Costly Multidimensional Screening*

Frank Yang†

November 1, 2021

Abstract

A screening instrument is *costly* if it is socially wasteful and *productive* otherwise. A principal screens an agent with multidimensional private information and quasilinear preferences that are additively separable across two components: a one-dimensional productive component and a multidimensional costly component. Can the principal improve upon simple one-dimensional mechanisms by also using the costly instruments? We show that if the agent has preferences between the two components that are positively correlated in a suitably defined sense, then simply screening the productive component is optimal. The result holds for general type and allocation spaces, and allows for nonlinear and interdependent valuations. We discuss applications to optimal regulation, labor market screening, and monopoly pricing.

Keywords: Multidimensional screening, costly instruments, mechanism design, selection markets, price discrimination, surplus destruction

*I thank Mohammad Akbarpour, Daniel Chen, Mira Frick, Nima Haghpanah, Andreas Haupt, Ravi Jagadeesan, Zi Yang Kang, Paul Milgrom, Mike Ostrovsky, Anne-Katrin Roesler, Ilya Segal, Andy Skrzypacz, Takuo Sugaya, Bob Wilson, Weijie Zhong, and seminar participants at Stanford University for helpful comments and suggestions.

†Graduate School of Business, Stanford University. Email: shuny@stanford.edu.

1 Introduction

Actions convey information. The effort to obtain credentials conveys information about the ability of a job applicant. The time spent waiting in line conveys information about the willingness to pay of a consumer. The undertaking of a lengthy patent application process conveys information about the confidence of an inventor. The endurance of physical activity conveys information about the health status of an individual.[1]

At the same time, as Stiglitz (2002) emphasizes, "There is a much richer set of actions which convey information beyond those on which traditional adverse selection models have focused."

These actions are often *costly* in that they are socially wasteful. However, because the preferences over these actions are correlated in some way with the private information that affects the allocation of productive assets, the informational content from these costly actions can be potentially useful for screening.[2] Under what conditions should we expect the costly instruments to be used in the design of optimal contracts?

The inherent difficulty of this problem is that it is multidimensional by nature. Not only are the costly instruments often complex and difficult to summarize in a single dimension but, more fundamentally, it is crucial to understand the interplay between the use of the costly instruments and other aspects of the screening contract. Stiglitz (2002) gives the following example: An insurance company "might realize that by locating itself on the fifth floor of a walk-up building, only those with a strong heart would apply. [...] More subtly, it might recognize that how far up it needs to locate itself depends on other elements of the strategy such as premium charged."

In this paper, we study the effectiveness of costly instruments in a general multidimensional screening model. The model consists of two components: (i) a standard one-dimensional productive component which the principal intrinsically cares about (such as insurance coverage), and (ii) a multidimensional costly component which the principal may utilize to help screening but destroys social surplus (such as walking up stairs). Our main result states that if the agent has preferences between the two components that are positively correlated in a suitably defined sense, then simply screening the one-dimensional productive component is optimal (and essentially uniquely optimal).

In the model, the principal designs a mechanism to assign the productive allocations in a one-dimensional space $\mathcal{X}$ and the costly actions in an *arbitrary* space $\mathcal{Y}$. Monetary

[1] The New York Times reports, "The [Wal-Mart's] memo suggests that the company could require all jobs to include some component of physical activity, like making cashiers gather shopping carts." *Wal-Mart's health care struggle is corporate America's, too*, The New York Times, October 29, 2005.

[2] Zeckhauser (2021) argues that socially wasteful ordeals play a prominent role in health care.

transfers are allowed. Both the principal and the agent have quasilinear preferences that are additively separable across the two components $\mathcal{X}$ and $\mathcal{Y}$. We say that the agent has preferences that are *positively correlated* between the two components if the type who has higher willingness to pay for the productive allocations tends to have higher willingness to pay for the costly actions in the stochastic dominance sense. We also provide a partial converse showing that for a given negative correlation structure, there exist utility functions such that the optimal contract must involve costly screening. More fundamentally, we also show that the main result holds even for environments in which monetary transfers can be imperfect.

