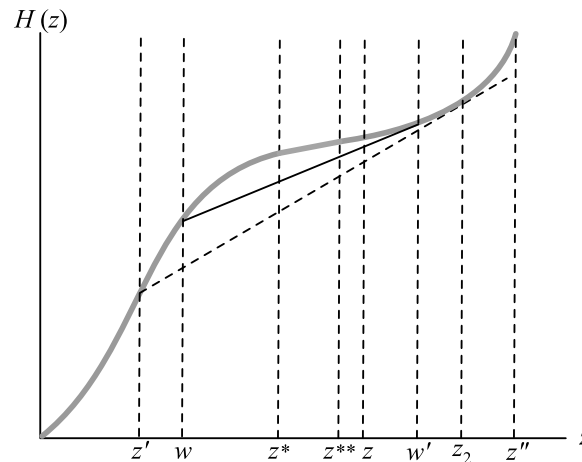


FIGURE 5. Illustration 1 for Proof of Proposition 5. Function  $H$  is strictly concave on  $[z', z^*]$ , linear on  $[z^*, z^{**}]$ , and strictly convex on  $[z^{**}, z'']$ . Optimal pooling *must* extend beyond  $z^{**}$  into the strictly convex region—at least to  $z_2$ . A quality with quantile  $z \in (z^{**}, z_2)$  satisfies the sufficient condition for pooling to be strictly optimal, as the chord joining  $H(w)$  and  $H(w')$  shows.

inequality in Lemma 5 holds, making pooling  $z$  strictly optimal. This is illustrated in Figure 5.

Now consider the mirrored problem  $(S, R) = (P_2, P_1)$ . Define  $P_1$  quantiles corresponding to the  $P_2$  quantiles above, replacing the notation  $z$  by  $y$ . Function  $H$  is strictly convex on  $[z^{**}, z'']$  or, equivalently,  $H^{-1}$  is strictly concave on  $(y^{**}, y'')$ , where  $y^{**} = H(z^{**})$  and  $y'' = H(z'')$ . By Lemma 5, quantiles  $[y^{**}, y'')$  of  $P_1$  belong to a pooling interval in any optimal categorization of problem  $(S, R) = (P_2, P_1)$ . However, by the same argument as in the previous paragraph, in any optimal categorization, the pooling interval containing  $[y^{**}, y'')$  must extend back to some  $y_1$  strictly smaller than  $y^*$ . Letting  $z_1 = H^{-1}(y_1)$ , we thus know that quantiles  $[z_1, z_2)$  must belong to a pooling interval in any optimal categorization in both problems, so the claim is proved.

*Step 2.* In light of Step 1, it only remains to show that the set of all  $p$  with *non-monotone* regular  $h$  is open and dense in  $\mathcal{P}$ . Call this set  $\mathcal{P}^0$ .

Pick  $p \in \mathcal{P}^0$ . There are intervals  $[a, b]$  and  $[c, d]$  on which  $h$  is strictly increasing and strictly decreasing, respectively. Define  $\delta \equiv \min\{h(b) - h(a), h(c) - h(d)\}/3 > 0$ . Because  $\mathcal{P}$  contains only strictly positive, continuous density pairs, there is  $\epsilon > 0$  such that if  $\|p - p'\| < \epsilon$  for some  $p'$  in  $\mathcal{P}$ , then  $\|h - h'\| < \delta$ , where  $h' = p'_1/p'_2$ . By the definition of  $\delta$  and uniform convergence, we see that  $h'$  cannot be monotone. Therefore,  $\mathcal{P}^0$  is open in  $\mathcal{P}$ .

