The Simple Economics of Optimal Persuasion*

Piotr Dworczak† and Giorgio Martini‡

First version: May 27, 2016
This version: November 30, 2017

Abstract

We study Bayesian persuasion problems in which the Sender's preferences depend only on the mean of posterior beliefs. We show that, given a price schedule for posterior means, the Sender faces a consumer-like choice problem, purchasing posterior means using the prior distribution as her endowment. Prices are determined in equilibrium of a Walrasian economy with the Sender as the only consumer and a production technology that garbles the state. Welfare theorems provide a verification tool for optimality of a persuasion scheme, and characterize the structure of prices that support the optimal solution. This price-theoretic approach yields a tractable solution method for persuasion problems with infinite state spaces. Moreover, we show that the approach extends to the general case with no restrictions on Sender's utility.

Keywords: Bayesian persuasion, information design, mean-preserving spreads, monotone partitions, Walrasian equilibrium.

---

*This paper was previously circulated under the title "A Duality Approach to Bayesian Persuasion." We would like to thank Anthony Lee Zhang for early discussions that were crucial for this project, and Isaìas Chaves, Matthew Gentzkow, Shota Ichihashi, Elliot Lipnowski, Emir Kamenica, Paul Milgrom, Michael Ostrovsky, Phil Reny, Andy Skrzypacz, Alexey Suzdaltsev, as well as three anonymous referees for helpful comments.

†University of Chicago, The Becker Friedman Institute, dworczak@uchicago.edu.

‡Stanford University, Graduate School of Business, gmartini@stanford.edu.

1 Introduction

Bayesian persuasion has become a canonical model of communication with commitment power following Kamenica and Gentzkow (2011).[1] However, the standard approach based on concavification of the value function has limited power when the state space is large. The concavification method alone is typically not sufficient to characterize the optimal signal, and fails to provide intuition for the structure of the underlying persuasion scheme. To overcome this difficulty, we develop a price-theoretic approach to Bayesian persuasion under the assumption that the Sender's preferences only depend on the mean of posterior beliefs.

We treat the Sender as a consumer who purchases posterior means at their prices, subject to a budget constraint, using the prior distribution as her endowment. Prices are determined in a Walrasian equilibrium of a "Persuasion Economy" which features posterior means as "goods", the Sender as the only consumer, and a single firm. The firm has the technology to garble the state, or "merge the goods", and maximizes profits.

We prove analogues of the two welfare theorems for this economy. In Theorem 1, we show that competitive equilibria are efficient (hence, since there is only one agent, optimal). This provides a verification tool for optimality of a candidate solution to the Sender's problem. In Theorem 2, under mild regularity assumptions we show that the Walrasian approach is also necessary: given any optimal persuasion scheme, it is always possible to find a price schedule that supports it as an equilibrium allocation.

Using the analogy to a Walrasian economy, we derive joint restrictions on equilibrium prices and allocations which provide insights about the structure of the optimal persuasion scheme. We show that the technological constraints of the persuasion economy imply that the price schedule must be convex. Moreover, whenever the price function is strictly convex, the state must be fully revealed. Pooling only happens in regions where prices are linear. These restrictions narrow down the set of candidate solutions to a relatively low-dimensional class in which the optimal solution can often be found using a simple graphical method.

Persuasion mechanisms used in practice often have a simple structure: information is either fully revealed or adjacent types are pooled together (e.g. coarse ratings used by bond rating agencies). In Theorem 3, we use our approach to derive a necessary and sufficient condition on the Sender's utility function under which a monotone partitional signal is optimal for any prior distribution of the state.

We further illustrate the usefulness of the price-theoretic approach to persuasion by solving two examples that could not be directly solved using previous methods. In the first application, an agent must be persuaded to exert effort on a project. The agent is rewarded with a fraction of the value of the project but only the principal knows how much the project is worth if successful.

[1] Other references containing important ideas include Aumann and Maschler (1995), Calzolari and Pavan (2006), Ostrovsky and Schwarz (2010) and Rayo and Segal (2010).

We prove that the principal should disclose the project's value when it is low, and pool high realizations into the lowest signal that induces maximal effort. In the second application, a financial analyst who possesses private information on the profitability of a risky asset wants to persuade an agent to invest in it; the optimal persuasion mechanism has a tractable structure in which the informativeness of the analyst's recommendation depends on the agent's degree of risk aversion. In both applications, a simple graphical analysis combined with our verification result is enough to characterize the optimal persuasion mechanism.

We show that our methods extend beyond the simple case in which a single Sender's utility depends solely on the posterior mean. In Section 7.1, we consider a model of competition in persuasion. There are multiple Senders with access to the same information, and they disclose signals simultaneously. Using our method, we characterize the set of equilibrium distributions of posterior means. In Section 7.2, we show that the analogy to Walrasian equilibrium holds for a general Bayesian persuasion problem, with no restrictions on the objective function of the Sender. Prices in this case are defined on the space of posterior beliefs rather than posterior means. We show that the supporting prices provide an alternative characterization of the Sender's value function, complementary to concavification.

We are not the first to study Bayesian persuasion in the case where payoffs only depend on the posterior mean. Kolotilin (2017) uses duality theory to characterize the optimal persuasion scheme in a related model, and introduces the idea of prices for messages in the context of an example. The graphical approach introduced in Gentzkow and Kamenica (2015) gives further insights about the structure of the problem. We defer discussion of the related literature to Section 8.

The rest of the paper is organized as follows. The next section introduces the persuasion problem formally. In Section 3, we prove our main result, and describe the connection to Walrasian equilibria. In Section 4, we analyze the structure of the optimal solution and equilibrium prices. Section 5 provides a necessary and sufficient condition for optimality of monotone partitional signals. In Section 6, we work through two applications of our methods. In Section 7 we discuss extensions to competition in persuasion and to general persuasion problem, and, finally, in Section 8 – related literature. Proofs and additional applications are collected in the Appendix.

2 Model

The state of nature is the realization of a real-valued random variable X with a cumulative distribution function F.[2] We assume that X has realizations in some non-degenerate bounded interval $[\underline{x}, \bar{x}]$, and its support includes the endpoints $\underline{x}$ and $\bar{x}$. Without loss of generality, we normalize $\underline{x} = 0$ and $\bar{x} = 1$. F is common knowledge between the two players, Sender and

[2] We assume throughout that there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which X is defined. The explicit probability space plays no further role in the analysis.

Receiver.

The Sender commits to an information structure which determines the signal that is sent to the Receiver. An information structure is a measurable mapping $\pi : [0, 1] \to \Delta(S)$, for some signal space S.[3]

Given an information structure π, every signal realization $s \in S$ induces a posterior belief over the distribution of X. We assume that the Sender's final utility depends on posterior beliefs only through the posterior mean. Formally, there exists a measurable function $u : [0, 1] \to \mathbb{R}$ such that $u(x)$ is the ex-post utility of the Sender when the induced posterior mean is x.

The assumption is satisfied when the Receiver's optimal action only depends on the expected state and when the Sender's preferences over actions depend linearly on the state (in particular, if they are state-independent). The Receiver's problem only influences the persuasion problem via the shape of the function u, and thus the Receiver will not play any role in the analysis.

Under this assumption, the expected value of an information structure π depends only on the distribution of posterior means that it induces. It is thus natural to optimize over distributions of posterior means directly. Given the prior F, a distribution of posterior means G is induced by some information structure if and only if F is a mean-preserving spread of G (Blackwell, 1953; Kolotilin, 2017; Gentzkow and Kamenica, 2015).[4]

The Sender's problem is thus

$$\max_G \int_0^1 u(x)dG(x) \tag{2.1}$$

subject to the constraint that F is a mean-preserving spread of G.

3 A price-theoretic approach to persuasion

Our first result provides a way to verify optimality of a candidate solution G by means of an auxiliary function p.

Theorem 1. *If there exist a cumulative distribution function G and a convex function $p : [0, 1] \to \mathbb{R}$, with $p(x) \geq u(x)$ for all $x \in [0, 1]$, that satisfy*

$$\operatorname{supp}(G) \subseteq \{x \in [0, 1] : u(x) = p(x)\}, \tag{3.1}$$

$$\int_0^1 p(x)dG(x) = \int_0^1 p(x)dF(x), \text{ and} \tag{3.2}$$

$$F \text{ is a mean-preserving spread of } G, \tag{3.3}$$

[3] The signal space S is endowed with a relevant σ-field, and π is measurable with respect to the σ-field on $[0, 1]$ with respect to which X is a random variable (which can be taken to be, for example, the Borel σ-field).

[4] There are many equivalent definitions of mean-preserving spreads (see e.g. Müller and Stoyan, 2002). We will use two: F is a mean-preserving spread of G if $\int_0^1 v(x)dF(x) \geq \int_0^1 v(x)dG(x)$ for all convex v; equivalently, if $\int_0^t F(x)dx \geq \int_0^t G(x)dx$ for all $t \in [0, 1]$ and $\int_0^1 F(x)dx = \int_0^1 G(x)dx$.

then G is a solution to problem (2.1).

Proof. See Appendix A.1.

We will be using Theorem 1 primarily as a tool to solve problem (2.1). However, the result also provides a useful interpretation of the optimal persuasion scheme as a Walrasian equilibrium allocation. The auxiliary function p describes equilibrium prices for posterior means. The Sender is a consumer endowed with F who selects a bundle G to maximize utility subject to a budget constraint, similar to conditions (3.1) – (3.2). A single firm transforms states into posterior means using a technology that satisfies the mean-preserving-spread condition (3.3). Subsection (3.1) formalizes this analogy.

3.1 The "Persuasion Economy"

We begin by describing the "Persuasion Economy". Let $[0,\, 1]$, the set of possible posterior means, be the space of goods. A bundle of goods, possibly including negative quantities, is denoted by a function $G : [0, 1] \to \mathbb{R}$, where $G(x)$ is the total mass of goods in $[0, x]$. In the special case where the bundle includes non-negative amounts of each good and the total mass of goods is 1, G is a cumulative distribution function. Let M denote the set of possible bundles.[5]

There is a single consumer with utility $u(x)$ for one unit of good x. Utility is linear in quantity and additively separable across goods: utility of bundle G is $\int_0^1 u(x)dG(x)$. The consumer can only consume non-negative amounts of each good, so the consumption set $M^+ \subset M$ consists of non-negative bundles: G must be non-decreasing. The consumer is initially endowed with a bundle F of goods.

There is a single competitive firm that chooses a production plan Z from the following production possibility set:

$$\mathcal{Z} = \left\{ Z \in M : \int_0^x Z(t)dt \le 0,\, \forall x \in [0,\, 1],\, \int_0^1 Z(t)dt = 0,\, Z(1) = 0 \right\}.$$

$\mathcal{Z}$ describes the technology of taking mean-preserving contractions. That is, the firm takes "states" as inputs, and produces posterior means as outputs. For example, a unit of good x_1 and a unit of good x_2 can be used to produce two units of good $(x_1 + x_2)/2$. The additional condition $Z(1) = 0$ says that the total mass of inputs is equal to the total mass of outputs, and can be interpreted as no free disposal (this rules out uninteresting equilibria). The consumer is the sole shareholder of the firm.

Definition. $(G,\, Z,\, p)$ is a *Walrasian equilibrium* if

[5] Formally, M is the set of signed, countably additive, regular, finite-variation measures on $[0,\, 1]$ with the Borel σ-algebra. With slight abuse of notation, we identify a measure $\mu \in M$ with its cumulative distribution $G(x) = \mu([0, x]) = \int_0^x d\mu(z)$.

1. The consumer maximizes utility given the budget constraint:

$$G \in \underset{\tilde{G} \in M^+}{\operatorname{argmax}} \int_0^1 u(x) d\tilde{G}(x) \text{ subject to } \int_0^1 p(x) d\tilde{G}(x) \leq \int_0^1 p(x) dF(x) + \pi, \tag{3.4}$$

   where π is the dividend from the firm.

2. The firm maximizes profits $\pi = \int_0^1 p(x) dZ(x)$:

$$Z \in \underset{\tilde{Z} \in \mathcal{Z}}{\operatorname{argmax}} \int_0^1 p(x) d\tilde{Z}(x). \tag{3.5}$$

3. Markets clear:

$$G = F + Z. \tag{3.6}$$

An equilibrium (G, Z, p) is not unique because prices p can be rescaled without affecting equilibrium properties. We argue that (G, Z, p) is a Walrasian equilibrium if and only if (G, p') satisfies conditions (3.1)-(3.3) of Theorem 1, where p' is an affine transformation of p.

1. In order for problem (3.5) to admit a solution, prices must be convex (otherwise, the firm can scale up production and achieve infinite profits). Furthermore, when prices are convex, the firm cannot make strictly positive profits (and can always achieve zero profits by not producing anything). Thus, the firm breaks even in equilibrium. By market clearing, the firm must produce $Z = G - F$. The zero-profit condition corresponds to (3.2).

2. To solve the consumer problem, we put a Lagrange multiplier λ on the consumer's budget constraint. Due to linearity, maximizing the Lagrangian is equivalent to $\text{supp}(G) \subseteq \text{argmax}_x\{u(x) - \lambda p(x)\}$. Moreover, the consumer chooses a bundle corresponding to a cdf in equilibrium: Market clearing and the no-free-disposal condition for the firm imply $G(1) = 1$. By letting $\beta \equiv \max_x\{u(x) - \lambda p(x)\}$, we can rescale prices to $p' = \lambda p + \beta$, and conclude that (3.1) and $p' \geq u$ both hold. Conversely, if we start with (G, p) satisfying (3.1)-(3.3), then G is a solution to the consumer's problem because the Lagrangian is maximized (with $\lambda = 1$), and the consumer's budget constraint is satisfied by (3.2).

3. Market clearing is implicit in equations (3.1) – (3.3). Condition (3.3) corresponds to the technological constraint of the firm.

3.2 Welfare Theorems for the Persuasion Economy

We can now interpret Theorem 1 as the First Welfare Theorem for the Persuasion Economy. If (G, Z, p) is an equilibrium, then G is Pareto efficient.[6] Efficiency in a one-consumer economy

[6]Even when u is non-positive (so that consumer's preferences are not necessarily locally insatiable) the First Welfare Theorem holds because the consumer always exhausts her budget in equilibrium. If there were a slack in

means that the allocation maximizes the consumer's welfare. Hence, G is optimal for the Sender.

Conversely, we might expect the Second Welfare Theorem to also hold for the Persuasion Economy: Do there always exist prices that support an optimal solution as an equilibrium allocation? We give a positive answer under mild regularity conditions on u (some assumption on u is clearly needed to guarantee existence of an optimal G).

Definition 1. Function u is *regular* if:

1. u is upper semi-continuous with at most finitely many one-sided jump discontinuities at interior points $y_1, ..., y_k \in (0, 1)$;[7]

2. u has bounded slope in each interval (y_i, y_{i+1});

3. for every convex function v such that $v \geq u$ pointwise, the set $\{x : v(x) > u(x)\}$ can be represented as a finite union of intervals.[8]

Theorem 2. *Suppose that u is regular. Then there exists an optimal solution G, and for every optimal solution G, there exists a convex and continuous $p : [0, 1] \to \mathbb{R}$ such that the pair (G, p) satisfies conditions* (3.1)–(3.3).

Proof. See Appendix A.2.

Because the Persuasion Economy has infinitely many goods, standard proofs of the Second Welfare Theorem cannot be adopted easily. Instead, to prove Theorem 2, we use duality techniques from the literature on optimization with stochastic dominance constraints (Dentcheva and Ruszczynski, 2003). We highlight one element of our proof that uses properties of Walrasian equilibria: Among other technical complications, the problem (2.1) does not satisfy the qualification constraint imposed by Dentcheva and Ruszczynski (2003). We construct a sequence of problems approximating (2.1) in which the qualification constraint holds, and show the appropriate convergence properties. However, this only gives us existence of one solution G supported by a price function p (the limit of the sequence). To prove existence of supporting prices for any G, we notice that our economy is quasi-linear: Because the consumer's utility is linear in quantities, there exists a numeraire good that can be treated as money. Gul and Stacchetti (1999) show that *any* equilibrium allocation in a quasi-linear economy is supported by *any* equilibrium price function. This implies that if G and G' are two solutions to problem (2.1), and (G, p) satisfy the conditions of Theorem 1, then so do (G', p). Thus, Theorem 2 holds for any solution.

the budget constraint, then the firm would be making negative profits, which cannot happen in equilibrium.

[7] Upper semi-continuity could also be derived as a consequence of the primitives, under the additional assumption that the Receiver chooses the Sender-preferred action when indifferent.

[8] If the function u is upper semi-continuous, then the set $\{x : v(x) > u(x)\}$ is always a union of intervals, but the union can in general require infinitely many elements. For example, $f(x) = \sin(\frac{1}{1-x})(1 - x)^2$ for $x < 1$ and $f(1) = 0$ is continuous, has bounded slope but is not regular.

