Screening Frontiers*

Frank Yang†

February 23, 2026

Abstract

A principal screens an agent with an arbitrary set of allocations $\mathcal{X}$. The agent's preferences over allocations are comonotonic. A subset of allocations $\mathcal{X}^\star \subseteq \mathcal{X}$ is a ***surplus-elasticity frontier*** if *(i)* any other allocation has a demand curve that is pointwise lower and less elastic than some allocation in $\mathcal{X}^\star$ and *(ii)* the allocations in $\mathcal{X}^\star$ can be ordered in terms of their demand curves such that a higher demand curve is more inelastic. We show that any surplus-elasticity frontier is an optimal menu. Moreover, if the incremental demand curves along the frontier are also ordered by their elasticities, then the frontier is optimal even among stochastic mechanisms. The result is agnostic to type distributions and redistributive welfare weights—the same frontier remains optimal for a broad class of objectives. As applications, we show how these results immediately yield new insights into optimal bundling, optimal taxation, sequential screening, selling information, and regulating a data-rich monopolist.

Keywords: Multidimensional screening, bundling, taxation, costly screening, sequential screening, selling information, monopoly regulation, redistributive allocation.

*I thank Ben Brooks, Jacob Leshno, Andrew McClellan, Daniel Rappoport, Doron Ravid, and the seminar audience at the University of Chicago for helpful discussions. Some results in this paper were also reported in my job market paper (Yang 2023), which has since been split into separate papers.

†Department of Economics, Harvard University. Email: fyang@fas.harvard.edu.

1 Introduction

Product features are increasingly rich. How should a multiproduct firm design its product mix? In this paper, we answer this question by considering a screening model with an arbitrary space of allocations $\mathcal{X}$. The only assumption we make is that the consumers can be ordered such that a higher type of consumer has a higher value for any allocation.

In this environment, we show that there are two key dimensions that the firm should optimize for: *(i)* the total surplus, and *(ii)* the demand elasticity.[1] Specifically, for each allocation, consider the ***demand curve*** generated when the product is sold alone. A menu of products $\mathcal{X}^\star \subseteq \mathcal{X}$ is called a ***surplus-elasticity frontier*** if:

(i) any other allocation has a demand curve that is pointwise lower and less elastic than some allocation in $\mathcal{X}^\star$;

(ii) the allocations in $\mathcal{X}^\star$ can be ordered in terms of their demand curves such that a higher demand curve is also more inelastic.

Our first main result shows that whenever such a frontier exists, it forms an optimal menu (Theorem 1). Our second main result shows that if the frontier has the additional property that the ***incremental demand curves***—the differences of consecutive demand curves—along the frontier are also ordered by their demand elasticities, then the frontier is optimal even among ***stochastic*** mechanisms (Theorem 2). As we show, for profit maximization, the optimal prices can then be determined by the demand profile method that simply prices the upgrades according to the incremental demand curves along the frontier, and the frontier is a ***minimal*** optimal menu under mild assumptions (Corollary 2).

Importantly, our approach is agnostic to the space of allocations and allows for objectives beyond profit maximization. As applications, we show that the abstract results immediately lead to a variety of new results in important screening problems such as optimal bundling, optimal taxation, sequential screening, selling information, and monopoly regulation.

The basic intuition behind our results is simple. The surplus of an allocation determines what to offer in the absence of private information. The demand elasticity, as we show, is equivalent to an appropriate notion of dispersion in the values and hence indicates the level of information rents to the agents. Perhaps surprisingly, it turns out that these two very familiar and empirically estimable concepts together can pin down the optimal menu in general mechanism design problems with arbitrary allocations.

[1] For expositional simplicity, we assume zero marginal costs; when there are positive costs, we can

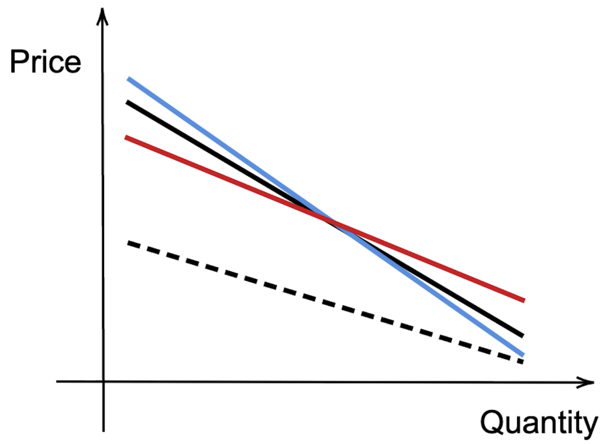

Figure 1: Illustration of Changes in Demand Curves. The dashed black line is the original demand curve. The solid lines are three demand curves that increase surplus but have different elasticity effects: Solid black line keeps the same elasticities (a *scaling*); solid red line increases elasticities (a *flattening*); solid blue line decreases elasticities (a *steepening*).

As a simple illustration of the surplus-elasticity frontier, consider the classic example of selling a single good with different lotteries. A well-known result by Myerson (1981) and Riley and Zeckhauser (1983) shows that a single posted price is optimal and lotteries are not needed. An alternative way to see it via Theorem 1 is to observe that different lotteries are simply ***scalings*** of the original demand curve and have no effect on its elasticity. Thus, the surplus-elasticity frontier consists of only the degenerate lottery of getting the good with probability 1, recovering the optimality of a posted price.

As another simple illustration, consider a monopolist who offers a product mix. Each product x has two different dimensions of features, say ***functionality*** and ***design***, $(\alpha, \beta) \in [0,1]^2$. For any type $t \in [1,2]$ consumer, their value is given by

$$v\big((\alpha,\beta),t\big) = (\alpha + t)^{\beta}.$$

Both features increase the consumers' values for the products, but the functionality dimension α ***flattens*** the demand curve whereas the design dimension β ***steepens*** the demand curve.[2] Figure 1 illustrates. Thus, the surplus-elasticity frontier consists of only

incorporate them by redefining the demand curves as net of production costs, with price elasticities defined with respect to the cost-adjusted demand curve.

[2]Indeed, the demand curve for each product is simply

$$P(x,q) = (\alpha + F^{-1}(1-q))^{\beta}$$

where F is the distribution of t, and hence increasing β rotates the demand curves clockwise, whereas increasing α rotates the demand curve counterclockwise.

$\{(1,\beta)\}_{\beta\in[0,1]}$. Theorem 1 says that the monopolist should offer a menu of products in which all the products have full functionality but they differ in the details of the design. Indeed, the incremental demand curves along the frontier $\{(1,\beta)\}_{\beta\in[0,1]}$ are also increasingly more inelastic as β increases, and thus by Theorem 2, $\{(1,\beta)\}_{\beta\in[0,1]}$ is optimal even if we allow for stochastic mechanisms, with optimal prices determined by upgrade pricing that prices against each incremental demand curve. The optimal prices depend on the distribution of consumer types t (e.g., the income distribution), but the same menu stays optimal—indeed, as we discuss, the surplus-elasticity frontier is ***agnostic*** to the type distribution.

Perhaps surprisingly, the surplus-elasticity frontier remains optimal even if the principal does not just care about the revenue—Theorem 1 turns out to hold identically even if the designer has redistributive preferences such as in optimal taxation or redistributive allocation problems, as long as the welfare weights are ***redistributive*** in that they are weakly decreasing in the vertical types. Whenever a surplus-elasticity frontier exists, the ***same*** frontier is optimal simultaneously for a broad set of objectives. In particular, in the above examples, the same menu would remain optimal even if the designer has redistributive preferences, and would remain optimal even among stochastic mechanisms.

Because the pointwise comparisons of the demand curves and their elasticities are both partial orders, the surplus-elasticity frontier need not exist in general. However, we can generalize our comparisons by comparing a given demand curve to a collection of possibly randomized demand curves. We say that a collection of (randomized) demand curves ***covers*** a given demand curve if they single-cross the given one from below individually and cover the graph of it collectively. Figure 2 illustrates. A menu of products $\mathcal{X}^\star$ is a ***generalized frontier*** if the elements can be ordered such that there exists a corresponding stochastically ordered set of lotteries $\mathcal{A}\in\Delta(\mathcal{X}^\star)$ such that *(i)* the demand curve of any element x can be covered by a collection of (randomized) demand curves from $\mathcal{A}$ and *(ii)* the demand curves of the elements in $\mathcal{X}^\star$ satisfy increasing differences. Our proof of Theorem 1 actually shows that any surplus-elasticity frontier is a generalized frontier and any generalized frontier must be robustly optimal. Thus, we can readily weaken the requirement of finding a surplus-elasticity frontier to finding a generalized frontier (Theorem 3). Moreover, the conclusion of Theorem 2 also applies—if the incremental demand curves are ordered by elasticities along the generalized frontier, then it is also optimal among stochastic mechanisms.

Besides the general economic insights, our main results turn out to be extremely helpful for identifying the optimal menus in a variety of important screening problems. We now explain our applications in detail.

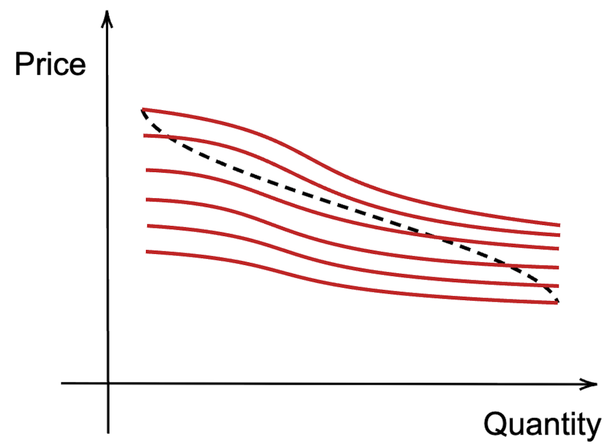

Figure 2: Illustration of Covering of a Demand Curve. The dashed black curve is the original demand curve. The collection of solid red demand curves *covers* the original one by single-crossing it from below individually and covering it collectively.

Optimal Bundling. Consider a designer who allocates bundles of a finite set of goods, where the consumers are vertically ordered and can have non-additive values. A common selling mechanism in such a setting is ***nested bundling*** where more expensive bundles include all the goods of the less expensive ones. An immediate application of Theorem 1 and Theorem 2 identifies a new condition (***robust nesting condition***) for when nested bundling is optimal (Proposition 5). Nested bundling becomes optimal precisely when there is a nested menu that can span the surplus-elasticity frontier. The robust nesting condition complements and illuminates the demand conditions in Yang (2025) which depend on the type distribution and hold for profit maximization. Indeed, Yang (2025) considers the partial order defined by *(i)* set inclusion and *(ii)* sold-alone quantities, and shows that if the undominated bundles with respect to this partial order are nested, then this nested menu is optimal for profit maximization under regularity conditions. This can be understood via our results since a larger bundle generates a pointwise higher demand curve and a higher sold-alone quantity corresponds to a demand curve with a larger elastic region. Moreover, by virtue of our main results, Proposition 5 shows that once there exists such a frontier, then the menu must be optimal for all type distributions and all redistributive welfare weights. This opens the door to apply results from the bundling literature to redistributive allocation problems. In particular, as another special case, we show that the ***stochastic ratio-monotonicity condition*** of Haghpanah and Hartline (2021) actually implies that pure bundling is optimal even if the designer is a social planner who has ***redistributive*** preferences that favor agents who have lower values for the grand bundle (Corollary 3)—in this case, even if the designer cares more about poorer agents, the optimal mechanism is to use a blunt instrument by simply lowering the price for the grand bundle rather than offering smaller bundles to target poorer agents. This is precisely because the grand bundle itself forms the surplus-elasticity frontier under the ratio-monotonicity condition—the best way to control information rents for the richer agents turns out to fully exclude the poorer agents.

Optimal Taxation. Consider the optimal nonlinear taxation problem with quasilinear preferences (Diamond 1998). Suppose the government has the ability to combine the ***tax instruments*** with ***costly screening*** to induce self-targeting in transfers. Would it be beneficial to do so? Nichols and Zeckhauser (1982) shows that in such an environment, it is indeed ***possible*** for the government to use socially wasteful ordeals to improve the net welfare because the ordeals can help with screening. Building on Nichols and Zeckhauser (1982), we further explore this question by applying our main results. We show that there are two key forces that determine whether an ordeal should be used in the presence of tax instruments (Proposition 6): *(i)* As anticipated by Nichols and Zeckhauser (1982), ordeals cannot be helpful if the higher ability types have lower costs from performing the ordeals; *(ii)* even if the direction of sorting is correct, ordeals still cannot be helpful if an appropriately defined ***take-up elasticity*** with respect to a small tax cut becomes lower when an ordeal is imposed. Intuitively, the reduction in surplus by an ordeal must be compensated via an increase in the screening effectiveness. In the absence of an ordeal, at the optimal tax schedule, the government would want to offer a marginal tax cut on the lower types which would also increase their labor supply but worry about the higher types shirking—an ordeal with ***high*** take-up elasticity is helpful when combined with such a marginal tax cut—it increases the labor supply sensitivity of the lower types and dampens the incentive for shirking by the higher types. An ordeal with ***low*** take-up elasticity cannot justify its distortion and hence results in taxation alone being optimal. As we discuss, the comparison here differs from the reforms that directly provide cash transfers conditional on the ordeals as studied in Yang, Dworczak, and Akbarpour (2024) and Rafkin, Solomon, and Soltas (2023), and highlights an appropriate notion of elasticity for assessing the screening gains from ***combining*** ordeals and tax instruments.

Sequential Screening. Consider the sequential screening problem introduced in Courty and Li (2000). A seller contracts with a buyer on the sale of a good after the buyer has some initial estimate t of their values θ but before they fully learn their values θ. This is a canonical example of ***dynamic mechanism design*** (Pavan, Segal, and Toikka 2014). In general, the optimal mechanism is sophisticated, involving a continuum of options, e.g., specifying different up-front fees for the possibility of having different refund options (Courty and Li 2000). In practice, however, even in many markets with buyer uncertainty, the selling mechanism is still very simple, often just a single ***nonrefundable posted price***. When is such a simple mechanism optimal in sequential screening? Applying our main results, we show that a nonrefundable posted price is optimal if higher ex-ante types not only have higher expected willingness to pay (in the sense of FOSD) but also have greater uncertainty about their ex-post valuations in the sense of the ***dispersive order on the log scale*** (Proposition 7). Perhaps surprisingly, this is the case even though the principal could instead offer refunds to better target high types and extract surplus through a higher up-front fee. The key intuition is that since ex-post values are FOSD increasing in the ex-ante types, such a refund would have a leakage that attracts the lower types, which will then involve too much refund and eventually lead to a lower revenue for the principal. Note that here a refund with a high up-front fee actually ***reduces*** the ex post surplus by reducing the consumption of consumers with lower ex post values, and hence it can be justified only by increasing the screening effectiveness by our main results. Yet under the log-dispersion order—which captures the elasticity of how ex post values vary with ex ante types—the screening effectiveness cannot be increased and hence a nonrefundable posted price is optimal. Even though sequential screening is a well-studied problem, we are not aware of any primitive condition under which a nonrefundable posted price is optimal. One reason is methodological: much of the dynamic mechanism design literature relies on the first-order approach under regularity assumptions that typically deliver fully separating allocations, so pooling/bunching mechanisms such as a nonrefundable posted price can be difficult to obtain without explicitly tracking the global incentive constraints.

Selling Information. Fix a finite state space. Consider a designer selling ***Blackwell experiments*** to an agent who has a decision problem to solve, just like in Bergemann, Bonatti, and Smolin (2018). However, unlike Bergemann, Bonatti, and Smolin (2018), suppose the agent has no private information about their prior belief about the state, but private information about their decision problems. Recall that every decision problem can be represented as a ***convex indirect utility function*** of the posterior beliefs. Therefore, each agent of type t is associated with a convex indirect utility function $u(\mu, t)$. We assume that the types are vertically ordered so that these indirect utility functions are increasing in the type t compared to the no-information outside option. As observed in Bergemann, Bonatti, and Smolin (2018), the space of Blackwell experiments is extremely rich as an allocation space. A commonly used class of signals is called ***truth-or-noise***

signals (Lewis and Sappington 1994)—such a signal either fully reveals the state with some probability α or draws a completely noisy signal according to the prior with probability $1-\alpha$. The truth-or-noise signals are very simple in that they are controlled by a one-dimensional index α, are Blackwell-monotone in α, and generate posterior belief distributions with sparse supports. When are they optimal? Applying our main results, we show that in any symmetric environment, selling a menu of truth-or-noise signals is optimal if the agent's indirect utility function $u(\cdot, t)$ is ***increasingly convex*** (in the Arrow-Pratt sense) in the posterior beliefs as the type increases (Proposition 8). This is the case where a higher vertical type (who has overall higher willingness to pay) tends to have decision problems that systematically depend ***more sensitively*** on the state. In this case, the optimal menu—among menus of arbitrary Blackwell experiments—turns out to be a menu of truth-or-noise signals with a higher price for a higher signal-to-noise ratio. We also show that in the opposite case where the agent's indirect utility function $u(\cdot, t)$ is ***increasingly concave*** in the posterior beliefs as the type increases, simply selling ***full information*** is optimal. This is the case where a higher vertical type tends to have decision problems that systematically depend ***less sensitively*** on the state. As this application shows, the existence of a screening frontier does not require a *specific* directional assumption but only the existence of *some* systematic structure of the agent's preferences.

Monopoly Regulation. Our last application studies the classic problem of monopoly regulation, just like in Baron and Myerson (1982). However, unlike Baron and Myerson (1982), suppose the monopolist has ***rich data*** about the consumers in the sense that the monopolist has a fine segmentation of the consumers and will engage in third-degree price discrimination in the absence of any regulation. Following Strack and Yang (2025), we model the data-rich monopolist as observing the consumers' values θ and designing any ***pricing rule*** that specifies a randomized price for each consumer θ. The monopolist privately observes their type t which determines their cost $c(\theta, t)$ for serving any consumer type θ. As in Strack and Yang (2025), this allows for interdependent costs which are very common in settings involving consumer data (e.g., credit markets and insurance markets). Consider a regulator who contracts with the monopolist on the pricing rules. As in Baron and Myerson (1982), the regulator wants to maximize a weighted sum of the consumer and producer surplus but does not observe the cost type t of the monopolist. The regulator hence will post a ***menu of pricing rules*** with their associated subsidies or taxes. We say that a mechanism is a ***uniform-pricing regulation*** if the menu consists of only deterministic pricing rules that do not depend on the consumers' type θ. Equivalently, such a regulation imposes the ***uniform pricing constraint*** for the firm and then simply taxes or subsidizes the firm according to the uniform price the firm chooses—under such a regulation, the determination of taxes/subsidies exactly collapses to the model of Baron and Myerson (1982). Of course, without the tax instrument, it is clear that a uniform pricing constraint increases consumer surplus since otherwise the consumers are fully exploited. However, with a tax instrument, are uniform pricing regulations optimal? If not, what regulations are optimal? Applying our main results, we show that if the ***total surplus*** $\theta - c(\theta, t)$ which depends on both consumer value θ and monopolist type t is ***log-submodular***, then uniform pricing regulations are optimal (Proposition 9). In the opposite case, we show that if the total surplus is strictly ***log-supermodular***, then uniform pricing regulations cannot be optimal. Indeed, in that case, we show that ***discriminatory pricing regulations*** are optimal—the regulator lets the monopolist fully utilize consumer data to price discriminate and design a suitable tax schedule to transfer the surplus to the consumers. As we discuss, our conditions capture the degree of adverse selection in the markets—in markets with ***advantageous selection*** or ***mild adverse selection***, we have uniform pricing regulations being optimal; in markets with ***severe adverse selection***, the optimal regulation must involve discriminatory pricing. Our results apply here because the problem can again be equivalently cast as a screening problem with a rich allocation space—the space of consumer-contingent pricing rules—by our main results, with a right comparison of elasticities with respect to the monopolist type (which is exactly what the log-submodular/log-supermodular surplus condition captures), the optimal mechanism is described by a screening frontier in our sense which turns out to be either uniform pricing regulations or discriminatory pricing regulations.

Related Literature. We view our main contributions as twofold. First, our conceptual contribution is to explicitly identify two intuitive channels, surplus and dispersion, measured by the levels and the elasticities of the sold-alone demand curves that together pin down the optimal menu in a general screening environment. Second, our methodological contribution is to show how various screening problems with rich allocation spaces can be characterized by identifying a suitable screening frontier that balances creating surplus and controlling information rents. We build on a large literature on multidimensional screening, most of which focuses on the optimal bundling problem. The model is traditionally set up as additive values with multidimensional types (McAfee, McMillan, and Whinston 1989; Rochet and Chone 1998; McAfee and McMillan 1988; Manelli and Vincent 2006; Daskalakis, Deckelbaum, and Tzamos 2017) and turns out to be analytically and computationally intractable (Rochet and Stole 2003; Daskalakis et al. 2014; Lahr and Niemeyer 2024). A recent line of work relaxes the additivity requirement and explores the consequences of ***comonotonic type space*** where the consumer types are totally ordered (Haghpanah and Hartline 2021; Ghili 2023; Yang 2025). Our model builds heavily on this line of literature by fully embracing the one-dimensional type space and fully relaxing the allocation space requirement. In contemporaneous work, Haghpanah and Siegel (2025) also study a model with an arbitrary allocation space but restrict attention to binary types, which allows them to study environments where types are not ordered. Relatedly, Pernoud and Yang (2025) characterize optimal mechanisms with non-ordered linear type spaces assuming additive values and provide a microfoundation for comonotonic types using equilibrium learning; Frick, Iijima, and Ishii (2024) study the monopolist's problem with precise consumer data and provide structural results with linear type spaces. Our proof method builds on and further develops the non-Myersonian approach in Yang (forthcoming). In particular, we generalize the ***downward sufficiency theorem*** in Yang (forthcoming) to allow for arbitrary redistributive welfare weights which, as our applications show, opens the door to apply results from multidimensional screening to recent redistributive allocation problems as well as classic optimal taxation and monopoly regulation problems.[3]

The remainder of the paper proceeds as follows. Section 2 presents the model. Section 3 presents the main results. Section 4 sketches the main proofs. Section 5 presents the applications. Section 6 concludes. Appendix A contains the omitted proofs.

2 Model

A principal screens an agent. The allocation space is a compact metric space $\mathcal{X} \ni x$. The agent has a one-dimensional type $t \in \mathcal{T} := [\underline{t}, \bar{t}]$, distributed according to a continuously differentiable F with a positive density f. The agent has quasilinear preferences, given by

$$v(x,t) - p$$

where p is the payment. We assume that the agent's value function $v(x,t)$ is strictly increasing and differentiable in t with $v_t(x,t)$ continuous on $\mathcal{X} \times \mathcal{T}$.