Next, we argue that  $\mathcal{P}^0$  is dense in  $\mathcal{P}$ . Pick  $p \in \mathcal{P} \setminus \mathcal{P}^0$ . We deform  $p$  locally so as to keep its associated  $h$  regular but make it non-monotone. In what follows we suppose that  $h$  is nondecreasing (the other case is proved similarly). Pick some quality level  $x^* \in (a, \bar{a})$ . Fix some small  $\epsilon > 0$ , and let  $k \equiv p_1(x^*)/p_2(x^*)$  and  $k' \equiv p_1(x^* + \epsilon)/p_2(x^* + \epsilon)$ .<sup>25</sup>

<sup>25</sup>Specifically,  $\epsilon > 0$  is smaller than  $k$  and also small enough so that  $x^* + 3\epsilon < \bar{x}$ .

Define a function  $p_1^\epsilon$  by

$$p_1^\epsilon(x) = \begin{cases} p_1(x) & \text{for } x \leq x^* \text{ or } x > x^* + 3\epsilon, \\ (k - x + x^*)p_2(x) & \text{for } x \in (x^*, x^* + \epsilon], \\ \frac{k - \epsilon}{k'} p_1(x) + (x - x^* - \epsilon)\mu p_2(x) & \text{for } x \in (x^* + \epsilon, x^* + 2\epsilon], \\ \frac{x - x^* - 2\epsilon}{\epsilon} [p_1(x^* + 3\epsilon) - p_1^\epsilon(x^* + 2\epsilon)] + p_1^\epsilon(x^* + 2\epsilon) & \text{for } x \in (x^* + 2\epsilon, x^* + 3\epsilon], \end{cases}$$

where  $\mu > 0$  is chosen so that

$$\int_{x^*}^{x^*+3\epsilon} p_1^\epsilon(x) dx = \int_{x^*}^{x^*+3\epsilon} p_1(x) dx. \tag{22}$$

Because  $h(x) = p_1(x)/p_2(x)$  is nondecreasing,  $p_1^\epsilon(x) < p_1(x)$  for  $x \in (x^*, x^* + \epsilon]$ .<sup>26</sup> Then  $p_1^\epsilon$  rises faster than  $p_1$  and intersects it from below, on  $(x^* + \epsilon, x^* + 2\epsilon]$ , achieving  $p_1^\epsilon(x^* + 2\epsilon) > p_1(x^* + 2\epsilon)$ ,<sup>27</sup> and  $p_1^\epsilon$  then adjusts to meet  $p_1$  again at point  $x^* + 3\epsilon$ . At the middle interval,  $p_1^\epsilon$  is determined by  $\mu > 0$ , picked so as to ensure that  $p_1^\epsilon$  is a bonafide density and integrates to 1. Note that there is a unique value of  $\mu$  such that (22) holds: the left-hand side of (22) is strictly increasing in  $\mu$ , smaller than the right-hand side as  $\mu \rightarrow 0$ , and larger than it as  $\mu \rightarrow \infty$ .

Now consider the prior pair  $p^\epsilon = (p_1^\epsilon, p_2^\epsilon)$ , with  $p_2^\epsilon = p_2$ , and let  $h^\epsilon(x) = p_1^\epsilon(x)/p_2^\epsilon(x)$ . We claim that  $h^\epsilon$  is regular and non-monotone for every  $\epsilon > 0$ . Obviously,  $h^\epsilon(x) = h(x)$  for all  $x \leq x^*$  and  $x \geq x^* + 3\epsilon$ . For any  $x \in (x^*, x^* + \epsilon)$ ,

$$h^\epsilon(x) = \frac{p_1^\epsilon(x)}{p_2^\epsilon(x)} = \frac{(k - x + x^*)p_2(x)}{p_2(x)} = k - x + x^*,$$

and so  $h^\epsilon$  strictly declines in  $x$  over this range. For  $x \in (x^* + \epsilon, x^* + 2\epsilon)$ ,

$$h^\epsilon(x) = \frac{p_1^\epsilon(x)}{p_2^\epsilon(x)} = \frac{k - \epsilon}{k'} \frac{p_1(x)}{p_2(x)} + (x - x^* - \epsilon)\mu = \frac{k - \epsilon}{k'} h(x) + (x - x^* - \epsilon)\mu,$$

which is strictly increasing in  $x$ , given that  $h$  is nondecreasing.