3.3 Further properties of Persuasion Economies

Before proceeding, we establish one more consequence of the analogy between optimal persuasion and Walrasian equilibrium. One of the proofs of existence of Walrasian equilibria features a mythical player, the Walrasian auctioneer, who chooses prices to maximize the value of the net trade. This can be used to provide insights about the structure of equilibrium prices. In the Persuasion Economy, the net trade is $G - F - Z$. Moreover, because $\int_0^1 p(x)dZ(x) = 0$ and $\int_0^1 p(x)dG(x)$ is proportional to $\int_0^1 u(x)dG(x)$ when the firm and the consumer are maximizing, our auctioneer focuses on minimizing the value of the endowment F.

Proposition 1. *Suppose that u is regular. If a price function $p^\star$ solves*

$$\min_p \int_0^1 p(x)dF(x) \text{ subject to } p \text{ being convex and } p \geq u, \tag{3.7}$$

then $p^\star$ is an equilibrium price function, and in particular satisfies (3.1) - (3.3) with any optimal solution G. Conversely, any price function satisfying (3.1) - (3.3) with some G is a solution to (3.7).

Proof. See Appendix A.3. □

Proposition 1 characterizes the equilibrium price function from Theorem 1 as a feasible price function that minimizes the value of the endowment of the consumer (the constraint $p \geq u$ is needed to obtain the particular normalization of the price function used in Theorem 1). It also provides a solution method for the Bayesian persuasion problem (2.1). First, solve (3.7) to find equilibrium prices. Then, find G to satisfy conditions (3.1) − (3.3). While this is always a valid approach under the regularity assumptions on u, solving (3.7) is non-trivial. We derive an alternative approach based on joint restrictions on (G, p) in the next section.

4 Structure of equilibrium prices and allocations

We first show how to construct supporting prices for the cases in which the solution to the persuasion problem is known and takes a simple form. We then characterize the general structure of supporting prices which allows us to apply Theorem 1 to solve non-trivial persuasion problems.

Corollary 1. *If there exists an affine function q such that $q(x) \geq u(x)$ for all x and $q(\mathbb{E}X) = u(\mathbb{E}X)$, then pooling (revealing nothing about X) is a solution to problem (2.1).*

In this case it is enough to take $p \equiv q$. Since p is affine, condition (3.2) is equivalent to equality of unconditional means of F and G. The distribution G that puts all mass on $\mathbb{E}X$ also satisfies conditions (3.1) and (3.3), and hence is optimal by Theorem 1. Corollary 1 applies not only when u is concave, but more generally when $-u$ has a supporting hyperplane at $\mathbb{E}X$.

Corollary 2. *If u is convex, then full revelation (always revealing X) is a solution to problem* (2.1)*.*

In this case we can take $p \equiv u$ and $G \equiv F$ for which conditions (3.1) – (3.3) hold trivially.

Supporting prices are also easy to find in persuasion problems where the Receiver chooses one of two actions. This has been solved using other methods (Gentzkow and Kamenica, 2015; Ivanov, 2015). For comparison, in Appendix A.4 we solve the two-action problem by constructing a piece-wise linear price function and applying Theorem 1.

In all previous examples the optimal G takes a simple form, and it is straightforward to use Theorem 1 to verify optimality. We now prove that conditions (3.1) – (3.3) impose tight joint restrictions on $(G,\, p)$ which greatly reduces the set of candidate solutions and supporting prices.

Proposition 2. *Suppose that F has full support and no mass points. Let p be a convex and continuous function that satisfies conditions* (3.1)–(3.3) *for some distribution G. Then, for any interval $[a, b] \subseteq [0, 1]$:*

(i) if p is strictly convex on $[a,\, b]$, then $p(x) = u(x)$ and $G(x) = F(x)$, for all $x \in [a, b]$; or,

(ii) if p is affine on $[a,\, b]$, and $[a,\, b]$ is a maximal interval on which p is affine, then $G(a) = F(a)$, $G(b) = F(b)$, $\int_a^b t\, dG(t) = \int_a^b t\, dF(t)$, and $p(c) = u(c)$ for at least one $c \in [a, b]$.

Furthermore, if u is regular, then there exists a coarsest partition $0 = x_0 < x_1 < ... < x_{n+1} = 1$ of $[0,\, 1]$ such that on each $[x_i,\, x_{i+1}]$ either (i) or (ii) holds.

Proof. See Appendix A.5.

The key property of the supporting price function can be understood in the context of the Persuasion Economy. In regions where the price function is strictly convex, the firm cannot engage in any non-trivial production as this would yield losses (given convexity of prices and the technological constraints). Thus $dF = dG$ (so that $dZ = 0$) in these regions which corresponds to full disclosure of the state. Non-trivial production can only take place in intervals where the price function is affine. On such intervals, the production of the firm is only constrained by the mean-preserving spread and no-free-disposal conditions. Figures 1 and 2 show an example u and F along with (G, p) that satisfy (3.1)–(3.3), and the structural properties given in Proposition 2.

Proposition 2 implies that it is enough to verify the mean-spread condition (3.3) separately for every maximal interval in which p is affine (the condition holds automatically conditional on realizations in intervals where p is strictly convex). Moreover, whenever $p(x) > u(x)$ for all $x \in [a,\, b]$, then p is piece-wise affine with at most one kink in $[a,\, b]$. Indeed, such $[a,\, b]$ can intersect at most two consecutive intervals $[x_i,\, x_{i+1}], [x_{i+1},\, x_{i+2}]$ because in every interval of the partition, p and u coincide at at least one point.

The conditions listed above often restrict the set of possible prices to a small class. Because the convex price function p can only coincide with u when u is convex, and p is piece-wise affine with

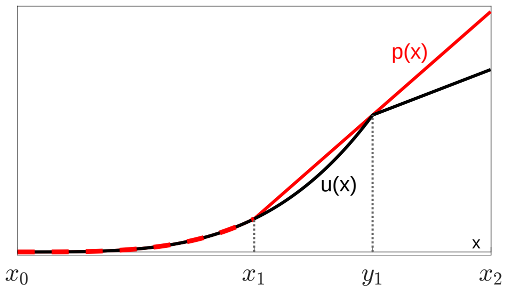

Figure 1: Utility $u(\cdot)$ and prices $p(\cdot)$.

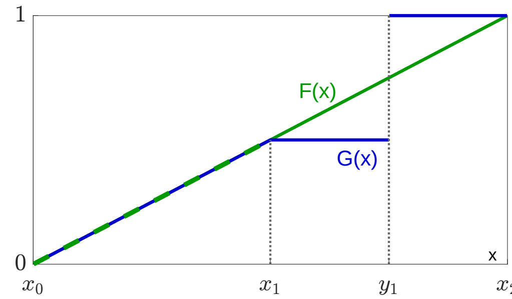

Figure 2: Prior CDF $F(\cdot)$ and CDF of posterior means $G(\cdot)$.

at most one kink between regions where $u \equiv p$, for relatively simple u, the set of potential price functions can be parameterized by a low-dimensional parameter. For any p in that set, points where $u \equiv p$ determine the support of G, and the supporting price function can be found by solving for the parameter which yields condition (3.3). We further aid the search for the solution in Section 5 where we provide sufficient conditions for when a monotone partitional signal is optimal. In Section 6, we illustrate the method by applying it to two persuasion problems.

5 Monotone partitional signals

Persuasion mechanisms that are seen in practice often only use *monotone partitional signals*: pooling, if present, is only between adjacent types. For example, many schools only release coarse information on student performance (Ostrovsky and Schwarz, 2010). Bond credit ratings also have a coarse structure (where very fine categories can be interpreted as full disclosure). These signal structures also appear in other models of communication: Crawford and Sobel (1982) show that all equilibria in their model feature monotone partitional signals.

Definition. A distribution of posterior means G is induced by a *monotone partitional signal* if there exists a finite partition of $[0, 1]$ into intervals $\{[x_i, x_{i+1}]\}_{i=1}^k$ such that for each i, either *(i)* $G \equiv F$ in $[x_i, x_{i+1}]$ (full revelation), or *(ii)* G puts all mass in $[x_i, x_{i+1}]$ on $\mathbb{E}[X \mid X \in [x_i, x_{i+1}]]$ (pooling).

In Section 8 we review papers that give various *sufficient* conditions for monotone partitional signals in related models of Bayesian persuasion. In our model, we prove that the following definition gives a sufficient and *necessary* condition on the objective function under which a monotone partitional signal is optimal regardless of the underlying prior distribution over states.

Definition 2. A function u is *affine-closed* if there do not exist $0 < x < y < 1$ and an affine function q such that:

1. $u(x) = q(x)$ and $u(y) = q(y)$;

2. $q(z) \geq u(z)$ for all $z \in [x, y]$;

3. $q(z) > u(z)$ for all $z \in \{w\} \cup (x - \epsilon,\, x) \cup (y,\, y + \epsilon)$ for some $\epsilon > 0$ and some $w \in (x,\, y)$.

To provide some intuition for the above definition, note first that adding an affine function to u does not change the optimal solution to our problem. Roughly speaking, an affine-closed function is defined by the property that $u+q$ has at most one local interior maximum for any affine function q. When this is not the case, that is, $u + q$ has two (or more) interior "peaks", we can find an affine function q' tangent to these peaks that would satisfy properties 1-3. Importantly, an affine-closed function can have local (one-sided) maxima at the endpoints. This intuition is precise when u itself is not affine in any interval. The definition is, however, more permissive when u is locally affine and can therefore have an interval of local maxima.

Convex and concave functions are always affine-closed. More complex examples of affine-closed functions are shown in Figures 3 and 4. In both cases there does not exist an affine function q with properties 1-3 listed in Definition 2. The affine function q_1 in Figure 3 does not satisfy property 1 because points of support must be interior (not 0 or 1). The affine function q_2 that supports u at x and y does not satisfy property 3, because there does not exist $w \in (x,\, y)$ such that $q_2(w) > u(w)$. Finally, the function in Figure 4 is affine-closed because the affine function q_3 cannot simultaneously satisfy properties 2 and 3, regardless of how we choose the support points x and y.

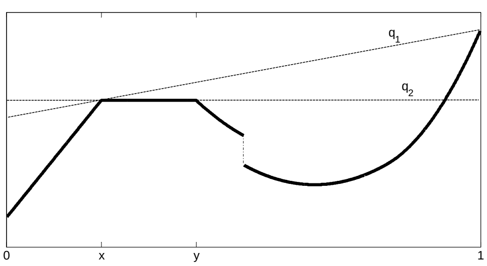

Figure 3: An affine-closed function (solid black line) and affine functions q_1 and q_2 (dotted lines)

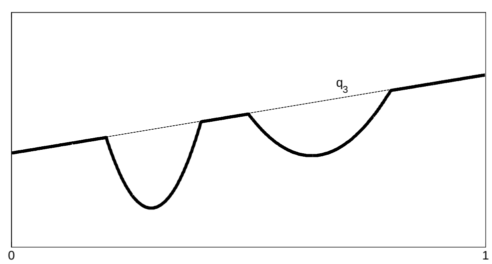

Figure 4: An affine-closed function (solid black line) and an affine function q_3 (dotted line)

Theorem 3. *Let u be regular. If u is affine-closed, then for any continuous and full-support prior F there exists an optimal solution G for problem (2.1) which is induced by a monotone partitional signal. Conversely, if u is not affine-closed, then there exists a (continuous) prior F such that no optimal G can be induced by a monotone partitional signal.*

Proof. See Appendix A.6.

In the proof, we use Theorem 2 to generate a solution G and a corresponding multiplier p. Starting from G, we construct a modified distribution of posteriors which is induced by a monotone partitional signal. Using Proposition 2 and the affine-closure property, we show that the multiplier p still supports the modified distribution, thus proving its optimality by Theorem 1. To prove the converse, we use the violation of affine-closure to construct a distribution F such that the optimal signal cannot have a monotone partitional structure.

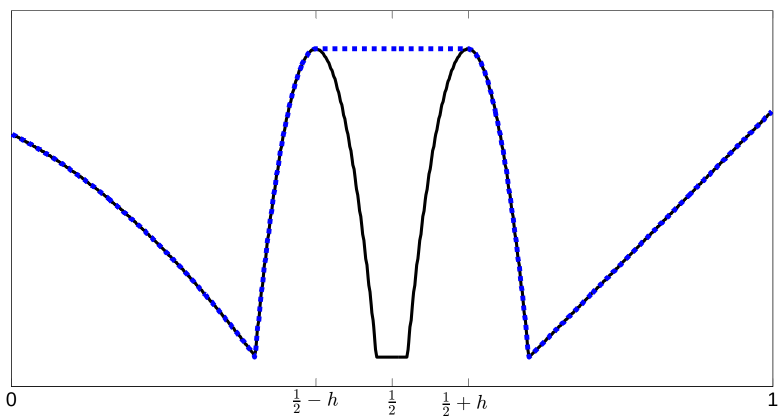

Figure 5: A non-affine-closed function (solid black line) and an affine-closed function (dotted blue line).

Further intuition behind the proof, and the importance of the affine-closure assumption can be understood using Figure 5. Consider the non-affine-closed function (black solid line). Suppose that F is the uniform distribution on $[0, 1]$. If $h \leq \frac{1}{4}$ in Figure 5, then by Theorem 1, the optimal posterior distribution of means has two atoms at the two peaks $\frac{1}{2} \pm h$ of the objective function (the multiplier p is a horizontal line tangent at the two peaks). Except for the non-generic case $h = \frac{1}{4}$, this distribution of posterior means cannot be induced by a monotone partitional signal. Consider instead the affine-closed function in Figure 5 (blue dotted line) as the objective. Although the same signal remains optimal, we can now modify it to obtain a monotone partition: We construct a signal which pools all realizations into one posterior mean $\frac{1}{2}$, and achieves the same payoff as the mixture between points $\frac{1}{2} - h$ and $\frac{1}{2} + h$. The affine closure property implies that the objective function coincides with the (locally affine) price function at the pooled mean $\frac{1}{2}$, allowing an optimal G to put an atom at $\frac{1}{2}$, in line with condition (3.1).

Continuity of F is needed for the first half of the theorem. Consider the extreme case where F puts all mass on two points. Then, there are only two distinct distributions G that are induced by monotone partitions: $G = F$ (full revelation) and $G = \delta_{\mathbb{E}[X]}$ (pooling at the prior mean). It is easy to construct (affine-closed) u for which neither of these is optimal.[9]

[9]However, for the case of discrete distributions, one can derive a version of Theorem 3 under a more permissive definition of a monotone partition that allows to "split" an atom into two "adjacent" signals.

The affine-closure property is distinct from the notion of a concave closure (as Figures 3 and 4 illustrate). Nevertheless, one can ask if there exists a well-defined notion of an affine closure $\hat{u}$ of the function u with the property that it produces the same value of the persuasion problem as u but the optimal signal is monotone partitional. The price function p from Theorem 1 satisfies all properties required of $\hat{u}$: it is affine-closed (because it is convex), and it yields the same value of the optimal persuasion problem as u (this follows from convexity of p and conditions (3.1) - (3.2)). The caveat is that p depends on the prior F. It is not possible to define affine closure $\hat{u}$ so that $\hat{u}$ is affine-closed and yields the same value of the persuasion problem as u for *every* F.

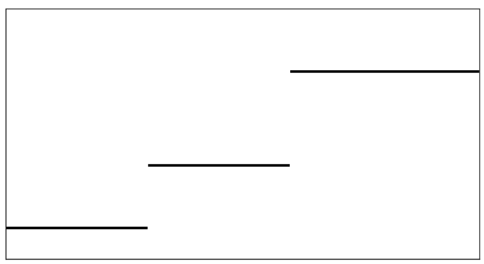

Figure 6: A non-affine-closed function

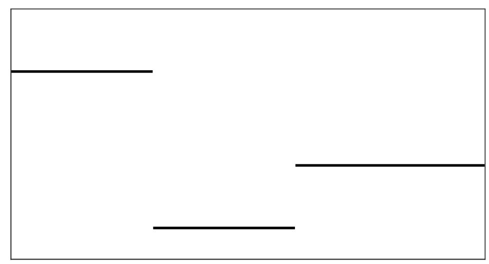

Figure 7: An affine-closed function

One natural example where the affine-closure assumption may not hold is when the Receiver chooses one of $n \geq 3$ actions. As noted in Gentzkow and Kamenica (2015), the optimal signal structure may fail to be monotone partitional in this case. Our method yields a complete characterization of cases when the existence of a monotone partitional signal is guaranteed. When the three actions are ordered as in Figure 6, u is not affine-closed. By Theorem 3, there exists a prior distribution for which no monotone partitional signal is optimal. On the other hand, when the actions are not ordered (both when the central action is the worst, as in Figure 7, and when it is the best), u is affine-closed and hence there always exists an optimal monotone partitional signal.