The principal wants to maximize a weighted average of welfare and revenue. By the revelation principle, it is without loss of generality to consider direct-revelation mecha-

[3] See also Doligalski et al. (2025) for applying multidimensional screening techniques to taxation problems. There is also a substantial literature in public finance that studies taxation with multiple instruments (e.g., Ferey, Lockwood, and Taubinsky 2024) or multidimensional heterogeneity (e.g., Golosov and Krasikov 2025).

nisms. A ***(direct-revelation) mechanism*** is a measurable map

$$(x, p) : \mathcal{T} \to \mathcal{X} \times \mathbb{R}$$

that satisfies the usual incentive compatibility (IC) and individual rationality (IR) constraints:

$$\begin{aligned} &v(x(t), t) - p(t) \geqslant v(x(\hat{t}), t) - p(\hat{t}) && \text{for all } t, \hat{t} \in \mathcal{T}\,; \\ &v(x(t), t) - p(t) \geqslant 0 && \text{for all } t \in \mathcal{T}\,. \end{aligned}$$

Let $\varnothing$ denote the ***outside option*** that gives all types 0 payoff. To simplify notation, we augment $\mathcal{X}$ with the outside option $\varnothing$. Two mechanisms are ***equivalent*** if they differ on a zero-measure set of types. Formally, the principal's objective is given by a weighted average of welfare and revenue:

$$\mathbb{E}\left[\lambda(t)\big(v(x(t), t) - p(t)\big)\right] + \mathbb{E}\left[p(t)\right],$$

where $\lambda(\cdot) \geqslant 0$ is a continuous ***welfare weight function***. We assume $\mathbb{E}[\lambda(t)] \leqslant 1$.[4] We assume that the welfare weights are ***redistributive*** in the sense that $\lambda(t)$ is non-increasing in t.[5] With constant λ, this objective captures any weighted average of surplus and revenue. With varying λ, this objective allows for optimal taxation (corresponding to $\mathbb{E}[\lambda(t)] = 1$) or redistributive allocation applications where the welfare weight is weakly higher for the lower types. When $\lambda = 0$, the principal is profit maximizing, in which case if there is a production cost $C(x)$ for providing allocation x, we can without loss of generality normalize it to be 0 by defining $\tilde{v}(x, t) = v(x, t) - C(x)$.[6]

A ***menu*** $\mathcal{Y}$ is any subset of $\mathcal{X}$ (which we assume also includes the outside option $\varnothing$).[7] A menu $\mathcal{Y}$ is ***optimal*** if there exists an optimal mechanism (x, p) such that $x(t) \in \mathcal{Y}$ for all types t. Since the space of allocations $\mathcal{X}$ is abstract, we can allow for stochastic mechanisms by directly encoding them in $\mathcal{X}$. However, some of our results have the feature that they only require comparisons for the elements in $\mathcal{X}$ to imply optimality even if the allocation space is extended to $\Delta(\mathcal{X})$. Thus, we explicitly say that a menu $\mathcal{Y}$ is ***optimal among stochastic mechanisms*** if there exists a mechanism (x, p) such that $x(t) \in \mathcal{Y}$ for

[4]If $\mathbb{E}[\lambda(t)] > 1$, the principal's problem would have unbounded value by scaling up a lump sum transfer to the agent.

[5]Since the agents' preferences are pinned down by t, the formulation is equivalent to having (λ, t) jointly distributed with $\mathbb{E}[\lambda \mid t]$ being non-increasing in t.

[6]In particular, all of our demand curves are defined as net of the production costs.

[7]To simplify notation, we omit the inclusion of $\varnothing$ from a menu whenever it is clear.

all types t and (x, p) is optimal among all IC and IR mechanisms mapping from $\mathcal{T}$ to $\Delta(\mathcal{X}) \times \mathbb{R}$ (assuming risk neutrality).

Let $P(x, q)$ be the ***demand curve*** for allocation x when x is sold alone:

$$P(x, q) := v(x, F^{-1}(1-q)).$$

Let $\eta(x, q)$ be the corresponding ***elasticity curve*** for allocation x:

$$\eta(x, q) := \left|\frac{\mathrm{d}\log P(x, q)}{\mathrm{d}\log q}\right|^{-1}.$$

Note that a higher η represents a more elastic demand curve. For any $x, x' \in \mathcal{X}$, we write

$$x \preceq_{\text{surplus}} x' \;\text{ if } P(x, q) \leqslant P(x', q) \text{ for all } q;$$

and

$$x \preceq_{\text{elasticity}} x' \;\text{ if } \eta(x, q) \leqslant \eta(x', q) \text{ for all } q.$$

We write $\prec_{\text{surplus}}$ or $\prec_{\text{elasticity}}$ if the above inequalities hold strictly for all $q \in (0, 1)$. Note that for the above elasticity comparison to be well defined, we implicitly assume that $P(x, q) > 0$, i.e., $v(x, t) > 0$ for any $x \neq \varnothing$. We will relax this assumption for our notion of generalized frontiers which replace elasticity comparisons with single-crossing comparisons of demand curves.

3 Main Results

3.1 Optimality of Surplus-Elasticity Frontier

A compact subset $\mathcal{X}^\star \subseteq \mathcal{X}$ is a ***surplus-elasticity frontier*** if:

(i) for any $x \notin \mathcal{X}^\star$, there exists $x' \in \mathcal{X}^\star$ such that

$$x \preceq_{\text{surplus}} x' \text{ and } x \preceq_{\text{elasticity}} x' \;;$$

(ii) there exists an index set $S \subset \mathbb{R}$ such that $\mathcal{X}^\star = \{x_s\}_{s \in S}$ and for all $s < s'$

$$x_s \prec_{\text{surplus}} x_{s'} \text{ and } x_s \succ_{\text{elasticity}} x_{s'}.$$

Our first main result says that any surplus-elasticity frontier is an optimal menu.

Theorem 1. *Any surplus-elasticity frontier is an optimal menu.*

The proof of Theorem 1 is in the appendix. We sketch the proof in Section 4.

We make a few remarks about this result here.

First, note that

$$x \preceq_{\text{surplus}} x' \iff v(x,t) \leqslant v(x',t) \text{ for all } t$$

and

$$x \preceq_{\text{elasticity}} x' \iff \frac{\mathrm{d}}{\mathrm{d}t}\log v(x,t) \geqslant \frac{\mathrm{d}}{\mathrm{d}t}\log v(x',t) \text{ for all } t.$$

Thus, the definition of a surplus-elasticity frontier does not depend on the type distribution F. Moreover, note that the definition of the frontier is also agnostic to the welfare weights $\lambda(t)$. Perhaps surprisingly, Theorem 1 asserts that the ***same*** frontier would be optimal across all type distributions F and all weakly decreasing welfare weights $\lambda(t)$.

Second, to further illustrate the intuition of the surplus-elasticity frontier, we connect the two orders to two stochastic orders. Let $\preceq_{\text{fosd}}$ denote the ***first-order stochastic dominance order***. Let $\preceq_{\text{disp}}$ denote the ***dispersive order*** (Müller and Stoyan 2002), i.e., two random variables $X \preceq_{\text{disp}} Y$ if $F_X^{-1}(q') - F_X^{-1}(q) \leqslant F_Y^{-1}(q') - F_Y^{-1}(q)$ for all $q < q' \in [0,1]$. Let V^x denote the random variable induced by $v(x,t)$. Note that

$$x \preceq_{\text{surplus}} x' \iff V^x \preceq_{\text{fosd}} V^{x'}$$

and

$$x \preceq_{\text{elasticity}} x' \iff \log V^x \succeq_{\text{disp}} \log V^{x'}.$$

In particular, the elasticity comparison is equivalent to the dispersion comparison according to the dispersive order on the log scale. Intuitively, this captures the level of information rents to the agents.[8]

Third, along the surplus-elasticity frontier, note that the agent's preferences satisfy ***increasing differences***: For any $s < s'$,

$$x_s \prec_{\text{surplus}} x_{s'} \text{ and } x_s \succ_{\text{elasticity}} x_{s'} \implies v(x_{s'},t) - v(x_s,t) \text{ is strictly increasing in } t.$$[9]

Thus, the pricing problem along the frontier can always be solved via well-known one-dimensional screening methods.

[8] See also Yang, Dworczak, and Akbarpour (2024) who show that the log-level dispersion is important in understanding the rent an agent can secure in one-dimensional linear screening problems.

[9] Indeed, for all $s < s'$, we can write $v_t(x_s,t) = \frac{v_t(x_s,t)}{v(x_s,t)} \cdot v(x_s,t) < \frac{v_t(x_{s'},t)}{v(x_{s'},t)} \cdot v(x_{s'},t) = v_t(x_{s'},t)$, where the strict inequality uses that the menu $\mathcal{X}^\star$ is a surplus-elasticity frontier and that $0 < v(x_s,t) \leqslant v(x_{s'},t)$.

Fourth, a particular special case is when the frontier consists of only one element, say $\overline{x}$. That is, element $\overline{x}$ generates more surplus than any other element in $\mathcal{X}$ and has a more elastic demand curve than that of any other element in $\mathcal{X}$. Theorem 1 says that it is optimal to simply offer this maximal-surplus element, i.e., no screening distortion is needed. This connects to the optimality of pure bundling as studied in Haghpanah and Hartline (2021). Indeed, note that

$$x \preceq_{\text{elasticity}} x' \iff v(x,t)/v(x',t) \text{ is nondecreasing in } t\,.$$

The main result of Haghpanah and Hartline (2021) establishes that if $v(b,t)/v(\overline{b},t)$ is nondecreasing in t for all bundles b, then selling only the grand bundle $\overline{b}$ is profit maximizing. This can be seen from Theorem 1 by taking $\mathcal{X}$ to be the space of lotteries over bundles, and observing that mixtures of deterministic bundles yield a randomized bundle with demand curve pointwise below that of the grand bundle, and pointwise more inelastic than that of the grand bundle given the ratio-monotonicity condition—the surplus-elasticity frontier consists of only the grand bundle. Perhaps surprisingly, Theorem 1 shows that the identical ratio-monotonicity condition of Haghpanah and Hartline (2021) actually implies the optimality of pure bundling (or equivalently, no screening distortion besides exclusion) for arbitrary redistributive welfare weights. We discuss more applications to optimal bundling in Section 5.

Fifth, in the above application to optimal bundling, we take the allocation space to be the space of lotteries over bundles, but in general, it turns out that we may be able to compare only elements in $\mathcal{X}$ and conclude the optimality of the menu in a much larger allocation space $\Delta(\mathcal{X})$. Moreover, in the above application, it is immediate that the single-option menu is minimal optimal in the sense that no option in the menu can be removed. In general, the minimal optimal menu would be a subset of the surplus-elasticity frontier, but as we show momentarily, there are conditions under which the surplus-elasticity frontier is a minimal optimal menu for profit maximization.

Ordered Surplus and Elasticities. In most of our applications, the existence of a surplus-elasticity frontier will be shown by construction, which is then an optimal menu by Theorem 1. However, when both the demand curves and the elasticity curves can be totally ordered themselves, the existence of surplus-elasticity frontier is guaranteed just like a standard Pareto frontier:

Corollary 1. *Consider any finite $\mathcal{X} = \{(a_i, b_i)\}_{i=1}^n \subset \mathbb{R}^2$. Suppose $(a,b) \prec_{\text{surplus}} (a',b')$ whenever $a < a'$, and $(a,b) \prec_{\text{elasticity}} (a',b')$ whenever $b < b'$. Then, the surplus-elasticity frontier* *exists and is optimal.*

Which Product Dimension to Distort? As another illustration of Theorem 1, we generalize the opening example about distortion of product features:

Proposition 1. *Let $\mathcal{X} = [0,1]^2$. Suppose that $v(x_1, x_2, t)$ is strictly increasing in (x_1, x_2), log-submodular in (x_1, t), and strictly log-supermodular in (x_2, t). Then, $\{(1, x_2)\}_{x_2 \in [0,1]}$ is an optimal menu.*

Proof. Since $v(x_1, x_2, t)$ is nondecreasing in x_1 and log-submodular in (x_1, t), note that

$$(x_1, x_2) \preceq_{\text{surplus}} (1, x_2), \qquad (x_1, x_2) \preceq_{\text{elasticity}} (1, x_2).$$

Moreover, since $v(x_1, x_2, t)$ is strictly increasing in x_2 and strictly log-supermodular in (x_2, t), note that

$$(1, x_2) \prec_{\text{surplus}} (1, x_2'), \qquad (1, x_2) \succ_{\text{elasticity}} (1, x_2').$$

for any $x_2 < x_2'$. It follows immediately that $\{(1, x_2)\}_{x_2 \in [0,1]}$ is an optimal menu by Theorem 1. □

Recall that in the opening example we have the utility function of the form $v(x_1, x_2, t) = (x_1 + t)^{x_2}$, which is log-submodular in (x_1, t) and log-supermodular in (x_2, t)—these notions exactly generalize the idea that an increase in feature x_1 (e.g., basic functionality) ***flattens*** the demand curves, while an increase in feature x_2 ***steepens*** the demand curve (e.g., product design), leading to no distortion of dimension x_1 in the surplus-elasticity frontier.[10]

3.2 Stochastic Optimality and Minimal Optimality

Now, to state the second main result, for any $x \prec_{\text{surplus}} x'$, we define the ***incremental demand curve*** as

$$P(x', q \mid x) := P(x', q) - P(x, q),$$

[10] In the case where $v(x_1, x_2, t)$ does not depend on x_2, Proposition 1 itself nests one-dimensional screening results on the profitability of price discrimination as special cases (e.g., Johnson and Myatt 2003; Anderson and Dana Jr 2009).

and the ***incremental elasticity curve*** as

$$\eta(x', q \mid x) = \left| \frac{\mathrm{d} \log P(x', q \mid x)}{\mathrm{d} \log q} \right|^{-1}.$$

A menu $\mathcal{Y}$ has the ***ordered incremental elasticity*** property if $\mathcal{Y} = \{x_s\}_{s \in S}$ for some index set $S \subset \mathbb{R}$ such that for all $s_1 < s_2 < s_3$ we have

$$\eta(x_{s_2}, q \mid x_{s_1}) > \eta(x_{s_3}, q \mid x_{s_2}) \text{ for all } q \in (0,1).$$

In the case of finite S, this condition can be verified by comparing only the ***adjacent elements***; similarly, in the case of continuous S where $P(x_s, q)$ is continuously differentiable in s, this condition can also be verified by comparing elasticities of the infinitesimal ***marginal demand curves*** $\frac{\partial}{\partial s} P(x_s, q)$.

We say that a surplus-elasticity frontier is a ***strong frontier*** if it has the ordered incremental elasticity property.

Theorem 2. *Any strong frontier is optimal among stochastic mechanisms.*

The proof of Theorem 2 is in the appendix. We sketch the proof in Section 4.

Demand Profile and Minimal Optimality. The pricing problem on a surplus-elasticity frontier can always be solved by one-dimensional methods. Moreover, it turns out that if the frontier is a strong frontier and the objective is profit maximization, then the pricing problem can always be solved via a very simple method, the ***demand profile method***, which simply prices each upgrade against the incremental demand curve. To see this connection, suppose that the strong frontier is ***differentiable*** in the sense that $P(x_s, q)$ is continuously differentiable in s. Without loss of generality, normalize the index $s \in [0,1]$. Let $\Delta P(s,q) = \partial_s P(x_s, q)$ denote the marginal demand curve for the upgrade at s. Since the agent's preferences have strict increasing differences on the frontier, by the demand profile method (Wilson 1993), an upper bound on the revenue of any mechanism with allocations restricted to $\mathcal{X}^\star$ can be obtained by consecutively ***upgrade pricing***:

$$\int \max_{q \in [0,1]} (\Delta P(s,q) \cdot q) \, \mathrm{d}s,$$

where for each s, the inner maximization problem solves a separate upgrade pricing problem on the marginal upgrade. The above pointwise objective after taking log is

$$\log \Delta P(s,q) + \log q$$

which is submodular in (s,q) given that $\mathcal{X}^\star$ has the ordered incremental elasticity property.[11] Thus, there exists a selection of pointwise optimizers $s \mapsto q^\star(s)$ that is non-increasing and hence can be implemented via the profile of upgrade prices to achieve the above upper bound, i.e., the demand profile method is valid on the strong frontier.

Now, say that the frontier is ***interior*** if any monopoly quantity of the upgrade pricing problem is interior, $\operatorname{argmax}\{\Delta P(s,q)\cdot q\} \subset (0,1)$, and say that it is ***minimal optimal*** if for any closed subset $\mathcal{X}' \subset \mathcal{X}^\star$ that is also optimal, $\mathcal{X}' = \{x_s\}_{s\in S'}$ where S' has Lebesgue measure 1. Combining the above argument with Theorem 2, we immediately obtain:

Corollary 2 (Demand Profile and Minimal Optimality)**.** *Suppose the objective is profit maximization (i.e., $\lambda \equiv 0$). For any strong and differentiable frontier, the demand profile method is valid. If, in addition, the frontier is interior, then the frontier is minimal optimal.*

Stochastic versus Deterministic Mechanisms. An immediate consequence of Theorem 2 is the following result on when randomization does not help with one-dimensional screening problems:

Proposition 2. *Suppose $\mathcal{X} = [0,1]$ and $v(x,t)$ is continuously differentiable and strictly increasing in x with $v(0,t) = 0$. Then, if $v_x(x,t)$ is* strictly log-supermodular *or* weakly log-submodular, *then a deterministic mechanism is optimal.*

Proof. If $\log(v_x(x,t))$ is strictly supermodular, then $\frac{v_x(x_2,t)}{v_x(x_1,t)}$ is strictly increasing in t for all $x_2 > x_1$, which implies that the marginal demand curves resulting from $v_x(x,t)$ must have ordered elasticities where higher x has a more inelastic marginal demand curve. This also implies that

$$\frac{v(x_2,t)}{v(x_1,t)} = \frac{\int_0^{x_2} v_x(s,t)\,\mathrm{d}s}{\int_0^{x_1} v_x(s,t)\,\mathrm{d}s}$$

is strictly increasing in t for all $x_2 > x_1$. Thus, $\mathcal{X}$ is a strong frontier and hence optimal in $\Delta(\mathcal{X})$ by Theorem 2. Similarly, if $\log(v_x(x,t))$ is weakly submodular, then $\frac{v_x(x_2,t)}{v_x(x_1,t)}$ is weakly

[11] While the ordered incremental elasticity property does not specify the incremental elasticities as decreasing or increasing in s, note that they must be decreasing in s for a surplus-elasticity frontier.

decreasing in t for all $x_2 > x_1$, which implies that

$$\frac{v(x_2,t)}{v(x_1,t)} = \frac{\int_0^{x_2} v_x(s,t)\,\mathrm{d}s}{\int_0^{x_1} v_x(s,t)\,\mathrm{d}s}$$

is weakly decreasing in t for all $x_2 > x_1$. Then, the strong frontier is simply $\{1\}$, and thus a menu of the single option $x = 1$ with a suitable price is optimal among all stochastic mechanisms by Theorem 2. □

In contrast to the known result on the optimality of deterministic mechanisms by Strausz (2003), this result does not depend on the type distribution nor the redistributive welfare weights, which can certainly violate the Myersonian regularity conditions and lead to bunching.

3.3 Generalized Frontier

In this section, we generalize the comparisons allowed in the surplus-elasticity frontier in two main ways: *(i)* we will allow a collection of randomized demand curves to dominate a single demand curve in an appropriate sense, and *(ii)* we will also allow the demand curves on the frontier not to be pointwise ordered.

To state the definition, we use the notation $P(a,q)$ to write the pointwise mixture of the demand curve induced by a lottery $a \in \Delta(\mathcal{X})$. For any partially ordered set $(\mathcal{Y}, \preceq)$, its induced ***stochastic dominance order*** $\preceq_{\mathrm{st}}$ is defined on $\Delta(\mathcal{Y})$ such that $a \preceq_{\mathrm{st}} a'$ if $\mathbb{E}_a[h(x)] \leqslant \mathbb{E}_{a'}[h(x)]$ for all monotone functions h on $\mathcal{Y}$.

We say that x is ***covered*** by $\mathcal{A} \subset \Delta(\mathcal{X})$ if for any $q \in [0,1]$, there exists $a \in \mathcal{A}$ such that $P(a,\cdot)$ weakly single-crosses $P(x,\cdot)$ from below at q.[12] We say $\mathcal{X}^\star \subset \mathcal{X}$ is a ***generalized frontier*** if $\mathcal{X}^\star$ can be totally ordered by some $\preceq$ and there exists some compact $\mathcal{A} \subset \Delta(\mathcal{X}^\star)$ totally ordered by $\preceq_{\mathrm{st}}$ such that

(i) Every x is covered by $\mathcal{A}$;

(ii) $P(x',q) - P(x,q)$ is strictly decreasing in q for all $x \preceq x' \in \mathcal{X}^\star$.

As we discuss in Section 4, the proof of Theorem 1 actually shows that *(i)* a surplus-elasticity frontier is a generalized frontier and *(ii)* a generalized frontier is sufficient for robust optimality. Thus, we immediately obtain the following result from the proofs of Theorem 1 and Theorem 2:

[12] A function g ***weakly single-crosses*** h ***from below at*** k if $g(s) \leqslant h(s)$ for all $s \leqslant k$ and $g(s) \geqslant h(s)$ for all $s \geqslant k$.

Theorem 3. *Any generalized frontier is an optimal menu. Moreover, if the generalized frontier has surplus ordered demand curves that satisfy the ordered incremental elasticity property, then it is optimal among stochastic mechanisms.*

Screening with Contracts. Many contracts can be viewed as one-dimensional functions (we discuss in detail in Section 5). Here, we describe an abstract setting where the allocation space is the space of bounded measurable functions, and the agent's utility is a linear functional. Consider the space of bounded measurable functions mapping from $[0,1]$ to $[0,1]$, endowed with the L_1 norm. Suppose that the valuation of each type t for each function x is linear: $v(x,t) := \int x(s)u(s,t)\,\mathrm{d}G(s)$, where x maps from $[0,1]$ to $[0,1]$, and $u(s,t)$ is a positive bounded measurable function that is continuously differentiable in t with u_t uniformly bounded, and G is absolutely continuous with full support.

Proposition 3. *Suppose $u(s,t)$ is strictly increasing in t. If u is log-supermodular, then $\{\mathbb{1}_{[0,k]}\}_k$ is an optimal menu. If u is log-submodular, then $\{\mathbb{1}_{[k,1]}\}_k$ is an optimal menu.*

Proof. Suppose u is log-supermodular. Fix any function $x(s)$ and any type t. For each fixed t, consider $x^\star(s;t) = \mathbb{1}_{s \leqslant k(t)}$ where $k(t)$ is chosen such that

$$\int x^\star(s;t)u(s,t)\,\mathrm{d}G(s) = \int x(s)u(s,t)\,\mathrm{d}G(s).$$

Such a $k(t)$ exists by the intermediate value theorem. Now observe that, by Karlin and Rubin (1956), since $x^\star(\cdot\,;t)$, by construction, weakly single-crosses $x(\cdot)$ from above at $k(t)$, and u is log-supermodular, we have that

$$\Delta(z) := \int \big(x^\star(s;t) - x(s)\big)u(s,z)\,\mathrm{d}G(s) \text{ weakly single-crosses 0 from above at } t,$$

which implies that the demand curve resulting from the allocation $x^\star(\cdot\,;t)$ must weakly single-cross that resulting from $x(\cdot)$ from below at the quantile of type t. Since this holds for all t, it also implies that the demand curve of $x(\cdot)$ is covered by $\{\mathbb{1}_{[0,k(t)]}\}_t$ according to our definition. This implies that $\mathcal{X}^\star := \{\mathbb{1}_{[0,k]}\}_k$ is a generalized frontier, because along the ordering given by k, we have that

$$\int_0^{k_2} u(s,t)\,\mathrm{d}G(s) - \int_0^{k_1} u(s,t)\,\mathrm{d}G(s) = \int_{k_1}^{k_2} u(s,t)\,\mathrm{d}G(s),$$

which is strictly increasing in t. Thus, taking $\mathcal{A}$ as the collection of Dirac measures supported on each of $\{1_{[0,k]}\}_k$ shows that $\mathcal{X}^\star$ is a generalized frontier.[13] Hence, by Theorem 3, $\{\mathbb{1}_{[0,k]}\}_k$ must be optimal. The case of u being log-submodular is exactly symmetric, because we can construct $x^\star(s;t) = \mathbb{1}_{s \geqslant k(t)}$ where $k(t)$ is defined similarly. By the same single-crossing argument, this then certifies that $\{\mathbb{1}_{[k,1]}\}_k$ is a generalized frontier and hence optimal by Theorem 3. □

Ordeal Mechanisms. Consider a utility function $u(c,y,t)$ where there exists some y_0 such that $u(c,y_0,t) \geqslant u(c,y,t) \geqslant 0$ for all $(c,y) \in \mathcal{X}$. When is it optimal to set $y = y_0$? This question can be interpreted as when we should expect ***ordeal mechanisms*** to be used in either profit-maximizing contexts, or more generally when there are redistributive concerns. We will discuss in detail an application to optimal taxation in Section 5, but let us first illustrate how to think about the problem using a generalized frontier.

Proposition 4. *Suppose $u(c,y_0,t)$ is strictly supermodular and log-supermodular in $(c,t) \in [0,1]^2$, and satisfies $u(0,y_0,t) = 0$. If for any $(c,y) \in \mathcal{X}$,*

$$\eta(c,y,q) \leqslant \eta(c,y_0,q) \text{ for all } q\,,$$

then $\{(c,y_0)\}_c$ is an optimal menu (i.e., the optimal menu involves no costly screening).