It follows that  $h^\epsilon$  has an isolated turn at  $x^* + \epsilon$  and is, therefore, regular. All that remains to do is take  $\epsilon \rightarrow 0$  and notice that  $p_1^\epsilon$  converges in the topology of uniform convergence to  $p_1$ . This shows that  $\mathcal{P}^0$  is dense in  $\mathcal{P}$ , and our proof is complete.  $\square$

**PROOF OF PROPOSITION 6.** For each pair of sender–receiver priors  $(S, R)$ , let  $\mathcal{A}(S, R)$  be the set of optimal categorizations. Now observe that for any  $(S, R)$ , there exists

<sup>26</sup>For all  $x \in (x^*, x^* + \epsilon]$ , the fact that  $h$  is nondecreasing tells us that  $p_1(x) \geq k p_2(x) > (k - x + x^*) p_2(x)$ .

<sup>27</sup>The integrability requirement (22) guarantees that  $p_1^\epsilon(x^* + 2\epsilon)$  must strictly exceed  $p_1(x^* + 2\epsilon)$ .

a maximally separating optimal categorization. To see this, take  $A^* \in \mathcal{A}(S, R)$  with  $\text{int}(\text{Sep}(\hat{A})) = \text{int}\{H = \hat{H}\}$ , where  $H = S \circ R^{-1}$ . Theorem 1 implies that for such  $A^*$ ,

$$A \in \mathcal{A}(S, R) \Rightarrow \text{int}(\text{Sep}(A)) \subset \text{int}(\text{Sep}(A^*)).$$

Let  $A$  and  $\hat{A}$  be maximally separating in  $\mathcal{A}(S, R)$  and  $\mathcal{A}(\hat{S}, R)$ , respectively. Suppose  $(a, a') \subset \text{int}(\text{Sep}(A))$ . By the definition of  $A$ ,  $(a, a')$  is an open interval over which  $H = S \circ R^{-1}$  coincides with its lower convex envelope. Equivalently, for every  $x \in (a, a')$  and  $y, z \in [\underline{a}, \bar{a}]$ , with  $y < x < z$ , defining  $\alpha \in (0, 1)$  such that  $x = \alpha y + (1 - \alpha)z$ ,

$$H(x) = H(\alpha y + (1 - \alpha)z) \leq \alpha H(y) + (1 - \alpha)H(z). \tag{23}$$

Now recall that  $\hat{S} \succsim_{\ell} S$  implies that there is some increasing and convex function  $\varphi$  such that  $\hat{S} = \varphi \circ S$ . This implies that for the same  $x, y, z \in [\underline{a}, \bar{a}]$  and  $\alpha \in (0, 1)$ ,

$$\begin{aligned} \varphi \circ H(\alpha y + (1 - \alpha)z) &\leq \varphi[\alpha H(y) + (1 - \alpha)H(z)] \\ &\leq \alpha \varphi \circ H(y) + (1 - \alpha) \varphi \circ H(z). \end{aligned} \tag{24}$$

Consequently,  $(a, a')$  is an open interval over which  $\hat{H} = \varphi \circ S \circ R^{-1}$  coincides with its lower convex envelope. Therefore, by the definition of  $\hat{A}$ ,  $(a, a') \subset \text{int}(\text{Sep}(\hat{A}))$ . This is true for every such  $(a, a')$ , and so  $\text{int}(\text{Sep}(A)) \subset \text{int}(\text{Sep}(\hat{A}))$ .  $\square$