6 Applications

In this section, we show that the methods developed in the paper can be successfully applied to solve persuasion problems with large action and state spaces. Concurrent work by the first author uses our results in a model studying optimal pre-trade transparency of financial over-the-counter markets. In Duffie, Dworczak, and Zhu (2016), dealers trade with customers who are uninformed about the common cost of providing an asset, and hence face uncertainty over the prices quoted by dealers. Because of costly search for the best price, entry by customers is limited. A social planner decides how much information about the common cost to reveal prior to trading. The distribution of the cost is continuous, resulting in a continuum of states. Our Theorem 1 is used to solve for the social planner's optimal information revelation scheme. Below, we provide two additional applications of our methods.

6.1 Motivating through strategic disclosure

A principal wants to motivate an agent to exert costly effort in order to complete a project.[10] The project, if completed successfully, has value X to the principal. We assume that the agent receives a fraction $\beta \in (0, 1)$ of the value of the project. The agent chooses effort level $e \in [0, \bar{e}]$, with $\bar{e} \leq 1$, where e is interpreted as probability that the project will be successfully completed. Choosing effort level e has disutility $c(e) = e^{\alpha}$, where $\alpha > 1$.[11]

The value X is distributed according to a continuous full-support distribution F on $[\underline{x}, \bar{x}]$. The principal observes the realization x of X, but the agent does not. The principal commits to a disclosure scheme in order to maximize her profits.

Given the belief of the agent that the expected value of the project is y, the chosen level of effort is equal to

$$e^{\star}(y) = \min\left\{ \left(\frac{\beta y}{\alpha}\right)^{\frac{1}{\alpha-1}}, \bar{e} \right\}.$$

Let $\bar{y} = (\alpha/\beta)\bar{e}^{\alpha-1}$ be the smallest expected value of the project such that $e^{\star}(\bar{y}) = \bar{e}$, i.e. the agent exerts maximal effort. To make the analysis interesting, we assume that $\underline{x} < \bar{y} < \bar{x}$.

The value to the principal from inducing the belief y of expected value of the project is given by

$$u(y) = (1-\beta)e^{\star}(y)y.$$

It is easy to check that u is strictly convex in $[\underline{x}, \bar{y}]$, and affine in $[\bar{y}, \bar{x}]$. The shape of the function $u(\cdot)$ is depicted in Figure 8.

Proposition 3. *If $\mathbb{E}X \geq \bar{y}$, it is optimal to reveal no information. Otherwise, let $x^{\star}$ be defined by $\mathbb{E}\left[X \mid X \geq x^{\star}\right] = \bar{y}$. Then, it is optimal to disclose x whenever $x < x^{\star}$, and to reveal only that $x \geq x^{\star}$ if $x \geq x^{\star}$.*

Proof. In the first case, by Corollary 1, it is optimal not to reveal anything.

In the second case, consider the price function

$$p(x) = \begin{cases} u(x) & \text{if } x < x^{\star} \\ u(x^{\star}) + \Delta(x - x^{\star}) & \text{if } x \geq x^{\star} \end{cases},$$

where

$$\Delta \equiv \frac{u(\bar{y}) - u(x^{\star})}{\bar{y} - x^{\star}}.$$

[10] A similar application (with binary state space) is studied in Section 6.2 of Kamenica and Gentzkow (2009).

[11] All the results in this section continue to hold under the assumption that $(c')^{-1}(e) \cdot e$ is convex, without any particular choice of functional form.

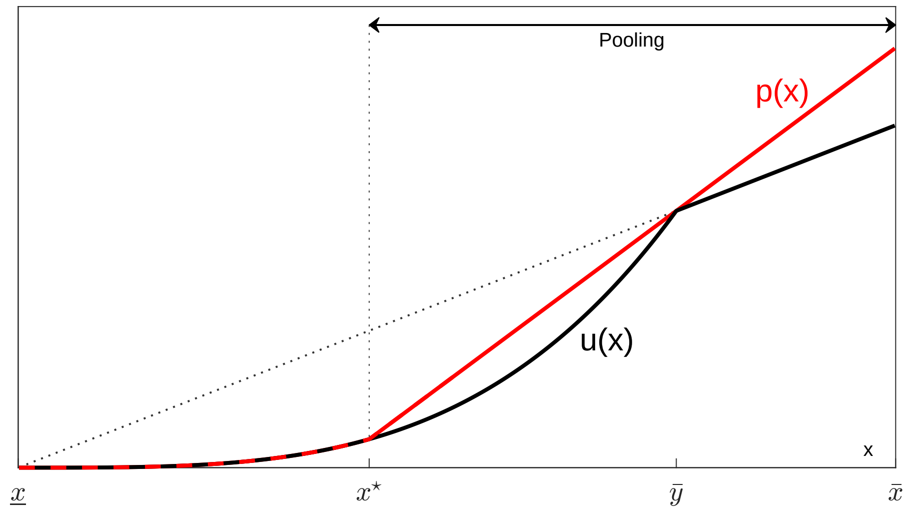

Figure 8: Motivating through strategic disclosure

Function p is convex. We then define the cdf G of posterior means as

$$G(x) = \begin{cases} F(x) & \text{if } x < x^\star \\ F(x^\star) + \mathbf{1}_{\{x \geq \bar{y}\}}(1 - F(x^\star)) & \text{if } x \geq x^\star \end{cases}.$$

That is, G coincides with F up to $x^\star$, and then puts an atom at $\bar{y} = \mathbb{E}\left[X \mid X \geq x^\star\right]$. Condition (3.1) holds because u and p coincide for $x \leq x^\star$ and $x = \bar{y}$. Condition (3.2) is satisfied because p is affine whenever $F \neq G$, and the conditional means of F and G are equal in that region. By the way we defined G, F is a mean-preserving spread of G, which verifies condition (3.3). By Theorem 1, G is optimal. □

Proposition 3 has the following economic interpretation. If the agent believes the value of the project to be high enough ex-ante ($\mathbb{E}X \geq \bar{y}$), he exerts maximal effort, so it is optimal for the principal to release no additional information. In the opposite case $\mathbb{E}X < \bar{y}$, the principal uses an “upper censorship” rule. She discloses the exact value of the project for low realizations but garbles the signal for high realizations by only informing the agent that the project deserves maximum effort (the realized value of X is above $x^\star$).

6.2 Investment recommendation

A risk-averse investor chooses how to divide her wealth w between a risk-free asset and a single risky asset, on which she can take a short or a long position. (A similar analysis is possible if only one of the two positions is available.) The amount invested in the risky asset either doubles in value or is entirely lost. To take any non-zero position in the risky asset, the investor must pay a fixed cost $c > 0$. The investor has prior belief $\frac{1}{2}$ that a long position will double her investment.

The investor consults a financial analyst, who has access to additional information about the payoff of the risky asset. Specifically, she knows the probability X that a long position will double the investment. (With probability $1 - X$ a short position will double the investment.) X is distributed according to F which is symmetric around the mean $\mathbb{E}[X] = \frac{1}{2}$ and admits a strictly positive density on $[0, 1]$. The analyst commits to a persuasion mechanism in order to influence the belief of the agent about the payoff of the risky asset. If the posterior belief is too close to the prior, because of the fixed cost the agent will optimally invest zero. The closer to 0 or 1 the belief, the more the agent is willing to invest.

We consider two different shapes of the analyst's utility u. In the first (Figure 9), as posterior belief approaches 0 and 1, the function is concave, meaning there are diminishing returns from inducing polarized beliefs. In the second (Figure 10), the function u is convex near 0 and 1. In both cases u is flat near $\frac{1}{2}$, where there is no investment. We provide a microfoundation of these shapes in Appendix A.7 assuming that the investor has CRRA preferences and the analyst's utility is proportional to the amount invested. The shape of u depends on the measure of relative risk aversion, with u becoming more convex as risk aversion increases. The linear boundary case is attained when the investor has log-utility. Similar shapes can also arise because of non-linearities in the analyst's fee structure.

In Proposition 4 we solve the analyst's persuasion problem under the assumption that

$$x_0 > \mathbb{E}\left[X \middle| X \leq \frac{1}{2}\right] \tag{6.1}$$

so that the analyst is, on average, sufficiently informed to always induce positive investment in the risky asset, when information is pooled in two signals (high and low). By symmetry of F and u, this implies $1 - x_0 < E[X|X \geq \frac{1}{2}]$.

Proposition 4. *In the concave case (Figure 9, with u formally defined in Appendix A.7), it is optimal to reveal whether $x < \frac{1}{2}$ or $x > \frac{1}{2}$. In the convex case (Figure 10, with u formally defined in Appendix A.7), there exist $a^\star < b^\star$ such that it is optimal to disclose x whenever $x \leq a^\star$ or $x \geq 1 - a^\star$, and to pool realizations (i) $x \in (a^\star, b^\star)$ at $x_0 = E[X|a^\star < X < b^\star]$; (ii) $x \in (1 - b^\star, 1 - a^\star)$ at $1 - x_0$; (iii) $x \in [b^\star, 1 - b^\star]$ at $\frac{1}{2}$.*

Proof. See Appendix A.7.

Proposition 4 can be interpreted as follows. If the agent is not highly risk averse (with the CRRA microfoundation, if the agent is less risk averse than log-utility; see Appendix A.7 for details), the analyst issues a binary recommendation "buy" or "sell", and the investor always invests a positive amount of wealth in the position recommended by the analyst. If the agent is

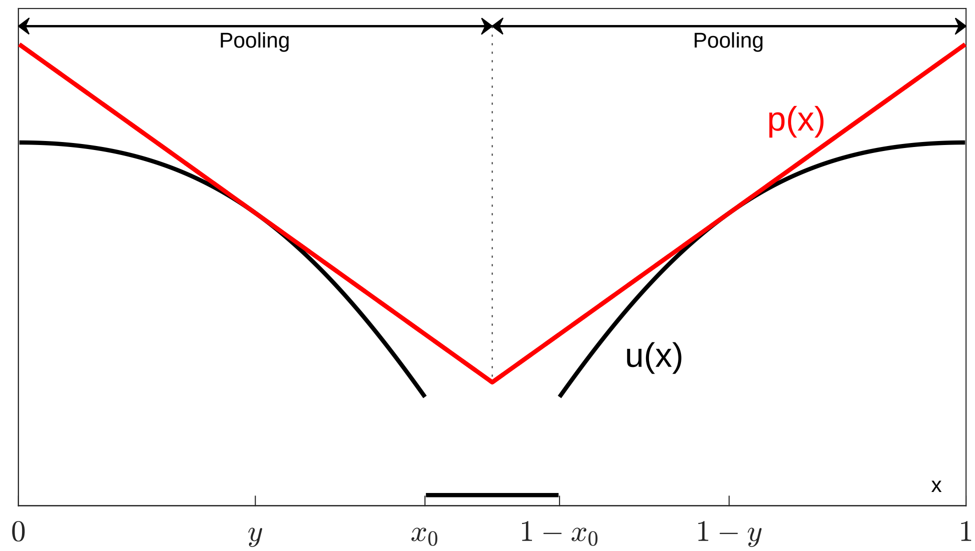

Figure 9: Portfolio recommendation: concave case

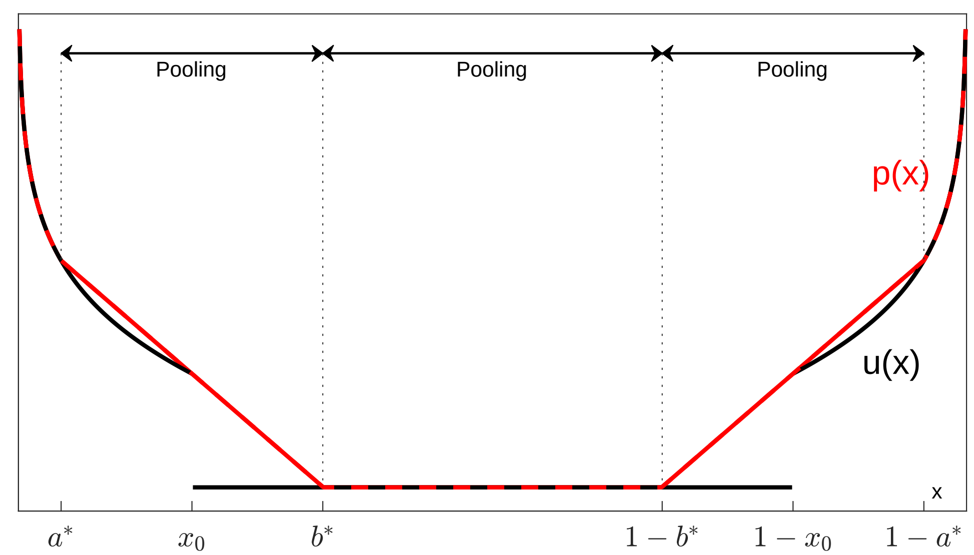

Figure 10: Portfolio recommendation: convex case

more risk averse, the recommendation of the analyst has a finer structure. If the analyst thinks the asset will go up (or down) with high probability, she provides full information about the assessed probability. If she is less confident, she issues a "weak recommendation" to either buy or sell. And if the realized x is close to the prior belief $\frac{1}{2}$, she does not recommend any of the positions ("hold"). In this last case, the agent refrains from investing.

7 Extensions

7.1 Competition in persuasion

In this subsection, we show that our methods extend to a setting where there are multiple Senders that have access to the same information. We adopt the competition in persuasion model of Gentzkow and Kamenica (2016) specialized to our setting. Formally, there are N Senders, and each Sender i maximizes the expectation of some function $u_i(x)$, where x is the mean of the posterior distribution (given the information revealed by all Senders). Senders choose the distributions of signals simultaneously, and can send arbitrarily correlated signals. We study Nash equilibria of this game.

By the results of Gentzkow and Kamenica (2016), a distribution H over posterior means is an equilibrium distribution of joint information revealed by all Senders if and only if it is *unimprovable* – no Sender i wants to reveal more information given H. This result allows us to provide a characterization of all equilibrium distributions under a mild regularity condition. We say that $v : [0,1] \to \mathbb{R}$ is *a convex translation* of $u : [0,1] \to \mathbb{R}$ if $v - u$ is convex.

Theorem 4. *If there exist: a cumulative distribution function H, a function $\hat{u}_i$ that is a convex translation of u_i, and a convex function $p_i : [0,1] \to \mathbb{R}$, $p_i \geq \hat{u}_i$, for each $i \in N$, such that*

$$\text{supp}(H) \subseteq \{x \in [0,1] : \hat{u}_i(x) = p_i(x)\}, \forall i \in N, \tag{7.1}$$

$$\int_0^1 p_i(x)dH(x) = \int_0^1 p_i(x)dF(x), \forall i \in N, \text{ and} \tag{7.2}$$

$$F \text{ is a mean-preserving spread of } H, \tag{7.3}$$

then H is an equilibrium distribution of posterior means.

Conversely, if each u_i is continuous and regular, then for any equilibrium distribution H, there exist convex p_i, and convex translations $\hat{u}_i$ of u_i, for each $i \in N$, such that conditions (7.1) – (7.3) hold.

We sketch the proof in Appendix A.8. Theorem 4 gives a number of straightforward corollaries. First, full disclosure is always an equilibrium: there always exists a convex translation of utility function $\hat{u}_i$ for each Sender i that is convex (we can then take $p_i = \hat{u}_i$), given the assumption that each u_i is regular and continuous. Second, if all utility functions are concave, then all distributions are equilibrium distributions – indeed, if u_i is concave, then $\hat{u}_i \equiv 0$ is a convex translation of u_i, and we can take $p_i \equiv 0$. Third, the set of equilibrium distributions shrinks when the utility of any Sender undergoes a convex translation – this reduces the set of convex translations of utility functions that can support a given candidate equilibrium distribution H. Fourth, Theorem 4 leads to an economically-meaningful algorithm for generating the set of equilibrium distributions: *(1)* Solve the individual optimization problem for all Senders that are more risk-loving (in the order induced by convex translations) than Sender i, and denote the set of obtained solutions by $\mathcal{H}_i$, for any $i \in N$; *(2)* The set of equilibrium distributions is exactly the intersection $\bigcap_{i \in N} \mathcal{H}_i$.

We can use Theorem 4 to solve a multi-Sender version of the application in Section 6.1. Suppose now that there are $N \geq 2$ Senders that observe X and compete to persuade the agent. We assume that utility u_i of each Sender i is: *(1)* strictly convex on $[0, c_i]$ for some $c_i \in (0, 1)$; *(2)* affine on $[c_i, 1]$; *(3)* continuous, regular, and not globally convex. That is, each u_i takes the shape analogous to the shape of u in Figure 8, but the locations c_i of the kink may be different across Senders. Without loss of generality, order the Senders so that $c_1 \geq c_2 \geq ... \geq c_N$. To limit the number of cases, assume that the prior mean satisfies $\mathbb{E}_F[X] < c_N$.