Proof. Fix any (c,y) and any type t. By the intermediate value theorem, there exists $c_t^\star \in [0,c]$ such that $u(c,y,t) = u(c_t^\star,y_0,t)$. Now, note that

$$\frac{u(c,y,s)}{u(c_t^\star,y_0,s)} = \frac{u(c,y_0,s)}{u(c_t^\star,y_0,s)} \cdot \frac{u(c,y,s)}{u(c,y_0,s)}$$

is nondecreasing in s, since the first term is nondecreasing in s by log-supermodularity of $u(c,y_0,s)$ in (c,s) and the second term is nondecreasing in s by the elasticity comparison given in the statement. It follows immediately that the demand curve $P(c_t^\star,y_0,q)$ single-crosses $P(c,y,q)$ from below with the crossing point corresponding to the quantile of type t. Since this holds for all t, it follows immediately that any (c,y) is covered by $\{(c_t^\star,y_0)\}_t$ and hence by $\{(c,y_0)\}_c$. Given that $u(c,y_0,t)$ has strict increasing differences in (c,t), it then follows immediately that $\{(c,y_0)\}_c$ is a generalized frontier and hence optimal by Theorem 3. □

In monopolistic pricing contexts, this means that costly screening can only be profitable if, with an ordeal requirement, the demand for some good becomes more ***elastic***.

[13] Formally, our model assumes $\mathcal{X}$ is compact but as the proof of Theorem 3 shows, this is not needed as long as $\mathcal{X}^\star$ is compact.

In the optimal taxation context, as we discuss in Section 5, the demand elasticity comparison reduces to ***take-up elasticity*** comparison.

4 Proof Sketches for the Main Results

We first sketch the proof for Theorem 1 assuming a finite allocation space $\mathcal{X}$, and then discuss the proofs of Theorem 2 and Theorem 3 at the end. The appendix provides details and generalizes the proofs to arbitrary $\mathcal{X}$ via approximation.

The proof strategy for Theorem 1 builds on Yang (forthcoming). It consists of three steps.

Step 1. First, let $\mathcal{X}^\star$ be ordered by $\preceq_{\text{surplus}}$. Fix any IC and IR mechanism (x,p). We show that there exists a stochastic mechanism (which we refer to as a ***reconstruction***) (a,p) where $a:\mathcal{T}\to\Delta(\mathcal{X}^\star)$ is a reconstructed allocation rule using only lotteries over elements in the frontier $\mathcal{X}^\star$, and p is the same payment rule such that:

- (a,p) is ***downward incentive compatible***, i.e., it satisfies all downward incentive constraints with $t>\hat{t}\in\mathcal{T}$ (and satisfies all the IR constraints) ;
- (a,p) generates ***identical payoffs*** to each type as in (x,p) (and hence also generates the same payoff to the principal) assuming truthful reporting by the agent ;
- $\text{Ran}(a)\subseteq\Delta(\mathcal{X}^\star)$ consists of ***stochastically ordered*** lotteries in the sense that $\text{Ran}(a)$ can be totally ordered by $\preceq_{\text{st}}$.

Step 2. Second, we show that there exists another stochastic mechanism $(\tilde{a},\tilde{p})$ where $\tilde{a}:\mathcal{T}\to\Delta(\mathcal{X}^\star)$ is a stochastic allocation rule further modifying the allocation rule a, and $\tilde{p}$ is a modified payment rule such that

- $(\tilde{a},\tilde{p})$ is ***fully incentive compatible*** ;
- $(\tilde{a},\tilde{p})$ generates a ***weakly higher payoff*** for the principal compared to (a,p) ;
- $\text{Ran}(\tilde{a})\subseteq\Delta(\mathcal{X}^\star)$ continues to consist of ***stochastically ordered*** lotteries.

In particular, this step shows that the downward incentive constraints are ***sufficient*** for the optimal solution. This step is of independent interest; as we discuss, it generalizes the ***downward sufficiency theorem*** in Yang (forthcoming) that holds for profit maximization in any one-dimensional screening problems to allow for redistributive welfare weights.

Step 3. Third, we show that there exists a ***deterministic*** mechanism $(x^\dagger, p^\dagger)$ where $x^\dagger : \mathcal{T} \to \mathcal{X}^\star$ is a deterministic allocation rule that uses only elements in the frontier $\mathcal{X}^\star$, and $p^\dagger$ is a modified payment rule such that

- $(x^\dagger, p^\dagger)$ is ***fully incentive compatible***;
- $(x^\dagger, p^\dagger)$ generates a ***weakly higher payoff*** for the principal compared to $(\tilde{a}, \tilde{p})$.

In particular, this step shows that in one-dimensional screening problems, the possibility of using lotteries that are ordered by stochastic dominance cannot improve the principal's objective. We refer to this result as ***purification lemma***; its proof is shown by a monotone coupling argument.

4.1 Reconstruction Lemma

For all t, let

$$x^+(t) := \inf\left\{x' \in \mathcal{X}^\star : v(x', t) \geqslant v(x(t), t)\right\}, \quad x^-(t) := \sup\left\{x' \in \mathcal{X}^\star : v(x', t) < v(x(t), t)\right\},$$

where the inf and sup are defined with respect to the total order $\preceq_{\text{surplus}}$ on $\mathcal{X}^\star$. By construction, if $x(t) \in \mathcal{X}^\star$, then $x^+(t) = x(t)$. Moreover, for all t, we have

$$v(x^+(t), t) \geqslant v(x(t), t) > v(x^-(t), t).$$

Now, for all t, let

$$\alpha(t) := \frac{v(x(t), t) - v(x^-(t), t)}{v(x^+(t), t) - v(x^-(t), t)} \in (0, 1].$$

The $(\mathcal{X}^\star, x, p)$-***reconstruction*** is a stochastic mechanism defined by:

- assigning each reported type t a lottery over $\{x^+(t), x^-(t)\}$ with probability $\alpha(t)$ to allocation $x^+(t)$ and probability $1 - \alpha(t)$ to allocation $x^-(t)$;
- keeping the payment $p(t)$ for each reported type t unchanged.

By construction, the $(\mathcal{X}^\star, x, p)$-reconstruction generates the same payoff to each agent type t as (x, p) does (and hence the same payoff to the principal), assuming truthful reporting. The next lemma shows that the reconstruction preserves all downward IC constraints.

Lemma 1 (Reconstruction)**.** *Let $\mathcal{X}^\star = \{x_1, \ldots, x_m\}$ be a surplus-elasticity frontier. Let (x, p) be any deterministic mechanism. Let (a, p) be the $(\mathcal{X}^\star, x, p)$-reconstruction. Then (a, p) satisfies all downward IC constraints: for all $\hat{t} < t \in \mathcal{T}$, we have*

$$v(a(t), t) - p(t) \geqslant v(a(\hat{t}), t) - p(\hat{t}),$$

where $v(a, t) := \mathbb{E}_{x \sim a}[v(x, t)]$.

The proof is in the appendix. The intuition can be understood as follows. The modified stochastic allocation rule $a(t)$ is constructed in such a way that for each type t, we replace its deterministic allocation $x(t)$ with a binary lottery supported on $\{x^+(t), x^-(t)\} \subseteq \mathcal{X}^\star$ such that type t would be ***indifferent*** between the new lottery and the original option. Now, fix any type t. For $x(t) \notin \mathcal{X}^\star$, we know that there exists $\hat{x} \in \mathcal{X}^\star$ that dominates $x(t)$ in both the surplus order and the elasticity order. Crucially, the construction of x^+, x^- ensures that x^+, x^- must also dominate $x(t)$ in terms of the elasticity order since by construction they generate less surplus compared to $\hat{x}$ and hence are more elastic by property *(ii)* of the surplus-elasticity frontier. This implies that the lottery $a(t)$ generates a demand curve that is more elastic than the demand curve generated by $x(t)$, and moreover crosses the demand curve of $x(t)$ exactly at the quantile corresponding to type t (by the indifference of type t). Note that pointwise elasticity ordering implies ***single-crossing*** in the demand curves and hence the demand curve of $a(t)$ must single-cross that of $x(t)$ from below. Then, for any type $t' > t$, the deviation payoff of mimicking type t must have weakly decreased given that the higher type t' would be located toward the left on the quantity axis compared to type t (and hence receives a lower deviating payoff of consuming the lottery $a(t)$ compared to $x(t)$ given the single-crossing property of the demand curves).

Now, to see that the reconstruction indeed yields $\mathrm{Ran}(a)$ that consists of stochastically ordered lotteries, note that

$$\mathrm{Ran}(a) \subseteq \mathcal{A} := \Big\{ a \in \Delta(\mathcal{X}^\star) : a \in \Delta\big(\{x_{j-1}, x_j\}\big) \text{ for some } x_j \in \mathcal{X}^\star \Big\},$$

which is totally ordered by $\preceq_{\mathrm{st}}$.

4.2 Generalized Downward Sufficiency Theorem

Note that for the allocations on the frontier $\mathcal{X}^\star$, the agent's preferences satisfy strict increasing differences. Moreover, since the set of lotteries we reconstructed in **Step 1** can be totally ordered by stochastic dominance $\preceq_{\mathrm{st}}$, the agent's preferences continue to satisfy strict increasing differences: Indeed, for any $a \preceq_{\text{st}} a'$ and $a \neq a'$, we can write

$$v_t(a,t) = \mathbb{E}_{x\sim a}[v_t(x,t)] < \mathbb{E}_{x\sim a'}[v_t(x,t)] = v_t(a',t),$$

where the strict inequality follows from that $v_t(x_s,t)$ is strictly increasing in s given the strict increasing differences on the frontier. Therefore, we may view the set of lotteries $\mathcal{A}$ used in the reconstruction step as a one-dimensional allocation space.

A key technical result, which may be of independent interest, is the following downward sufficiency theorem that generalizes that in Yang (forthcoming) to allow for any redistributive welfare weights:

Theorem 4 (Downward Sufficiency). *Suppose that $v(a,t)$ is continuous and has strict increasing differences on $\mathcal{A} \times \mathcal{T}$, where $\mathcal{A}, \mathcal{T}$ are two compact subsets of $\mathbb{R}$. Then, for any type distribution, any redistributive welfare weights, and any $(a,p): \mathcal{T} \to \mathcal{A} \times \mathbb{R}$ that satisfies the IR constraints and the downward IC constraints, there exists $(\tilde{a},\tilde{p}): \mathcal{T} \to \mathcal{A} \times \mathbb{R}$ such that*

(i) $(\tilde{a},\tilde{p})$ *satisfies the IR constraints;*

(ii) $(\tilde{a},\tilde{p})$ *satisfies both the upward and downward IC constraints;*

(iii) $(\tilde{a},\tilde{p})$ *yields a weakly higher objective value.*

The proof is in the appendix. We sketch the proof for the finite-type case here.

As discussed in Yang (forthcoming), the above result does not hold if one imposes only ***local*** downward incentive constraints; however, perhaps surprisingly, the set of ***global*** downward incentives together is sufficient for determining the optimal solution in one-dimensional screening problems even with redistributive welfare weights.

Suppose $\mathcal{T} = \{t_1, \ldots, t_n\}$. Let a_i denote the allocation assigned to type t_i and p_i denote the payment assigned to type t_i. Let $\mu(t_i)$ and $\lambda(t_i)$ denote the type distribution and welfare weights, respectively. Suppose, without loss, that μ has full support. In the finite-type space, the problem we are considering is the following:

$$\max_{(a,p)} \sum_i \mu(t_i)\Big(\lambda(t_i)(v(a_i,t_i) - p_i) + p_i\Big) \qquad \textbf{(Downward-IC)}$$

subject to the downward IC constraints and IR constraints:

$$v(a_i,t_i) - p_i \geqslant v(a_j,t_i) - p_j \qquad \text{for all } i > j;$$
$$v(a_i,t_i) - p_i \geqslant 0 \qquad \text{for all } i.$$

The existence of a solution to the above problem can be shown by compactness arguments (recall that $\mathbb{E}[\lambda(t)] \leqslant 1$ and $\lambda(t)$ is non-increasing); see the appendix. As shown in Yang (forthcoming), every allocation rule $a \in \mathcal{A}^n$ is implementable with some transfers $p \in \mathbb{R}^n$ to be downward incentive compatible, i.e., feasible for the program (**Downward-IC**).

We prove Theorem 4 in two steps. In the first step, we show that every non-monotone $a \in \mathcal{A}^n$ can be weakly improved. Thus, it is without loss of optimality to focus on monotone $a \in \mathcal{A}^n$. In the second step, we show that for monotone $a \in \mathcal{A}^n$, the local downward incentive constraints must be binding, and hence with the optimal transfer subject to the downward incentive constraints, we in fact get a mechanism that is fully incentive compatible.

Step (A). Consider the set of all optimal solutions to (**Downward-IC**) program. For a given optimal solution a, let $o(a) \in \{1, \ldots, n\}$ be the first index at which $a_i > a_{i+1}$; if no such index exists, we set $o(a) = n$. This is the first index at which the allocation rule starts to decrease. We claim that there must exist a monotone optimal solution, i.e., $o(a) = n$. Suppose for contradiction that all optimal solutions satisfy $o(a) < n$.

Fix any optimal solution a with the largest index $o(a)$ among those. Let $i = o(a) < n$. First, we consider what happens if we replace (a_{i+1}, p_{i+1}) that is assigned to type t_{i+1} with (a_i, p_i) while keeping everything else fixed. Call the new allocation and payment rules $(\hat{a}, \hat{p})$. By construction, we have

$$a_1 \leqslant a_2 \leqslant \cdots \leqslant a_i \text{ and } a_i > a_{i+1}\,.$$

By definition $\{a_j : j < i\}$ were assigned to types t_j below type t_i, and hence obtainable by type t_i; the fact that t_i finds it downward incentive compatible to consume (a_i, p_i) implies that type t_i must prefer (a_i, p_i) over (a_j, p_j) for all $j < i$. At the same time, because $a_j \leqslant a_i$, by the increasing difference property $v(a, t)$, for all $j < i$, we have

$$v(a_i, t_{i+1}) - v(a_j, t_{i+1}) \geqslant v(a_i, t_i) - v(a_j, t_i) \geqslant p_i - p_j\,,$$

where the second inequality is due to IC$[i \to j]$ under (a, p). It follows that IC$[(i+1) \to j]$ constraints are satisfied for all $j < i + 1$ under the modified allocation and payment rule $(\hat{a}, \hat{p})$. Moreover, note that for any $k > (i + 1)$,

$$v(a_k, t_k) - p_k \geqslant v(a_i, t_k) - p_i = v(\hat{a}_{i+1}, t_k) - \hat{p}_{i+1}\,,$$

where the first inequality is due to IC$[k \to i]$ under (a, p) and the equality holds by construction. Therefore, IC$[k \to (i+1)]$ constraints are satisfied for all $i+1 < k$ under the modified allocation and payment rule $(\hat{a}, \hat{p})$. Since we have only changed the allocation and payment of type t_{i+1}, it follows that $(\hat{a}, \hat{p})$ is downward incentive compatible (it also satisfies all IR constraints given that $v(a,t)$ is nondecreasing in t). Then, the fact that (a,p) is optimal for (**Downward-IC**) implies that $(\hat{a}, \hat{p})$ cannot yield a strictly higher objective value for the principal. Thus, we must have

$$\lambda(t_{i+1})\big(v(a_{i+1}, t_{i+1}) - p_{i+1}\big) + p_{i+1} \geqslant \lambda(t_{i+1})\big(v(a_i, t_{i+1}) - p_i\big) + p_i\,,$$

or equivalently

$$\lambda(t_{i+1})\Big(v(a_{i+1}, t_{i+1}) - v(a_i, t_{i+1}) + p_i - p_{i+1}\Big) \geqslant p_i - p_{i+1}\,,$$

which implies that

$$\lambda(t_i)\Big(v(a_{i+1}, t_{i+1}) - v(a_i, t_{i+1}) + p_i - p_{i+1}\Big) \geqslant p_i - p_{i+1}\,,$$

since $\lambda(t_i) \geqslant \lambda(t_{i+1})$ by assumption and $v(a_{i+1}, t_{i+1}) - v(a_i, t_{i+1}) + p_i - p_{i+1} \geqslant 0$ given IC$[(i+1) \to i]$ under (a,p).

Then, since $a_{i+1} < a_i$, by the strict increasing difference property of $v(a,t)$, we have

$$v(a_{i+1}, t_{i+1}) - v(a_i, t_{i+1}) < v(a_{i+1}, t_i) - v(a_i, t_i)\,.$$

Thus, since $\lambda(t_i) \geqslant 0$, we must have

$$\lambda(t_i)\Big(v(a_{i+1}, t_i) - v(a_i, t_i) + p_i - p_{i+1}\Big) \geqslant p_i - p_{i+1}\,,$$

or equivalently

$$\lambda(t_i)\Big(v(a_{i+1}, t_i) - p_{i+1}\Big) + p_{i+1} \geqslant \lambda(t_i)\big(v(a_i, t_i) - p_i\big) + p_i\,,$$

and therefore, the principal's objective must have increased when changing the original (a,p) to $(\tilde{a}, \tilde{p})$ where the new mechanism is identical to the original mechanism except that we replace the allocation and payment (a_i, p_i) of type t_i with (a_{i+1}, p_{i+1}). We argue the mechanism $(\tilde{a}, \tilde{p})$ satisfies all IR and all downward IC constraints. First, note that by the above arguments, we have

$$v(a_{i+1}, t_i) - v(a_i, t_i) + p_i - p_{i+1} \geqslant v(a_{i+1}, t_{i+1}) - v(a_i, t_{i+1}) + p_i - p_{i+1} \geqslant 0\,,$$

and thus

$$v(a_{i+1}, t_i) - p_{i+1} \geqslant v(a_i, t_i) - p_i \geqslant 0\,.$$

Therefore, $(\tilde{a}, \tilde{p})$ satisfies all IR constraints. Moreover, for any $j < i$, we have

$$v(a_{i+1}, t_i) - p_{i+1} \geqslant v(a_i, t_i) - p_i \geqslant v(a_j, t_i) - p_j\,,$$

where the second inequality uses IC$[i \to j]$ under (a, p). Therefore, IC$[i \to j]$ constraints hold for all $j < i$ under $(\tilde{a}, \tilde{p})$. For any $k > i$, we have

$$v(a_k, t_k) - p_k \geqslant v(a_{i+1}, t_k) - p_{i+1} = v(\tilde{a}_i, t_k) - \tilde{p}_i\,,$$

where the first inequality uses IC$[k \to (i+1)]$ under (a, p), and the equality follows by construction. Therefore, the mechanism $(\tilde{a}, \tilde{p})$ is downward incentive compatible, and hence a feasible solution to (**Downward-IC**) program. Moreover, as we have argued, under $(\tilde{a}, \tilde{p})$ weakly improves the principal's objective—therefore, $(\tilde{a}, \tilde{p})$ must also be an optimal solution to (**Downward-IC**).

Now, let

$$r := \min\left\{j : a_j > a_{i+1}\right\}.$$

By construction $r \leqslant i$, and

$$a_1 \leqslant a_2 \leqslant \cdots \leqslant a_{r-1} \leqslant a_{i+1} < a_r \leqslant \cdots \leqslant a_i\,.$$

If $r < i$, then we must have that $o(\tilde{a}) = i - 1$. The identical argument above applies to $(\tilde{a}, \tilde{p})$, and yields an optimal solution $(\tilde{a}^{(1)}, \tilde{p}^{(1)})$ to program (**Downward-IC**) with $\tilde{a}^{(1)}_{i-1} = \tilde{a}^{(1)}_i = a_{i+1}$. Repeating this process $(i - r)$ times yields an optimal solution $(\tilde{a}^{(i-r)}, \tilde{p}^{(i-r)})$ to program (**Downward-IC**) with

$$\tilde{a}^{(i-r)}_r = \tilde{a}^{(i-r)}_{r+1} = \cdots = \tilde{a}^{(i-r)}_i = a_{i+1}\,,$$

with $\tilde{a}^{(i-r)}_j = a_j$ for all $j \notin \{r, \ldots, i\}$.

Then, it follows immediately that for all $1 \leqslant j \leqslant i$,

$$\tilde{a}^{(i-r)}_j \leqslant \tilde{a}^{(i-r)}_{j+1}\,,$$

where we use the notation $\tilde{a}^{(0)} = \tilde{a}$. Therefore,

$$o(\tilde{a}^{(i-r)}) \geqslant i + 1 > o(a)\,,$$

contradicting that, by construction, $o(a)$ is the largest such index among all optimal solutions to program (**Downward-IC**).

Step (B). By **Step (A)**, it is then without loss of optimality to assume an optimal solution (a, p) where a is monotone. We claim that there exists $\tilde{p}$ under which $(a, \tilde{p})$ is also an optimal solution with IC$[(i+1) \to i]$ binding for all i. Suppose for contradiction that there is no such $\tilde{p}$. Then, for any payment rule $\tilde{p}$ such that $(a, \tilde{p})$ is an optimal solution to (**Downward-IC**), we have

$$\Delta[\tilde{p}] := \sum_{i=1}^{n-1} \big(v(a_{i+1}, t_{i+1}) - \tilde{p}_{i+1}\big) - \big(v(a_i, t_{i+1}) - \tilde{p}_i\big) > 0.$$

Fix a payment rule p such that (a, p) is optimal and the above sum is minimized among all such $\tilde{p}$. Clearly, the above sum must be strict for payment rule p as well.

Fix some i such that IC$[(i+1) \to i]$ is slack under (a, p). That is,

$$\big(v(a_{i+1}, t_{i+1}) - p_{i+1}\big) - \big(v(a_i, t_{i+1}) - p_i\big) > 0.$$

We claim that IC$[k \to j]$ must be slack for all $k > i$ and all $j \leqslant i$. Indeed, by IC$[k \to (i+1)]$, we have

$$\big(v(a_k, t_k) - p_k\big) - \big(v(a_{i+1}, t_k) - p_{i+1}\big) \geqslant 0.$$

Moreover, by the increasing differences property of $v(a, t)$ and the slack of IC$[(i+1) \to i]$, we have

$$\big(v(a_{i+1}, t_k) - p_{i+1}\big) - \big(v(a_i, t_k) - p_i\big) > 0.$$

Similarly, by the increasing differences property of $v(a, t)$ and IC$[i \to j]$, we have

$$\big(v(a_i, t_k) - p_i\big) - \big(v(a_j, t_k) - p_j\big) \geqslant 0.$$

Now, summing the previous three inequalities gives

$$\underbrace{\big(v(a_k, t_k) - p_k\big) - \big(v(a_j, t_k) - p_j\big)}_{=:\delta_{kj}} > 0,$$

and hence IC$[k \to j]$ holds with slack. Now, let

$$\delta := \min\big\{\delta_{kj} : k > i,\, j \leqslant i\big\} > 0.$$

Let

$$q := \sum_{k>i} \mu(t_k) \in (0,1).$$

Consider a modified mechanism $(a, \tilde{p})$ where for all k,

$$\tilde{p}_k = \begin{cases} p_k + \delta/2 - q \cdot \delta/2 & \text{if } k > i\,; \\ p_k - q \cdot \delta/2 & \text{otherwise}. \end{cases}$$

We first note that $(a, \tilde{p})$ is downward incentive compatible. Indeed, by construction, it suffices to check IC$[k \to j]$ where $k > i$ and $j \leqslant i$. But for any such IC constraint, as shown above, there exists at least δ slack, and hence lowering the deviation payoff gap of type t_k mimicking t_j by $\delta/2$ preserves IC$[k \to j]$. Moreover, $(a, \tilde{p})$ also satisfies all IR constraints because IC$[j \to 1]$ constraints hold for all j, IR[1] holds, and $v(a,t)$ is nondecreasing in t. Now, we claim that $(a, \tilde{p})$ yields a weakly higher objective for the principal. Indeed, by construction

$$\sum_k \mu(t_k)\tilde{p}_k = -q \cdot \delta/2 + \sum_{k \leqslant i} \mu(t_k) p_k + \sum_{k>i} \mu(t_k)(p_k + \delta/2) = \sum_k \mu(t_k) p_k.$$

Moreover, since $q < 1$ by construction, note that for all k, we must have

$$\sum_{j=k}^{n} \mu(t_j) p_j \leqslant \sum_{j=k}^{n} \mu(t_j) \tilde{p}_j.$$

The change in principal's objective is given by

$$\sum_k \Big(\mu(t_k)\tilde{p}_k - \mu(t_k) p_k\Big)(1 - \lambda(t_k)).$$

Note that $1 - \lambda(t_k)$ is non-decreasing in k, and therefore, using summation by parts (Abel's identity), we have

$$\begin{aligned} \sum_k \mu(t_k) p_k (1 - \lambda(t_k)) &= (1 - \lambda(t_1)) \sum_{k=1}^{n} \mu(t_k) p_k + \sum_{k=2}^{n} \Big(\sum_{j=k}^{n} \mu(t_j) p_j\Big)(\lambda(t_{k-1}) - \lambda(t_k)) \\ &\leqslant (1 - \lambda(t_1)) \sum_{k=1}^{n} \mu(t_k) \tilde{p}_k + \sum_{k=2}^{n} \Big(\sum_{j=k}^{n} \mu(t_j) \tilde{p}_j\Big)(\lambda(t_{k-1}) - \lambda(t_k)) \\ &= \sum_k \mu(t_k) \tilde{p}_k (1 - \lambda(t_k)). \end{aligned}$$

where the inequality is due to *(i)* $\sum_k \mu(t_k)\tilde{p}_k = \sum_k \mu(t_k)p_k$ and *(ii)* $\sum_{j=k}^n \mu(t_j)p_j \leqslant \sum_{j=k}^n \mu(t_j)\tilde{p}_j$ for all k. Therefore, $(a,\tilde{p})$ must also be an optimal solution to (**Downward-IC**). Moreover,

$$\Delta[\tilde{p}] = \Delta[p] - \delta/2 < \Delta[p],$$

which is a direct contradiction since p is chosen to minimize $\Delta[p]$ to begin with.