**PROOF OF LEMMA 4.** Part (i). Let  $\ell(a)$  be an incentive-compatible learning function. First, note that for an initial choice of  $\ell(\underline{a}) \in [0, \frac{A_{\ell}(\underline{a}) - \underline{a}}{c(\underline{a}) - \lambda}]$ , (9) is satisfied for the lowest ability student. By Lemma 3,  $\ell$  is nondecreasing. Next, take  $a$  and  $a'$  in the same separating interval. From (1) and (9), we have

$$\frac{1}{c(a') - \lambda} \geq \frac{\ell(a) - \ell(a')}{a - a'} \geq \frac{1}{c(a) - \lambda}.$$

Take  $a' \rightarrow a$  to obtain (10) on any separating interval. Now take  $a \in [a_{k-1}, a_k)$  for some  $k > 1$ . Use (9) to see that

$$\ell(a) + \frac{A_{\ell}(a_k) - A_{\ell}(a)}{c(a) - \lambda} \leq \ell(a_k) \leq \ell(a) + \frac{A_{\ell}(a_k) - A_{\ell}(a)}{c(a_k) - \lambda}$$

and send  $a \rightarrow a_k$  to get (11). Conversely, with  $\ell(\underline{a}) \in [0, \frac{A_{\ell}(\underline{a}) - \underline{a}}{c(\underline{a}) - \lambda}]$ ,  $\ell(a)$  constant on pooling intervals, and given (10) and (11), we must conclude that  $\ell$  is an incentive-compatible learning function.

Part (ii). Let  $A \in \mathcal{A}_R$ . Given  $\underline{\ell} \in [0, \frac{A(\underline{a}) - \underline{a}}{c(\underline{a}) - \lambda}]$ , define a function  $\ell$  with  $\ell(\underline{a}) = \underline{\ell}$ , with (10) holding on separating intervals of  $A$ , with  $\ell$  constant on pooling intervals of  $A$ , and satisfying (11), with  $A_{\ell} = A$ , at every left edge  $a_k$  of every interval. Standard arguments for differential equations ensure that  $\ell$  is well defined and unique. By Part (i),  $\ell$  is incentive-compatible.  $\square$

**PROOF OF PROPOSITION 7.** The proof will rely on the following lemmas.

LEMMA 6. For  $A \in \mathcal{A}$  and  $\underline{\ell} \in [0, \frac{A(\underline{a})-\underline{a}}{c(\underline{a})-\lambda}]$ , let  $\ell$  be the unique associated learning function as given by Lemma 4. Let  $dA$  and  $d\ell$  be the Stieltjes measures associated with  $A$  and  $\ell$ , respectively. Then  $d\ell$  is absolutely continuous with respect to  $dA$  and

$$\frac{d\ell}{dA}(a) = \frac{1}{c(a) - \lambda}$$

is the Radon–Nikodym derivative of  $d\ell$  with respect to  $dA$ .

PROOF. For any set  $B \subset [\underline{a}, \bar{a}]$ ,  $dA(B) = 0$  trivially implies  $d\ell(B) = 0$  since  $A$  and  $\ell$  are constant or strictly increasing in exactly the same intervals. Hence  $d\ell$  is absolutely continuous with respect to  $dA$ , and so there exists a Radon–Nikodym derivative between the two measures. Now let  $[b, b'] \subset [\underline{a}, \bar{a}]$ . Then  $[b, b']$  is made up of countably many intervals  $[c, c'] \in \mathcal{C}$  on which both  $A$  and  $\ell$  are continuous and differentiable, along with at most countably many points of discontinuity  $d \in \mathcal{D}$ . It follows that

$$\begin{aligned} d\ell[b, b'] &= \ell(c) - \ell(b) = \sum_{(c, c') \in \mathcal{C}} \int_c^{c'} \ell'(x) dx + \sum_{d \in \mathcal{D}} [\ell(d) - \ell^\uparrow(d)] \\ &= \sum_{(c, c') \in \mathcal{C}} \int_c^{c'} \frac{1}{c(x) - \lambda} A'(x) dx + \sum_{d \in \mathcal{D}} \frac{1}{c(d) - \lambda} (A(d) - A^\uparrow(d)) \\ &= \int_b^{b'} \frac{1}{c(x) - \lambda} dA(x), \end{aligned}$$

where the third equality uses Lemma 4.