Proposition 5. *H is an equilibrium distribution if and only if H is feasible, fully reveals the state below some $a \leq c_1$, and has no mass on (a, c_1). In particular, the least informative equilibrium is the distribution that would be optimal for Sender 1 if she were the only Sender.*

Proof. See Appendix A.9. □

7.2 Concavification versus prices in a general Bayesian persuasion problem

In this subsection, we provide an analog of Theorem 1 in the general Bayesian persuasion problem (Kamenica and Gentzkow, 2011), where the Sender's utility depends on the distribution of posterior beliefs (and not necessarily on posterior means only). We then illustrate the connection between the concavification approach and our approach based on prices.

Let Θ be a compact metric space, and suppose that the Sender's utility $V : \Delta(\Theta) \to \mathbb{R}$ is upper-semi continuous. Let $\mu_0 \in \Delta(\Theta)$ be the prior distribution. The general Bayesian persuasion problem is

$$\max_{\tau \in \Delta(\Delta(\Theta))} \int_{\Delta(\Theta)} V(\mu) d\tau(\mu) \tag{7.4}$$

subject to the Bayes-plausibility constraint

$$\int_{\Delta(\Theta)} \mu d\tau(\mu) = \mu_0. \tag{7.5}$$

Theorem 5. *Suppose that there exists $\tau^\star \in \Delta(\Delta(\Theta))$ and a convex functional $P : \Delta(\Theta) \to \mathbb{R}$, such that $P \geq V$, and*

$$supp(\tau^\star) \subseteq \{\mu \in \Delta(\Theta) : V(\mu) = P(\mu)\}, \tag{7.6}$$

$$\int_{\Delta(\Theta)} P(\mu) d\tau^\star(\mu) = \int_{\Theta} P(\delta_\theta) d\mu_0(\theta), \tag{7.7}$$

$$\int_{\Delta(\Theta)} \mu\tau^\star(\mu) = \mu_0, \tag{7.8}$$

where δ_θ is a measure that puts all mass on $\{\theta\}$. Then, $\tau^\star$ is a solution to problem (7.4) – (7.5).

Proof. Identical to the proof of Theorem 1 (omitted). □

Theorem 5 is analogous to Theorem 1 except that the convex price functional P is defined on beliefs rather than on posterior means – this is natural given that in the general Bayesian persuasion problem the preferences of the Sender depend on the entire belief profile.

In the competitive equilibrium interpretation, the consumer is endowed with the prior μ_0 whose market value under prices P is given by $\int_\Theta P(\delta_\theta) d\mu_0(\theta)$. The consumer purchases posterior beliefs μ at prices $P(\mu)$ to maximize her quasi-linear utility – condition (7.6) can be equivalently written as $\text{supp}(\tau^\star) \subseteq \text{argmax}_{\mu\in\Delta(\Theta)}\{V(\mu) - P(\mu)\}$. Given the technological constraint on what posterior beliefs can be produced from the endowment μ_0, the firm breaks even (condition 7.7) and the markets clear (condition 7.8).

When the function V is convex,[12] the supporting prices are given by $P(\mu) = \int_\Theta V(\delta_\theta) d\mu(\theta)$. P is affine, and coincides with V on the set of extreme points of $\Delta(\Theta)$. Thus, $P \geq V$, and the three conditions of Theorem 5 follow immediately with $\tau^\star$ corresponding to full disclosure. When the function V is concave, let $\Phi_{\mu_0}(\mu)$ be the (affine) function whose graph is the supporting hyperplane of V at μ_0. With $P(\mu) = \Phi_{\mu_0}(\mu)$, $P \geq V$, and conditions (7.6) – (7.8) hold trivially when $\tau^\star(\mu_0) = 1$, i.e. $\tau^\star$ corresponds to no disclosure. Finally, suppose that the preferences of the Sender depend only on the posterior mean, $V(\mu) = u(\mathbb{E}_\mu(\theta))$, and let (G, p) satisfy conditions (3.1) – (3.3). Then, take $P(\mu) = p(\mathbb{E}_\mu(\theta))$. Conditions (7.6) – (7.8) with $\tau^\star$ corresponding to G follow from the corresponding conditions of Theorem 1.

Theorem 1 allows us to draw a connection between our approach and the concavification approach of Kamenica and Gentzkow (2011).

Claim 1. *Suppose that $P^\star$ is the convex price functional satisfying the conditions of Theorem 5 (under the prior μ_0 and with some optimal distribution $\tau^\star$). Then*

$$co(V)(\mu_0) = \min\left\{\int_\Theta P(\delta_\theta) d\mu_0(\theta) |\, P : \Delta(\Theta) \to \mathbb{R},\ P \text{ is convex},\ P \geq V\right\}. \tag{7.9}$$

[12] In fact, a weaker property than convexity is sufficient: For any $\mu \in \Delta(\Theta)$, $V(\mu) \leq \int_\Theta V(\delta_\theta) d\mu(\theta)$.

Moreover, the minimum is attained at $P = P^\star$:

$$co(V)(\mu_0) = \int_\Theta P^\star(\delta_\theta) d\mu_0(\theta). \tag{7.10}$$

Proof. See Appendix A.10. □

Claim 1 shows the dual relationship between the concave closure of V (which is the value function for the problem (7.4) – (7.5)) and the convex price functional P from Theorem 5. The value of the persuasion problem is equal to the value of the endowment at the equilibrium prices. Equation (7.10) can be interpreted in the following way. If V is replaced with its concave closure co(V) (which is pointwise greater) as the objective function of the Sender, it is optimal to reveal *nothing* and the value of the persuasion problem stays the same. Analogously, if V is replaced with $P^\star$ (which is pointwise greater) as the objective function of the Sender, it is optimal to reveal *everything* and the value of the persuasion problem stays the same.[13] In this sense, $P^\star$ can be seen as a (prior-dependent) "convex closure" of V.

8 Related literature

The model of Bayesian persuasion where the Sender's preferences only depend on the posterior mean has received some attention in the literature. Along with their analysis of the general model, Kamenica and Gentzkow (2011) apply the concavification approach to the same setting as ours. The concavification of the value function over posterior means reveals whether the Sender benefits from persuasion, but does not explicitly characterize the value of the problem or the optimal signal. In contrast, our approach directly establishes the structure of the optimal signal. Gentzkow and Kamenica (2015) focus on the same setting as ours and characterize the set of feasible distributions over posterior means using a graphical method. Their method is then used to solve simple persuasion problems in which the Receiver chooses between two or three actions, and hence the Sender's preferences over posterior means are a step function. In contrast, our techniques apply to an almost arbitrary objective function of the Sender, and in particular allow for a continuum of both states and actions.

Kolotilin (2017) provides a strong duality theorem for a model of persuasion in which the Receiver is privately informed and chooses between two actions.[14] He also studies a case of his model that is mathematically equivalent to our setting.[15] Our Theorems 1 and 2 bear similarity to the strong duality result of Kolotilin but are different across two important dimensions. First, due to a different representation of the mean-preserving spread constraint, the Lagrange multiplier

[13]The last claim follows from conditions (7.6) - (7.7).

[14] Related models of persuasion with a privately informed receiver (but not using duality techniques) are considered by Kolotilin, Li, Mylovanov, and Zapechelnyuk (2017) and Guo and Shmaya (2017).

[15]We impose much weaker regularity conditions, e.g., allow for atoms in F, and do not require u to be differentiable which is important for the applications in Section 6.

derived by Kolotilin is two-dimensional. Our price function supporting the optimal solution is one-dimensional.[16] Together with the tight characterization of prices provided in Section 4, this makes our method significantly easier to apply. Second, the price function permits the interpretation of the persuasion problem as a Walrasian equilibrium. This provides useful intuition about the structure of the solution and allows us to extend the method to competition in persuasion and to a general Bayesian persuasion problem.

With regard to our Section 5, other papers have found sufficient conditions for monotone partitional signals in related models. In the general setting of Kamenica and Gentzkow (2011), Mensch (2017) gives a sufficient condition based on supermodularity of the Sender's and Receiver's preferences over action and state. Our condition is not comparable to his, since we take a reduced-form approach where (Sender's) preference is over posterior means directly. Ivanov (2015) studies a setting in which the preferences of the Sender depend not only on the posterior mean but also on the order of posterior means induced by a signal. Such a setting captures communication problems more general than Bayesian persuasion. Ivanov provides a sufficient condition for optimality of a monotone partitional signal in his setting. However, that condition becomes trivial when projected to our setting – it says that the objective function is convex. Finally, in a setting analogous to ours, Kolotilin (2017) provides a characterization of *interval revelation schemes*. Because interval revelation schemes, as defined by Kolotilin (2017), belong to the class of monotone partitional signals, his condition is sufficient. We give a condition that is both necessary and sufficient under milder regularity conditions on u.

Daskalakis, Deckelbaum, and Tzamos (2017) study the problem of finding a profit-maximizing mechanism for a multi-product monopolist. Their primal problem involves optimizing over convex functions, and they show that a suitable dual formulation obtains using mean-preserving spreads; our setting is specular, with mean-preserving spreads in the primal and convex functions in the dual. The approach we use to prove strong duality (Theorem 2) is however very different: Daskalakis et al. use methods from optimal transport, whereas we use results from infinite-dimensional linear programming. As Daskalakis et al. note, it is often difficult to establish existence of interior points in the feasible set, which is necessary for strong duality; we bypass this difficulty by studying the limit of a sequence of perturbed problems where the interior is non-empty. It is an interesting open question whether optimal transport methods can be applied to persuasion problems.

[16] While the proof of Theorem 2 uses duality techniques, our price function is not equal to the Lagrange multiplier; see the proof in Appendix A.2.

References

Aumann, Robert J and Michael Maschler. *Repeated games with incomplete information*. MIT press, 1995.

Blackwell, David. Equivalent Comparisons of Experiments. *The Annals of Mathematical Statistics*, 24(2):265–272, 1953.

Bonnans, J. F. and Alexander Shapiro. *Perturbation analysis of optimization problems*. Springer, 2000.

Calzolari, Giacomo and Alessandro Pavan. On the optimality of privacy in sequential contracting. *Journal of Economic Theory*, 130(1):168–204, September 2006.

Crawford, Vincent P and Joel Sobel. Strategic Information Transmission. *Econometrica*, 50(6): 1431–51, November 1982.

Daskalakis, Constantinos, Alan Deckelbaum, and Christos Tzamos. Strong Duality for a Multiple-Good Monopolist. *Econometrica*, 85(3):735–767, 2017.

Dentcheva, Darinka and Andrzej Ruszczynski. Optimization with stochastic dominance constraints. *SIAM Journal on Optimization*, 14(2):548–566, 2003.

Duffie, Darrell, Piotr Dworczak, and Haoxiang Zhu. Benchmarks in search markets. *Journal of Finance*, forthcoming, 2016.

Gentzkow, Matthew and Emir Kamenica. A Rothschild-Stiglitz approach to Bayesian persuasion. *American Economic Review Papers & Proceedings*, 2015.

Gentzkow, Matthew and Emir Kamenica. Competition in persuasion. *The Review of Economic Studies*, 84(1):300–322, 2016.

Gul, Faruk and Ennio Stacchetti. Walrasian Equilibrium with Gross Substitutes. *Journal of Economic Theory*, 87(1):95–124, 1999.

Guo, Yingni and Eran Shmaya. The Interval Structure of Optimal Disclosure. 2017.

Ivanov, Maxim. Optimal Signals in Bayesian Persuasion Mechanisms with Ranking. 2015.

Kamenica, Emir and Matthew Gentzkow. Bayesian Persuasion. *NBER Working Paper No. 15540*, 2009.

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

Kolotilin, Anton. Optimal Information Disclosure: A Linear Programming Approach. *Theoretical Economics*, forthcoming, 2017.

Kolotilin, Anton, Ming Li, Tymofiy Mylovanov, and Andriy Zapechelnyuk. Persuasion of a Privately Informed Receiver. *Econometrica*, Forthcoming, 2017.

Mensch, Jeffrey. Monotone persuasion. 2017.

Müller, A and D Stoyan. *Comparison Methods for Stochastic Models and Risks*. 2002.

Ostrovsky, Michael and Michael Schwarz. Information Disclosure and Unraveling in Matching Markets. *American Economic Journal: Microeconomics*, 2(2):34, 2010.

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

A Proofs and additional material

A.1 Proof of Theorem 1[17]

Let (G, p) satisfy conditions (3.1)–(3.3). To show that G is a solution to the Sender's problem, it is enough to show that $\int_0^1 u(x)dG(x) \geq \int_0^1 u(x)dH(x)$, for any H such that F is a mean-preserving spread of H.

By (3.1),

$$\int_0^1 (u(x) - p(x))dG(x) \geq \int_0^1 (u(x) - p(x))dH(x).$$

Rearranging,

$$\begin{aligned}\int_0^1 u(x)dG(x) - \int_0^1 u(x)dH(x) &\geq \int_0^1 p(x)dG(x) - \int_0^1 p(x)dH(x) \\ &\overset{(1)}{=} \int_0^1 p(x)dF(x) - \int_0^1 p(x)dH(x) \\ &\overset{(2)}{\geq} \int_0^1 p(x)dF(x) - \int_0^1 p(x)dF(x) = 0\end{aligned}$$

where equality (1) follows from (3.2) and inequality (2) holds because $-p(x)$ is concave, and F is a mean-preserving spread of H, by assumption. Therefore $\int_0^1 u(x)dG(x) \geq \int_0^1 u(x)dH(x)$. $\square$

A.2 Proof of Theorem 2

We prove the theorem in three steps. In the first step, we make the additional assumption that u is continuous, and study a perturbed problem. The perturbation allows us to show that a generalized Slater condition holds, which results in existence of the appropriate Lagrange multiplier. In the second step, we show that the solution to the perturbed problem provides the correct approximation of the solution to the unperturbed problem. By taking the limit, we obtain the statement of Theorem 2 for the special case of a continuous u. In the third step, we relax the assumption that u is continuous, again using an approximation approach. We relegate some technical lemmas to Appendix A.11.

We make two preliminary observations. First, a solution to the Sender's problem always exists because the objective function is upper semi-continuous, and the set of feasible points is compact in the weak* topology. Second, to show that a supporting price function p exists for all optimal solutions, it is enough to prove that it exists for one solution. This is because if p supports solution G, then is also supports any other solution G'. To see this, consider $H = G'$ in

[17] We thank Shota Ichihashi for showing us this direct proof. Our previous proof employed duality techniques from the literature on optimization with stochastic dominance constraints, which we use in the proof of Theorem 2.

the proof of Theorem 1. If G' is optimal, then all inequalities must hold as equalities, and thus $(G',\, p)$ satisfy conditions (3.1) – (3.2). Condition (3.3) holds because G' is feasible.

Step 1. In the first step, we use the proof technique developed by Dentcheva and Ruszczynski (2003). Dentcheva and Ruszczynski provide a duality theory for optimization problems with stochastic dominance constraints. They study a case where the constraint takes the form of second-order stochastic dominance. Our constraint additionally incorporates equality of unconditional means, resulting in a mean-preserving spread condition.

Assume that u is continuous. Consider the following perturbed problem, where the mean-preserving spread condition is only imposed on the interval $[\epsilon,\, 1-\epsilon]$, instead of on $[0,\, 1]$.

$$\max_{G\in\Delta([0,\,1])} \int_0^1 u(x)dG(x) \tag{A.1}$$

$$\text{s.t.} \int_0^x F(t)dt \geq \int_0^x G(t)dt \text{ for all } x \in [\epsilon,\, 1-\epsilon], \tag{A.2}$$

$$\int_x^1 G(t)dt \geq \int_x^1 F(t)dt \text{ for all } x \in [\epsilon,\, 1-\epsilon]. \tag{A.3}$$

Conditions (A.2) and (A.3) enforced on the entire interval $[0,\, 1]$ would be jointly equivalent to F being a mean-preserving spread of G. Note that although the perturbed problem (A.1) only requires the mean-preserving spread condition on $[\epsilon,\, 1-\epsilon]$, all distributions are defined on $[0,\, 1]$. We take $\epsilon > 0$ small enough so that $\epsilon < \mathbb{E}[F] < 1-\epsilon$.

Let $\mathcal{C}_\epsilon := \mathcal{C}([\epsilon,\, 1-\epsilon])$ denote the space of continuous functions on $[\epsilon,\, 1-\epsilon]$. Define K_ϵ as the cone of continuous non-negative functions on $[\epsilon,\, 1-\epsilon]$, that is,

$$K_\epsilon := \{g \in \mathcal{C}_\epsilon :\, g(x) \geq 0,\, \forall x \in [\epsilon,\, 1-\epsilon]\}.$$

Let $\Phi :\, \Delta([0,\, 1]) \to \mathcal{C}_\epsilon \times \mathcal{C}_\epsilon$ be defined by

$$(\Phi(G))_i(x) = \begin{cases} \int_0^x F(t)dt - \int_0^x G(t)dt & \text{if } i = 1, \\ \int_x^1 G(t)dt - \int_x^1 F(t)dt & \text{if } i = 2, \end{cases}$$

for any $x \in [\epsilon,\, 1-\epsilon]$. Conditions (A.2) and (A.3) are now equivalent to $\Phi(G) \in K_\epsilon \times K_\epsilon$.