Therefore, there exists an optimal solution (a,p) to (**Downward-IC**) such that (i) a is monotone and (ii) every local downward incentive constraint binds under (a,p); hence, by a standard argument, (a,p) must satisfy all upward incentive compatibility constraints as well, concluding the proof.

4.3 Purification Lemma

The final step is to purify the stochastic mechanism in **Step 2** to a deterministic mechanism. It turns out that whenever the stochastic mechanism uses ordered lotteries in the stochastic dominance sense and the principal has quasilinear preferences, there exists a purification. Formally, say that the principal has ***generalized quasilinear preferences*** if the (ex post) objective value of assigning allocation x to type t with payment p is given by

$$V(x,t) + W(t) \cdot p$$

for some continuous functions V and W. Clearly, the principal we consider has generalized quasilinear preferences.

Lemma 2 (Purification)**.** *Let $(\mathcal{Y}, \preceq)$ be any compact set of ordered allocations for which the agent's preferences satisfy strict increasing differences. Consider any mechanism $(a,p) : \mathcal{T} \to \mathcal{A} \times \mathbb{R}$ where $\mathcal{A} \subseteq \Delta(\mathcal{Y})$ is totally ordered by $\preceq_{\mathrm{st}}$. Then, for any generalized quasilinear preferences of the principal, there exists a deterministic mechanism $(x^\dagger, p^\dagger) : \mathcal{T} \to \mathcal{Y} \times \mathbb{R}$ such that $(x^\dagger, p^\dagger)$ yields a weakly higher objective value than (a,p).*

The proof is in the appendix. We sketch the proof here. Since (a,p) is a fully IC mechanism, by the Envelope theorem, we can write

$$p(t) = v(a(t),t) - \Big(\int_{\underline{t}}^{t} v_t(a(z),z)\,\mathrm{d}z + U(\underline{t}) \Big),$$

where U is the indirect utility function. Now, note that the principal's objective is given by

$$\mathbb{E}\left[\widetilde{V}(a(t),t) - W(t)\int_{\underline{t}}^{t} v_t(a(z),z)\,\mathrm{d}z\right] - \mathbb{E}[W(t)]U(\underline{t}),$$

where $\widetilde{V}(a,t) = V(a,t) + W(t)v(a,t)$. Moreover, since $\mathcal{A}$ is totally ordered by $\preceq_{\text{st}}$, for any $a \preceq_{\text{st}} a'$ and $a \neq a'$, we have

$$\mathbb{E}_{x\sim a}[v(x,t_2) - v(x,t_1)] < \mathbb{E}_{x\sim a'}[v(x,t_2) - v(x,t_1)],$$

given that $v(x,t)$ has strict increasing differences on $\mathcal{Y} \times \mathcal{T}$.[14] Therefore, $v(a,t)$ has strict increasing differences on $\mathcal{A} \times \mathcal{T}$. Since (a,p) is fully IC, this implies that $a(\cdot)$ is stochastically monotone in the sense that $a(t) \preceq_{\text{st}} a(t')$ for all $t < t'$. Then, by Strassen's theorem (see Lemma 1 in Yang forthcoming), there exists a ***monotone coupling*** in the sense that for all t,

$$x(t,\varepsilon) \sim a(t) \in \Delta(\mathcal{Y})$$

where $\varepsilon \in \mathcal{E}$ is some random variable independent of t and $x(\cdot,\varepsilon)$ is monotone for all ε. This implies that the above objective is equivalent to

$$\mathbb{E}_{t,\varepsilon}\left[\widetilde{V}(x(t,\varepsilon),t) - W(t)\int_{\underline{t}}^{t} v_t(x(z,\varepsilon),z)\,\mathrm{d}z\right] - \mathbb{E}[W(t)]U(\underline{t}).$$

Therefore, the objective is bounded from the above by

$$\sup_{\varepsilon\in\mathcal{E}} \mathbb{E}_t\left[\widetilde{V}(x(t,\varepsilon),t) - W(t)\int_{\underline{t}}^{t} v_t(x(z,\varepsilon),z)\,\mathrm{d}z\right] - \mathbb{E}[W(t)]U(\underline{t}),$$

which is in fact ***attainable*** by using the allocation rule given by $x(t,\varepsilon^\dagger)$ if the maximization problem above admits a solution—in particular, $x(t,\varepsilon^\dagger)$ is monotone in t and hence ***implementable*** with the transfers given by the Envelope theorem (up to a lump sum transfer which is pinned down by $U(\underline{t})$). Even if the above maximization does not admit a solution, an approximation argument suffices to yield such a deterministic mechanism. This then gives a deterministic mechanism that generates a weakly higher payoff for the principal under any generalized quasilinear preferences by the principal.

[14] Indeed, to see the above must hold with strict inequality, note that by Strassen's theorem a' can be obtained from a via a coupling (Y,Y') where $Y \leq Y'$ and $\mathbb{P}(Y < Y') > 0$.

4.4 Proof Sketches for Theorem 2 and Theorem 3

Building on the proof of Theorem 1, we sketch the proofs for Theorem 2 and Theorem 3.

Proof Sketch for Theorem 2. For the proof sketch, we assume a finite $\mathcal{X}$. Fix any stochastic mechanism $(a, p) : \mathcal{T} \to \Delta(\mathcal{X}) \times \mathbb{R}$. Without loss of generality, we can write $a(t) = x(t, \varepsilon)$ where ε is a randomization device. We reconstruct an alternative stochastic mechanism $(\tilde{a}, \tilde{p}) : \mathcal{T} \to \Delta(\mathcal{X}^\star) \times \mathbb{R}$, following a similar procedure as before. In particular, we keep the transfer rule unchanged, and we modify the allocation rule as follows: For any ε, we construct $\tilde{a}(t, \varepsilon)$ exactly as in Lemma 1. By treating the randomization of ε and the randomization of $\tilde{a}(t, \varepsilon)$ given ε as a compound lottery, we obtain a stochastic allocation rule $(\tilde{a}, \tilde{p}) : \mathcal{T} \to \Delta(\mathcal{X}^\star) \times \mathbb{R}$. In general, this stochastic allocation rule need not use stochastically ordered lotteries. However, we will show that there exists a further modification of $\tilde{a}$ such that it uses only stochastically ordered lotteries and satisfies all the downward IC constraints.

The second step of modification works as follows. Fix the reconstructed $(\tilde{a}, \tilde{p})$. We keep the transfer rule again unchanged, and further modify the allocation rule as follows: For every type t, we look for a lottery

$$\hat{a}(t) \in \mathcal{A} = \bigcup_j \Delta\big(\{x_{j-1}, x_j\}\big) \subseteq \Delta(\mathcal{X}^\star),$$

where $\{x_j\}_j := \mathcal{X}^\star$ with x_j ordered by the surplus ordering, such that

$$v(\hat{a}(t), t) = v(\tilde{a}(t), t).$$

It turns out that such a lottery always exists. Moreover, under the ordered incremental elasticity condition, we always have that the demand curve generated by $\hat{a}$ single-crosses that generated by $\tilde{a}$ from below. Then, just like in the proof of Theorem 1, this implies that all downward incentive constraints continue to hold under the modified mechanism $(\hat{a}, \hat{p})$. Now the newly modified mechanism uses stochastically ordered lotteries in $\mathcal{A}$, and hence the rest of the proof follows identically to that of Theorem 1.

Proof Sketch for Theorem 3. To see why a generalized frontier is an optimal menu, fix any mechanism (x, p). Fix some type t. Consider the assigned allocation $\hat{x} = x(t)$. By assumption, we know that there exists some $\hat{a} \in \mathcal{A}$ such that the demand curve generated by $\hat{a}$ weakly single-crosses that generated by $\hat{x} \in \mathcal{X}$ from below at the quantile of type t. Then, consider the allocation rule defined by such a reconstruction $\hat{a}(t)$. Then, as in the proof of Theorem 1, the mechanism $(\hat{a}, p)$ is downward incentive compatible and individually rational. Moreover, it uses only lotteries that are stochastically ordered and supported on the elements in $\mathcal{X}^\star$. It follows immediately by **Step 2** and **Step 3** of the proof of Theorem 1 that it can be further weakly improved by a deterministic mechanism $(\tilde{x}, \tilde{p})$ that is fully incentive compatible and uses only allocations in the generalized frontier $\mathcal{X}^\star$, concluding the proof.

Now to see why the same ordered incremental elasticity condition implies stochastic optimality, note that just like in the proof of Theorem 2, we can reconstruct a stochastic mechanism for each realization of the randomization device ε in the original mechanism—since at each type t, for each ε, the reconstructed demand curve single-crosses the original demand curve (fixing the realized ε) at the quantile of type t, the pointwise mixture of the reconstructed demand curves must also single-cross the pointwise mixture of the original demand curves (mixing over ε) at the quantile of type t. The rest then follows immediately by the proof of Theorem 2.

5 Applications

5.1 Nested Bundling: Robustness and Redistribution

Suppose that there are n goods $\{1, \ldots, n\} =: G$. A ***bundle*** $b \in 2^G$ is a set of goods. Let $\mathcal{X} = 2^G$ be the allocation space. A ***menu*** $B \subset 2^G$ is then a set of bundles. A consumer's valuation $v(b, t)$ is monotone in b in the set-inclusion order and strictly increasing in t.

A menu is ***nested*** if it can be totally ordered by set inclusion. As argued in Yang (2025), selling nested menus (***nested bundling***) where more expensive bundles include all the goods of the less expensive ones appears to be common in practice. An immediate consequence of Theorem 1 and Theorem 2 is the following new characterization for when nested bundling is optimal:

Proposition 5 (Robust Nesting)**.** *For any nested menu B, if:*

(i) for any $b_1 \subset b_2 \in B$

$$\frac{\mathrm{d}}{\mathrm{d}t} \log v(b_1, t) < \frac{\mathrm{d}}{\mathrm{d}t} \log v(b_2, t) \text{ for all } t \tag{1}$$

(ii) for any $b_1 \notin B$, there exists $b_2 \in B$ such that $b_1 \subset b_2$, and

$$\frac{\mathrm{d}}{\mathrm{d}t} \log v(b_1, t) \geqslant \frac{\mathrm{d}}{\mathrm{d}t} \log v(b_2, t) \text{ for all } t \tag{2}$$

then menu B is optimal for all type distributions F and redistributive welfare weights λ. Moreover, if condition (i) is strengthened to be:

(i') Let $B = \{b_i\}$ where $b_1 \subset \cdots \subset b_m$. For any $i = 2, \ldots, m-1$,

$$\frac{\mathrm{d}}{\mathrm{d}t} \log \Big(v(b_i, t) - v(b_{i-1}, t) \Big) < \frac{\mathrm{d}}{\mathrm{d}t} \log \Big(v(b_{i+1}, t) - v(b_i, t) \Big) \text{ for all } t \tag{1'}$$

then menu B is optimal among stochastic mechanisms for all type distributions F and redistributive welfare weights λ.

Proof. Under conditions *(i)* and *(ii)*, B is a surplus-elasticity frontier (recall the discussion about the elasticity order after Theorem 1) and hence optimal by Theorem 1. If in addition we have condition *(i')*, then B is a strong frontier, and hence optimal among all stochastic mechanisms by Theorem 2. □

The ***robust nesting condition*** in Proposition 5 complements and illuminates the nesting condition derived in Yang (2025), which depends on the type distribution and holds for profit maximization. Indeed, Yang (2025) considers the partial order defined by *(i)* set inclusion and *(ii)* sold-alone quantities, and shows that if the undominated elements with respect to this partial order are nested, then this nested menu is optimal for profit maximization under regularity conditions. Proposition 5 shows that the dominance criterion can be broadly viewed as comparing demand curves in terms of their levels and their elasticities—whenever an appropriate frontier exists, then it becomes optimal. Indeed, a larger bundle generates a pointwise higher demand curve because of free disposal, and a higher sold-alone quantity corresponds to a demand curve with a larger elastic region.

Moreover, by virtue of our main results, Proposition 5 shows that once there exists such a frontier, then the menu must be optimal for all type distributions and all redistributive welfare weights. Indeed, in practice, nested bundling or tiered pricing appears more broadly beyond profit-maximizing contexts—public provision of goods often uses tiered pricing as well (e.g., public housing, healthcare, and childcare programs). With redistributive concerns in mind, nested bundling mechanisms become optimal by *(i)* extracting more surplus from the richer agents who get larger bundles to supplement the consumption by the poorer agents (e.g., by lowering the price), and *(ii)* distorting the consumption of poorer agents by allocating smaller bundles to make it harder for the richer agents to mimic the poorer agents. Indeed, when a nested menu B satisfies both criteria by spanning the surplus-elasticity frontier—i.e., satisfying the robust nesting condition in Proposition 5—it then becomes optimal.

Redistributive Bundling. In the special case where $B = \{\overline{b}\}$ and $\overline{b} = \{1, \dots, n\}$, Proposition 5 immediately recovers the ratio-monotonicity condition of Haghpanah and Hartline (2021) and shows that their condition actually gives the optimality of pure bundling even with redistributive concerns. More generally, we can relax the one-dimensional types here as in Haghpanah and Hartline (2021). Say that ***stochastic ratio-monotonicity*** holds if $v^b/v^{\overline{b}}$ is stochastically nondecreasing in $v^{\overline{b}}$ for a random vector $\mathbf{v} := (v^b)_{b \in 2^G}$ of the bundle values. Let the welfare weights and bundle values $(\lambda, \mathbf{v})$ be jointly distributed.

Corollary 3. *Suppose $v^{\overline{b}}$ is a sufficient statistic for the redistributive welfare weights in that $\mathbb{E}[\lambda \mid \mathbf{v}] = \mathbb{E}[\lambda \mid v^{\overline{b}}]$ which is continuous, non-increasing in $v^{\overline{b}}$. Then, under the stochastic ratio-monotonicity condition, pure bundling is optimal among all stochastic mechanisms.*

Corollary 3 broadens the scope to which Haghpanah and Hartline (2021)'s condition applies and shows that in many redistributive contexts, it could be optimal to simply offer a ***broad access*** to resources as a single option without further screening of the recipient. This is perhaps surprising because the principal cares more about exactly the types who tend to have ***lower*** values for the grand bundle, and yet the principal does not provide any smaller bundle for them to consume—the only instrument the principal uses is to lower the price for the grand bundle and expand the market access to help (as λ becomes more redistributive). This is precisely because the grand bundle itself forms the surplus-elasticity frontier under the ratio-monotonicity condition—the best way to control information rents for the richer agents turns out to fully exclude the poorer agents.

5.2 Optimal Taxation and Ordeals

We consider the optimal taxation problem with quasilinear preferences (Diamond 1998). In addition to the tax instrument, the government can also utilize ordeal mechanisms to induce self-targeting. As shown in Nichols and Zeckhauser (1982), ordeals that purely destroy surplus can potentially increase the total welfare achievable under nonlinear taxation by helping with screening. When should we expect an ordeal to be useful? We apply the main results to study this question.

In this environment, an agent of type t has the payoff given by $l - \psi(l, y, t) - T$, where l is the ***labor income***, y is the ***ordeal***, T is the ***tax***. The function ψ is the ***disutility*** function, which we assume is strictly decreasing in ability type t. The government maximizes a weighted welfare

$$\mathbb{E}[\lambda(t)(l(t) - \psi(l(t), y(t), t) - T(t))]$$

subject to IC and the ***budget balance*** constraint: $\mathbb{E}\left[T(t)\right] \geqslant E$, where $E \geqslant 0$ is an exogenous public expenditure. To see how this problem is nested in our framework, let κ be the multiplier on the budget balance constraint. Then, writing out the Lagrangian and normalizing the objective by κ, we have

$$\mathbb{E}\Big[\frac{\lambda(t)}{\kappa}\big(l(t)-\psi(l(t),y(t),t)-T(t)\big)\Big]+\mathbb{E}\Big[T(t)\Big].$$

Let $\tilde{\lambda}(t)=\frac{\lambda(t)}{\kappa}$. Note that if $\mathbb{E}[\tilde{\lambda}(t)]<1$ or $\mathbb{E}[\tilde{\lambda}(t)]>1$, then there exists no solution to the Lagrangian that is budget feasible. Thus, by strong duality, $\mathbb{E}[\tilde{\lambda}(t)]=1$. Moreover, as standard, even though the original taxation problem has no IR constraint, we can normalize the lowest type to have 0 payoff in the above program, and then uniformly adjust the tax schedule level after solving the above program to satisfy the budget constraint. This then becomes a special case of the main model. As before, we assume that $\lambda(t)$ is redistributive in the sense that $\lambda(t)$ is non-increasing in t.

In this case, the allocation space $\mathcal{X}\subset\mathbb{R}^2_+$ consists of elements (l,y) and we assume that $\{(l,0)\}_{l\in[\underline{l},\bar{l}]}\subset\mathcal{X}$ and

$$\psi(l,0,t)\leqslant\psi(l,y,t)$$

for all $(l,y)\in\mathcal{X}$, so $y=0$ represents the absence of costly screening. We assume that $\psi(l,0,t)$ is strictly submodular in (l,t), continuous in l, and satisfies $\psi(0,0,t)=0$.

Proposition 6. *Suppose that either of the following two conditions holds:*

(a) Either, $\psi(l,y,t)-\psi(l,0,t)$ is non-increasing in t;

(b) or, $l-\psi(l,0,t)$ is log-supermodular in (l,t), and $l-\psi(l,y,t)\geqslant 0$ with

$$\frac{l-\psi(l,y,t)}{l-\psi(l,0,t)}\ \textit{is non-decreasing in } t$$

for all $(l,y)\in\mathcal{X}$.

Then, $\{(l,0)\}_{l\in[\underline{l},\bar{l}]}$ is an optimal menu (i.e., the optimal menu involves no costly screening).

Proof. Under condition (a), note that fixing any mechanism $t\mapsto(l(t),y(t),T(t))$ we can replicate the same truth-telling payoff by setting $\tilde{y}(t)=0$ and $\widetilde{T}(t)=T(t)+\big(\psi(l(t),y(t),t)-\psi(l(t),0,t)\big)$. Moreover, under condition (a), note that after this reconstruction, the deviation payoff of any type t misreporting a lower type $\hat{t}<t$ must weakly decrease. Thus, the modified mechanism must be downward incentive compatible (as defined in Section 4). Then, by the generalized downward sufficiency theorem (Theorem 4), it follows immediately that there exists a full IC mechanism that involves no costly screening and weakly improves on the original mechanism, proving the result.

Under condition (b), this is an immediate consequence of Proposition 4 (i.e., the menu $\{(l,0)\}_{l\in[\underline{l},\bar{l}]}$ is a generalized frontier). $\square$

Condition (b) can be equivalently stated using take-up elasticities with respect to a marginal tax reduction. Indeed, consider two thought experiments. First, consider the case where there is no ordeal involved. Fix some income level l, and measure the ***take-up elasticity*** by offering a small reduction in tax liability for all income above l—which will increase the percentage of population achieving income level l. Second, consider the case where ordeal y is involved, and measure the take-up elasticity by offering a small reduction in tax liability for all income above l as well but only under the requirement that the ordeal must be performed. Formally, let

$$\eta^{\text{take-up}}(l,y,q) := -\frac{\mathrm{d}}{\mathrm{d}\log(T)}\log\mathbb{P}\Big(l-\psi(l,y,t)-T\geqslant 0\Big)\,|_{T^\star(l,y,q)},$$

where $T^\star(l,y,q)$ is the tax level for income l such that the population earning above l equals q. Then, it follows immediately that the

$$\frac{l-\psi(l,y,t)}{l-\psi(l,0,t)} \text{ is non-decreasing in } t \iff \eta^{\text{take-up}}(l,y,q)\leqslant\eta^{\text{take-up}}(l,0,q).$$

i.e., a tax reduction leads to a smaller aggregate labor response with the ordeal than without the ordeal. Therefore, Proposition 6 shows that for ordeals to be *useful*, the following two conditions must hold: *(i)* the additional disutility resulting from an ordeal should be ***increasing*** in the ability type t rather than decreasing, and *(ii)* the ordeal should also ***increase*** take-up elasticity with respect to the benefit. The first part is consistent with the takeaway from Yang (forthcoming) that an ordeal is only useful if the agents' preferences over the ordeal and the productive labor supply are suitably negatively correlated. Proposition 6 shows that this conclusion holds even in the optimal taxation context. Moreover, Proposition 6 shows that, for an ordeal to be useful, it is not enough for the agents' preferences to be negatively correlated—the reduction in surplus must be compensated via an increase in the screening effectiveness, captured by an appropriate notion of elasticity measures. To gain some intuition, note that, in the absence of an ordeal, at the optimal tax schedule, the government would want to offer a marginal tax cut on the lower types which would also increase their labor supply but worry about the higher types shirking—an ordeal with ***high*** take-up elasticity is helpful when combined with such a marginal tax cut—it increases the labor supply sensitivity of the lower types and dampens the incentive for shirking by the higher types.

The take-up elasticity derived here is related to, but differs subtly from, those derived in Yang, Dworczak, and Akbarpour (2024) and Rafkin, Solomon, and Soltas (2023)—indeed, the take-up elasticity here is with respect to a marginal tax reduction which increases overall labor responses—if the reform were to be a ***direct cash transfer*** conditional on the ordeal, then a ***low*** take-up elasticity with respect to the cash benefit is helpful because it allows the government to transfer a fixed budget of cash with limited rent dissipation. To reconcile the contrast, note that an inelastic take-up rate with respect to cash benefit typically arises when the ordeal cost rises steeply with ability, which would result in the consumption utility net of screening cost to rise slowly with ability—and thus an elastic take-up rate with respect to a tax cut. Thus, the key insight in Proposition 6 is consistent with the insights derived in one-dimensional redistributive screening problems as well as perturbation analysis in public finance, and it highlights an appropriate notion of elasticity for assessing the screening gains from ***combining*** ordeal and tax instruments.