Because the intervals  $[b, b'] \in [\underline{a}, \bar{a}]$  generate the Borel  $\sigma$ -algebra in  $[\underline{a}, \bar{a}]$ , we conclude that  $\frac{d\ell}{dA}(x) = \frac{1}{c(x) - \lambda}$  is the Radon–Nikodym derivative of  $d\ell$  with respect to  $dA$ .  $\square$

LEMMA 7. *Integration by Parts.* If  $P$  is a  $Q$ -integrable function on  $[\underline{a}, \bar{a}]$ , then  $Q$  is  $P$ -integrable on  $[\underline{a}, \bar{a}]$  and

$$\int_{\underline{a}}^{\bar{a}} P(x) dQ(x) = P(\underline{a}) \int_{\underline{a}}^{\bar{a}} dQ(x) + \int_{\underline{a}}^{\bar{a}} \int_{\underline{a}}^x dQ(y) dP(x).$$

PROOF. If  $P$  is  $Q$ -integrable, then the standard integral by parts formula yields

$$\int_{\underline{a}}^{\bar{a}} P(x) dQ(x) = P(\bar{a})Q(\bar{a}) - P(\underline{a})Q(\underline{a}) - \int_{\underline{a}}^{\bar{a}} Q(x) dP(x). \tag{25}$$

Rearrange (25) to get

$$\int_{\underline{a}}^{\bar{a}} P(x) dQ(x) = \left[ P(\underline{a}) + \int_{\underline{a}}^{\bar{a}} dP(x) \right] Q(\bar{a}) - P(\underline{a})Q(\underline{a}) - \int_{\underline{a}}^{\bar{a}} Q(x) dP(x)$$

$$\begin{aligned}
 &= P(\underline{a}) \int_{\underline{a}}^{\bar{a}} dQ(x) + \int_{\underline{a}}^{\bar{a}} \left( \int_{\underline{a}}^{\bar{a}} dQ(y) \left( - \int_{\underline{a}}^x dQ(y) \right) \right) dP(x) \\
 &= P(\bar{a})Q(\bar{a}) - P(\underline{a})Q(\underline{a}) - \int_{\underline{a}}^{\bar{a}} Q(x) dP(x). \quad \square \\
 &\quad \square
 \end{aligned}$$

*Case 1.* First assume  $\lambda > \sigma \int_{\underline{a}}^{\bar{a}} c(a) dR_0(a)$ , so that  $\ell^*(\underline{a}) = \frac{A(\underline{a}) - \underline{a}}{c(\underline{a}) - \lambda}$ .

Set  $P = A$  and  $Q = R_0$  in Lemma 7. Because  $R_0$  is continuous and of bounded variation, the relevant integral is defined and

$$\int_{\underline{a}}^{\bar{a}} A(a) dR_0(a) = A(\underline{a}) \int_{\underline{a}}^{\bar{a}} dR_0(a) + \int_{\underline{a}}^{\bar{a}} \int_a^{\bar{a}} dR_0(x) dA(a). \tag{26}$$

Next, setting  $P = \ell$ , setting  $dQ(x) = [\lambda - \sigma c(x)] dR_0(x)$  in Lemma 7, and noting again that  $Q$  is continuous and of bounded variation, we see that

$$\begin{aligned}
 &\int_{\underline{a}}^{\bar{a}} [\lambda - \sigma c(a)] \ell(a) dR_0(a) \\
 &= \ell(\underline{a}) \int_{\underline{a}}^{\bar{a}} [\lambda - \sigma c(a)] dR_0(a) + \int_{\underline{a}}^{\bar{a}} \int_a^{\bar{a}} [\lambda - \sigma c(x)] dR_0(x) d\ell(a). \tag{27}
 \end{aligned}$$