The operator Φ is concave with respect to the product cone $K_\epsilon \times K_\epsilon$. By the Riesz representation theorem, the space dual to $\mathcal{C}_\epsilon$ is the space $\mathbf{rca}([\epsilon,\, 1-\epsilon])$ of regular countably additive measures on $[\epsilon,\, 1-\epsilon]$ having finite variation. We define a Lagrangian $\Lambda :\, \Delta([0,\, 1]) \times \mathbf{rca}([\epsilon,\, 1-\epsilon]) \times \mathbf{rca}([\epsilon,\, 1-\epsilon]) \to \mathbb{R}$,

$$\Lambda(G,\, \mu_1,\, \mu_2) = \int_0^1 u(x)dG(x) + \int_\epsilon^{1-\epsilon} (\Phi(G))_1(x)d\mu_1(x) + \int_\epsilon^{1-\epsilon} (\Phi(G))_2(x)d\mu_2(x).$$

We now show that the generalized Slater condition holds for the problem (A.1). By Bonnans and Shapiro (2000) (Proposition 2.106), it is enough to show that there exists $\tilde{G} \in \Delta([0,\, 1])$ such that

$$\Phi(\tilde{G}) \in \mathrm{int}(K_\epsilon \times K_\epsilon). \tag{A.4}$$

That is, we have to find a distribution (cdf) $\tilde{G}$ supported on $[0,\, 1]$ such that $\Phi(\tilde{G})$ is a Cartesian product of two functions that are both in K_ϵ, and are bounded away from zero on $[\epsilon,\, 1-\epsilon]$. Let $\tilde{G}(x) = \mathbf{1}_{\{x \geq \mathbb{E}[F]\}}$. The function $(\Phi(\tilde{G}))_i(x)$, for $i = 1,\, 2$, is equal to 0 at $x = 0$ and $x = 1$. Since 0 and 1 are in the support of F, $(\Phi(\tilde{G}))_i(x)$ is strictly positive in the interior of $[0, 1]$. Therefore, using the structure of $\tilde{G}$, it must be bounded away from zero on $[\epsilon,\, 1-\epsilon]$. Thus, the generalized Slater condition holds.

By Bonnans and Shapiro (2000) (Theorem 3.4), we conclude that if G is a solution to the problem (A.1) (we have already argued that a solution exists), then there exist non-negative measures $\mu_1^\star,\, \mu_2^\star \in \mathbf{rca}([\epsilon,\, 1-\epsilon])$ such that

$$\Lambda(G,\, \mu_1^\star,\, \mu_2^\star) = \max_{\hat{G} \in \Delta([0,\, 1])} \Lambda(\hat{G},\, \mu_1^\star,\, \mu_2^\star), \tag{A.5}$$

and

$$\int_\epsilon^{1-\epsilon} (\Phi(G))_i(x) d\mu_i^\star(x) = 0,\; i = 1,\, 2. \tag{A.6}$$

Using an argument analogous to the one used by Dentcheva and Ruszczynski (2003), we can associate with each measure $\mu_i^\star$ a function $p_i^\star : [0,\, 1] \to \mathbb{R}$,

$$p_1^\star(x) = \begin{cases} \int_x^{1-\epsilon} \mu_1^\star([\tau,\, 1-\epsilon]) d\tau, & x < 1-\epsilon \\ 0, & x \geq 1-\epsilon, \end{cases}$$

$$p_2^\star(x) = \begin{cases} 0, & x < \epsilon \\ \int_\epsilon^x \mu_2^\star([\epsilon,\, \tau]) d\tau, & x \geq \epsilon, \end{cases}$$

where each $\mu_i^\star$ is extended to $[0, 1]$ by putting zero mass beyond the interval $[\epsilon,\, 1-\epsilon]$. By the properties of $\mu_i^\star$, the function $p_1^\star$ is non-increasing and convex, and $p_2^\star$ is non-decreasing and convex.

We have (by changing the order of integration)

$$\int_\epsilon^{1-\epsilon} \left(\int_0^x G(t) dt \right) d\mu_1^\star(x) = \int_0^1 \left(\int_\epsilon^{1-\epsilon} \mathbf{1}_{\{t \leq x\}} d\mu_1^\star(x) \right) G(t) dt = \int_0^1 \mu_1^\star([t,\, 1-\epsilon]) G(t) dt.$$

Using the definition of $p_i^\star$, we can write

$$\int_\epsilon^{1-\epsilon} \left(\int_0^x G(t)dt \right) d\mu_1^\star(x) = -\int_0^1 G(t)dp_1^\star(t).$$

Similarly, we have

$$\int_\epsilon^{1-\epsilon} \left(\int_x^1 G(t)dt \right) d\mu_2^\star(x) = \int_0^1 G(t)dp_2^\star(t).$$

Using integration by parts, we get

$$\int_0^1 G(t)dp_i^\star(t) = p_i^\star(1) - \int_0^1 p_i^\star(x)dG(x).$$

Therefore, the complementary-slackness condition (A.6) becomes

$$\int_0^1 p_i^\star(x)dG(x) = \int_0^1 p_i^\star(x)dF(x),\ i = 1,\ 2.$$

Finally, define $p(x) = p_1^\star(x) + p_2^\star(x)$. Then, $p(x)$ is convex, and obviously satisfies

$$\int_0^1 p(x)dG(x) = \int_0^1 p(x)dF(x). \tag{A.7}$$

Finally, condition (A.5) implies that

$$G \in \text{argmax}_G \left\{ \int_0^1 u(x)dG(x) - \int_0^1 p(x)dG(x) \right\}. \tag{A.8}$$

We can always add a constant to p without changing any of it properties. Because of property (A.8), we can normalize p so that condition (3.1) holds. That is, $p \geq u$ and

$$\text{supp}(G) \subseteq \{x \in [0,\ 1] :\ u(x) = p(x)\}. \tag{A.9}$$

Therefore, for the perturbed problem, we have shown that there exists a convex $p :\ [0,\ 1] \to \mathbb{R}$ such that conditions (A.7) and (A.9) both hold.

Step 2. In this step, we show that we can take the limit of the perturbed problems from Step 1, and obtain a solution to the original (unperturbed) problem (still maintaining the assumption that u is continuous).

Consider a sequence of problems (A.1) - (A.3) defined by taking $\epsilon = 1/n$ for $n = 1, 2, \ldots$. We obtain a sequence $(G_n,\ p_n)$ of pairs satisfying conditions (A.7), (A.9), as well as (A.2) and (A.3).

We will first show a simplified version of the proof using the following assumption.

Assumption 1. The sequence of functions (p_n) converges uniformly on $[0, 1]$ to some convex function p.

In Appendix A.11, we present a full version of the proof without Assumption 1. This adds technical complications because the sequence p_n can potentially explode near the endpoints of the interval $[0, 1]$. In Appendix A.11, we show that we can still establish the result by appropriately modifying the functions p_n, and using the special structure of the problem.

Under Assumption 1, we have a well defined limit p of the sequence p_n. The sequence (G_n), seen as a sequence of probability measures, lives in a compact set (in the weak* topology). Because the space of measures is metrizable, compactness is equivalent to sequential compactness, and thus we can choose a converging subsequence. Without loss of generality, G_n converges in the weak* topology to some distribution $G \in \Delta([0, 1])$.

We have thus defined the limiting pair (G, p). We want to prove that (G, p) satisfies conditions (3.1)–(3.3) on $[0, 1]$.

First, as $n \to \infty$, conditions (A.2) and (A.3) imply that F is a mean-preserving spread of G on the entire interval $[0, 1]$. This establishes condition (3.3).

Second, we note that[18]

$$\mathrm{supp}(G) \subseteq \limsup_n \{\mathrm{supp}(G_n)\}.$$

Given that condition (A.9) holds for each n, and because p_n converges to p,

$$\limsup_n \{\mathrm{supp}(G_n)\} \subseteq \limsup_n \{x \in [0, 1] : u(x) = p_n(x)\} \subseteq \{x \in [0, 1] : u(x) = p(x)\}.$$

We conclude that

$$\mathrm{supp}(G) \subseteq \{x \in [0, 1] : u(x) = p(x)\}, \tag{A.10}$$

establishing condition (3.1).

Third, we will show that (3.2) holds. We will argue that $\int_0^1 p_n(x) dG_n(x) \to \int_0^1 p(x) dG(x)$. To see this, note that

$$\int_0^1 p_n dG_n - \int_0^1 p dG = \int_0^1 (p_n - p) dG_n - \int_0^1 p(dG - dG_n).$$

The second integral converges to zero by definition of convergence of G_n to G in the weak*

[18]The lim sup of a sequence of sets A_n is defined as

$$\limsup_n A_n = \{x : \exists (x_n)_n \text{ s.t. } x_n \in A_n, \forall n, \text{ and } x_n \to x\}.$$

topology. For the first integral, we have

$$\int_0^1 (p_n - p)dG_n \leq \sup_{x \in [0,\,1]} \{|p_n(x) - p(x)|\}$$

which converges to zero because p_n converges to p uniformly on [0, 1], by Assumption 1. Similarly, we have

$$\lim_n \int_0^1 p_n dF = \int_0^1 p dF.$$

Combining these two results with condition (A.7), we get

$$\int_0^1 p(x)dF(x) = \int_0^1 p(x)dG(x),$$

which is what we wanted to prove.

This finishes the proof of Step 2, i.e., we have shown that for a continuous u conditions (3.1) - (3.3) hold with $(G,\, p)$. By Theorem 1, G is the optimal solution.

Step 3. We now prove Theorem 2 without the additional assumption that u is continuous. As stated in the regularity assumption (Definition 1), u has finitely many one-sided jump discontinuities at y_1, ..., $y_k \in (0, 1)$.

First, we construct a continuous approximation of u. Fix $\epsilon > 0$. We only modify the function u in an ϵ-neighborhood of each y_i. For small enough ϵ, these neighborhoods are disjoint. Take any $i = 1$, ..., k, and suppose without loss of generality that $\lim_{y \uparrow y_i} u(y) < u(y_i) = \lim_{y \downarrow y_i} u(y)$. We denote by $\bar{u}_\epsilon$ the continuous function constructed by replacing u in each such neighborhood $[y_i - \epsilon,\, y_i]$[19] with an affine majorant described above.

Because $\bar{u}_\epsilon$ is a continuous function with a bounded slope, by Steps 1 and 2, there exists an optimal distribution G_ϵ and a Lagrange multiplier p_ϵ such that $(G_\epsilon,\, p_\epsilon)$ satisfies conditions (3.1)–(3.3).[20]

Lemma 1. *The sequence p_ϵ constructed by taking $\epsilon = 1/n$ and $n \to \infty$ has a subsequence that converges to some convex continuous p uniformly on* [0, 1].

Proof. See Appendix A.11.

The lemma gives us a limit p of the sequence p_ϵ. By the same argument as in the first step of the proof, a subsequence of solutions G_ϵ as $\epsilon \to 0$ also converges to some distribution $G \in \Delta([0,\, 1])$.

[19] Or $[y_i,\, y_i + \epsilon]$ if the function u "jumps down", that is, its value is locally lower to the right of y_i.

[20] Note that we only need $\bar{u}_\epsilon$ to have a bounded slope for a fixed ϵ. We do not claim that the slope is bounded uniformly in ϵ.

The last step of the proof is to show that (G, p) satisfy conditions (3.1)–(3.3). This is immediate by the same arguments as used in the first step of the proof, and the fact that u and $\bar{u}_\epsilon$ coincide on the support of G_ϵ for small enough ϵ.[21]

A.3 Proof of Proposition 1

To prove the first part, let G be an optimal solution to problem (2.1), and let p be the supporting price function (whose existence follows from Theorem 2). Let $p^\star$ be the solution to (3.7). We have

$$0 \overset{(1)}{\geq} \int_0^1 (u(x) - p^\star(x))dG(x) \overset{(2)}{=} \int_0^1 p(x)dG(x) - \int_0^1 p^\star(x)dG(x) \tag{A.11}$$

$$\overset{(3)}{=} \int_0^1 p(x)dF(x) - \int_0^1 p^\star(x)dG(x) \overset{(4)}{\geq} \int_0^1 p(x)dF(x) - \int_0^1 p^\star(x)dF(x) \overset{(5)}{\geq} 0, \tag{A.12}$$

where (1) follows from the fact that $p^\star \geq u$, (2) and (3) follow from the fact that (G, p) satisfy conditions (3.1) - (3.3), (4) follows from convexity of $p^\star$ and the fact that F is a MPS of G, and (5) follows from the fact that $p^\star$ solves (3.7). The above sequence of inequalities must be satisfied with equality which yields immediately that $p^\star$ is a supporting price function for G.

To prove the converse part, it is enough to use the same sequence of inequalities: If (G, p) satisfy (3.1) - (3.3), then all inequalities except for (5) hold for any feasible function $q = p^\star$. We get that

$$\int_0^1 p(x)dF(x) - \int_0^1 q(x)dF(x) \leq 0$$

for any feasible q, and thus p solves (3.7).

A.4 Receiver has two actions

A simple example where our methods can be applied is when the Receiver chooses one of two actions. Suppose that the Receiver takes the Sender-preferred action if and only if her posterior mean is greater than or equal to x_0 (so that indifferences are broken in Sender's favor). We normalize Sender's utility to 0 and 1 for the two actions. This example can be solved by other methods (see both Gentzkow and Kamenica, 2015 and Ivanov, 2015).

Proposition 6. *Assume that F has no atoms. Let u be the non-decreasing step function*

$$u(x) = \begin{cases} 0 & \text{if } 0 \leq x < x_0 \\ 1 & \text{if } x_0 \leq x \leq 1 \end{cases}.$$

[21] Formally, this follows from the fact that p_ϵ have a uniformly bounded slope, as shown in the proof of Lemma 1.

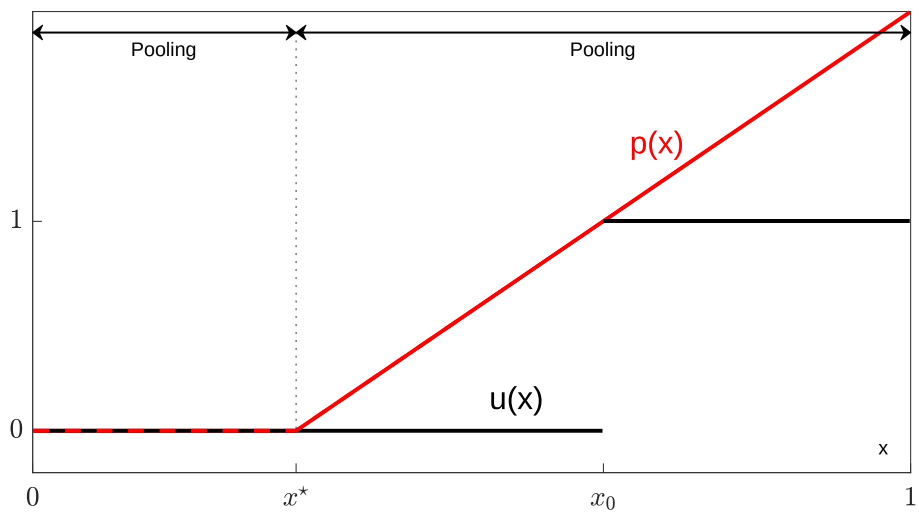

Figure 11: Two-action case

If $\mathbb{E}X \geq x_0$, then the optimal mechanism reveals nothing. If $\mathbb{E}X < x_0$, then the optimal mechanism reveals whether x is below or above $x^\star$, where $x^\star$ satisfies $\mathbb{E}[X|X \geq x^\star] = x_0$.

Proof. If $\mathbb{E}X \geq x_0$, the objective function u is superdifferentiable at $\mathbb{E}X$, so by Corollary 1, it is optimal to reveal nothing.

Now assume $\mathbb{E}X < x_0$. Let $x^\star$ be defined by $\mathbb{E}[X|X \geq x^\star] = x_0$. Consider the piece-wise affine p given by

$$p(x) = \begin{cases} 0 & 0 \leq x < x^\star \\ \dfrac{x - x^*}{x_0 - x^*} & x^\star \leq x \leq 1 \end{cases},$$

and the cdf G given by

$$G(x) = \begin{cases} 0 & 0 \leq x < E[X|X < x^\star] \\ F(x^\star) & E[X|X < x^\star] \leq x < x_0 \\ 1 & x_0 \leq x \leq 1 \end{cases}.$$

The function p is convex by construction. Condition (3.1) holds because u and p coincide for $x \leq x^\star$ and $x = x_0$. Condition (3.2) is satisfied because p is piece-wise affine, and the conditional means of F and G are equal in both regions, by construction. By the way we defined G, F is a mean-preserving spread of G, which verifies condition (3.3). Thus, by Theorem 1, G is optimal. □

A.5 Proof of Proposition 2

Lemma 2. *If (G,p) satisfy conditions (3.2), (3.3), and p is convex and continuous, then*

$$\int_0^1 \left(\int_0^x F(t)dt - \int_0^x G(t)dt\right) dp'(x) = 0, \tag{A.13}$$

interpreted as a Riemann-Stieltjes integral with respect to the measure induced by the non-decreasing function p'.