5.3 Sequential Screening

We consider the sequential screening problem as studied in Courty and Li (2000). The setting is as follows. There is one good. The agent has an ***ex ante type*** t, which is correlated with the ***ex post type*** θ which is their actual value for consuming the good. The contracting happens at the stage where the agent observes their ex ante type t but does not observe their ex post type θ. Let $g(\theta \mid t)$ denote the conditional density of ex post type θ conditional on the ex ante type t. As in Courty and Li (2000), we assume that $\theta \mid t$ is ***FOSD-increasing*** in t. By the revelation principle, the principal can design a ***direct dynamic mechanism*** $(t,\theta) \mapsto \big(q(t,\theta), p(t,\theta)\big)$ where q is the allocation probability and p is the payment, satisfying the following constraints:

$$
\begin{aligned}
\theta q(t,\theta) - p(t,\theta) &\geqslant \theta q(t,\hat{\theta}) - p(t,\hat{\theta}) && \text{for all } t,\theta,\hat{\theta}\,; \\
\int \Big(\theta q(t,\theta) - p(t,\theta)\Big) g(\theta \mid t)\,\mathrm{d}\theta &\geqslant \int \Big(\theta q(\hat{t},\theta) - p(\hat{t},\theta)\Big) g(\theta \mid t)\,\mathrm{d}\theta && \text{for all } t,\hat{t}\,; \\
\int \Big(\theta q(t,\theta) - p(t,\theta)\Big) g(\theta \mid t)\,\mathrm{d}\theta &\geqslant 0 && \text{for all } t\,,
\end{aligned}
$$

where the first constraint is the second-period IC, the second constraint is the first-period IC, and the last constraint is the first-period IR constraint.

Now, consider a very simple mechanism that sells the good in period 1 at a posted price p, which we refer to as a ***nonrefundable posted price***. Recall that $\preceq_{\text{disp}}$ denotes the

dispersive order, i.e., $X \preceq_{\text{disp}} Y$ if $F_X^{-1}(q') - F_X^{-1}(q) \leqslant F_Y^{-1}(q') - F_Y^{-1}(q)$ for all $q < q' \in [0,1]$. We say that the types have ***increasing dispersion*** if for all $t < t'$

$$\log(\theta) \mid t \preceq_{\text{disp}} \log(\theta) \mid t'.$$

Proposition 7. *Under increasing dispersion, a nonrefundable posted price is optimal.*

Proof. We formulate this problem in our framework in the following way. As in Eso and Szentes (2007), note that we can write $\theta = G^{-1}(\varepsilon \mid t)$ where ε is a uniform $[0,1]$ random variable independent of t. In particular, the model is equivalent to the one where the second-period type is indexed by ε instead of θ. Under this reparameterization, the ex post utility function is given by

$$G^{-1}(\varepsilon \mid t) q(t,\varepsilon) - p(t,\varepsilon).$$

In this parameterization, note that the second-period IC implies that $q(t,\cdot)$ is monotone in ε. The first-period IC constraints would involve double deviations under this parameterization, but we keep the following subset of IC and IR constraints:

$$\int G^{-1}(\varepsilon \mid t) q(t,\varepsilon)\,\mathrm{d}\varepsilon - \tilde{p}(t) \geqslant \int G^{-1}(\varepsilon \mid t) q(\hat{t},\varepsilon)\,\mathrm{d}\varepsilon - \tilde{p}(\hat{t}) \qquad \text{for all } t, \hat{t};$$
$$\int G^{-1}(\varepsilon \mid t) q(t,\varepsilon)\,\mathrm{d}\varepsilon - \tilde{p}(t) \geqslant 0 \qquad \text{for all } t,$$

where $\tilde{p}(t) = \mathbb{E}_\varepsilon[p(t,\varepsilon)]$. We relax all other constraints and hence directly choose $(q, \tilde{p})$ to solve the above problem. Now, note that since $q(t,\cdot)$ is monotone, up to a measure 0 set of ε's, we can write

$$q(t,\varepsilon) = \int \mathbb{1}_{\varepsilon \geqslant 1-k} \mu_t(\mathrm{d}k),$$

where $\mu_t \in \Delta([0,1])$. Then,

$$\int G^{-1}(\varepsilon \mid t) q(t,\varepsilon)\,\mathrm{d}\varepsilon = \int_0^1 \left(\int_{1-k}^1 G^{-1}(\varepsilon \mid t)\,\mathrm{d}\varepsilon \right) \mu_t(\mathrm{d}k),$$

Letting

$$u(k,t) := \int_{1-k}^1 G^{-1}(\varepsilon \mid t)\,\mathrm{d}\varepsilon,$$

then we have

$$\int_0^1 u(k,t) \mu_t(\mathrm{d}k).$$

Therefore, μ_t is equivalent to a lottery over the allocations $k \in [0,1]$ with utility functions given by $u(k,t)$ which is strictly increasing in k and t (since $\theta \mid t$ is FOSD-increasing in t). Thus, equivalently, in the relaxed problem, we can write the allocation space as

$$\mathcal{X} = \Delta([0,1]),$$

with utility function given by $v(\mu,t) = \mathbb{E}_{k\sim\mu}[u(k,t)]$. The result follows immediately by Proposition 2 by noting that

$$\log(u_k(k,t)) = \log G^{-1}(1-k \mid t)$$

is weakly submodular if $\log(\theta) \mid t$ is weakly increasing in the dispersive order. Indeed, by the proof of Proposition 2, in the weakly submodular case, offering $k=1$ alone, at some price $p^\star$ (depending on the type distribution and welfare weights), is optimal in the relaxed problem. Evidently, offering a nonrefundable posted price $p^\star$ yields the same expected payoff for every type t and the same revenue as in the relaxed problem—thus, it is optimal. □

By virtue of our main results, Proposition 7 holds for all type distributions of t and all redistributive welfare weights. Even though sequential screening is a well-studied problem, we are not aware of any primitive condition under which a nonrefundable posted price is optimal. One reason is methodological: much of the dynamic mechanism design literature relies on the first-order approach under regularity assumptions that typically deliver fully separating allocations, so pooling/bunching mechanisms such as a nonrefundable posted price can be difficult to obtain without explicitly tracking the global incentive constraints.

Proposition 7 asserts that nonrefundable ticket sales are optimal when higher ex-ante types not only have higher expected willingness to pay (FOSD) but also have greater uncertainty about their ex-post valuations in the sense of the log-dispersion order. Perhaps surprisingly, this is the case even though the principal could instead offer refunds to better target high types and extract surplus through a higher up-front fee. The key friction is leakage. Since ex-post values are FOSD increasing in the ex-ante types, such a refund would have a leakage that attracts the lower types, which will then involve too much refund and eventually lead to a lower revenue for the principal. Note that here a refund with a high up-front fee actually ***reduces*** the ex post surplus by reducing the consumption of consumers with lower ex post values, and hence it can be justified only by increasing the screening effectiveness by our main results. Yet under the log-dispersion order—which captures the elasticity of how ex post values vary with ex ante types—the screening effectiveness cannot be increased and hence a nonrefundable posted price is optimal.

To illustrate our log-dispersion condition, consider the following simple examples:

Example 1. Let ε be some random variable independent of t and

$$\log\theta = \alpha \cdot t + \sigma(t) \cdot \varepsilon, \qquad \textbf{(Log-linear)}$$

where $\alpha > 0$. It can be easily verified that the types have increasing dispersion if $\sigma(t)$ is nondecreasing in t. As another example, consider

$$\theta = \alpha \cdot t + \sigma(t) \cdot \varepsilon, \qquad \textbf{(Linear)}$$

where $\alpha > 0$ is a constant. It can be easily verified that the types have increasing dispersion if $\sigma(t)/t$ is nondecreasing in t. □

5.4 Selling Information

Consider any finite ***state space*** Ω with n states. The seller has information about the state ω given by ***Blackwell experiments*** about the state. We assume a uniform prior $\mu_0 \in \Delta(\Omega)$. Without loss of generality, we can identify the experiments with the ***posterior distributions*** $\gamma \in \Delta(\Delta(\Omega))$ that satisfy ***Bayes plausibility***:

$$\int \mu \, \mathrm{d}\gamma(\mu) = \mu_0.$$

The buyer privately observes their type t which determines their private decision problems. Without loss of generality, we can define the buyer's ***indirect utility function*** from the decision problem at any posterior μ by

$$u(\mu, t) : \Delta(\Omega) \times \mathcal{T} \to \mathbb{R}$$

where $u(\cdot, t)$ is convex in μ. The buyer's valuation for an experiment γ is then given by

$$v(\gamma, t) = \int u(\mu, t) \, \mathrm{d}\gamma(\mu) - u(\mu_0, t).$$

The types are ordered by the level of their induced utility functions compared to the no-information outside option. Specifically, we assume that $u(\mu, t) - u(\mu_0, t)$ is strictly increasing in t for $\mu \neq \mu_0$, and thus without loss of generality we may normalize $u(\mu_0, t) = 0$.[15] For two functions h_1, h_2 defined on $\Delta(\Omega)$, we say that h_1 is ***less convex*** than h_2, if

$$h_2 = g \circ h_1$$

where g is a nondecreasing convex function, and ***strictly less convex*** if g is strictly convex. We say that the environment is ***symmetric*** if $u(\mu, t)$ is permutation symmetric with respect to μ.

We say that an experiment is a ***truth-or-noise*** experiment if the signal either fully reveals the state or is an i.i.d. draw from the prior μ_0 (Lewis and Sappington 1994). Formally, for any $\omega \in \Omega$ and any $\alpha \in [0, 1]$, let

$$\mu_\alpha^\omega := (1-\alpha)\mu_0 + \alpha\delta_\omega.$$

Then, let γ_α denote the posterior distribution that is supported on $\{\mu_\alpha^\omega\}_\omega$ with the probability mass being $\gamma_\alpha(\mu_\alpha^\omega) = \mu_0(\omega)$ for all ω. The truth-or-noise experiments are very simple in that they are controlled by a one-dimensional index α, are Blackwell-monotone in α, and generate posterior belief distributions with sparse supports. When are they optimal?

Proposition 8. *Consider a symmetric environment. If*

$$u(\cdot, t) \textit{ is increasingly } \text{concave} \textit{ in } t$$

then selling full information is optimal. If

$$u(\cdot, t) \textit{ is strictly increasingly } \text{convex} \textit{ in } t,$$

then selling truth-or-noise experiments $\{\gamma_\alpha\}_{\alpha\in[0,1]}$ is optimal.

Proof. Let $\mathcal{X}$ be the set of all Bayes-plausible posterior distributions. To prove the first part, suppose u is increasingly concave in t. We show that the fully informative experiment $\{\gamma^\star\}$ is the surplus-elasticity frontier. Clearly, it generates a demand curve that is pointwise maximal. Now we check whether the demand curve it generates is more elastic compared to the one generated by some other γ. Fix any two types $t_1 < t_2$. Then,

$$u(\mu, t_2) = g(u(\mu, t_1))$$

[15] Note that we can simply redefine $\tilde{u}(\mu, t) = u(\mu, t) - u(\mu_0, t)$.

for a concave function g. Moreover, since $u(\mu_0, t) = 0$ by our normalization, $g(0) = 0$. Together, these imply that $g(z)/z$ is non-increasing in z for all $z > 0$. By symmetry, $u(\delta_\omega, t) = u(\delta_{\omega'}, t)$ for all ω, ω'. Therefore, the fully informative experiment yields a deterministic payoff for type t_1, say $u_1^\star$. Clearly, $u_1^\star \geqslant u(\mu, t_1)$ for all μ. Then, by the previous observation, $g(u_1^\star)/u_1^\star \leqslant g(u(\mu, t_1))/u(\mu, t_1)$. Thus,

$$\frac{v(\gamma, t_2)}{v(\gamma, t_1)} = \frac{\int g(u(\mu, t_1))\,\mathrm{d}\gamma}{\int u(\mu, t_1)\,\mathrm{d}\gamma} \geqslant \frac{g(u_1^\star)/u_1^\star \int u(\mu, t_1)\,\mathrm{d}\gamma}{\int u(\mu, t_1)\,\mathrm{d}\gamma} = \frac{g(u_1^\star)}{u_1^\star} = \frac{v(\gamma^\star, t_2)}{v(\gamma^\star, t_1)},$$

which proves the result by Theorem 1.

To prove the second part, suppose u is strictly increasingly convex in t. We show that $\mathcal{X}^\star := \{\gamma_\alpha\}$ is a generalized frontier. Note that for any truth-or-noise experiment γ_α, we have by construction

$$v(\gamma_\alpha, t) = \sum_\omega \mu_0(\omega) u\big((1-\alpha)\mu_0 + \alpha\delta_\omega, t\big).$$

By symmetry, we have that for any ω and ω',

$$u\big((1-\alpha)\mu_0 + \alpha\delta_\omega, t\big) = u\big((1-\alpha)\mu_0 + \alpha\delta_{\omega'}, t\big) =: h(\alpha, t).$$

Note that $h(\alpha, t)$ is strictly increasing in t and has strict increasing differences since for any $s > t$, there exists a strictly convex g with $g(z) \geqslant z$ and $g(z)/z$ is strictly increasing such that

$$h(\alpha,s) - h(\alpha,t) = \big(g(h(\alpha,t))/h(\alpha,t) - 1\big) \cdot h(\alpha,t) < \big(g(h(\alpha',t))/h(\alpha',t) - 1\big) \cdot h(\alpha',t) = h(\alpha',s) - h(\alpha',t)$$

for any $\alpha' > \alpha$. Now fix any γ and fix any type t. By the intermediate value theorem, there exists some $\alpha^\star$ such that

$$v(\gamma_{\alpha^\star}, t) = h(\alpha^\star, t) = \int u(\mu, t)\,\mathrm{d}\gamma(\mu).$$

Then, note that for any $s > t$, since u is increasingly convex, there exists some convex g such that

$$v(\gamma_{\alpha^\star}, s) = h(\alpha^\star, s) = g(h(\alpha^\star, t)) \leqslant \int g(u(\mu, t))\,\mathrm{d}\gamma(\mu) = v(\gamma, s),$$

where the inequality uses that g is convex. Clearly, the opposite inequality holds for $s < t$. Since this holds for all t, it implies that $\mathcal{X}^\star$ is a generalized frontier. Indeed, any γ is covered by $\mathcal{A}$ defined as the collection of Dirac measures supported on each of $\{\gamma_\alpha\}_\alpha$.

This completes the proof by Theorem 3. □

Proposition 8 shows that in any symmetric environment, if the types who have a higher overall willingness to pay for the information also have decision problems that depend ***more sensitively*** on the state, then a menu of truth-or-noise signals is optimal; on the other hand, if the types who have a higher overall willingness to pay for the information also have decision problems that depend ***less sensitively*** on the state, then simply selling the full information is optimal.

5.5 Regulating a Data-Rich Monopolist

We consider a model of monopoly regulation just like in Baron and Myerson (1982), but allowing the monopolist to observe rich data about the consumers' valuations.

There is a monopolist and a unit mass of consumers. Consumers are indexed by their ***values*** $\theta \in [\underline{\theta}, \overline{\theta}]$ for the monopolist's good, distributed according to G. The monopolist has a private ***cost type*** $t \in [\underline{t}, \overline{t}]$, distributed according to F, independent of θ. The cost for the monopolist to serve consumer type θ can be ***interdependent***, denoted by $c(\theta, t)$. We assume $c(\theta, t)$ is strictly decreasing in t so that a higher t represents a lower cost type. Following Strack and Yang (2025), we model the ***data-rich monopolist*** as observing the consumers' valuations θ and engaging in third-degree price discrimination in the absence of any regulation; in practice, θ can be thought of as a rich enough ***tag*** or ***segmentation*** such that the residual uncertainty over consumer valuations is small conditional on the tag. In particular, in the absence of any regulation, the monopolist extracts the ***total surplus*** $S(\theta, t) := \theta - c(\theta, t)$, which we assume is non-negative. As in Strack and Yang (2025), the monopolist can use any ***pricing rule*** $p(\theta, r)$ which specifies the posted price for θ consumer depending on the realization of a randomization device r. Given a pricing rule p, the seller's profit is

$$\Pi(p, t) := \mathbb{E}_{\theta, r}\left[\Big(p(\theta, r) - c(\theta, t)\Big)\mathbb{1}_{\theta \geqslant p(\theta, r)}\right]$$

and the consumer θ's surplus is

$$\mathrm{CS}(p, \theta) := \mathbb{E}_r\left[\big(\theta - p(\theta, r)\big)_+\right]$$

Let $\mathcal{P}$ denote the set of all pricing rules.

A regulator contracts with the monopolist on the pricing rules. The regulator does not observe the monopolist's cost type t and hence will post a menu of pricing rules and associated subsidies/taxes $\{(p,T)\}$. By the revelation principle, without loss of generality, the regulator uses a ***(direct, incentive-compatible) mechanism***

$$(p,T) : [\underline{t}, \overline{t}] \to \mathcal{P} \times \mathbb{R}$$

that satisfies the IC and IR constraints of the monopolist

$$\begin{aligned} &\Pi(p(t),t) - T(t) \geqslant \Pi(p(\hat{t}),t) - T(\hat{t}) && \text{for all } t, \hat{t}\,; \\ &\Pi(p(t),t) - T(t) \geqslant 0 && \text{for all } t\,. \end{aligned}$$

Following Baron and Myerson (1982), the regulator maximizes a weighted sum of consumer and producer surplus:

$$\mathbb{E}[\mathrm{CS}(p(t),\theta) + T(t)] + \alpha \cdot \mathbb{E}[\Pi(p(t),t) - T(t)],$$

where $\alpha \in [0,1]$. The above objective balances the budget as in Baron and Myerson (1982) since any tax to the monopolist is rebated to the consumers and vice versa.

We say that a mechanism is a ***uniform pricing regulation*** if the assigned pricing rules $p(\theta, r; t)$ do not depend on any consumer data θ and randomization device r for all t. That is, the regulator simply imposes uniform pricing and taxes/subsidies the monopolist depending on the uniform price. Note that if the regulator were to use a uniform pricing regulation, then the determination of the optimal taxes/subsidies exactly collapses to the model of Baron and Myerson (1982). However, when are uniform pricing regulations optimal? Our next result answers this question. Before it, we also introduce a different class of mechanisms. We say that a mechanism is a ***discriminatory pricing regulation*** if the assigned pricing rules are deterministic and either specify full discriminatory pricing or no trade, i.e., $p(\theta, r; t) \in \{\theta, \overline{\theta} + 1\}$, for all types t. In this case, the consumer surplus is entirely generated via lump-sum transfers by the regulator.

Proposition 9. *If the total surplus $S(\theta, t)$ is log-submodular, then uniform pricing regulations are optimal. If $S(\theta, t)$ is log-supermodular, then discriminatory pricing regulations are optimal.*

Proof. Fix any mechanism (p,T). Let $\widetilde{T}(t) = \mathbb{E}_\theta[\mathrm{CS}(p(t),\theta)] + T(t)$. Then, note that the mechanism using the modified transfer rule satisfies all IC and IR constraints with respect to the monopolist's modified utility that takes into account the consumer surplus:

$$\Pi(p,t) + \mathbb{E}_\theta[\mathrm{CS}(p,\theta)] - \widetilde{T}\,.$$

After this relabeling, the objective then becomes

$$\mathbb{E}\Big[\mathrm{CS}(p(t),\theta)+T(t)\Big]+\alpha\cdot\mathbb{E}\Big[\Pi(p(t),t)-T(t)\Big]=\mathbb{E}\Big[\widetilde{T}(t)\Big]+\alpha\cdot\mathbb{E}\Big[\Pi(p(t),t)+\mathbb{E}_\theta[\mathrm{CS}(p(t),\theta)]-\widetilde{T}(t)\Big],$$

which is a special case of the objective in our main model with $\alpha \in [0,1]$ weight on the agent's net utility (which is the monopolist's modified utility). Now, note that

$$\Pi(p,t)+\mathbb{E}_\theta[\mathrm{CS}(p,\theta)]=\int\Big(\theta-c(\theta,t)\Big)\,\mathbb{E}_r[\mathbb{1}_{\theta\geqslant p(\theta,r)}]\,\mathrm{d}G(\theta)=\int S(\theta,t)x(\theta)\,\mathrm{d}G(\theta),$$

where $x:[\underline{\theta},\overline{\theta}]\to[0,1]$ is defined by $x(\theta):=\mathbb{E}_r[\mathbb{1}_{\theta\geqslant p(\theta,r)}]$. Moreover, conversely, any $x:[\underline{\theta},\overline{\theta}]\to[0,1]$ can be written as $\mathbb{E}_r[\mathbb{1}_{\theta\geqslant p(\theta,r)}]$ for some pricing rule p. Therefore, by Proposition 3, the menu $\mathcal{X}^\star=\{\mathbb{1}_{[k,\overline{\theta}]}\}_k$ is optimal if S is log-submodular, and $\mathcal{X}^\star=\{\mathbb{1}_{[\underline{\theta},k]}\}_k$ is optimal if S is log-supermodular.

The first part of the result follows immediately by noting that in the log-submodular case, the menu is equivalent to a uniform pricing regulation (with the k's being the uniform prices). Now, to see the second part, let $k(t)$ be the assignment for type t given the optimal prices for the menu $\mathcal{X}^\star=\{\mathbb{1}_{[\underline{\theta},k]}\}_k$, and write $p(\theta,r;t)=\mathbb{1}_{\theta\leqslant k(t)}\theta+\mathbb{1}_{\theta>k(t)}(\overline{\theta}+1)$ which then gives a discriminatory pricing regulation that implements menu $\mathcal{X}^\star$. □

To understand Proposition 9, consider a simple example:

Example 2. Suppose $c(\theta,t)=h(\theta)w(t)$. Then

$$\partial_{\theta t}\log S(\theta,t)=\frac{\theta h'(\theta)-h(\theta)}{(\theta-h(\theta)w(t))^2}\cdot(-w'(t)).$$

Since $-w'(t)>0$, the above expression has the same sign as $\theta h'(\theta)-h(\theta)$, i.e., whether $h(\theta)/\theta$ is increasing or decreasing with θ. Note that $h(\theta)/\theta$ would be decreasing when either $h(\theta)$ is decreasing or $h(\theta)$ does not increase as fast as θ. □

As illustrated in Example 2, Proposition 9 shows that uniform-pricing regulations are optimal in markets with either ***advantageous selection*** or ***mild adverse selection***. Moreover, Proposition 9 shows that in markets with ***severe adverse selection***, uniform pricing regulations *cannot* be optimal—instead, the regulator must rely on discriminatory pricing regulations. To gain some intuition, consider a market with extreme adverse selection, then there is a large degree of market failure in which case uniform pricing leads to market unraveling—consumer data are extremely helpful here for restoring efficiency, and thus the optimal mechanism must let the monopolist utilize their consumer data and then tax the monopolist's profits directly to transfer it to the consumers.

6 Conclusion

We consider a general screening problem with an arbitrary set of allocations $\mathcal{X}$. The agent's preferences over allocations are comonotonic. We compare allocations by their induced demand curves in sold-alone markets, in particular, their surplus level and their demand elasticities. We show that if a surplus-elasticity frontier exists, then it is an optimal menu. Moreover, if the incremental demand curves along the frontier are also ordered by their elasticities, then the frontier is optimal even among stochastic mechanisms. The results are agnostic to type distributions and redistributive welfare weights—in particular, the same frontier remains optimal for a broad class of objectives. As applications, we derive new results in multiproduct pricing, optimal bundling, optimal taxation, sequential screening, selling information, and monopoly regulation.

References

Anderson, E. T. and J. D. Dana Jr (2009): "When is Price Discrimination Profitable?" *Management Science*, 55(6), 980–989.

Baron, D. P. and R. B. Myerson (1982): "Regulating a Monopolist with Unknown Costs," *Econometrica*, 50(4), 911–930.

Bergemann, D., A. Bonatti, and A. Smolin (2018): "The design and price of information," *American economic review*, 108, 1–48.

Carroll, G. (2017): "Robustness and Separation in Multidimensional Screening," *Econometrica*, 85(2), 453–488.

Courty, P. and H. Li (2000): "Sequential Screening," *The Review of Economic Studies*, 67, 697–717.

Daskalakis, C., A. Deckelbaum, and C. Tzamos (2014): "The Complexity of Optimal Mmechanism Design," in *Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms*, SIAM, 1302–1318.