Recall that  $\ell(\underline{a}) = A(\underline{a})/[c(\underline{a}) - \lambda]$ . Use this in (27) and invoke Lemma 6 to get

$$\begin{aligned}
 &\int_{\underline{a}}^{\bar{a}} [\lambda - \sigma c(a)] \ell(a) dR_0(a) \\
 &= (A(\underline{a}) - \underline{a}) \int_{\underline{a}}^{\bar{a}} \frac{\lambda - \sigma c(a)}{c(\underline{a}) - \lambda} dR_0(a) + \int_{\underline{a}}^{\bar{a}} \int_a^{\bar{a}} \frac{\lambda - \sigma c(x)}{c(a) - \lambda} dR_0(x) dA(a) \\
 &= K + A(\underline{a}) \int_{\underline{a}}^{\bar{a}} \frac{\lambda - \sigma c(a)}{c(\underline{a}) - \lambda} dR_0(a) + \int_{\underline{a}}^{\bar{a}} \int_a^{\bar{a}} \frac{\lambda - \sigma c(x)}{c(a) - \lambda} dR_0(x) dA(a), \tag{28}
 \end{aligned}$$

where  $K = -\underline{a} \int_{\underline{a}}^{\bar{a}} \frac{\lambda - \sigma c(a)}{c(\underline{a}) - \lambda} dR_0(a)$ . Combining (26) and (28),

$$\begin{aligned}
 &\int_{\underline{a}}^{\bar{a}} [A(a) + (\lambda - \sigma c(a)) \ell(a)] dR_0(a) \\
 &= K + A(\underline{a}) \int_{\underline{a}}^{\bar{a}} \frac{c(\underline{a}) - \sigma c(a)}{c(\underline{a}) - \lambda} dR_0(a) + \int_{\underline{a}}^{\bar{a}} \int_a^{\bar{a}} \frac{c(a) - \sigma c(x)}{c(a) - \lambda} dR_0(x) dA(a) \\
 &= K + A(\underline{a}) [1 - S(\underline{a})] + \int_{\underline{a}}^{\bar{a}} [1 - S(a)] dA(a), \tag{29}
 \end{aligned}$$

where  $S$  is defined by (15):

$$S(a) = R_0(a) + \int_a^{\bar{a}} \frac{\sigma c(x) - \lambda}{c(a) - \lambda} dR_0(x).$$

Note that  $S$  is continuous,  $S(\underline{a})$  is finite, and  $S(\bar{a}) = 1$ . Also remark that  $R_0(\underline{a}) = 0$ , so that  $S(\underline{a})$  is strictly negative.

We claim that  $S$  has bounded variation. Define  $\Delta^+(x) \equiv \max\{\lambda - \sigma c(x), 0\}$  and  $\Delta^\dagger(x) \equiv -\min\{\lambda - \sigma c(x), 0\}$ . Then, by (15),

$$\begin{aligned} S(a) &= R_0(a) + \int_a^{\bar{a}} \frac{\Delta^+(x)}{c(a) - \lambda} dR_0(x) - \int_a^{\bar{a}} \frac{\Delta^\dagger(x)}{c(a) - \lambda} dR_0(x) \\ &= R_0(a) + \int_a^{\bar{a}} \frac{\Delta^+(x)}{c(a) - \lambda} dR_0(x) - \int_a^a \frac{\Delta^+(x)}{c(a) - \lambda} dR_0(x) \\ &\quad - \int_a^{\bar{a}} \frac{\Delta^\dagger(x)}{c(a) - \lambda} dR_0(x) + \int_a^a \frac{\Delta^\dagger(x)}{c(a) - \lambda} dR_0(x). \end{aligned}$$