Proof. Because p is convex, it is absolutely continuous in the interior of the domain, and continuous at the endpoints by assumption. We can use integration by parts for the Riemann-Stieltjes integral:

$$\int_0^1 p(x)dG(x) = [p(x)G(x)]_0^1 - \int_0^1 G(x)dp(x) = p(1) - \int_0^1 p'(x)G(x)dx,$$

where the second equality uses the fact that $dp(x) = p'(x)dx$ by absolute continuity of p. Next, we have

$$\int_0^1 p'(x)G(x)dx = \int_0^1 p'(x)d\left(\int_0^x G(t)dt\right),$$

and we can use integration by parts for the Riemann-Stieltjes integral again to obtain

$$\begin{aligned}\int_0^1 p'(x)G(x)dx &= \left[p'(x)\left(\int_0^x G(t)dt\right)\right]_0^1 - \int_0^1 \left(\int_0^x G(t)dt\right) dp'(x)\\ &= p'(1)\left(\int_0^1 G(t)dt\right) - \int_0^1 \left(\int_0^x G(t)dt\right) dp'(x).\end{aligned}$$

Because G was arbitrary, the same transformations are true for $G = F$, and hence condition (3.2) is equivalent to

$$p'(1)\left(\int_0^1 G(t)dt\right) - \int_0^1 \left(\int_0^x G(t)dt\right) dp'(x) = p'(1)\left(\int_0^1 F(t)dt\right) - \int_0^1 \left(\int_0^x F(t)dt\right) dp'(x).$$

By condition (3.3), F and G have the same mean, and thus

$$p'(1)\left(\int_0^1 G(t)dt\right) = p'(1)\left(\int_0^1 F(t)dt\right)$$

which ends the proof. □

By condition (3.3), F is a mean-preserving spread of G which implies that G second-order stochastically dominates F. Thus,

$$\int_0^x F(t)dt \geq \int_0^x G(t)dt, \, \forall x \in [0,\, 1]. \tag{A.14}$$

Because p is convex, p' is non-decreasing, and thus p' induces a positive measure. Therefore, condition (A.13) is satisfied if and only if $\int_0^x F(t)dt = \int_0^x G(t)dt$ for p'-almost all $x \in [0,\, 1]$. That is, the equality has to hold on every set that has positive measure under p', in particular for each x at which there is a jump in p', and for every interval on which p is strictly convex. We conclude that:

1. $\int_0^x F(t)dt = \int_0^x G(t)dt$ in every interval $[a,\, b] \subset [0,\, 1]$ in which p is strictly convex;
2. p is affine in every interval $[a,\, b] \subset [0,\, 1]$ such that $\int_0^t F(x)dx > \int_0^t G(x)dx$ for all $t \in [a,\, b]$;
3. $\int_0^x F(t)dt = \int_0^x G(t)dt$ for each x at which p has a jump in the first derivative.

To strengthen the conclusion of points 1 and 3 above, we prove the following lemma.

Lemma 3. *If $x \in (0,\, 1)$ is such that $\int_0^x F(t)dt = \int_0^x G(t)dt$, then $F(x) = G(x)$.*

Proof. We will prove the contrapositive: if $F(x) \neq G(x)$ for some $x \in (0, 1)$, then $\int_0^x F(t)dt \neq \int_0^x G(t)dt$.

Fix $x \in (0, 1)$ and first suppose that $F(x) < G(x)$. Since F and G are right-continuous, there exists a $z > x$ such that $F(t) < G(t)$ for every $t \in (x, z)$. Then, since $\int_0^z F(t)dt \geq \int_0^z G(t)dt$ holds by equation (A.14), we have

$$\int_0^x F(t)dt = \int_0^z F(t)dt - \int_x^z F(t)dt > \int_0^z G(t)dt - \int_x^z G(t)dt = \int_0^x G(t)dt.$$

If instead $F(x) > G(x)$, since G is nondecreasing and F has full support and no atoms, there exists a $w < x$ such that $F(t) > G(t)$ for all $t \in (w, x)$. Then

$$\int_0^x F(t)dt = \int_0^w F(t)dt + \int_w^x F(t)dt > \int_0^w G(t)dt + \int_w^x G(t)dt = \int_0^x G(t)dt. \qquad \square$$

By Lemma 3, $F(x) = G(x)$ in every interval in which p is strictly convex, and for every x at which p has a jump in the first derivative.

Take any maximal interval $[a,\, b]$ in which p is affine (that is, p is not affine on any other $[c,\, d]$ which contains $[a,\, b]$). By maximality, we must have $F(a) = G(a)$, $F(b) = G(b)$, and $\int_a^b F(t)dt = \int_a^b G(t)dt$ as we would otherwise violate the observation in the previous paragraph. Moreover, there exists $x_0 \in [a,\, b]$ such that $u(x_0) = p(x_0)$ because otherwise, by condition (3.1), the function G would be constant on $[a,\, b]$ (while F is strictly increasing because the distribution has full support).

Now, take any interval $[a,\, b]$ where p is strictly convex. Then, $F(x) = G(x)$ for all $x \in [a,\, b]$, and because F has full support, $[a,\, b] \subseteq \mathrm{supp}(G)$. Because G and p satisfy condition (3.1), we must have $u(x) = p(x)$ for all $x \in [a,\, b]$.

Suppose that $p(x) > u(x)$ in some interval $[a,\, b]$. Because F has full support, the function $\int_0^x F(t)dt$ is strictly convex. Because G satisfies condition (3.1), it is not supported in $[a,\, b]$, and thus $\int_0^x G(t)dt$ is affine on $[a,\, b]$. Using inequality (A.14), we conclude that $\int_0^x F(t)dt$ and $\int_0^x G(t)dt$ coincide at at most one point in $(a,\, b)$, call it $z^\star$. From condition 2 above, we obtain that p is affine in $(a,\, z^\star)$ and $(z^\star,\, b)$. Thus, whenever $p(x) > u(x)$ in some interval, $p(x)$ is piece-wise affine with at most one kink in that interval.

We can now recursively define the partition $0 = x_0 < x_1 < \ldots < x_{n+1} = 1$. Given x_i, we define

$$x_{i+1} = \inf\{\alpha > x_i : p \text{ is affine on } [x_i,\, \alpha] \text{ or } p \text{ is strictly convex on } [x_i,\, \alpha]\}.$$

We first prove that $x_{i+1} > x_i$ and that the partition is finite. By regularity of u, the set $\{x : p(x) > u(x)\}$ is a finite union of intervals. In every such interval, as proven above, p is piece-wise affine with at most one kink. The complement set $\{x : p(x) = u(x)\}$ is also a finite union of intervals, and u is convex in each such interval. By regularity, each such interval can be decomposed into a finite union of intervals in which p is either affine or strictly convex.[22] Thus, there are finitely many candidate points for any x_i, and thus such a partition is well defined and finite.

By construction, the partition is the coarsest one such that u is either affine or strictly convex on each interval of the partition. Because each element of the partition is a maximal interval in which p is either strictly convex or affine, the properties listed in Proposition 2 follow directly from above observations.

A.6 Proof of Theorem 3

Let u be regular (Definition 1) and affine-closed (Definition 2), and let F be continuous. By Theorem 2, we can find G and p that satisfy conditions (3.1)–(3.3). From G we will build an optimal distribution of posterior means H which is induced by a monotone partitional signal. By Proposition 2, there are finitely many intervals $[x_i, x_{i+1}]$ in which p is either strictly convex or affine. On each of these intervals $[a, b] = [x_i, x_{i+1}]$ we will verify that

$$\mathrm{supp}(H) \cap [a, b] \subseteq \{x \in [a, b] : u(x) = p(x)\}, \tag{A.15}$$

[22] Proof: suppose it is not possible, i.e. there are infinitely many intervals in which u is alternately affine and strictly convex. Then, we can define a piece-wise affine and globally convex function v which coincides with u exactly in intervals where u is affine. This would violate regularity.

$$\int_a^b p(x)dH(x) = \int_a^b p(x)dF(x), \text{ and} \tag{A.16}$$

$$F|_{[a,b]} \text{ is a mean-preserving spread of } H|_{[a,b]}. \tag{A.17}$$

If (A.15), (A.16), and (A.17) hold on each interval, then for (H, p), (3.1), (3.2) and (3.3) hold on $[0, 1]$. By Theorem 1, this is sufficient to verify optimality of H.

Consider each interval $[a, b] = [x_i, x_{i+1}]$ in turn, beginning with $[0, x_1]$. If $[a, b]$ is an interval where p is strictly convex, set $H(x) = G(x)$ for all $x \in [a, b]$. By Proposition 2, $u = p$ on $[a, b]$, hence (A.15) is satisfied. Since $F = G = H$ on $[a, b]$, (A.16) and (A.17) are automatically satisfied. The signal which induces H is full revelation of X on $[a, b]$, which is part of a monotone partitional signal (Definition 5).

If instead p is affine on $[a, b]$, let $y = \mathbb{E}[X|a \leq X \leq b]$. By Proposition 2, the mean of G conditional on $[a, b]$ is equal to y. If $u(y) = p(y)$, modify G by specifying that H puts all mass in the interval $[a, b]$ on y (pooling in the interval $[a, b]$). Formally, $H(x) = G(a)$ for $x \in [a, y)$ and $H(x) = G(b)$ for $x \in [y, b]$. Condition (A.15) holds because $u(y) = p(y)$. Condition (A.16) holds because p is affine on $[a, b]$ and H and F have the same conditional mean. Finally, condition (A.17) holds because $F|_{[a,b]}$ is a mean-preserving spread of $G|_{[a,b]}$, $G|_{[a,b]}$ is a mean-preserving spread of $H|_{[a,b]}$, and the mean-preserving spread relationship is transitive.

The remaining case is when p is affine on $[a, b]$ and $u(y) < p(y)$. Let $A := \{x \in [a, b] : u(x) = p(x)\}$. The support of G restricted to $[a, b]$ is a subset of A, by (3.1). Since u is upper semi-continuous and $u \leq p$, A is a closed subset of $[a, b]$. Since $y \notin A$ by assumption and y is the conditional mean of G on $[a, b]$, the support of G (hence its superset A) must contain points in $[a, y)$ and in $(y, b]$. Write A as the disjoint union $A = A_L \sqcup A_R$, where $A_L \subset [a, y)$ and $A_R \subset (y, b]$ are closed and nonempty.

Next we show that affine-closure implies that at least one of A_L and A_R is a closed interval that extends to a or b respectively; that is, either $A_L = [a, c]$, or $A_R = [d, b]$, or both, for some $c, d \in (0, 1)$. Suppose neither were true. Write the closed set A_L as a union of disjoint closed intervals, choose any one of these intervals that has its left endpoint not equal to a, and define α as its left endpoint. Similarly we can define $\beta < b$ as the right endpoint of an interval of A_R. By construction, the definition of affine-closure applies to u and p at α and β: they belong to A, so $u(\alpha) = p(\alpha)$ and $u(\beta) = p(\beta)$; $p \geq u$ at all points in $[a, b]$; and $p > u$ in a left-neighborhood and right-neighborhood of α and β respectively. Thus, by affine-closure, $u(x) = p(x)$ for all $\alpha \leq x \leq \beta$. In particular this holds for $x = y$, which contradicts our previous assumption that $u(y) < p(y)$.

From now on suppose that $A_L = [a, c]$ for some $c < y$; the symmetric case $A_R = [d, b]$ follows from the same argument. We now construct H on the interval $[a, b]$. Let $\delta := \min A_R$, and define ω as the smallest solution to $\mathbb{E}[X|\omega \leq X \leq b] = \delta$. The solution exists because $\mathbb{E}[X|\omega \leq X \leq b]$, as a function of ω, is nondecreasing, continuous (F has no mass points), and ranges from $y < \delta$ at $\omega = a$ to $b > \delta$ at $\omega = b$.

Now consider the following monotone partition: $[a, \omega]$, $(\omega, b]$. The H that it induces has a mass point at $\gamma := \mathbb{E}[X|a \leq X \leq \omega]$ and one at δ. As it is a monotone partition of F, it satisfies (A.17).

Finally, we check that (3.1) holds. The required equality $u(\delta) = p(\delta)$ holds by construction, so the only thing left is to check that $u(\gamma) = p(\gamma)$. Since $A_L = [a, c]$, it is enough to check that $\gamma \leq c$. Suppose instead that $\gamma > c$. We will derive a contradiction by showing that this implies that F is not a mean-preserving spread of G in the interval $[a, b]$, contradicting Proposition 2.

When $\gamma > c$, all mass that G puts to the left of y must be to the left of γ. We show that this mass is smaller than what H puts: $G(\gamma) < H(\gamma)$. If instead $G(\gamma) \geq H(\gamma)$, then $G(z) \geq H(z)$ for all $z \in [a, \omega]$ with the inequality strict for at least some $z < \gamma$, since $H(a) = G(a)$ (because $H = G$ at all endpoints of the convex regions to the left of a, by construction) and both G and H are constant for $z \in (\gamma, \omega]$. Therefore $\int_a^\omega G(z)dz > \int_a^\omega H(z)dz$. The right-hand side evaluates to $\int_a^\omega H(z)dz = F(\omega)(\omega - \gamma) = \omega F(\omega) - \int_a^\omega z dF(z) = \int_a^\omega F(z)dz$, where the last equality is integration by parts. Putting things together, $\int_a^\omega G(z)dz > \int_a^\omega F(z)dz$, which contradicts the assumption that F is a mean-preserving spread of G.

Next, note that $\int_\omega^b H(z)dz = F(\omega)(\delta - \omega) + (b - \delta) = bF(b) - \omega F(\omega) - (1 - F(\omega))\delta = \int_\omega^b F(z)dz$, where the last equality is again integration by parts. Furthermore, $1 = H(z) \geq G(z)$ for $z \in [\delta, b]$, and $H(z) > G(z)$ for $z \in [\gamma, \delta)$ (because both G and H are constant on $[\gamma, \delta)$ and $H(\gamma) > G(\gamma)$). Therefore $\int_\omega^b G(z)dz < \int_\omega^b H(z)dz = \int_\omega^b F(z)dz$. Since F is a mean-preserving spread of G, $\int_a^\omega G(z)dz \leq \int_0^\omega F(z)dz$. But then $\int_0^b G(z)dz = \int_a^\omega G(z)dz + \int_\omega^b G(z)dz < \int_a^\omega F(z)dz + \int_\omega^b F(z)dz = \int_a^b F(z)dz$, contradicting the assumption that F is a mean-preserving spread of G (which implies $\int_a^b G(z)dz = \int_0^b F(z)dz$).

This concludes the proof that (3.1) holds on $[a, b]$. Note that H is well-defined, as by construction $H = G$ on all endpoints of the intervals $[a, b]$. We conclude that H is optimal, and by construction it is induced by a monotone partitional signal.

We now prove the converse. If u is not affine-closed, then there exist $x, y \in (0, 1)$, $x < y$, and an affine function q such that: $u(x) = q(x)$, $u(y) = q(y)$; $q(z) \geq u(z)$ for all $z \in (x, y)$; there exists $w \in (x, y)$ such that $q(w) > u(w)$; and there exists $\varepsilon > 0$ such that $q(z) > u(z)$ for all $z \in (x - \varepsilon, x) \cup (y, y + \varepsilon)$, where ε is chosen so that $x - \varepsilon > 0$ and $y + \varepsilon < 1$.

Consider the distribution F that puts weight $\alpha > 0$ uniformly on $[x - \varepsilon, x - \varepsilon/2]$ and weight $1 - \alpha$ uniformly on $[y + \varepsilon/2, y + \varepsilon]$, where α is chosen so that $\mathbb{E}[F] = w$, where $q(w) > u(w)$ holds.