——— (2017): "Strong Duality for a Multiple-Good Monopolist," *Econometrica*, 85(3), 735–767.

Diamond, P. A. (1998): "Optimal income taxation: an example with a U-shaped pattern of optimal marginal tax rates," *American Economic Review*, 83–95.

Doligalski, P., P. Dworczak, M. Akbarpour, and S. D. Kominers (2025): "Optimal redistribution via income taxation and market design," Tech. rep., Working Paper.

Eso, P. and B. Szentes (2007): "Optimal Information Disclosure in Auctions and the Handicap Auction," *Review of Economic Studies*, 74(3), 705–731.

Ferey, A., B. B. Lockwood, and D. Taubinsky (2024): "Sufficient statistics for nonlinear tax systems with general across-income heterogeneity," *American Economic Review*, 114, 3206–3249.

Frick, M., R. Iijima, and Y. Ishii (2024): "Multidimensional screening with rich consumer data," *arXiv preprint arXiv:2411.06312*.

Fuchino, S. and S. Plewik (1999): "On a Theorem of E. Helly," *Proceedings of the American Mathematical Society*, 127(2), 491–497.

Ghili, S. (2023): "A Characterization for Optimal Bundling of Products with Non-Additive Values," *American Economic Review: Insights*.

Golosov, M. and I. Krasikov (2025): "The optimal taxation of couples," *The Quarterly Journal of Economics*, qjaf021.

Haghpanah, N. and J. Hartline (2021): "When is Pure Bundling Optimal?" *Review of Economic Studies*, 88(3), 1127–1156.

Haghpanah, N. and R. Siegel (2025): "Screening Two Types," in *Proceedings of the 26th ACM Conference on Economics and Computation*, 921–921.

Johnson, J. P. and D. P. Myatt (2003): "Multiproduct Quality Competition: Fighting Brands and Product Line Pruning," *American Economic Review*, 93(3), 748–774.

Karlin, S. and H. Rubin (1956): "The theory of decision procedures for distributions with monotone likelihood ratio," *The Annals of Mathematical Statistics*, 272–299.

Lahr, P. and A. Niemeyer (2024): "Extreme Points in Multi-Dimensional Screening," *arXiv preprint arXiv:2412.00649*.

Lewis, T. R. and D. E. Sappington (1994): "Supplying information to facilitate price discrimination," *International Economic Review*, 309–327.

Madarász, K. and A. Prat (2017): "Sellers with Misspecified Models," *Review of Economic Studies*, 84(2), 790–815.

Manelli, A. M. and D. R. Vincent (2006): "Bundling as an Optimal Selling Mechanism for a Multiple-Good Monopolist," *Journal of Economic Theory*, 127(1), 1–35.

McAfee, R. P. and J. McMillan (1988): "Multidimensional Incentive Compatibility and Mechanism Design," *Journal of Economic Theory*, 46(2), 335–354.

McAfee, R. P., J. McMillan, and M. D. Whinston (1989): "Multiproduct Monopoly, Commodity Bundling, and Correlation of Values," *Quarterly Journal of Economics*, 104.

Müller, A. and D. Stoyan (2002): *Comparison Methods for Stochastic Models and Risks*, New York: Wiley.

Myerson, R. B. (1981): "Optimal Auction Design," *Mathematics of Operations Research*, 6(1), 58–73.

Nichols, A. and R. Zeckhauser (1982): "Targeting Transfers through Restrictions on Recipients," *American Economic Review*, 72, 372–377.

Pavan, A., I. Segal, and J. Toikka (2014): "Dynamic Mechanism Design: A Myersonian Approach," *Econometrica*, 82(2), 601–653.

Pernoud, A. and F. Yang (2025): "Bundling against Learning," *arXiv preprint arXiv:2509.16396*.

Rafkin, C., A. Solomon, and E. Soltas (2023): "Self-targeting in US transfer programs," *Available at SSRN 4495537*.

Riley, J. and R. Zeckhauser (1983): "Optimal Selling Strategies: When to Haggle, When to Hold Firm," *Quarterly Journal of Economics*, 98(2), 267–289.

Rochet, J.-C. and P. Chone (1998): "Ironing, Sweeping, and Multidimensional Screening," *Econometrica*, 66(4), 783–826.

Rochet, J. C. and L. A. Stole (2003): "The Economics of Multidimensional Screening," in *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress, Volume 1*, Cambridge: Cambridge University Press.

Strack, P. and K. H. Yang (2025): "Non-Discriminatory Personalized Pricing," *arXiv preprint arXiv:2506.20925*.

Strassen, V. (1965): "The Existence of Probability Measures with Given Marginals," *Annals of Mathematical Statistics*, 36(2), 423–439.

Strausz, R. (2003): "Deterministic Mechanisms and the Revelation Principle," *Economics Letters*, 79(3), 333–337.

Wilson, R. B. (1993): *Nonlinear Pricing*, Oxford: Oxford University Press.

Winkler, G. (1988): "Extreme Points of Moment Sets," Tech. Rep. 4.

Yang, F. (2023): "Nested bundling," *Working Paper*, https://web.stanford.edu/~shuny/papers/bundling.pdf.

——— (2025): "Nested bundling," *American Economic Review*, 115 (9), 2970–3013.

——— (forthcoming): "Costly Multidimensional Screening," *Review of Economic Studies*, rdaf105.

Yang, F., P. Dworczak, and M. Akbarpour (2024): "Comparison of screening devices," Tech. rep.

Online Appendix to "Screening Frontiers"

Frank Yang

A Omitted Proofs

A.1 Proof of Theorem 1

The proof follows closely the outline in Section 4. We divide the proof into five steps. Section A.1.1 derives some preliminary inequalities. Section A.1.2 proves the reconstruction lemma. Section A.1.3 proves the generalized downward sufficiency theorem. Section A.1.4 proves the purification lemma. Section A.1.5 completes the proof first for finite allocation space and then extends the result to arbitrary allocation space.

A.1.1 Preliminary Inequalities

Lemma 3. *Let $\mathcal{X}^\star = \{x_1, \dots, x_m\}$ be a surplus-elasticity frontier, where a higher index corresponds to a pointwise higher demand curve. For any $x \notin \mathcal{X}^\star$, let $x_k \in \mathcal{X}^\star$ such that*

$$x \preceq_{\text{elasticity}} x_k .$$

Then, for all $i < j \leqslant k$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\log(v(x,t) - v(x_i,t)) > \frac{\mathrm{d}}{\mathrm{d}t}\log(v(x_j,t) - v(x_i,t))$$

for all

$$t \in \mathcal{T}' := \left\{s : v(x_j,s) > v(x,s) > v(x_i,s)\right\}.$$

Moreover, if $\mathcal{T}' \neq \varnothing$, then it must be an interval.

Proof. Since $\mathcal{X}^\star$ is a surplus-elasticity frontier, for all $i < j \leqslant k$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\log(v(x,t)) \geqslant \frac{\mathrm{d}}{\mathrm{d}t}\log(v(x_k,t)) \geqslant \frac{\mathrm{d}}{\mathrm{d}t}\log(v(x_j,t)) > \frac{\mathrm{d}}{\mathrm{d}t}\log(v(x_i,t)) \text{ for all } t .$$

By the mediant inequality, this implies the following two inequalities: *(i)* for all $i < j \leqslant k$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\log(v(x_j,t) - v(x_i,t)) > \frac{\mathrm{d}}{\mathrm{d}t}\log(v(x_j,t)) > \frac{\mathrm{d}}{\mathrm{d}t}\log(v(x_i,t)) \geqslant 0 \text{ for all } t.$$

and *(ii)* for all $i < j \leqslant k$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\log(v(x,t)) \geqslant \frac{\mathrm{d}}{\mathrm{d}t}\log(v(x_j,t)) \geqslant \frac{\mathrm{d}}{\mathrm{d}t}\log(v(x_j,t)-v(x,t)) \text{ for all } t \text{ s.t. } v(x_j,t) > v(x,t).$$

Combining these two inequalities gives *(iii)* for all $i < j \leqslant k$

$$\frac{\mathrm{d}}{\mathrm{d}t}\log(v(x_j,t)-v(x_i,t)) > \frac{\mathrm{d}}{\mathrm{d}t}\log(v(x_j,t)) \geqslant \frac{\mathrm{d}}{\mathrm{d}t}\log(v(x_j,t)-v(x,t)) \text{ for all } t \text{ s.t. } v(x_j,t) > v(x,t).$$

Finally, by combining the mediant inequality and inequality *(iii)*, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\log(v(x,t)-v(x_i,t)) > \frac{\mathrm{d}}{\mathrm{d}t}\log(v(x_j,t)-v(x_i,t)) > \frac{\mathrm{d}}{\mathrm{d}t}\log(v(x_j,t)-v(x,t))$$

for all

$$t \in \mathcal{T}' = \left\{s : v(x_j,s) > v(x,s) > v(x_i,s)\right\},$$

proving the inequality. Now, note that

$$\frac{\mathrm{d}}{\mathrm{d}t}\log(v(x,t)) \geqslant \frac{\mathrm{d}}{\mathrm{d}t}\log(v(x_j,t)) \text{ for all } t$$

implies that for all $t < t'$,

$$v(x,t) \geqslant v(x_j,t) \implies v(x,t') \geqslant v(x_j,t').$$

Therefore, the set $\{s : v(x_j,s) > v(x,s)\}$ is of the form $[\underline{t},t')$ or $[\underline{t},\overline{t}]$. Similarly, the set $\{s : v(x,s) > v(x_i,s)\}$ is of the form $(t'',\overline{t}]$ or $[\underline{t},\overline{t}]$. Therefore, if $\mathcal{T}' \neq \varnothing$, then $\mathcal{T}'$ must be an interval. □

A.1.2 Proof of Lemma 1

Note that by construction, for all t, we have

$$\begin{aligned} v(a(t),t) &= \frac{v(x(t),t)-v(x^-(t),t)}{v(x^+(t),t)-v(x^-(t),t)} \cdot v(x^+(t),t) + \left[1 - \frac{v(x(t),t)-v(x^-(t),t)}{v(x^+(t),t)-v(x^-(t),t)}\right] \cdot v(x^-(t),t) \\ &= v(x(t),t). \end{aligned}$$

Thus, $v(a(t),t) - p(t) = v(x(t),t) - p(t)$. Hence, if type t reports truthfully, then its payoff is unchanged in the reconstruction compared to that in (x,p). Because the mechanism (x,p) satisfies the downward IC constraints, to show that (a,p) satisfies the downward IC constraints, it suffices to show that the deviating payoff of type t misreporting to be $\hat{t} < t$

is weakly lower under (a, p) compared to the deviating payoff under (x, p).

Since the payment rule is unchanged, it suffices to show that for $\hat{t} < t$,

$$v(a(\hat{t}), t) = \alpha(\hat{t}) v(x^+(\hat{t}), t) + (1 - \alpha(\hat{t})) v(x^-(\hat{t}), t) \leqslant v(x(\hat{t}), t). \tag{A.1}$$

To ease notation, fix $\hat{t} < t$ and write $\hat{x}$, $\hat{x}^+$, and $\hat{x}^-$ for $x(\hat{t})$, $x^+(\hat{t})$, and $x^-(\hat{t})$ respectively. By construction, $\hat{x}^- \prec_{\text{surplus}} \hat{x}^+$. We can write the deviating payoff as

$$\begin{aligned} \alpha(\hat{t}) v(x^+(\hat{t}), t) + (1 - \alpha(\hat{t})) v(x^-(\hat{t}), t) &= \alpha(\hat{t}) \big(v(\hat{x}^+, t) - v(\hat{x}^-, t) \big) + v(\hat{x}^-, t) \\ &= \frac{v(\hat{x}, \hat{t}) - v(\hat{x}^-, \hat{t})}{v(\hat{x}^+, \hat{t}) - v(\hat{x}^-, \hat{t})} \big(v(\hat{x}^+, t) - v(\hat{x}^-, t) \big) + v(\hat{x}^-, t). \end{aligned}$$

Note that if $\hat{x} \in \mathcal{X}^\star$, then $\hat{x} = \hat{x}^+$ and hence (A.1) clearly holds. From now on, suppose $\hat{x} \notin \mathcal{X}^\star$. Then, since the menu $\mathcal{X}^\star$ is a surplus-elasticity frontier, by the definition of $\hat{x}^+$, we must have $\hat{x}^+ \preceq_{\text{surplus}} x_k \in \mathcal{X}^\star$ for some $x_k \succ_{\text{surplus}} \hat{x}$ such that

$$\frac{\mathrm{d}}{\mathrm{d}s} \log v(\hat{x}, s) \geqslant \frac{\mathrm{d}}{\mathrm{d}s} \log v(x_k, s) \geqslant \frac{\mathrm{d}}{\mathrm{d}s} \log v(\hat{x}^+, s) > \frac{\mathrm{d}}{\mathrm{d}s} \log v(\hat{x}^-, s) \text{ for all } s.$$

This implies that for all $s < s'$,

$$v(\hat{x}, s) \geqslant v(\hat{x}^+, s) \implies v(\hat{x}, s') \geqslant v(\hat{x}^+, s'). \tag{A.2}$$

Note that if $v(\hat{x}, t) \geqslant v(\hat{x}^+, t)$, then (A.1) clearly holds as

$$\alpha(\hat{t}) \big(v(\hat{x}^+, t) - v(\hat{x}^-, t) \big) + v(\hat{x}^-, t) \leqslant v(\hat{x}^+, t) \leqslant v(\hat{x}, t).$$

Henceforth, suppose $v(\hat{x}, t) < v(\hat{x}^+, t)$. Then, by (A.2), we have $v(\hat{x}, \hat{t}) < v(\hat{x}^+, \hat{t})$. By definition of $\hat{x}^-$, we have $v(\hat{x}^-, \hat{t}) < v(\hat{x}, \hat{t}) < v(\hat{x}^+, \hat{t})$. But for any $s < s'$ we also have

$$v(\hat{x}, s) > v(\hat{x}^-, s) \implies v(\hat{x}, s') > v(\hat{x}^-, s'). \tag{A.3}$$

Thus, we have $v(\hat{x}^-, t) < v(\hat{x}, t) < v(\hat{x}^+, t)$. Therefore, we have

$$\{\hat{t}, t\} \subseteq \mathcal{T}' := \big\{ s \in \mathcal{T} : v(\hat{x}^-, s) < v(\hat{x}, s) < v(\hat{x}^+, s) \big\}.$$

Since $\mathcal{T}' \neq \varnothing$, by Lemma 3, $\mathcal{T}'$ must be an interval. Moreover, by Lemma 3, we have

$$\frac{\mathrm{d}}{\mathrm{d}s} \log(v(\hat{x}, s) - v(\hat{x}^-, s)) > \frac{\mathrm{d}}{\mathrm{d}s} \log(v(\hat{x}^+, s) - v(\hat{x}^-, s))$$

for all $s \in \mathcal{T}'$, which implies that

$$\frac{\mathrm{d}}{\mathrm{d}s} \log \left[\frac{v(\hat{x}, s) - v(\hat{x}^-, s)}{v(\hat{x}^+, s) - v(\hat{x}^-, s)} \right] > 0$$

for all $s \in \mathcal{T}'$. Thus, we have

$$g(s) := \frac{v(\hat{x}, s) - v(\hat{x}^-, s)}{v(\hat{x}^+, s) - v(\hat{x}^-, s)}$$

is strictly increasing on the interval $\mathcal{T}'$. But then since $\hat{t} < t \in \mathcal{T}'$, we have

$$\begin{aligned} \alpha(\hat{t})v(x^+(\hat{t}), t) + (1 - \alpha(\hat{t}))v(x^-(\hat{t}), t) &= \frac{v(\hat{x}, \hat{t}) - v(\hat{x}^-, \hat{t})}{v(\hat{x}^+, \hat{t}) - v(\hat{x}^-, \hat{t})} \big(v(\hat{x}^+, t) - v(\hat{x}^-, t) \big) + v(\hat{x}^-, t). \\ &\leqslant \frac{v(\hat{x}, t) - v(\hat{x}^-, t)}{v(\hat{x}^+, t) - v(\hat{x}^-, t)} \big(v(\hat{x}^+, t) - v(\hat{x}^-, t) \big) + v(\hat{x}^-, t). \\ &= v(\hat{x}, t) - v(\hat{x}^-, t) + v(\hat{x}^-, t) = v(\hat{x}, t), \end{aligned}$$

which proves (A.1). The claim follows.

A.1.3 Proof of Theorem 4

Completion for the finite-type case. To complete the proof of Theorem 4 in the finite-type case, it suffices to show that the finite (**Downward-IC**) program admits a solution. Fix any $a \in \mathcal{A}$. We consider the problem of maximizing over $p \in \mathbb{R}^n$, or equivalently over $u \in \mathbb{R}^n$ where u_i is the net utility of type i. Since fixing a, the problem of maximizing over u is a linear program, once we show that there always exists a solution for every a, it follows that the value function over a would be continuous and admit an optimal solution by a standard compactness argument. For any $i > j$, let

$$\Delta_{ji}(a) := v(a_j, t_i) - v(a_j, t_j) \geqslant 0.$$

Then IC$[i \to j]$ can be equivalently written as

$$u_i \geqslant u_j + \Delta_{ji}(a).$$

IR$[i]$ is simply $u_i \geqslant 0$. Let $\mathcal{U}(a)$ denote the feasible polyhedron. The objective can be written as

$$\sum_i \mu_i(\lambda_i u_i + v(a_i, t_i) - u_i) = \sum_i \mu_i v(a_i, t_i) + \sum_i \mu_i(\lambda_i - 1)u_i.$$

Clearly the LP over u is feasible. By the fundamental theorem of LP, to show that it admits an optimal solution, it suffices to show that it cannot be unbounded. To show that, it suffices to show that the objective is bounded along any recession direction. Consider any $d \in \mathbb{R}^n$ such that $u + \alpha d \in \mathcal{U}(a)$ for all $\alpha \geqslant 0$. We claim that $d_j \leqslant d_i$ for all $j < i$. Indeed, IC$[i \to j]$ gives that

$$(u_i - u_j) + \alpha(d_i - d_j) \geqslant \Delta_{ji}(a),$$

for which to hold for all $\alpha \geqslant 0$, we must have $d_i \geqslant d_j$. Moreover, IR constraints imply that $d_i \geqslant 0$ for all i. Now, note that for any $d \in \mathcal{U}(a)$ such that d_i is non-decreasing in i, by Abel's identity, we have

$$\sum_i \mu_i(\lambda_i - 1)d_i = \sum_k \Big(\sum_{i=k}^n \mu_i(\lambda_i - 1)\Big)(d_k - d_{k-1}),$$

with $d_0 := 0$. Note that since $\mathbb{E}[\lambda] \leqslant 1$ and λ_i is non-increasing in i, we must have

$$\sum_{i=k}^n \mu_i(\lambda_i - 1) \leqslant 0$$

for all k. Therefore, since d_i is non-decreasing in i, we have

$$\sum_i \mu_i(\lambda_i - 1)d_i = \sum_k \Big(\sum_{i=k}^n \mu_i(\lambda_i - 1)\Big)(d_k - d_{k-1}) \leqslant 0.$$

It follows that the objective value must be non-increasing along any recession direction d. Thus, the objective value is bounded and hence the LP admits an optimal solution.

Approximation argument for general case. We prove the generalized downward sufficiency theorem for the general case where $\mathcal{T}, \mathcal{A}$ are any compact subsets of $\mathbb{R}$. Since $\lambda(t)$ is continuous in t, and $v(a,t)$ is continuous on $\mathcal{A} \times \mathcal{T}$, the approximation arguments are identical to those given in Appendix A.1.3 of Yang (forthcoming), provided that one shows the full IC program for the general-type case admits a solution with uniformly bounded transfers. The existence proof is identical to that given in Lemma 7 of Yang (forthcoming) via Helly's selection theorem, except that we need to show that even with welfare weights λ, the transfers can be without loss of optimality restricted to satisfy a uniform upper and lower bound.

We argue that it is without loss of optimality to restrict attention to payment rules

$p : \mathcal{T} \to [\underline{v} - (\overline{v} - \underline{v}), \overline{v}]$, where

$$\underline{v} := \min_{(a,t) \in \mathcal{A} \times \mathcal{T}} v(a,t) \qquad \overline{v} := \max_{(a,t) \in \mathcal{A} \times \mathcal{T}} v(a,t).$$

Note that by IR constraints, we immediately have

$$p(t) \leqslant v(a(t), t) \leqslant \overline{v}.$$

Now to see the lower bound, we first claim that for the lowest type $\underline{t}$ we must have $p(\underline{t}) \geqslant \underline{v}$. Indeed, suppose otherwise. Then, $p(\underline{t}) < \underline{v}$, which means that the IR constraint for the lowest type must be slack. Let $\varepsilon := \underline{v} - p(\underline{t})$. Change the original payment rule p to $\tilde{p}$ where $\tilde{p} = p + \varepsilon$ pointwise. Then, this new mechanism is clearly IC, and satisfies IR constraint for the lowest type, which together also implies that it satisfies all IR constraints due to that $v(a,t)$ is non-decreasing in t. Note that the change in the objective is given by $\varepsilon - \mathbb{E}[\lambda]\varepsilon \geqslant 0$ since $\mathbb{E}[\lambda] \leqslant 1$, proving the claim.

Now using the IC constraint from $\underline{t}$ to any type t, we get that

$$v(a(\underline{t}), \underline{t}) - p(\underline{t}) \geqslant v(a(t), \underline{t}) - p(t)$$

and thus

$$p(t) \geqslant v(a(t), \underline{t}) - v(a(\underline{t}), \underline{t}) + p(\underline{t}) \geqslant (\underline{v} - \overline{v}) + \underline{v},$$

proving the claim. Therefore, it is without loss of optimality to restrict attention to payment rules $p : \mathcal{T} \to [\underline{v} - (\overline{v} - \underline{v}), \overline{v}]$. Thus, the full IC program with strict increasing differences preferences always admits a solution by the proof of Lemma 7 in Yang (forthcoming). By the same approximation arguments given in Appendix A.1.3 of Yang (forthcoming), it follows immediately that Theorem 4 holds.