For example, in labor market screening, a higher ability applicant tends to find both accomplishing the job easier (productive component) and education less costly (costly component); in monopoly regulation, a more efficient firm typically has a lower cost of operation (productive component) and finds it easier to pass performance inspections (costly component). Our result then implies that a monopsonistic firm should *not* make its offers contingent on the costly signals from an applicant despite the fact that the firm prefers a higher ability applicant; a regulator's tax and subsidy schedule should *not* depend on the performance of a regulated firm on the inspections despite the fact that the regulator prefers a more efficient firm. In the case of selling insurance, however, an individual with a higher risk tends to have a higher willingness to pay for the insurance (productive component) and find it *harder* to pass various health tests (costly component). Only in cases like this can the principal potentially benefit from the costly instruments.

To understand how these seemingly contradictory implications follow from the same underlying principle, we emphasize a key assertion of our result: The effectiveness of costly instruments is driven by the *agent's* instead of the principal's preferences. A naïve intuition for the result is that since monetary transfers are allowed, instead of using the costly instruments the principal can simply use money, which does not destroy social surplus. This reasoning, albeit intuitive, is not what drives the result. In particular, how the agent's preferences are correlated between the two components does not appear in this reasoning.

The correct intuition comes from a deeper understanding of both the functionality of costly instruments and the structure of one-dimensional screening problems. A costly instrument can be thought of as a special kind of currency whose value depends on the private information of the agent. Using these special kinds of currency, the principal can loosen some incentive constraints but is forced to tighten other incentive constraints. In particular, we show that if the agent's preferences are positively correlated between the two components, then the costly instruments loosen *upward* incentive constraints and tighten *downward* incentive constraints. Next, we show that the set of downward incentive constraints is *sufficient* for any standard one-dimensional screening problem satisfying a single-crossing condition. Therefore, loosening the upward incentive constraints with costly instruments cannot benefit the principal under positive correlation.

From the literature on monopoly regulation, it is well known that the information about a monopolist's production costs can be helpful for designing regulatory policies (Laffont and Tirole, 1986). Indeed, if a regulator observes non-manipulable data correlated (positively or negatively) with the monopolist's production costs, then the optimal mechanism is contingent on that information. In our application to monopoly regulation (see Section 6.1), by contrast, the data can be manipulated by the monopolist at a cost. Our result demonstrates that once the information has to be elicited from the agent, whether it is used in the optimal mechanism depends on *how* the agent's preferences are correlated between the productive allocations and the costly actions.

From the literature on selection markets, it is well known that costly signals are important for markets involving adverse selection (Spence 1973a, 1974). The major difference in our application to labor market screening is that we study monopsonistic firms and allow for contracts that screen with both work allocations and costly signals (see Section 6.2). To make this clear, in Appendix B.1, we study a competitive screening model analogous to the main model and show how costly screening can emerge in equilibrium. In light of this analysis, our result demonstrates that whether costly instruments appear in a market depends crucially on the distribution of market power.[3,4]

Besides the literal interpretation, our framework also delivers insights into settings that do not explicitly involve costly instruments. For example, consider a multiple-good monopolist selling different qualities of bundles. A common selling strategy in streaming services is to offer the bundle of all content at a range of monthly fees depending on the level of quality. When is such a strategy optimal? We argue that it is useful to view selling the bundle of all goods as the productive component, and selling smaller bundles *instead of* the grand bundle as the costly instruments for screening values of the grand bundle. Using this perspective, in Section 5.2, we generalize a recent finding of Haghpanah and Hartline (2021), who derive sufficient conditions for the optimality of pure bundling, to a multiple-good monopoly setting that allows for *both* probabilistic bundling and quality

[3]Monoposony and labor market power have recently received resurgent interest from empirical studies of labor markets. See e.g. Azar et al. (2020), Dube et al. (2020), and Prager and Schmitt (2021).

[4]In his pioneering work, Spence (1974) also compares monopolistic and competitive screening in the labor market context. In contrast to us, he studies a one-dimensional problem where the wage schedule can only depend on the signal level. He gives local optimality conditions and finds, in a similar spirit to us, that a monopolist induces people to underinvest in the signal. Subsequent to this finding, he writes, "One might have guessed the reverse. [...] I do not, however, find this an intuitively obvious result."

discrimination. The optimal mechanism in our setting generally involves price discrimination but does so only along the vertical (quality) dimension.