That is, let F have density

$$f(z) = \begin{cases} 0 & \text{if } z < x - \varepsilon \\ \frac{2\alpha}{\varepsilon} & \text{if } x - \varepsilon \le z \le x - \frac{\varepsilon}{2} \\ 0 & \text{if } x - \frac{\varepsilon}{2} < z < y + \frac{\varepsilon}{2} \\ \frac{2(1-\alpha)}{\varepsilon} & \text{if } y + \frac{\varepsilon}{2} \le z \le y + \varepsilon \\ 0 & \text{if } z > y + \varepsilon \end{cases}.$$

Given such a prior F, the Sender cannot do strictly better than choosing a distribution of posterior means G that has support limited to x and y. To see this, first note that the mean-preserving spread condition implies that supp $G \subseteq [x - \varepsilon, y + \varepsilon]$. Suppose the Sender's utility were q. Since q is linear, all feasible distributions of posterior means are optimal; in particular, so is one with support equal to $\{x, y\}$. Since $q \ge u$ on $[x - \varepsilon, y + \varepsilon]$, this gives an upper bound to the utility attainable under u. The upper bound is attained if and only if all mass is concentrated at points where $q(z) = u(z)$. In particular, the distribution

$$G(z) = \begin{cases} 0 & \text{if } 0 \le z < x \\ \alpha & \text{if } x \le z < y \\ 1 & \text{if } y \le z \le 1 \end{cases}$$

is feasible and attains the upper bound.

To conclude it is enough to show that no monotone partitional signal achieves the upper bound. First consider the trivial partition that pools all realizations at the prior mean $w = \mathbb{E}[F]$. By our choice of α, $u(w) < q(w)$, hence this is not optimal. Next consider monotone partitions with two or more signals. Since all prior mass is concentrated on $[x-\varepsilon, x-\varepsilon/2]$ and $[y+\varepsilon/2, y+\varepsilon]$, at least one of the intervals in the partition has a conditional posterior mean in $[x-\varepsilon, x-\varepsilon/2]$ or $[y+\varepsilon/2, y+\varepsilon]$. But by construction $u(z) < q(z)$ on such intervals. Hence the monotone partition is not optimal. We conclude that no monotone partition is optimal for such a prior F. $\square$

A.7 Material for Section 6.2 and proof of Proposition 4

We first present the parametrization of preferences that underlies Figures 9 and 10. Assume that the investor has CRRA utility v with Arrow-Pratt measure of relative risk aversion $\eta > 0$ over final wealth z; that is, $v(z) = \frac{z^{1-\eta}-1}{1-\eta}$ if $\eta \neq 1$, and $v(z) = \ln z$ if $\eta = 1$. Given a posterior belief $x > \frac{1}{2}$ that the long position is profitable ($x < \frac{1}{2}$ is symmetric), the investor chooses the level of investment $y > 0$ in the risky asset that maximizes $xv(w+y-c)+(1-x)v(w-y-c)$, or decides not to invest ($y = 0$) and receives the outside option $v(w)$. Let $y^\star(x)$ be the optimal investment as a function of belief x.

Assume the analyst's payoff is proportional to the amount invested. Then (omitting the irrelevant proportionality constant) $u(x) \equiv y^\star(x)$ can be shown to be

$$u(x) = \begin{cases} -(w-c)\frac{1-\left(\frac{1-x}{x}\right)^{\frac{1}{\eta}}}{1+\left(\frac{1-x}{x}\right)^{\frac{1}{\eta}}} & \text{if } x \leq x_0 \\ 0 & \text{if } x_0 < x < 1-x_0 \\ (w-c)\frac{1-\left(\frac{1-x}{x}\right)^{\frac{1}{\eta}}}{1+\left(\frac{1-x}{x}\right)^{\frac{1}{\eta}}} & \text{if } x \geq 1-x_0 \end{cases}$$

where x_0 is the threshold belief at which the investor is indifferent between investing or not: $v(w) = x_0 v(w + y^\star(x_0) - c) + (1-x_0)v(w - y^\star(x_0) - c)$. If $\eta < 1$ (less risk averse than log utility) $u(x)$ is strictly concave in $[0, x_0]$ and $[1-x_0, 1]$; if $\eta > 1$ (more risk averse than log utility) then $u(x)$ is strictly convex in the same regions. For all $\eta > 0$, u is affine-closed. By Theorem 3, there exists a monotone partitional optimal signal. Since the multiplier $p(x)$ must be convex, hence continuous, $p(x)$ cannot coincide with $u(x)$ in a neighborhood of the discontinuities x_0 and $1-x_0$. Thus, by Proposition 2, x_0 and $1-x_0$ must be contained in pooling regions. These insights lead us to consider the multipliers depicted in Figure 9 and Figure 10. Define $w(x)$ as

$$w(x) = \left(-\mathbf{1}_{\{x \leq \frac{1}{2}\}} + \mathbf{1}_{\{x > \frac{1}{2}\}}\right)(w-c)\frac{1-\left(\frac{1-x}{x}\right)^{\frac{1}{\eta}}}{1+\left(\frac{1-x}{x}\right)^{\frac{1}{\eta}}}.$$

This coincides with u for $x \leq x_0$ and $x \geq 1-x_0$. For $\eta < 1$, w is strictly concave on $\left[0, \frac{1}{2}\right]$ and $\left[\frac{1}{2}, 1\right]$. For $\eta > 1$, w is (globally) strictly convex.

Proof of Proposition 4. In the concave case ($\eta < 1$), define $y = \mathbb{E}[X | X \leq \frac{1}{2}]$ and consider the piece-wise linear, convex function

$$p(x) = \begin{cases} u'(y)(x-y) + u(y) & \text{if } x \leq \frac{1}{2} \\ u'(1-y)(x-(1-y)) + u(1-y) & \text{if } x > \frac{1}{2} \end{cases}$$

By construction, p is tangent to u at y and $1-y$, and has a kink at $\frac{1}{2}$. Furthermore, since and u coincide at y and $1-y$, and w is concave, $p(x) \geq w(x) \geq u(x)$ for all $x \in [0,1]$ and $p(x) > u(x)$ for $x \notin \{y, 1-y\}$. Consider the distribution of posterior means

$$G(x) = \begin{cases} 0 & \text{if } x < y \\ \frac{1}{2} & \text{if } y \leq x < 1-y \\ 1 & \text{if } x \geq 1-y \end{cases}$$

that puts atoms of size $F\left(\frac{1}{2}\right)=\frac{1}{2}$ on $y=\mathbb{E}[X|X\leq\frac{1}{2}]$ and on $1-y=\mathbb{E}[X|X\geq\frac{1}{2}]$. Conditions (3.1)–(3.3) are satisfied, hence, by Theorem 1, G is optimal.

Now consider the convex case ($\eta>1$). Consider the set of lines passing through $(x_0,u(x_0))$ with slopes m that satisfy $\frac{u(x_0)-u(0)}{x_0-0}<m<u'(x_0)$. Since $u(x)\leq w(x)$ and w is strictly convex, each such line intersects $u(x)$ at two points besides x_0: $a(m)\in(0,x_0)$ and $b(m)\in(x_0,\frac{1}{2})$. Functions $a(\cdot)$ and $b(\cdot)$ are continuous and strictly increasing. Let $t(m)=\mathbb{E}[X|X\in[a(m),b(m)]]$. Since F is continuous and strictly increasing by assumption, $t(\cdot)$ is also continuous and strictly increasing. By assumption (6.1), $t\left(\frac{u(x_0)-u(0)}{x_0-0}\right)<\mathbb{E}\left[X|X\in\left[0,\frac{1}{2}\right]\right]<x_0$. By construction, $t(u'(x_0))=\mathbb{E}[X|X\in[x_0,b(u'(x_0))]]>x_0$. By the intermediate value theorem, there exists a (unique) $m^\star$ such that $t(m^\star)=x_0$. Define $a^\star:=a(m^\star)$, $b^\star:=b(m^\star)$. Now consider the following p and G:

$$p(x)=\begin{cases}u(x) & \text{if } x<a^\star\\ m^\star(x-a^\star)+u(a^\star) & \text{if } a^\star\leq x<b^\star\\ u(x) & \text{if } b^\star\leq x<1-b^\star\\ -m^\star(x-(1-a^\star))+u(1-a^\star) & \text{if } 1-b^\star\leq x<1-a^\star\\ u(x) & \text{if } x\geq 1-a^\star\end{cases},$$

$$G(x)=\begin{cases}F(x) & \text{if } x<a^\star\\ F(a^\star)+\mathbf{1}_{\{x\geq x_0\}}(F(b^\star)-F(a^\star)) & \text{if } a^\star\leq x<b^\star\\ F(b^\star)+\mathbf{1}_{\{x\geq\frac{1}{2}\}}(F(1-b^\star)-F(b^\star)) & \text{if } b^\star\leq x<1-b^\star\\ F(1-b^\star)+\mathbf{1}_{\{x\geq 1-x_0\}}(F(1-a^\star)-F(1-b^\star)) & \text{if } 1-b^\star\leq x<1-a^\star\\ F(x) & \text{if } x\geq 1-a^\star\end{cases}.$$

G reveals x when $x<a^\star$ or $x\geq 1-a^\star$. The remaining intervals $[a^\star,b^\star]$, $(b^\star,1-b^\star)$, and $[1-b^\star,1-a^\star]$ are pooled at x_0, $\frac{1}{2}$ and $1-x_0$, respectively. Conditions (3.1)–(3.3) are satisfied by construction, hence, by Theorem 1, G is optimal. □

A.8 Proof of Theorem 4

Because the proof is similar to the proof of Theorem 1 and Theorem 2, we omit the details and only highlight the differences.

First, verifying that H is unimprovable amounts to checking that H is a solution to the following optimization problem, for each Sender i:

$$\max_G \int_0^1 u_i(x)dG(x) \tag{A.18}$$

subject to

$$F \text{ is a mean-preserving spread of } G; \tag{A.19}$$

$$G \text{ is a mean-preserving spread of } H, \tag{A.20}$$

where the second condition reflects the constraint that Sender i can only disclose additional information but cannot "hide" the information revealed by other Senders. We can write down the analog of Theorem 1 for this problem: If there exist convex functions p_i and q_i, and a distribution G such that (A.19) – (A.20) hold, and

$$\text{supp}(G) \subseteq \text{argmax}_{x\in[0,\,1]}\{u_i(x) - p_i(x) + q_i(x)\} \tag{A.21}$$

$$\int_0^1 p_i(x)dG(x) = \int_0^1 p_i(x)dF(x), \tag{A.22}$$

$$\int_0^1 q_i(x)dG(x) = \int_0^1 q_i(x)dH(x), \tag{A.23}$$

then G is optimal for the problem (A.18) – (A.20). The additional price function q_i and the additional constraint (A.23) are a consequence of adding the constraint that G is a mean-preserving spread of H. This result is proven in the same way as Theorem 1. (Condition (A.21) is equivalent to assuming $p_i(x) - q_i(x) \geq u_i(x)$ and $\text{supp}(G) \subseteq \{x \in [0,\, 1] :\ u_i(x) = p_i(x) - q_i(x)\}$.)

With this verification tool, we can write down the conditions under which H is an optimal solution to the problem (A.18) – (A.20):

$$\text{supp}(H) \subseteq \text{argmax}_{x\in[0,\,1]}\{u_i(x) - p_i(x) + q_i(x)\} \tag{A.24}$$

$$\int_0^1 p_i(x)dH(x) = \int_0^1 p_i(x)dF(x), \tag{A.25}$$

where condition (A.23) is omitted because it holds vacuously. Because q_i only appears in condition (A.24), we can define $\hat{u}_i = u_i + q_i$ which leads to a function that is a convex translation of u_i. This proves the first part of Theorem 4: If for every i the conditions of Theorem 4 hold, we can define $q_i = \hat{u}_i - u_i$, and conditions (A.24) – (A.25) will hold. Condition (A.24) is equivalent to (7.1) because prices can be normalized appropriately.

To prove the second part, it is enough to prove existence of convex functions q_i and p_i such that conditions (A.21) – (A.23) hold. This can be done analogously as in the proof of Theorem 2, where the additional function q_i is derived in the same way as the function p_i from the additional constraint (A.20).

There are two differences in the proof. First, the generalized Slater condition (A.4) is not guaranteed to hold for the problem in which G is "sandwiched" between H and F in the mean-preserving spread order. This can be circumvented by studying a slightly different perturbation than in the proof of Theorem 2: The mean-preserving spread condition is only imposed on $[\epsilon,\, 1-\epsilon]$ and additionally G is only required to be a mean-preserving spread of H_ϵ (instead of H) with the property that H_ϵ converges to H in the weak* topology as $\epsilon \to 0$, and F is a **strict** mean-preserving spread of H_ϵ on $[\epsilon,\, 1-\epsilon]$ in the sense that

$$\int_0^x F(t)dt > \int_0^x H_\epsilon(t)dt,\ \forall x \in [\epsilon,\, 1-\epsilon],$$

$$\int_x^1 H_\epsilon(t)dt > \int_x^1 F(t)dt,\ \forall x \in [\epsilon,\, 1-\epsilon].$$

Such an approximation H_ϵ exists because the support of F contains 0 and 1 by assumption. The perturbed problem satisfies the Slater condition (A.4), and the rest of the proof is fully analogous.

Second, it is not possible to take the final step in the proof of Theorem 2, i.e. extend the conclusion to discontinuous utility functions. This is because we cannot guarantee that the sequence of prices for the perturbed problem will have a converging subsequence in this case. The existence of constraint (A.23) might force the prices for the perturbed problem to have an arbitrarily high slope, and thus the sequence might diverge. This is not just a problem with the proof. The conclusion of Theorem 4 is false in the case of discontinuous utility functions. To see this, take an example with two Senders, in which the utility function u_1 of the first Sender is strictly convex, and the utility function u_2 of the second Sender has a discontinuity in the interior of $[0,\, 1]$. The unique equilibrium distribution corresponds to full disclosure, so if the result were true, there would exist a convex translation $\hat{u}_2$ of u_2 such that $\hat{u}_2$ would coincide with some convex p_2 on $[0,\, 1]$ (this is necessary to support a fully-revealing solution). But this is impossible because any $\hat{u}_2$ would still have a discontinuity in the interior of $[0,\, 1]$, and every convex function is continuous on the interior of its domain.

Therefore, Theorem 4 is only established for the case of regular, continuous utility functions.

A.9 Proof of Proposition 5

Lemma 4. *Suppose that $\hat{u}_i$ is a convex translation of u_i. If p_i is a convex function satisfying all conditions of Theorem 4 (with some distribution H), then*

$$p_i(x) = \begin{cases} \hat{u}_i(x) & x \leq a_i \\ \hat{u}_i(a_i) + (x - a_i)\hat{u}_i(b_i) & a_i \leq x \leq b_i \\ \hat{u}_i(x) & b_i \leq x \end{cases},$$

for some $a_i \leq c_i \leq b_i$.

Proof. Because the prior mean is smaller than c_i, $\hat{u}_i$ has to coincide with p_i for at least some points in $[0, c_i)$, by condition (7.1). Because $\hat{u}_i$ is convex on $[0, c_i]$ (because u_i is convex), the only possibility is that $\hat{u}_i$ coincides with p_i on some subinterval $[\underline{d}, a_i]$ of $[0, c_i]$ (this follows from the properties proved in Proposition 2). Moreover, we must have $\underline{d} = 0$ as otherwise the mean-preserving spread condition would be violated on $[0, \underline{d}]$, again by Proposition 2. By the same reasoning, if $\hat{u}_i$ coincides with p_i anywhere in the interval $(c_i, 1]$, then it has to coincide on a subinterval of the form $[b_i, 1]$. Finally, by Proposition 2 and due to the shape of $\hat{u}_i$, the price function has to be affine in the remaining interval $[a_i, b_i]$ because it can touch $\hat{u}_i$ only at c_i, if at all. $\square$

Lemma 4 restricts the set of distributions that can emerge as optimal for any optimization problem of Sender i. Because $\hat{u}_i$ is strictly convex on $[0, a_i]$, by Proposition 2, the state is fully revealed in that interval. In the interval $[a_i, c_i]$, there can only be an atom in the posterior distribution at c_i (if p_i touches $\hat{u}_i$ at c_i). Finally, because $\hat{u}_i$ may be affine on $[c_i, 1]$, the posterior distribution can have mass anywhere in this interval, as long as the mean-preserving spread condition is satisfied. This leads us to consider the following class $\mathcal{H}_i$ of posterior distributions for Sender i: any $H \in \mathcal{H}_i$ reveals the state fully on $[0, a_i]$ for some $a_i \leq c_i$, has no mass on (a_i, c_i), and satisfies $\mathbb{E}_H[X|X \geq a_i] = \mathbb{E}_F[X|X \geq a_i]$. By the above reasoning, the class $\mathcal{H}_i$ contains all candidate distributions that may be unimprovable for Sender i. We will show that indeed each $H \in \mathcal{H}_i$ is unimprovable for i.

To this end, given any $H \in \mathcal{H}_i$, we will construct a $\hat{u}_i$ and p_i to satisfy conditions (7.1) – (7.3). The function $\hat{u}_i$ is constructed by adding a wedge function $w(x) = \alpha(x - c_i)^+$ to u_i. Consider a p_i function that coincides with u_i on $[0, a_i]$, and is affine on $[a_i, 1]$, tangent to u_i at c_i. Such a function exists because u_i has bounded slope and is not convex (there is a kink at c_i). Then, choose α in the wedge function w so that $\hat{u}_i = u_i + w$ coincides with p_i on $[c_i, 1]$. It is now a routine check that all conditions (7.1) – (7.3) hold.