The first term on the right-hand side of this equation is a cdf, nondecreasing in  $a$ . Consider each of the four integrals (without the sign that precedes them). Each integrand is a nonnegative-valued function (because  $c(a) > \lambda$ , and  $\Delta^+$  and  $\Delta^\dagger$  are each nonnegative), and each is nondecreasing in  $a$  (because  $c(a)$  declines in  $a$ ). Therefore, each integral is nondecreasing in  $a$ . It follows that  $S$  can be written as the sum/difference of five nondecreasing functions and, consequently, is of bounded variation. Therefore, integration with respect to  $S$  is well defined. Define  $P = A$  and  $Q(a) = 1 - S(a)$ , and apply Lemma 7 yet again to (29) to obtain (14).

The remainder now follows by applying Theorem 1 to the induced problem  $(S, R)$ , solving for an optimal  $A^*$ , setting  $\underline{\ell} = \ell^*$ , and then backing out the optimal learning function via Lemma 4.

Case 2. Now assume  $\lambda \leq \sigma \int_a^{\bar{a}} c(a) dR_0(a)$ , so that  $\ell^*(\underline{a}) = 0$ .

Equations (26) and (27) still hold as in Case 1, but now note that  $\ell(\underline{a}) = 0$  and so by setting  $S(\underline{a}) = 0$ , we can rewrite (27) as

$$\int_a^{\bar{a}} [\lambda - \sigma c(a)] \ell(a) dR_0(a) = -A(\underline{a})S(\underline{a}) + \int_a^{\bar{a}} \int_a^{\bar{a}} \frac{\lambda - \sigma c(x)}{c(a) - \lambda} dR_0(x) dA(a). \quad (30)$$

Finally, combine (26) and (30) to again get

$$\int_a^{\bar{a}} [A(a) + (\lambda - \sigma c(a)) \ell(a)] dR_0(a) = A(\underline{a})[1 - S(\underline{a})] + \int_a^{\bar{a}} [1 - S(a)] dA(a). \quad (31)$$

We prove that  $S$  has bounded variation just as before. Since  $S$  is left-continuous and only discontinuous at  $\underline{a}$ , integration with respect to  $S$  is still well defined. Define  $P = A$  and  $Q(a) = 1 - S(a)$ , apply Lemma 7 yet again to (31), and set  $K = 0$  to obtain (14).  $\square$

#### REFERENCES

- Alonso, Ricardo and Odilon Camara (2016), “Bayesian persuasion with heterogeneous priors.” *Journal of Economic Theory*, 165, 672–706. [1409, 1415, 1416, 1421]
- Board, Simon (2009), “Monopolistic group design with peer effects.” *Theoretical Economics*, 4, 89–125. [1416]