A.1.4 Proof of Lemma 2

We follow the same notation as in the main text. By the argument given in the main text, $a(\cdot)$ must be stochastically monotone according to the order $\preceq_{\text{st}}$. Then, by Strassen (1965) (see Lemma 1 in Yang forthcoming), there exists some random variable $\varepsilon \in \mathcal{E}$ that is independent of t, where $\mathcal{E}$ is some measurable space, and some function $x : \mathcal{T} \times \mathcal{E} \to \mathcal{Y}$ such that *(i)* $x(\cdot, \varepsilon)$ is deterministically monotone according to $\preceq$ for all ε, and *(ii)* for every $t \in \mathcal{T}$, we have

$$x(t, \varepsilon) \sim a(t) \in \Delta(\mathcal{Y}).$$

Then, by the Envelope theorem, under the mechanism (a,p), the principal's objective is given by

$$\begin{aligned}
&\mathbb{E}\left[\widetilde{V}(a(t),t) - W(t)\int_{\underline{t}}^{t} v_t(a(z),z)\,\mathrm{d}z\right] - \mathbb{E}[W(t)]U(\underline{t})\\
&= \mathbb{E}_{t,\varepsilon}\left[\widetilde{V}(x(t,\varepsilon),t) - W(t)\int_{\underline{t}}^{t} v_t(x(z,\varepsilon),z)\,\mathrm{d}z\right] - \mathbb{E}[W(t)]U(\underline{t})\\
&\leqslant \sup_{\varepsilon\in\mathcal{E}}\mathbb{E}_t\left[\widetilde{V}(x(t,\varepsilon),t) - W(t)\int_{\underline{t}}^{t} v_t(x(z,\varepsilon),z)\,\mathrm{d}z\right] - \mathbb{E}[W(t)]U(\underline{t}).
\end{aligned}$$

Let

$$K := \sup_{\varepsilon\in\mathcal{E}}\mathbb{E}_t\left[\widetilde{V}(x(t,\varepsilon),t) - W(t)\int_{\underline{t}}^{t} v_t(x(z,\varepsilon),z)\,\mathrm{d}z\right]$$

and

$$K_\varepsilon := \mathbb{E}_t\left[\widetilde{V}(x(t,\varepsilon),t) - W(t)\int_{\underline{t}}^{t} v_t(x(z,\varepsilon),z)\,\mathrm{d}z\right]$$

Note that, by construction, there exists a sequence ε_j such that as $j\to\infty$, we have

$$K_{\varepsilon_j}\to K.$$

Now, consider the sequence $x(\cdot,\varepsilon_j)$. This is a sequence of monotone functions taking values in $\mathcal{Y}$ with the ordering given by $\preceq$. Since $\mathcal{Y}$ is compact, by Helly's selection theorem for monotone functions on linearly ordered sets (Fuchino and Plewik 1999, Theorem 7), there exists a subsequence $\{x(\cdot,\varepsilon_{j_k})\}_k$ that converge pointwise.[16] For all t, let

$$x^\dagger(t) := \lim_{k\to\infty} x(t,\varepsilon_{j_k}).$$

Clearly, we have that $x^\dagger(\cdot)$ is monotone. Moreover, since $\mathcal{Y}$ is compact and v_t is continuous, by the dominated convergence theorem, we have

$$\mathbb{E}\left[\widetilde{V}(x^\dagger(t),t) - W(t)\int_{\underline{t}}^{t} v_t(x^\dagger(z),z)\,\mathrm{d}z\right] = \mathbb{E}\left[\lim_{k\to\infty}\left(\widetilde{V}(x(t,\varepsilon_{j_k}),t) - W(t)\int_{\underline{t}}^{t} v_t(x(z,\varepsilon_{j_k}),z)\,\mathrm{d}z\right)\right]$$

[16] In particular, note that given the strict increasing differences on $(\mathcal{Y},\preceq)$, the elements in Y can be equivalently ordered by $\psi(x) := v(x,\bar{t}) - v(x,\underline{t})$, which defines a continuous bijection $\psi:\mathcal{Y}\to\psi(\mathcal{Y})$. Since $\mathcal{Y}$ is compact, ψ is a homeomorphism. Thus, $\mathcal{Y}$ can be equivalently viewed as a subset of $\mathbb{R}$, which combined with Helly's selection theorem and the continuous inverse ψ^{-1}, implies a pointwise convergence subsequence.

$$= \lim_{k\to\infty} \mathbb{E}\left[\left(\widetilde{V}(x(t,\varepsilon_{j_k}),t) - W(t)\int_{\underline{t}}^{t} v_t(x(z,\varepsilon_{j_k}),z)\,\mathrm{d}z\right)\right]$$
$$= K$$
$$\geqslant \mathbb{E}\left[\widetilde{V}(a(t),t) - W(t)\int_{\underline{t}}^{t} v_t(a(z),z)\,\mathrm{d}z\right].$$

Since $x^{\dagger}(\cdot)$ is monotone, the Envelope theorem implied transfer rule implements $x^{\dagger}(\cdot)$ given increasing differences. Moreover, by setting the level of the transfers so that the lowest type $\underline{t}$ gets a net payoff of $U(\underline{t})$, we have thus constructed a new mechanism $(x^{\dagger}, p^{\dagger})$ that weakly improves on (a,p).

A.1.5 Completion of the Proof via Approximation

By the argument in the main text and preceding lemmas, we have proved Theorem 1 for the case where $\mathcal{X}$ is a finite set. We now generalize the result to the case where $\mathcal{X}$ is any compact metric space.

Let the optimal value of the principal's objective be denoted by

$$V(\mathcal{Y}) := \sup_{(x,p):\text{IC, IR, Ran}(x)\subseteq\mathcal{Y}} \text{Objective}(x,p),$$

for any $\mathcal{Y} \subset \mathcal{X}$. Since $\mathbb{E}[\lambda] \leqslant 1$, it is easy to see that $V(\mathcal{Y})$ is well defined for any $\mathcal{Y}$.

Building on Madarász and Prat (2017), we prove the following finite approximation lemma, which would be the key to generalizing the result:

Lemma 4 (Finite Approximation)**.** *For any compact metric $\mathcal{X}$, there exists a sequence $\mathcal{X}^{(n)} \subseteq \mathcal{X}$ where each $\mathcal{X}^{(n)}$ is finite such that*

$$\lim_{n\to\infty} V(\mathcal{X}^{(n)}) = V(\mathcal{X}).$$

We delay the proof of this lemma to the end. Assuming the above lemma, we now extend Theorem 1 to any compact metric $\mathcal{X}$.

Completion of the proof of Theorem 1. We first show that

$$V(\mathcal{X}^{\star}) = V(\mathcal{X}),$$

and then show that there exists an optimal solution. Fix the approximating sequence from Lemma 4. Fix some n. Let m enumerate the options in $\mathcal{X}^{(n)}$. For any $x_m \in \mathcal{X}^{(n)} \backslash \mathcal{X}^{\star}$, by construction of $\mathcal{X}^\star$, there exists some $\tilde{x}_m \in \mathcal{X}^\star$ such that

$$x_m \preceq_{\text{surplus}} \tilde{x}_m \text{ and } x_m \preceq_{\text{elasticity}} \tilde{x}_m.$$

For any $x_m \in \mathcal{X}^{(n)} \cap \mathcal{X}^\star$, let $\tilde{x}_m = x_m$. Now, consider the collection

$$\mathcal{X}_n^\star := \{\tilde{x}_m\}_m \subseteq \mathcal{X}^\star.$$

Note that, by construction, $\mathcal{X}_n^\star$ is a surplus-elasticity frontier of

$$\tilde{\mathcal{X}}^{(n)} = \mathcal{X}^{(n)} \cup \{\tilde{x}_m\}_m$$

which is finite. Moreover, by construction, we also have $V(\tilde{\mathcal{X}}^{(n)}) \geqslant V(\mathcal{X}^{(n)})$ and hence by Lemma 4,

$$V(\mathcal{X}) \geqslant \lim_{n\to\infty} V(\tilde{\mathcal{X}}^{(n)}) \geqslant \lim_{n\to\infty} V(\mathcal{X}^{(n)}) = V(\mathcal{X}),$$

showing that $V(\tilde{\mathcal{X}}^{(n)})$ converges to $V(\mathcal{X})$ as well. Now, apply Theorem 1 to $\tilde{\mathcal{X}}^{(n)}$. By Theorem 1, we immediately have that

$$V(\mathcal{X}_n^\star) = V(\tilde{\mathcal{X}}^{(n)}),$$

and therefore

$$V(\mathcal{X}) \geqslant V(\mathcal{X}^\star) \geqslant \lim_{n\to\infty} V(\mathcal{X}_n^\star) = V(\mathcal{X}),$$

proving the claim.

We now show that an optimal mechanism using allocations in $\mathcal{X}^\star$ exists. By the proof given in Section A.1.3, for any n, since the agent's preferences have strict increasing differences on $\mathcal{X}_n^\star$, there exists an optimal mechanism for the full IC program when the allocation space is restricted to $\mathcal{X}_n^\star$. Let $\left(x^{(n)}, p^{(n)}\right)$ be a sequence of the optimal mechanisms. Let $\preceq$ denote the linear order on $\mathcal{X}^\star$ indexed by s (i.e., the surplus order on the frontier). By construction, $x^{(n)} : \mathcal{T} \to \mathcal{X}^\star$ is a monotone function with respect to $\preceq$. Since $\mathcal{X}^\star$ is compact, using Helly's selection theorem for monotone functions on linearly ordered sets (Fuchino and Plewik 1999, Theorem 7), we can obtain a subsequence $x^{(n_j)}$ that converges pointwise as j increases.[17] Let

$$u^{(n)}(t) = v(x^{(n)}(t), t) - p^{(n)}(t)$$

be the equilibrium payoff of type t. Clearly, $u^{(n)}$ is also a monotone function for all n. Moreover, as we have shown before, the payment rule can be without loss of optimality restricted attention to a uniform upper bound and lower bound (independent of n). Therefore, $u^{(n)}$ are uniformly bounded monotone functions. Applying the same Helly's selection theorem again to the sequence $u^{(n_j)}$, we immediately obtain a subsequence $(x^{(n_{j_k})}, u^{(n_{j_k})})$ that converges pointwise. Equivalently, the sequence $(x^{(n_{j_k})}, u^{(n_{j_k})}, p^{(n_{j_k})})$ converges pointwise to some (x, u, p). Since $v(x,t)$ is continuous on $\mathcal{X} \times \mathcal{T}$, it follows immediately that (x, p) satisfies all IC and IR constraints. Finally, note that the objective value of this mechanism is given by

$$\mathbb{E}\big[\lambda(t)u(t) + p(t)\big] = \lim_{k\to\infty} \mathbb{E}\big[\lambda(t)u^{(n_{j_k})}(t) + p^{(n_{j_k})}(t)\big] = \lim_{k\to\infty} V(\mathcal{X}^\star_{n_{j_k}}) = V(\mathcal{X}),$$

where the first equality follows by the dominated convergence theorem. Thus, (x, p) is an optimal mechanism that uses allocations only in the surplus-elasticity frontier $\mathcal{X}^\star$, completing the proof of Theorem 1.

Proof of Lemma 4. We explicitly construct a sequence of finite allocation spaces $\mathcal{X}^{(n)}$ that satisfies the desired property. For any $n \in \mathbb{N}$, partition $\mathcal{T}$ into n subintervals:

$$\underline{t} = t_0 < t_1 < \cdots < t_n = \overline{t}, \qquad t_{k+1} - t_k = \frac{\overline{t} - \underline{t}}{n} =: \Delta_n.$$

Let $\mathcal{T}^{(n)} := \{t_1, \ldots, t_n\}$ and let $\mu^{(n)} \in \Delta(\mathcal{T}^{(n)})$ be defined by: for all k

$$\mu^{(n)}_k := \mathbb{P}\big(t \in [t_{k-1}, t_k)\big) = F(t_k) - F(t_{k-1}).$$

Recall that F is the continuous CDF of t. Let $\mu \in \Delta(\mathcal{T})$ be the distribution given by F. Let $\lambda^{(n)}_k := \lambda(t_k)$ for all k.

For any n, consider the same screening problem we have, but with the type space and

[17] In particular, note that the strict surplus order on the frontier $\mathcal{X}^\star$ can be equivalently represented by $\psi(x) := v(x, t_0)$ for some fixed t_0, which defines a continuous bijection $\psi : \mathcal{X}^\star \to \psi(\mathcal{X}^\star)$. Since $\mathcal{X}^\star$ is compact, ψ is a homeomorphism. Thus, $\mathcal{X}^\star$ can be equivalently viewed as a subset $\mathcal{Y} \subset \mathbb{R}$, which combined with Helly's selection theorem and the continuous inverse ψ^{-1}, implies a pointwise convergence subsequence.

type distribution given by $(\mathcal{T}^{(n)}, \mu^{(n)})$. In particular, the objective is given by

$$\sum_k \mu_k^{(n)} \Big(\lambda_k^{(n)} \big(v(x^{(n)}(t_k), t_k) - p^{(n)}(t_k) \big) + p^{(n)}(t_k) \Big) .$$

Fixing any allocation rule $x^{(n)}$, the optimization problem for $p^{(n)}$ is a finite-dimensional LP, which by the same argument in Section A.1.3, admits an optimal solution. The LP value function is continuous in $x^{(n)} \in \mathcal{X}^n$, which combined with the continuity property of $v(x,t)$ implies that an optimal solution exists for this program by a standard compactness argument. Let $(x^{(n)}, p^{(n)})$ be an optimal solution to this program. Now define

$$\mathcal{X}^{(n)} := \mathrm{Ran}(x^{(n)}) \subseteq \mathcal{X} ,$$

which is a finite subset of $\mathcal{X}$ for any n. We claim that

$$\lim_{n \to \infty} V(\mathcal{X}^{(n)}) = V(\mathcal{X}) .$$

To prove it, let $W(\mathcal{T}^{(n)}, \mu^{(n)})$ be the optimal value for the same screening problem under the type space $\mathcal{T}^{(n)} \subseteq \mathcal{T}$ and the type distribution $\mu^{(n)} \in \Delta(\mathcal{T}^{(n)})$. For two distributions $\mu_1, \mu_2 \in \Delta(\mathcal{T})$, μ_1 and μ_2 are said to be δ-***close*** if $\mathcal{T}$ can be partitioned into disjoint measurable sets $S_1, \ldots, S_k$ such that $||t - t'||_{\sup} < \delta$ for any t, t' in the same cell S_j and $\mu_1(S_j) = \mu_2(S_j)$ for each S_j (Madarász and Prat 2017). Note that by construction, $\mu^{(n)}$ and μ are Δ_n-close for any n.

Now, we make use of the following result:

Lemma 5 (Madarász and Prat 2017, Carroll 2017)**.** *For any $\varepsilon > 0$, there exists $\delta > 0$ such that for any mechanism (x, p), there exists another mechanism $(\tilde{x}, \tilde{p})$ such that (i) for any two δ-close distributions μ, μ',*

$$\mathbb{E}_{\mu'}[\tilde{p}(t)] > \mathbb{E}_{\mu}[p(t)] - \varepsilon ,$$

and (ii)

$$\{\tilde{x}(t)\}_t \subseteq \mathrm{Closure}(\{x(t)\}_t) ,$$

and (iii)

$$\tilde{u}(t) \geqslant u(t) \;\; \textit{for all } t \in \mathcal{T}$$

where $\tilde{u}, u$ are the indirect utility functions given $(\tilde{x}, \tilde{p})$ and (x, p) respectively.

Now, fix any $\varepsilon > 0$. Let $\delta > 0$ be given by Lemma 5. Fix any mechanism (x, p) defined on the type space $\mathcal{T}$. Let $(\tilde{x}, \tilde{p})$ be given by Lemma 5. Fix any n such that $\Delta_n \leqslant \delta$. By the

Envelope theorem, $u(t)$ is continuous and hence $\lambda(t)u(t)$ is continuous on the compact set $[\underline{t}, \bar{t}]$, which by the Heine-Cantor theorem implies that $\lambda(t)u(t)$ is uniformly continuous. Let $\omega(\cdot)$ denote its modulus of continuity. By Lemma 5, we have

$$\mathbb{E}_{\mu^{(n)}}[\tilde{p}(t)] > \mathbb{E}_{\mu}[p(t)] - \varepsilon$$

and for all $t \in \mathcal{T}$,

$$\tilde{u}(t) \geqslant u(t).$$

The restriction of $(\tilde{x}, \tilde{p})$ to the type space of $\mathcal{T}^{(n)}$ defines a valid mechanism for the type space $\mathcal{T}^{(n)}$ with the objective value given by exactly

$$\begin{aligned}
\mathbb{E}_{\mu^{(n)}}\left[\tilde{p}(t) + \lambda(t)\tilde{u}(t)\right] &> \mathbb{E}_{\mu}\left[p(t)\right] - \varepsilon + \mathbb{E}_{\mu^{(n)}}\left[\lambda(t)\tilde{u}(t)\right] \\
&\geqslant \mathbb{E}_{\mu}\left[p(t)\right] - \varepsilon + \mathbb{E}_{\mu^{(n)}}\left[\lambda(t)u(t)\right] \\
&\geqslant \mathbb{E}_{\mu}\left[p(t)\right] - \varepsilon + \mathbb{E}_{\mu}\left[\lambda(t)u(t)\right] - \omega(\Delta_n) \\
&= \mathbb{E}_{\mu}\left[p(t) + \lambda(t)u(t)\right] - \varepsilon - \omega(\Delta_n).
\end{aligned}$$

Therefore,

$$W(\mathcal{T}^{(n)}, \mu^{(n)}) \geqslant \mathbb{E}_{\mu}\left[p(t) + \lambda(t)u(t)\right] - \varepsilon - \omega(\Delta_n).$$

Since this holds for all mechanisms (x, p), we immediately have that

$$W(\mathcal{T}^{(n)}, \mu^{(n)}) \geqslant V(\mathcal{X}) - \varepsilon - \omega(\Delta_n).$$

Therefore, we immediately have

$$\liminf_{n \to \infty} W(\mathcal{T}^{(n)}, \mu^{(n)}) \geqslant V(\mathcal{X}) - \varepsilon.$$

Taking $\varepsilon \to 0$ then gives

$$\liminf_{n \to \infty} W(\mathcal{T}^{(n)}, \mu^{(n)}) \geqslant V(\mathcal{X}).$$

Now we claim that

$$\liminf_{n \to \infty} V(\mathcal{X}^{(n)}) \geqslant \liminf_{n \to \infty} W(\mathcal{T}^{(n)}, \mu^{(n)}).$$

Fix any $\varepsilon > 0$. Let $\delta > 0$ be given by Lemma 5. Fix any n such that $\Delta_n \leqslant \delta$. Fix any optimal mechanism $(x^{(n)}, p^{(n)})$ defined on $\mathcal{T}$ with type distribution given by $\mu^{(n)} \in \Delta(\mathcal{T})$, such that it also satisfies $\mathrm{Ran}(x^{(n)}) \subseteq \mathcal{X}^{(n)}$. Since $\lambda(t)$ is continuous on the compact set $\mathcal{T}$, λ is also uniformly continuous with some modulus of continuity $\omega_\lambda(\cdot)$. Indeed, by the Envelope theorem we have that for all $|t-s| \leqslant \Delta$,

$$|\lambda(t)u^{(n)}(t) - \lambda(s)u^{(n)}(s)| \leqslant ||\lambda||_\infty \sup_{x,t} |v_t(x,t)|\Delta + ||u^{(n)}||_\infty \sup_{|t-s|\leqslant\Delta} |\lambda(t) - \lambda(s)|$$

Since $\sup_{x,t} |v_t(x,t)| \leqslant M$ for some constant M, and $||u^{(n)}||_\infty \leqslant M'$ for some M' (since, as we argued, we can take $u(\underline{t}) = 0$ without loss and the derivative is uniformly bounded), we can further bound the above by

$$|\lambda(t)u^{(n)}(t) - \lambda(s)u^{(n)}(s)| \leqslant \underbrace{||\lambda||_\infty M\Delta + M'\omega_\lambda(\Delta)}_{=:\omega(\Delta)}.$$

Then, $\omega(\Delta)$ is a modulus of continuity of $\lambda u^{(n)}$ that holds uniformly over n. Let $(\tilde{x}, \tilde{p})$ be given by Lemma 5 against $(x^{(n)}, p^{(n)})$. By Lemma 5, we have

$$\mathbb{E}_\mu[\tilde{p}(t)] > \mathbb{E}_{\mu^{(n)}}[p^{(n)}(t)] - \varepsilon$$

and for all $t \in \mathcal{T}$,

$$\tilde{u}(t) \geqslant u^{(n)}(t),$$

and $\text{Ran}(\tilde{x}) \subseteq \text{Closure}(\text{Ran}(x^{(n)})) \subseteq \mathcal{X}^{(n)}$. Therefore, by construction, we have

$$\begin{aligned}
V(\mathcal{X}^{(n)}) &\geqslant \mathbb{E}_\mu\left[\tilde{p}(t) + \lambda(t)\tilde{u}(t)\right] \\
&\geqslant \mathbb{E}_{\mu^{(n)}}[p^{(n)}(t)] - \varepsilon + \mathbb{E}_\mu[\lambda(t)u^{(n)}(t)] \\
&\geqslant \mathbb{E}_{\mu^{(n)}}[p^{(n)}(t)] - \varepsilon + \mathbb{E}_{\mu^{(n)}}[\lambda(t)u^{(n)}(t)] - \omega(\Delta_n) \\
&= W(\mathcal{T}^{(n)}, \mu^{(n)}) - \varepsilon - \omega(\Delta_n).
\end{aligned}$$

Therefore, we immediately have that

$$\liminf_{n\to\infty} V(\mathcal{X}^{(n)}) \geqslant \liminf_{n\to\infty} W(\mathcal{T}^{(n)}, \mu^{(n)}) - \varepsilon,$$

and hence taking $\varepsilon \to 0$ and using our previous conclusion gives

$$\liminf_{n\to\infty} V(\mathcal{X}^{(n)}) \geqslant \liminf_{n\to\infty} W(\mathcal{T}^{(n)}, \mu^{(n)}) \geqslant V(\mathcal{X}).$$

Evidently, $V(\mathcal{X}^{(n)}) \leqslant V(\mathcal{X})$ and thus $\limsup_n V(\mathcal{X}^{(n)}) \leqslant V(\mathcal{X})$. Thus,

$$\lim_{n\to\infty} V(\mathcal{X}^{(n)}) = V(\mathcal{X})$$

proving Lemma 4.

A.2 Proof of Theorem 2

Similar to the proof of Theorem 1, we prove Theorem 2 first assuming $\mathcal{X}$ is finite and then generalize it to arbitrary compact metric $\mathcal{X}$ via approximation.

A.2.1 Proof under a Finite Allocation Space

Suppose $\mathcal{X}$ is finite. Let $\mathcal{X}^\star$ be a strong frontier. Let $\mathcal{X}^\star = \{x_1, \dots, x_m\}$ where $x_i \prec_{\text{surplus}} x_j$ for all $i < j$ with $x_1 = \varnothing$ being the outside option. Note that for any i, we have

$$\frac{v(x_{i+1}, t) - v(x_i, t)}{v(x_i, t) - v(x_{i-1}, t)}$$

is strictly increasing in t by the ordered incremental elasticity property, which is equivalent to: for any $t_1 < t_2$ we have

$$\frac{v(x_{i+1}, t_2) - v(x_i, t_2)}{v(x_{i+1}, t_1) - v(x_i, t_1)}$$

is strictly increasing in i. We claim that for any $t_1 < t_2$, there exists a convex function g such that

$$v(x, t_2) = g(v(x, t_1)) \text{ for all } x \in \mathcal{X}^\star.$$

To see this, fix any $t_1 < t_2 \in (\underline{t}, \overline{t})$. Let $z_i = v(x_i, t_1)$ and $y_i = v(x_i, t_2)$ for all i. Since the elements in $\mathcal{X}^\star$ are strictly ordered by the surplus order, we know that both z_i and y_i are strictly increasing in i. Define $g(z_i) = y_i$ on $\{z_i\}_{i=1}^m$ and extend it via linear interpolation on $[z_1, z_m]$. Then we have $0 = y_1 = g(z_1) = g(0)$, and on each segment $[z_i, z_{i+1}]$,

$$g'(z) = \frac{g(z_{i+1}) - g(z_i)}{z_{i+1} - z_i} = \frac{v(x_{i+1}, t_2) - v(x_i, t_2)}{v(x_{i+1}, t_1) - v(x_i, t_1)} > 0.$$

Since the above slope is increasing in i, it follows immediately that g is convex, proving the claim.

Now, fix any stochastic mechanism. Without loss, we can write it as $t \mapsto (x(t, \varepsilon), p(t, \varepsilon))$ where $x(t, \varepsilon) \in \mathcal{X}, p(t, \varepsilon) \in \mathbb{R}$ and ε is a randomization device following an independent uniform distribution. Now, for each ε, applying the construction in Lemma 1 to the deterministic mechanism defined by $(x(\cdot,\varepsilon),p(\cdot,\varepsilon))$, we obtain a reconstructed stochastic mechanism $(a(\cdot,\varepsilon),p(\cdot,\varepsilon))$ where $a(\cdot,\varepsilon)\in\Delta(\mathcal{X}^\star)$ that keeps the same objective value, maintains the truth-telling payoff for each type, and weakly decreases any downward deviation payoff. Now, since the original mechanism is downward IC (when averaging over ε), it follows immediately that for all $\hat{t}<t$, we have

$$\begin{aligned}\mathbb{E}_\varepsilon\Big[v(a(t,\varepsilon),t)-p(t,\varepsilon)\Big] &= \mathbb{E}_\varepsilon\Big[v(x(t,\varepsilon),t)-p(t,\varepsilon)\Big]\\ &\geqslant \mathbb{E}_\varepsilon\Big[v(x(\hat{t},\varepsilon),t)-p(\hat{t},\varepsilon)\Big]\\ &\geqslant \mathbb{E}_\varepsilon\Big[v(a(\hat{t},\varepsilon),t)-p(\hat{t},\varepsilon)\Big],\end{aligned}$$

i.e., the downward IC constraints continue to hold under the compound stochastic mechanism $t\mapsto(a(t,\varepsilon),p(t,\varepsilon))$. Note that for each t, the compound lottery $a(t,\varepsilon)$ is equivalent to a simple lottery $\tilde{a}(t)\in\Delta(\mathcal{X}^\star)$. Let $\tilde{p}(t)=\mathbb{E}_\varepsilon[p(t,\varepsilon)]$ for all t. Then it follows immediately that $t\mapsto(\tilde{a}(t),\tilde{p}(t))\in\Delta(\mathcal{X}^\star)\times\mathbb{R}$ satisfies all downward IC and all IR constraints and gives the same expected payoff to all types t and to the principal, assuming truthful reporting.