We are far from being the first to study screening problems involving costly instruments. This question goes back at least as far as Spence (1973b). In that paper, he proposes the idea that willingness to spend time can be and is used as a screening device in various contexts. Even though he does not give a formal treatment of this problem, he points out its importance: "Let me put the matter another way. The argument is that non-price signaling and screening in economic and social contexts deserve more attention, in spite of the fact that they are frequently inefficient." Since then, however, there is still no general framework for understanding when costly instruments are effective or ineffective in the design of optimal contracts.

Several previous papers have analyzed mechanism design with a costly instrument when monetary transfers are *not* feasible. That line of work studies the design of mechanisms for surplus maximization with multiple agents engaging in a one-dimensional costly activity (Hartline and Roughgarden 2008, Condorelli 2012, Chakravarty and Kaplan 2013).[5] We study a single-agent contract design problem allowing for both monetary transfers and multidimensional costly instruments. The principal in our model does not maximize social surplus and may have arbitrary interdependent valuations.

Conceptually, Acemoglu and Wolitzky (2011) ask a question similar to ours. They study how labor coercion, which is socially wasteful, can benefit the principal in the design of employment contracts. The agent in their setting, however, has no private information but takes unobservable actions. So their focus is on the moral hazard instead of the adverse selection aspect of contract design. Labor coercion in their model is a punishment on the agent for rejecting a contract, and directly influences the agent's reservation utility. We study a screening problem. The costly instruments in our model do not affect the agent's reservation utility but are used in combination with other elements of the contract through the agent's self-selection.

There is a substantial literature on multidimensional screening. The structure of multidimensional screening differs significantly from its single-dimensional counterpart and remains elusive to characterize despite much research over the past decades (Rochet and Stole, 2003). Much of the work focuses on the linear multiple-good monopoly problem.[6] In that problem, the principal cares only about transfers and the agent has linear utility functions. When there is a single good, the optimal mechanism is simply a posted price

[5] See also Akbarpour et al. (2020) and Malladi (2020) for related problems.

[6] A notable exception is Carroll (2017) who allows the agent to have arbitrary preferences within each additive component; however, he focuses on characterizing the worst-case optimum with known marginals.

(Myerson 1981, Riley and Zeckhauser 1983). However, as soon as there is more than one good, seemingly simple special cases remain poorly understood. Significant progress has been made in developing duality approaches to certify optimality of candidate mechanisms (Rochet and Chone 1998, Daskalakis et al. 2017; Cai et al. 2016, Carroll 2017). In response to the analytical difficulty, several recent papers study either approximately optimal mechanisms (Li and Yao 2013, Babaioff et al. 2014, Cai et al. 2016, Hart and Nisan 2017), or worst-case optimal mechanisms (Carroll 2017, Koçyiğit et al. 2021, Che and Zhong 2021, Deb and Roesler 2021).

In contrast to past work, we consider a multidimensional screening model in which all dimensions except one are surplus destructive. The multiple-good monopoly problem can be viewed as a special case of our framework by redefining the allocation space. In this sense, our model offers an alternative perspective — it separates the screening effect from the surplus-generating effect of the additional dimensions. After shutting down the surplus-generating effect, we show that simple conditions, like positive correlation of preferences, are sufficient to eliminate the screening effect of the additional dimensions, and thereby characterize the exact Bayesian optimal mechanism in our setting. Our result holds for general type spaces, general allocation spaces, general utility functions. Importantly, motivated by applications — including regulation, labor markets, and insurance markets — our model also allows for interdependent preferences.

Part of our proof technique builds on the approach of path decomposition that partitions the multidimensional type space into one-dimensional paths (Wilson 1993, Armstrong 1996, Eso and Szentes 2007, Haghpanah and Hartline 2021). The method in Haghpanah and Hartline (2021) is most closely related to ours. Our method shares with theirs the focus on Strassen-type decompositions and downward incentive constraints. Our method differs in that we take a primal instead of dual approach. They use a delicate construction of the dual variables to certify the optimality of a particular mechanism (pure bundling) with the linear programming duality method in Cai et al. (2016) and Carroll (2017).[7] Our problem in general is not linear or even convex. The optimum generally requires ironing on an infinite menu of options. Our proof makes no appeal to any duality approach. Instead, we use a "shift" argument and tackle the primal problem directly. This approach, when applicable, bypasses the difficulty in constructing dual variables, especially when one has no particular candidate mechanism in mind.