- Boleslavsky, Raphael and Kyungmin Kim (2020), “Bayesian persuasion and moral hazard.” Working paper. [1426]
- Che, Yeon-Koo and Navin Kartik (2009), “Opinions as incentives.” *Journal of Political Economy*, 117, 815–860. [1415]
- Curello, Gregorio and Ludvig Sinander (2022), “The comparative statics of persuasion.” Working paper. [1420, 1421]
- de Clippel, Geoffroy and Xu Zhang (2020), “Non-Bayesian persuasion.” Working paper. [1415]
- Dworczak, Piotr and Giorgio Martini (2019), “The simple economics of optimal persuasion.” *Journal of Political Economy*, 127, 1993–2048. [1409]
- Galperti, Simone (2019), “Persuasion: The art of changing worldviews.” *American Economic Review*, 109, 996–1031. [1415]
- Goldstein, Itay and Yaron Leitner (2018), “Stress tests and information disclosure.” *Journal of Economic Theory*, 177, 34–69. [1416]
- Ivanov, Maxim (2021), “Optimal monotone signals in Bayesian persuasion mechanisms.” *Economic Theory*, 72, 955–1000. [1409]
- Jewitt, Ian and Daniel Quigley (2022), “Conjugate persuasion.” Working paper. [1422]
- Kamenica, Emir and Matthew Gentzkow (2011), “Bayesian persuasion.” *American Economic Review*, 101, 2590–2615. [1409, 1415]
- Kartik, Navin, Frances Xu Lee, and Wing Suen (2017), “Investment in concealable information by biased experts.” *The RAND Journal of Economics*, 48, 24–43. [1415]
- Kartik, Navin, Frances Xu Lee, and Wing Suen (2021), “Information validates the prior: A theorem on Bayesian updating and applications.” *American Economic Review: Insights*, 3, 165–182. [1410, 1415]
- Kleiner, Andreas, Benny Moldovanu, and Philipp Strack (2021), “Extreme points and majorization: Economic applications.” *Econometrica*, 89, 1557–1593. [1409, 1415]
- Kolotilin, Anton (2018), “Optimal information disclosure: A linear programming approach.” *Theoretical Economics*, 13, 607–636. [1409]
- Kolotilin, Anton, Roberto Corrao, and Alexander Wolitzky (2022), “Persuasion with non-linear preferences.” Working paper. [1422]
- Kolotilin, Anton and Hongyi Li (2021), “Relational communication.” *Theoretical Economics*, 16, 1391–1430. [1409]
- Kolotilin, Anton, Tymofiy Mylovanov, and Andriy Zapechelnyuk (2019), “Censorship as optimal persuasion.” Working paper. [1427]
- Kolotilin, Anton and Andriy Zapechelnyuk (2019), “Persuasion meets delegation.” Working paper. [1416]



- Lizzeri, Alessandro (1999), “Information revelation and certification intermediaries.” *RAND Journal of Economics*, 30, 214–231. [1426]
- Mensch, Jeffrey (2021), “Monotone persuasion.” *Games and Economic Behavior*, 130, 521–542. [1409]
- Mussa, Michael and Sherwin Rosen (1978), “Monopoly and product quality.” *Journal of Economic Theory*, 18, 301–317. [1414]
- Myerson, Roger (1981), “Optimal auction design.” *Mathematics of Operations Research*, 6, 58–73. [1408, 1409, 1414]
- Ostrovsky, Michael and Michael Schwarz (2010), “Information disclosure and unraveling in matching markets.” *American Economic Journal: Microeconomics*, 2, 34–63. [1426]
- Rayo, Luis (2013), “Monopolistic signal provision.” *The BE Journal of Theoretical Economics*, 13, 27–58. [1408, 1410, 1414, 1415, 1416, 1418]
- Rayo, Luis and Ilya Segal (2010), “Optimal information disclosure.” *Journal of Political Economy*, 118, 949–987. [1409, 1416, 1421, 1422]
- Rodina, David and John Farragout (2016), “Inducing effort through grades.” Working paper. [1426]
- Saeedi, Maryam and Ali Shourideh (2020), “Optimal rating design.” Working paper. [1426]
- Shaked, Moshe and J. George Shanthikumar (2007), “Stochastic orders.” *Springer Science & Business Media*. [1417]
- Van den Steen, Eric (2004), “Rational overoptimism (and other biases).” *American Economic Review*, 94, 1141–1151. [1415]
- Van den Steen, Eric (2009), “Authority versus persuasion.” *American Economic Review, Papers and Proceedings*, 99, 448–453. [1415]
- Van den Steen, Eric (2010), “On the origin of shared beliefs (and corporate culture).” *RAND Journal of Economics*, 41, 617–648. [1415]
- Van den Steen, Eric (2011), “Overconfidence by Bayesian-rational agents.” *Management Science*, 57, 884–896. [1415]
- Yaari, Menahem (1987), “The dual theory of choice under risk.” *Econometrica*, 95–115. [1422]

---

Co-editor Simon Board handled this manuscript.

Manuscript received 13 September, 2021; final version accepted 14 December, 2022; available online 13 January, 2023.