In the final step, in order to find the set of equilibrium distributions, we only have to take the intersection of $\mathcal{H}_i$ over $i \in N$. This intersection is equal to $\mathcal{H}_1$ because these sets are nested (by assumption, c_i is highest for Sender 1).

So the set of equilibrium distributions contains exactly H such that H is feasible, fully reveals the state below some $a \leq c_1$, and has no mass on (a, c_1).

The least informative equilibrium $\underline{H}$ is thus as follows: a is chosen so that the conditional mean of F conditional on $X \geq a$ is equal to c_1. Then, $\underline{H}$ reveals the state below a and puts a mass point at c_1. Hence, this is just the individually-optimal persuasion scheme of Sender 1.

A.10 Proof of Claim 1

If $\tau^\star$ is the optimal solution for some prior μ_0, and $P : \Delta(\Theta) \to \mathbb{R}$ is convex and $P \geq V$, then

$$\int_{\Delta(\Theta)} V(\mu) d\tau^\star(\mu) \overset{(1)}{\leq} \int_{\Delta(\Theta)} P(\mu) d\tau^\star(\mu) \overset{(2)}{\leq} \int_{\Delta(\Theta)} \int_\Theta P(\delta_\theta) d\mu(\theta) d\tau^\star(\mu)$$
$$\overset{(3)}{=} \int_\Theta P(\delta_\theta) \int_{\Delta(\Theta)} \mu(\theta) d\tau^\star(\mu) d\theta \overset{(4)}{=} \int_\Theta P(\delta_\theta) d\mu_0(\theta), \quad \text{(A.26)}$$

where (1) follows from $V \leq P$, (2) from the fact that P is convex, (3) is by change of integration, and (4) follows from condition (7.8). On the other hand,

$$\int_{\Delta(\Theta)} V(\mu) d\tau^\star(\mu) = \int_{\Delta(\Theta)} P_{\mu_0}(\mu) d\tau^\star(\mu) = \int_\Theta P_{\mu_0}(\delta_\theta) d\mu_0(\theta),$$

where the first equality follows from condition (7.6) and the second from condition (7.7).

A.11 Technical Appendix

In this Appendix, we fill the gaps in the proof of Theorem 2.

A.11.1 Step 2 without Assumption 1

For a fixed ϵ, consider a pair (G, p) from Step 1 of the proof of Theorem 2 from Appendix A.2. We can assume without loss of generality that p is non-negative.[23] We first show how to modify the function p to ensure that its slope is uniformly bounded (this will be important in the next step of the proof). When we consider the sequence p_n, the slope of p_n could diverge to infinity close to the endpoints, upsetting uniform convergence. We show in the lemma below that we can control the slope by using the properties of the pair (G, p) established in Step 1.

[23] The optimization problem is unaffected by adding a constant to the utility function u, so we can assume without loss of generality that $u(x) \geq 0$ for all $x \in [0, 1]$, and then non-negativity of p follows from $p \geq u$.

Lemma 5. *Consider problem (A.1) - (A.3) for a fixed $\epsilon > 0$. There exists a convex function q which satisfies (A.7) and (A.9) and has a slope uniformly bounded by*

$$\max\left\{\frac{c}{\int_0^{\underline{x}}(\underline{x}-x)dF(x)}, \frac{c}{\int_{\bar{x}}^1(x-\bar{x})dF(x)}\right\},$$

where c is a constant that does not depend on ϵ, and

$$\underline{x} := \inf\{x \in [0, 1] : u(x) = p(x)\},$$

$$\bar{x} := \sup\{x \in [0, 1] : u(x) = p(x)\}.$$

Proof. In the case when either $\underline{x} = 0$ or $\bar{x} = 1$, there is nothing to prove because the bound is equal to ∞. We assume otherwise, and focus on showing the bound by $c/\int_0^{\underline{x}}(\underline{x}-x)dF(x)$ on $[0, 1/2]$ (an analogous argument establishes the other bound on $[1/2, 1]$).

Recall that u is assumed continuous, and that by the regularity assumption u has a slope uniformly bounded by $M < \infty$. This also implies that u is bounded: $\|u\|_\infty < \infty$. Because p is convex, $p \geq u$, and p coincides with u at $\underline{x}$ and $\bar{x}$, p inherits the bound M on the slope in the interval $[\underline{x}, \bar{x}]$.

Next, note that condition (A.9) is not affected by modifying $p(x)$ on $[0, \underline{x})$ or $(\bar{x}, 1]$. Because G puts no mass on $[0, \underline{x}) \cup (\bar{x}, 1]$ by property (A.9), the value of the integral $\int_0^1 p(x)dG(x)$ is also unaffected. Therefore, we can replace p on $[0, \underline{x})$ by some other function q and preserve conditions (A.7) and (A.9) as long as

$$\int_0^{\underline{x}} p(x)dF(x) = \int_0^{\underline{x}} q(x)dF(x). \tag{A.27}$$

Consider a function q that coincides with p on $[\underline{x}, 1]$, and is affine otherwise: $q(x) = p(\underline{x}) + \Delta(\underline{x} - x)$, for some $\Delta > 0$, for $x \in [0, \underline{x}]$. Choose Δ to satisfy equation (A.27). Because p is convex, it is clear that such Δ has to be larger than the slope of p at $\underline{x}$, so q remains convex. Moreover, we have

$$\begin{aligned}\|u\|_\infty \geq \int_0^1 u(x)dG(x) \overset{(1)}{=} \int_0^1 p(x)dG(x) \overset{(2)}{=} \int_0^1 p(x)dF(x) \overset{(3)}{\geq} \int_0^{\underline{x}} p(x)dF(x)\\ \overset{(4)}{=} \int_0^{\underline{x}} q(x)dF(x) \overset{(5)}{\geq} \Delta\int_0^{\underline{x}}(\underline{x}-x)dF(x),\end{aligned} \tag{A.28}$$

where (1) follows from (A.9), (2) follows from (A.7), (3) follows from the fact that p is non-negative, and (4) and (5) from the definition of q. We conclude that

$$\Delta \leq \frac{\|u\|_\infty}{\int_0^{\underline{x}}(\underline{x}-x)dF(x)}.$$

Therefore, the slope of q is bounded by

$$\frac{\max\{\|u\|_\infty,\, M\}}{\int_0^{\underline{x}}(\underline{x}-x)dF(x)}.$$

This finishes the proof.[24] $\square$

We come back to the proof of Step 2 without Assumption 1. Recall that we have a sequence $(G_n,\, p_n)$ satisfying (A.2), (A.3), (A.7) and (A.9), with each p_n convex. Moreover, by Lemma 5, we can modify the functions p_n so that p_n has a slope bounded uniformly by

$$\max\left\{\frac{c}{\int_0^{\underline{x}_n}(\underline{x}_n-x)dF(x)},\, \frac{c}{\int_{\bar{x}_n}^1(x-\bar{x}_n)dF(x)}\right\},$$

where c does not depend on n, and $\underline{x}_n$ and $\bar{x}_n$ are defined by Lemma 5:

$$\underline{x}_n := \inf\{x\in[0,\,1]:\, u(x)=p_n(x)\},$$

$$\bar{x}_n := \sup\{x\in[0,\,1]:\, u(x)=p_n(x)\}.$$

We can assume without loss (by passing to a subsequence if necessary) that both $\underline{x}_n$ and $\bar{x}_n$ converge to some $\underline{x}$ and $\bar{x}$, respectively. If $\underline{x}>0$ and $\bar{x}<1$, then, for sufficiently high n, all p_n have a slope uniformly bounded by

$$\max\left\{\frac{c}{\int_0^{\underline{x}/2}(\underline{x}/2-x)dF(x)},\, \frac{c}{\int_{(\bar{x}+1)/2}^1(x-(\bar{x}+1)/2)dF(x)}\right\},$$

using the assumption that 0 and 1 are in the support of F. Consider the opposite case when either *(i)* $\underline{x}=0$ or *(ii)* $\bar{x}=1$. Then, for a sufficiently small $\delta>0$, all p_n have a uniformly bounded slope on $[\delta,\,1-\delta]$, for sufficiently high n. This is because each $p_n\geq u$, p_n is convex, and thus p_n has a slope bounded by the slope of u (which is bounded by M by the regularity assumption).

We can thus conclude that for every (small enough) $\delta>0$, p_n have a uniformly bounded slope on $[\delta,\,1-\delta]$. Thus, the sequence of functions p_n is uniformly bounded on $[\delta,\,1-\delta]$. This follows from the fact that each p_n is convex, has a uniformly bounded slope, and the domain $[\delta,\,1-\delta]$ is compact. A uniformly bounded sequence of convex functions is Lipshitz continuous with a common Lipshitz constant L. In particular, the sequence $(p_n)_n$ is equi-continuous on $[\delta,\,1-\delta]$. By the Arzela-Ascoli Theorem, p_n has a uniformly converging subsequence on every interval $[\delta,\,1-\delta]$. Therefore, a subsequence of p_n converges to some continuous convex p on $(0,\,1)$,

[24] We have not verified that such q satisfies $q\geq u$ on $[0,\,\underline{x}]$. However, this does not pose a problem for the remainder of the proof because we will only need the bound derived in the lemma to hold for $\underline{x}$ sufficiently close to 0. If q does not satisfy $q\geq u$, we can always replace $\underline{x}$ with some smaller $\underline{x}'>0$, and the rest of the proof remains the same.

uniformly on each compact subset of $(0, 1)$. We can complete the definition of p by specifying that p is continuous at 0 and at 1 (the properties of p at any single point do not play a role).

Just as in Step 2 of the proof from Appendix A.2, we prove that G_n converges in the weak* topology to some $G \in \Delta([0, 1])$, and that the limiting pair (G, p) satisfies conditions (3.1) and (3.3). A separate argument is needed to show condition (3.2) because now we only have convergence of p_n to p uniformly on every compact subset of $(0, 1)$ but not necessarily on $[0, 1]$.

Define, for each n, the smallest convex function q_n that coincides with p_n on $[\underline{x}_n, \bar{x}_n]$. Note that on $[\underline{x}_n, \bar{x}_n]$ the slope of p_n is bounded by the slope of u (which is bounded by M by assumption), so q_n can be constructed by linearly extending p_n beyond $[\underline{x}_n, \bar{x}_n]$ with the slope equal to the relevant derivative of p_n at the endpoints $\underline{x}_n$ and $\bar{x}_n$. Obviously, we have $p_n \geq q_n$. By construction, q_n has a uniformly bounded slope on $[0, 1]$ (bounded by M), so by the same argument as above, it has a uniformly convergent subsequence to some function q. Therefore (indexing the subsequence by n again), $\int_0^1 q_n(x)dG_n(x) \to \int_0^1 q(x)dG(x)$. To see this, note that

$$\int_0^1 q_n dG_n - \int_0^1 q dG = \int_0^1 (q_n - q)dG_n - \int_0^1 q(dG - dG_n);$$

The second integral converges to zero by the definition of convergence of G_n to G in the weak* topology. For the first integral, we have

$$\int_0^1 (q_n - q)dG_n \leq \sup_{x \in [0, 1]} \{|q_n(x) - q(x)|\}$$

which converges to zero because q_n converges to q uniformly on $[0, 1]$.

We have

$$\begin{aligned}\int_0^1 p(x)dF(x) \overset{(1)}{\geq} \int_0^1 p(x)dG(x) \overset{(2)}{\geq} \int_0^1 q(x)dG(x) = \lim_n \int_0^1 q_n(x)dG_n(x) \\ \overset{(3)}{=} \lim_n \int_0^1 p_n(x)dG_n(x) \overset{(4)}{=} \lim_n \int_0^1 p_n(x)dF(x) \overset{(5)}{\geq} \lim_n \int_\delta^{1-\delta} p_n(x)dF(x) \\ \overset{(6)}{=} \int_\delta^{1-\delta} p(x)dF(x) \overset{(7)}{\geq} \int_0^1 p(x)dF(x) - \varepsilon(\delta),\end{aligned} \quad \text{(A.29)}$$

where (1) follows because p is convex and F is a mean-preserving spread of G, (2) follows because the inequality $p_n(x) \geq q_n(x)$ is preserved in the limit, (3) follows because, by definition, p_n and q_n coincide on the support of G_n, (4) follows from (A.7), (5) follows for any $\delta > 0$ from non-negativity of p_n, (6) follows because p_n converges to p uniformly on every compact subset of $(0, 1)$, and (7) is true for some $\varepsilon(\delta)$ which goes to zero as $\delta \to 0$. Because δ (and hence $\varepsilon(\delta)$) can be arbitrarily small, we must have:

$$\int_0^1 p(x)dF(x) = \int_0^1 p(x)dG(x),$$

which is what we wanted to prove. This finishes the proof of Step 2, i.e. we have shown conditions (3.1) - (3.3) for the case of a continuous u (and hence the optimality of G, by Theorem 1).

A.11.2 Proof of Lemma 1

First, we note that Lemma 2 stated in the proof of Proposition 2 holds even without the additional assumptions on F made in Proposition 2 (the proof of Lemma 2 does not use these assumptions). Thus, we can use properties $1 - 3$ of (G, p) stated in Appendix A.5 after Lemma 2 even though Proposition 2 itself may not hold without the additional assumptions on F.

We prove that the functions in the sequence p_ϵ are uniformly bounded. Suppose not. Then there exists a subsequence of p_ϵ (which we take to be the original sequence to simplify notation) such that $\lim_{\epsilon \to 0} \|p_\epsilon\|_\infty = \infty$. By the assumption that u has a uniformly bounded slope in the intervals where it is continuous, the only possibility is that p_ϵ has an affine piece whose slope diverges to infinity and which touches u at one of the points of discontinuity y_i. Because there are finitely many points of discontinuity of u, we can choose a divergent subsequence (which we take to be the sequence itself) in which each p_ϵ touches u with an affine piece at the same discontinuity point y_i.

Using the properties of u and $\bar{u}_\epsilon$, we claim that for small enough ϵ the affine piece of p_ϵ that touches u at y_i must have the following properties: *(i)* p_ϵ is affine on $[\delta_\epsilon, 1]$ for some $\delta_\epsilon \leq y_i$, and is not affine on any interval that strictly contains $[\delta_\epsilon, 1]$, and *(ii)* p_ϵ touches $\bar{u}_\epsilon$ only at y_i in the interval $[y_i, 1]$, and *(iii)* $\delta_\epsilon \to y_i$ as $\epsilon \to 0$. Note that condition *(i)* takes the above form because we have assumed that the function u jumps up at y_i (if it jumped down, we would work with the interval $[0, \delta_\epsilon]$ instead of $[\delta_\epsilon, 1]$).

We argue why properties *(i)* - *(iii)* hold. Properties *(i)* and *(ii)* hold because, by the choice of the sequence p_ϵ, the affine piece of p_ϵ that touches u at y_i has a divergent slope. Because u has a uniformly bounded slope whenever it is continuous, and only finitely many (interior) discontinuities, for small enough ϵ, p_ϵ cannot touch $\bar{u}_\epsilon$ to the right of y_i. Because 1 belongs to the support of F, $\int_0^1 F(x)dx = \int_0^1 G(x)dx$, and condition (3.1) implies that G has no mass to the left of y_i, it follows that $\int_0^t F(x)dx > \int_0^t G(x)dx$ for all $t > y_i$. Property 2 of (G, p) stated in Appendix A.5 then implies that p_ϵ is affine to the right of y_i. We can define $\delta_\epsilon \leq y_i$ by requiring that $[\delta_\epsilon, 1]$ is the maximal interval in which p_ϵ is affine (no interval on which p_ϵ is affine strictly contains it). Property *(iii)* then follows from the fact that the slope of the affine piece of p_ϵ on $[\delta_\epsilon, 1]$ goes to plus infinity but p_ϵ touches $\bar{u}_\epsilon$ at y_i.

We are ready to obtain a contradiction. By the properties $1 - 3$ of (G, p) stated in Appendix

A.5, and the definition of δ_ϵ, F is a mean-preserving spread of G_ϵ in the interval $[\delta_\epsilon, 1]$. Because p_ϵ does not touch u to the right of y_i, G_ϵ must put all mass on $[\delta_\epsilon, y_i]$, and hence it has to be that $\mathbb{E}[X|X \geq \delta_\epsilon] \in [\delta_\epsilon, y_i]$. In the limit as $\epsilon \to 0$, we obtain $\mathbb{E}[X|X \geq y_i] = y_i$ which is a contradiction because 1 belongs to the support of F and $y_i < 1$.

The obtained contradiction implies that p_ϵ are uniformly bounded. Because each p_ϵ is convex, the family is equi-continuous, and, as before, we can use the Arzelà-Ascoli theorem to conclude that a subsequence of p_ϵ converges to some convex continuous p.