The remainder of the paper proceeds as follows. Section 2 presents our model. Section 3 presents the main result and a partial converse. Section 4 presents the proof of the

[7]Constructing appropriate dual variables is a major difficulty in duality-based approaches. See Bergemann et al. (2021) for a recent development following Haghpanah and Hartline (2021)'s duality approach.

main result. Section 5 presents a stronger result in a special case of the model and its application to multiple-good monopoly pricing. Section 6 discusses additional applications. Section 7 concludes.

2 Model

A principal wants to screen an agent. The agent has private information summarized by a multidimensional type $\theta = (\theta^A, \theta^B)$, where $\theta^A \in \Theta^A \subseteq \mathbb{R}$ and $\theta^B \in \Theta^B \subseteq \mathbb{R}^N$ for a finite N; for convenience, sometimes we also refer to θ^A as θ^0 and θ^B as $(\theta^1, \ldots, \theta^N)$. We use the superscripts A, B to indicate the productive and costly components, respectively.

Both Θ^A and Θ^B are assumed to be compact. Let $\Theta := \Theta^A \times \Theta^B$ denote the type space; let $\Delta(\Theta)$ denote the space of Borel probability measures on Θ, equipped with the weak-* topology. The agent's type is drawn from a commonly known distribution $\gamma \in \Delta(\Theta)$.

The space of productive allocations $\mathcal{X} \ni x$ is a compact subset of $\mathbb{R}$; the space of costly instruments $\mathcal{Y} \ni y$ is an arbitrary measurable space.

Both the principal and the agent have quasilinear preferences that are additively separable across the two components: The principal's (ex post) payoff is given by

$$v^A(x, \theta^A) + v^B(y, \theta^B) + t,$$

and the agent's payoff is given by

$$u^A(x, \theta^A) + u^B(y, \theta^B) - t,$$

where t stands for transfers. The utility functions for the productive component u^A, v^A are assumed to be continuous on $\mathcal{X} \times \Theta^A$; those for the costly component u^B, v^B are allowed to be any bounded measurable functions on $\mathcal{Y} \times \Theta^B$. The principal has *interdependent preferences* if v^A or v^B depends on the agent's type.

The (ex post) surplus functions for the two components are denoted by

$$s^A(x, \theta^A) := u^A(x, \theta^A) + v^A(x, \theta^A), \quad s^B(y, \theta^B) := u^B(y, \theta^B) + v^B(y, \theta^B).$$

The defining feature of the costly component is that any allocation is socially wasteful under complete information: for all $y \in \mathcal{Y}$ and all $\theta^B \in \Theta^B$,

$$s^B(y, \theta^B) \leq 0. \tag{1}$$

There is an element $y_0 \in \mathcal{Y}$ (with $\{y_0\}$ measurable) representing *no costly screening*:

$$v^B(y_0, \theta^B) = u^B(y_0, \theta^B) = 0\,.$$

We say the instruments are *strictly costly* if (1) holds strictly for all $y \neq y_0$ and all θ^B.

A *mechanism* is a measurable map

$$(x, y, t) : \Theta \to \mathcal{X} \times \mathcal{Y} \times \mathbb{R}$$

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

$$\begin{aligned} &u^A(x(\theta), \theta^A) + u^B(y(\theta), \theta^B) - t(\theta) \geq u^A(x(\hat{\theta}), \theta^A) + u^B(y(\hat{\theta}), \theta^B) - t(\hat{\theta}) && \text{for all } \theta, \hat{\theta} \in \Theta\,; \\ &u^A(x(\theta), \theta^A) + u^B(y(\theta), \theta^B) - t(\theta) \geq 0 && \text{for all } \theta \in \Theta\,. \end{aligned}$$

Let $\mathcal{M}(\Theta)$ denote the space of mechanisms. The principal wants to solve

$$\sup_{(x,y,t) \in \mathcal{M}(\Theta)} \mathbb{E}[v^A(x(\theta), \theta^A) + v^B(y(\theta), \theta^B) + t(\theta)].$$

A mechanism (x, y, t) *involves no costly screening* if $y(\theta) = y_0$ for all θ and (x, t) does not depend on θ^B, in which case the mechanism screens only the productive component. A mechanism (x, y, t) *almost surely involves no costly screening* if $y(\theta) \overset{a.s.}{=} y_0$.[8]

We make the following assumptions on the productive component.