Now, we perform one more round of reconstruction. Fix any t. The lottery $\tilde{a}(t)\in\Delta(\mathcal{X}^\star)$ induces a lottery in the utility space $\{z_i\}_{i=1}^m$ where $z_i=v(x_i,t)$. Let $\bar{z}:=\mathbb{E}_{x\sim\tilde{a}(t)}[v(x,t)]$. Then $\bar{z}\in[z_i,z_{i+1}]$ for some i. In particular, there exists a unique lottery $\alpha(t)\in\Delta(\{x_i,x_{i+1}\})$ such that $\mathbb{E}_{x\sim\alpha(t)}[v(x,t)]=\bar{z}$. Moreover, it is easy to see that $v^\alpha\preceq_{\text{mps}}v^{\tilde{a}}$ where $v^\alpha,v^{\tilde{a}}$ are the utility lottery for t from α and $\tilde{a}$ respectively, and $\preceq_{\text{mps}}$ is the mean-preserving spread order. For any $t'>t$, by our previous observation, we can write

$$v(x_i,t')=g(v(x_i,t))\text{ for all }x_i\in\mathcal{X}^\star,$$

where g is a convex function. It follows immediately that

$$\mathbb{E}_{\alpha(t)}[v(x_i,t')]=\mathbb{E}_{\alpha(t)}[g(v(x_i,t))]=\mathbb{E}[g(v^\alpha)]\leqslant\mathbb{E}[g(v^{\tilde{a}})]=\mathbb{E}_{\tilde{a}(t)}[v(x_i,t')],$$

where the inequality is due to that $v^\alpha\preceq_{\text{mps}}v^{\tilde{a}}$ and g is convex. Therefore, the mechanism defined by $t\mapsto(\alpha(t),\tilde{p}(t))$ gives the same payoff to all types t and hence the principal, assuming truthful reporting—and moreover any deviation payoff of type t' deviating to type $t<t'$ is weakly lowered and hence $(\alpha,\tilde{p})$ continues to satisfy all downward IC constraints and all IR constraints.

Now, observe that for any t, we have

$$\alpha(t) \in \mathcal{A} := \left\{ a \in \Delta(\mathcal{X}^\star) : a \in \Delta\big(\{x_{j-1}, x_j\}\big) \text{ for some } x_j \in \mathcal{X}^\star \right\},$$

which is the totally ordered set of lotteries used in the proof of Theorem 1. Applying the identical arguments in the proof of Theorem 1—using Theorem 4 and Lemma 2—we immediately obtain Theorem 2 for any finite $\mathcal{X}$.

A.2.2 Completion of the Proof via Approximation

The argument is similar to the approximation argument given for Theorem 1. Let the optimal value of the principal's objective be denoted by

$$V(\Delta(\mathcal{Y})) := \sup_{(a,p):\text{IC, IR, Ran}(a)\subseteq\Delta(\mathcal{Y})} \text{Objective}(a,p),$$

for any $\mathcal{Y} \subset \mathcal{X}$. Since $\mathbb{E}[\lambda] \leqslant 1$, it is easy to see that $V(\Delta(\mathcal{Y}))$ is well defined for any $\mathcal{Y}$.

Parallel to the finite approximation lemma (Lemma 4), we will prove a stochastic finite approximation lemma:

Lemma 6 (Stochastic Finite Approximation)**.** *For any compact metric $\mathcal{X}$, there exists a sequence $\mathcal{X}^{(n)} \subseteq \mathcal{X}$ where each $\mathcal{X}^{(n)}$ is finite such that*

$$\lim_{n\to\infty} V(\Delta(\mathcal{X}^{(n)})) = V(\Delta(\mathcal{X})).$$

Again, we delay the proof of this lemma to the end. Assuming the above lemma, we now extend Theorem 2 to any compact metric $\mathcal{X}$.

Completion of the proof of Theorem 2. We show that

$$V(\mathcal{X}^\star) = V(\Delta(\mathcal{X})).$$

Note that by Theorem 1, we already have the existence of an optimal mechanism among all deterministic mechanisms. Fix the approximating sequence from Lemma 6. The argument proceeds as before. Fix some n. Let m enumerate the options in $\mathcal{X}^{(n)}$. For any $x_m \in \mathcal{X}^{(n)} \backslash \mathcal{X}^\star$, by construction of $\mathcal{X}^\star$, there exists some $\tilde{x}_m \in \mathcal{X}^\star$ such that

$$x_m \preceq_{\text{surplus}} \tilde{x}_m \text{ and } x_m \preceq_{\text{elasticity}} \tilde{x}_m.$$

For any $x_m \in \mathcal{X}^{(n)} \cap \mathcal{X}^\star$, let $\tilde{x}_m = x_m$. Now, consider the collection

$$\mathcal{X}_n^\star := \{\tilde{x}_m\}_m \subseteq \mathcal{X}^\star .$$

Note that, by construction, $\mathcal{X}_n^\star$ is a surplus-elasticity frontier of

$$\tilde{\mathcal{X}}^{(n)} = \mathcal{X}^{(n)} \cup \{\tilde{x}_m\}_m$$

which is finite. Moreover, since $\mathcal{X}_n^\star \subseteq \mathcal{X}^\star$, $\mathcal{X}_n^\star$ continues to have the ordered incremental elasticity property and hence is a strong frontier of $\tilde{\mathcal{X}}^{(n)}$. Now, applying Theorem 2 to $\tilde{\mathcal{X}}^{(n)}$, we obtain

$$V(\mathcal{X}_n^\star) = V(\Delta(\tilde{\mathcal{X}}^{(n)})).$$

Moreover, by construction, we also have $V(\Delta(\tilde{\mathcal{X}}^{(n)})) \geqslant V(\Delta(\mathcal{X}^{(n)}))$ and hence by Lemma 6

$$V(\Delta(\mathcal{X})) \geqslant \lim_{n\to\infty} V(\Delta(\tilde{\mathcal{X}}^{(n)})) \geqslant \lim_{n\to\infty} V(\Delta(\mathcal{X}^{(n)})) = V(\Delta(\mathcal{X})),$$

showing that $V(\Delta(\tilde{\mathcal{X}}^{(n)}))$ converges to $V(\Delta(\mathcal{X}))$ as well. Therefore, we have

$$V(\Delta(\mathcal{X})) \geqslant V(\mathcal{X}^\star) \geqslant \lim_{n\to\infty} V(\mathcal{X}_n^\star) = \lim_{n\to\infty} V(\Delta(\tilde{\mathcal{X}}^{(n)})) = V(\Delta(\mathcal{X})),$$

proving Theorem 2.

Proof of Lemma 6. The high-level argument is similar to that of Lemma 4. We explicitly construct a sequence of finite allocation spaces $\mathcal{X}^{(n)}$ that satisfies the desired property. For any $n \in \mathbb{N}$, partition $\mathcal{T}$ into n equal-length subintervals as before. Let Δ_n denote the length of each subinterval. Let $\mathcal{T}^{(n)} := \{t_1, \ldots, t_n\}$ and let $\mu^{(n)} \in \Delta(\mathcal{T}^{(n)})$ be defined by: for all k

$$\mu_k^{(n)} := \mathbb{P}\big(t \in [t_{k-1}, t_k)\big) = F(t_k) - F(t_{k-1}).$$

Let $\mu \in \Delta(\mathcal{T})$ be the distribution given by F. Let $\lambda_k^{(n)} := \lambda(t_k)$ for all k.

For any n, consider the same screening problem with stochastic allocation space $\Delta(\mathcal{X})$, but with the type space and type distribution given by $(\mathcal{T}^{(n)}, \mu^{(n)})$. In particular, the objective is given by

$$\sum_k \mu_k^{(n)} \Big(\lambda_k^{(n)} \big(v(a^{(n)}(t_k), t_k) - p^{(n)}(t_k) \big) + p^{(n)}(t_k) \Big).$$

Note that since we are allowing for stochastic mechanisms, the above program is a linear program in $(a^{(n)}, p^{(n)}) \in \Delta(\mathcal{X})^n \times \mathbb{R}^n$. As argued before, without loss of optimality we can restrict attention to $p^{(n)} \in [-K, K]^n$ for a large enough universal constant K. We claim that there exists an optimal solution $(a^{(n)}, p^{(n)})$ to the above program where $\{a^{(n)}(t)\}_{t \in \mathcal{T}^{(n)}} \subseteq \Delta(\mathcal{X}^{(n)})$ where $\mathcal{X}^{(n)}$ is a finite set.

To see this, we enrich the space of optimizing variables: Consider choosing a joint lottery across allocations and payments and across types:

$$\beta \in \Delta((\mathcal{X} \times [-K, K])^n),$$

where the i-th dimension marginal β_i prescribes the stochastic allocation and stochastic payment for a reported t_i type. Evidently, the above objective is a linear functional of β with constraints given by $n \times n$ many IC constraints on β, and n many IR constraints on β. Each IC constraint is a moment constraint on β; each IR constraint is also a moment constraint on β. Since $(\mathcal{X} \times [-K, K])^n$ is compact, we have that $\Delta((\mathcal{X} \times [-K, K])^n)$ is compact in the weak-* topology. Then, by Bauer's maximum principle, there exists an optimal solution $\beta^{(n)}$ that is an extreme point of the set of measures constrained by $n^2 + n$ moment constraints. By Winkler (1988), it follows immediately that $\beta^{(n)}$ is finitely supported—in particular, it implies that $\beta_i^{(n)}$ is finitely supported for each type t_i. For each type t_i, take the allocation-marginal $a_i^{(n)}$ of $\beta_i^{(n)}$ and let $\mathcal{X}_i^{(n)} \subseteq \mathcal{X}$ be its finite support. Also take the expectation $p_i^{(n)}$ of the payment-marginal of $\beta_i^{(n)}$. It then follows that $(a^{(n)}, p^{(n)})$ is an optimal solution to our finite-type program. Moreover, let

$$\mathcal{X}^{(n)} = \bigcup_{i=1}^{n} \mathcal{X}_i^{(n)}$$

which is finite. It then follows immediately that $\{a^{(n)}(t)\}_{t \in \mathcal{T}^{(n)}} \subseteq \Delta(\mathcal{X}^{(n)})$, proving the claim.

Now, we claim that

$$\lim_{n \to \infty} V(\Delta(\mathcal{X}^{(n)})) = V(\Delta(\mathcal{X})).$$

This step is very similar to that given in Lemma 4 and in particular relies on Lemma 5. In particular, let $W_\Delta(\mathcal{T}^{(n)}, \mu^{(n)})$ be the optimal value for the stochastic screening problem under the type space $\mathcal{T}^{(n)} \subseteq \mathcal{T}$ and the type distribution $\mu^{(n)} \in \Delta(\mathcal{T}^{(n)})$. Since any stochastic mechanism (a, p) defined on $\mathcal{T}$ also induces a uniformly continuous indirect utility function u, replacing $\mathcal{X}$ with $\Delta(\mathcal{X})$ in the argument given in Lemma 4, we immediately obtain

$$\liminf_{n \to \infty} W_\Delta(\mathcal{T}^{(n)}, \mu^{(n)}) \geq V(\Delta(\mathcal{X})).$$

Moreover, since $\sup_{x,t} |v_t(x,t)| \leqslant M$ for some constant M, we have

$$\sup_{a \in \Delta(\mathcal{X}),\, t \in \mathcal{T}} \left| \mathbb{E}_{x \sim a}[v_t(x,t)] \right| \leqslant M$$

as well. Then, fix any optimal mechanism $(a^{(n)}, p^{(n)})$ defined on $\mathcal{T}$ with type distribution given by $\mu^{(n)} \in \Delta(\mathcal{T})$, such that it also satisfies $\mathrm{Ran}(a^{(n)}) \subseteq \Delta(\mathcal{X}^{(n)})$. Note that the induced indirect utility function $u^{(n)}$ is uniformly continuous with a modulus of continuity $\omega(\cdot)$ that holds uniformly across n. Replacing $\mathcal{X}$ with $\Delta(\mathcal{X})$ in the argument given in Lemma 4, we also immediately obtain

$$\liminf_{n \to \infty} V(\Delta(\mathcal{X}^{(n)})) \geqslant \liminf_{n \to \infty} W_\Delta(\mathcal{T}^{(n)}, \mu^{(n)}).$$

Therefore, it follows immediately that

$$V(\Delta(\mathcal{X})) \geqslant \limsup_{n \to \infty} V(\Delta(\mathcal{X}^{(n)})) \geqslant \liminf_{n \to \infty} V(\Delta(\mathcal{X}^{(n)})) \geqslant \liminf_{n \to \infty} W_\Delta(\mathcal{T}^{(n)}, \mu^{(n)}) \geqslant V(\Delta(\mathcal{X})),$$

showing that they are all equal and hence proving Lemma 6.

A.3 Proof of Corollary 2

The demand profile method is valid by the argument given in the main text. We show that $\mathcal{X}^\star$ is minimal optimal if $\operatorname{argmax}\{\Delta P(s,q) \cdot q\} \subset (0,1)$. Suppose for contradiction that there exists a closed subset $S \subset [0,1]$ such that $[0,1] \backslash S$ has a positive Lebesgue measure and S is an optimal menu. Let (x,p) be the optimal mechanism supporting this optimal menu S. Equivalently, we can represent (x,p) via a non-increasing function $s \mapsto q(s)$ by

$$q(s) := \mathbb{P}\big(x(t) \geqslant s\big).$$

and write the revenue obtained by (x,p) by

$$\int \big(\Delta P(s, q(s)) \cdot q(s)\big) \, \mathrm{d}s.$$

Since S has a measure strictly less than 1, there exists an open interval $(s_1, s_2) \subseteq [0,1] \backslash S$. Then, by definition of q, there exists a constant $\bar{q}$ such that $q(s) = \bar{q}$ for all $s \in (s_1, s_2)$. Since we have argued that the upper bound via pointwise upgrade pricing can be obtained:

$$\int \max_{q \in [0,1]} (\Delta P(s,q) \cdot q) \, \mathrm{d}s,$$

it follows that for almost all $s \in [0,1]$, we have

$$q(s) \in \operatorname*{argmax}_{q \in [0,1]} \Big(\Delta P(s,q) \cdot q \Big) ,$$

because otherwise (x,p) cannot be optimal. Since the frontier is interior, this implies that for almost all $s \in (s_1, s_2)$, we have

$$\frac{\mathrm{d}}{\mathrm{d}\log q} \log \Delta P(s,q) \mid_{q=\bar{q}} = -1 ,$$

which is an immediate contradiction since the ordered incremental elasticity property implies that $\frac{\mathrm{d}}{\mathrm{d}\log q} \log \Delta P(s,q)$ is strictly decreasing in s for all $q \in (0,1)$.

A.4 Proof of Theorem 3

Optimality. We first show that a generalized frontier $\mathcal{X}^\star$ is an optimal menu. Since $\mathcal{X}^\star$ is a generalized frontier, there exists an order $\preceq$ on $\mathcal{X}^\star$ and a set of lotteries $\mathcal{A} \subseteq \Delta(\mathcal{X}^\star)$ that are totally ordered by $\preceq_{\mathrm{st}}$ such that the agent's preferences have strict increasing differences on $\mathcal{X}^\star$ according to the order $\preceq$ and every element x can be covered by $\mathcal{A}$. Now, fix any mechanism (x,p). Fix any type t. By the definition of a covering, we know that there exists some $a \in \mathcal{A}$ such that

$$v(a,t) = v(x(t),t) ,$$

since the point $\big(1 - F(t), v(x(t),t)\big)$ is in the graph of the demand curve $P(x(t),\cdot)$ and hence must be in the graph of the demand curve $P(a,\cdot)$ for some $a \in \mathcal{A}$. Moreover, by the definition of a covering, we can take such an a with $P(a,\cdot)$ weakly single-crossing $P(x(t),\cdot)$ from below, and hence for all $t' > t$, we have

$$v(a,t') \leqslant v(x(t),t') .$$

We will use $a(t)$ as the reconstructed allocation rule following this procedure for all t, but before that we first ensure that there exists a measurable selection $t \mapsto a(t)$ following this procedure. Formally, let

$$\Gamma(t) := \Big\{ a \in \mathcal{A} : v(a,t) = v(x(t),t) \text{ and } v(a,t') \leqslant v(x(t),t') \text{ for all } t' \geqslant t \Big\} .$$

By the earlier argument, we observe that for any t, $\Gamma(t) \subseteq \Delta(\mathcal{X}^\star)$ is a non-empty compact set. Moreover the graph of Γ,

$$\text{Graph}(\Gamma) = \left\{(t,a) \in \mathcal{T} \times \mathcal{A} : v(a,t) = v(x(t),t) \text{ and } \max_{t' \in [t,\bar{t}]} \{v(a,t') - v(x(t),t')\} \leqslant 0\right\},$$

is Borel measurable given that $x(\cdot)$ is Borel measurable. Then by the Kuratowski–Ryll-Nardzewski measurable selection theorem, there exists a measurable selection $t \mapsto a(t)$ such that $a(t) \in \Gamma(t)$ for all t. Fix this measurable selection and let the reconstruction be defined by (a,p). Then, (a,p) satisfies all IR and all downward IC constraints, and yields the same objective value to the principal, assuming truthful reporting.

Since $\{a(t)\}_{t\in\mathcal{T}} \subseteq \mathcal{A}$ which is totally ordered by $\preceq_{\text{st}}$, by the same argument as in the proof of Theorem 1, we immediately have that for any $a \preceq_{\text{st}} a'$ and $a \neq a'$, we have

$$\mathbb{E}_{x\sim a}[v(x,t_2) - v(x,t_1)] < \mathbb{E}_{x\sim a'}[v(x,t_2) - v(x,t_1)],$$

given that $v(x,t)$ has strict increasing differences on $\mathcal{X}^\star \times \mathcal{T}$. Indeed, to see the above must hold with strict inequality, note that by Strassen's theorem a' can be obtained from a via a coupling (Y,Y') where $Y \preceq Y'$ and $\mathbb{P}(Y \prec Y') > 0$.

Therefore, as in the proof of Theorem 1, we can apply Theorem 4 to (a,p) and obtain a fully IC and IR mechanism $(\tilde{a},\tilde{p})$ where $\text{Ran}(\tilde{a}) \subseteq \mathcal{A}$ and $(\tilde{a},\tilde{p})$ yields a weakly higher objective than (a,p). Then, as in the proof of Theorem 1, we can apply Lemma 2 to $(\tilde{a},\tilde{p})$ and obtain a fully IC and IR deterministic mechanism $(x^\dagger,p^\dagger)$, where $\text{Ran}(x^\dagger) \subseteq \mathcal{X}^\star$, that further improves the objective. This completes the proof.

Stochastic Optimality. Now, suppose that the generalized frontier $\mathcal{X}^\star$ has pointwise ordered demand curves that satisfy the ordered incremental elasticity property. We show that the frontier is optimal among all stochastic mechanisms. Fix any stochastic mechanism. As in the proof of Theorem 2, we can write it equivalently as $t \mapsto (x(t,\varepsilon),p(t))$ where ε is a randomization device drawn independently from t. For each realization of ε, we can apply the same reconstruction argument just like before. Then, as in the proof of Theorem 2, we can obtain a modified stochastic mechanism $t \mapsto (a(t,\varepsilon),p(t))$—in particular, by the same measurable selection argument as above, $a(t,\varepsilon)$ can be made jointly measurable in (t,ε)—that satisfies all IR constraints and all downward IC constraints and weakly improves the principal's objective, under truthful reporting. Writing out the compound lotteries as simple lotteries over $\mathcal{X}^\star$, we can equivalently write the downward-IC stochastic mechanism as $t \mapsto (\tilde{a}(t),p(t))$ where $\tilde{a}(t) \in \Delta(\mathcal{X}^\star)$ for all t.

Now, consider $\mathcal{X}^\star$ itself as an allocation space. Clearly, given the incremental elasticity property, $\mathcal{X}^\star$ is a strong frontier for $\mathcal{X}^\star$ itself. By inspection, the proof of Theorem 2 not only establishes that any strong frontier is optimal among all stochastic mechanisms, it actually proves a stronger claim: This holds among all downward-IC stochastic mechanisms—formally, for any strong frontier $\mathcal{Y}^\star$ of $\mathcal{Y}$, there exists a full-IC, deterministic mechanism (x,p) such that $\mathrm{Ran}(x) \subseteq \mathcal{Y}^\star$ that is optimal among all downward IC and stochastic mechanisms (i.e., that use elements in $\Delta(\mathcal{Y})$). Since $\mathcal{X}^\star$ is a strong frontier of $\mathcal{X}^\star$, then we immediately obtain that $\mathcal{X}^\star$ is optimal among all downward-IC stochastic mechanisms—in particular, there exists a full-IC, deterministic mechanism $(x^\dagger, p^\dagger)$ where $\mathrm{Ran}(x^\dagger) \subseteq \mathcal{X}^\star$ such that the objective value under $(x^\dagger, p^\dagger)$ is weakly higher than that under $(\tilde{a}, p)$ and hence weakly higher than the original stochastic mechanism $t \mapsto (x(t,\varepsilon), p(t))$. Since the original stochastic mechanism is chosen arbitrarily, this then completes the proof of Theorem 3.

A.5 Proof of Corollary 3

Note that for any mechanism $(a,p) : \mathbf{v} \mapsto (a(\mathbf{v}), p(\mathbf{v}))$, the objective is given by

$$\begin{aligned}\mathbb{E}\Big[\lambda\big(a(\mathbf{v})\cdot\mathbf{v} - p(\mathbf{v})\big) + p(\mathbf{v})\Big] &= \mathbb{E}\Big[\mathbb{E}[\lambda \mid \mathbf{v}]\big(a(\mathbf{v})\cdot\mathbf{v} - p(\mathbf{v})\big) + p(\mathbf{v})\Big] \\ &= \mathbb{E}\Big[\mathbb{E}[\lambda \mid v^{\overline{b}}]\big(a(\mathbf{v})\cdot\mathbf{v} - p(\mathbf{v})\big) + p(\mathbf{v})\Big].\end{aligned}$$

Let t be a random variable following the same distribution as $v^{\overline{b}}$. Also write $\lambda(t) := \mathbb{E}[\lambda \mid v^{\overline{b}} = t]$ which is a continuous non-increasing function by assumption. By Strassen's theorem (see Lemma 1 in Yang forthcoming), under the stochastic ratio-monotonicity, we can write

$$\mathbf{v} \overset{d}{=} \Big(t, \big(v(b,t;\varepsilon)\big)_{b\neq\overline{b}}\Big),$$

for some independent random variable ε and jointly measurable functions $v(b,t;\varepsilon)$ such that $v(b,t;\varepsilon)/v(\overline{b},t;\varepsilon)$ is non-decreasing in t. Now, for any ε, in the relaxed problem where the principal observes the realized ε, by Proposition 5, pure bundling is optimal among all stochastic mechanisms (formally, Proposition 5 assumes differentiability but ratio-monotonicity suffices by the proof of Theorem 2). Moreover, for any ε, the optimal price is determined by maximizing the objective

$$\mathbb{E}[\lambda(t)\big(x(t)t - p(t)\big) + p(t)],$$

which does not depend on ε (since ε is independent of t). Thus, there exists a single pure-bundling mechanism that does not depend on ε and solves the principal's relaxed problem when observing ε. This mechanism is implementable even without observing ε, and hence it must be optimal in the original problem, completing the proof.