Assumption A1 (Productive Component)**.**

(1.1) $u^A(x, \theta^A)$ is nondecreasing in θ^A.

(1.2) $u^A(x, \theta^A)$ has strict increasing differences: for any $x < \hat{x}$, $\theta^A < \hat{\theta}^A$,

$$u^A(\hat{x}, \theta^A) - u^A(x, \theta^A) < u^A(\hat{x}, \hat{\theta}^A) - u^A(x, \hat{\theta}^A).$$

(1.3) $s^A(x, \theta^A)$ has weak single-crossing differences: for any $x < \hat{x}$, $\theta^A < \hat{\theta}^A$,

$$s^A(\hat{x}, \theta^A) - s^A(x, \theta^A) > 0 \implies s^A(\hat{x}, \hat{\theta}^A) - s^A(x, \hat{\theta}^A) > 0\,.$$

To state our notion of positive correlation of preferences between the two components, we introduce some notation. Let $\preceq_{st}$ denote the usual stochastic order for $\mathbb{R}^N$-valued ran-

[8] As usual we use notation $\overset{a.s.}{=}$ for almost sure equality and $\overset{d}{=}$ for equality in distribution.

dom variables, i.e. $X \preceq_{st} Y$ if $\mathbb{E}[f(X)] \leq \mathbb{E}[f(Y)]$ for all bounded nondecreasing (measurable) functions $f : \mathbb{R}^N \to \mathbb{R}$. Let $\theta^B|\theta^A$ denote the regular conditional distribution of θ^B given θ^A.[9]

Assumption A2 (Positive Correlation of Preferences)**.**

(2.1) $u^B(y, \theta^B)$ is nondecreasing in θ^B.

(2.2) $\theta^B|\theta^A \preceq_{st} \theta^B|\hat{\theta}^A$ for all $\theta^A < \hat{\theta}^A$ in the support.

2.1 Discussion of Assumptions

Mechanism Space. As formally defined, our model restricts attention to deterministic mechanisms. However, when there are finitely many pure allocations, we may be able to redefine the allocation spaces to be the probabilities. We take this approach in Section 5.2 where we show how probabilistic bundling by a multiple-good monopolist is nested in this framework.

In the model, there is no feasibility constraint across the allocation spaces $\mathcal{X}$ and $\mathcal{Y}$. However, for any subset $\mathcal{S} \subseteq \mathcal{X} \times \mathcal{Y}$, we may constrain the feasible allocations by requiring $(x, y) : \Theta \to \mathcal{S}$. Provided that $\mathcal{X} \times \{y_0\} \subseteq \mathcal{S}$, our main result is unaffected by such constraints (since the optimum in the original problem would still be feasible).

Monetary Transfers. In the model, money is perfectly transferable. However, by rescaling, our model accommodates some environments in which monetary transfers can be imperfect. Suppose for any mechanism (x, y, t) the principal's payoff is given by

$$\mathbb{E}[v^A(x(\theta), \theta^A) + v^B(y(\theta), \theta^B) + \alpha t(\theta)],$$

where α is any positive constant (that may represent, for example, adjustment for tax). We can factor out α and see that the principal's problem is equivalent to that with the scaled objective

$$\mathbb{E}\left[\frac{1}{\alpha} v^A(x(\theta), \theta^A) + \frac{1}{\alpha} v^B(y(\theta), \theta^B) + t(\theta)\right].$$

If v^A satisfies the weak increasing differences condition[10] and $v^B = 0$, then this problem automatically fits our model. These assumptions may hold in, for example, monopoly pricing with costly signals (see Section 5.1) and labor market screening (see Section 6.2).

[9] See e.g. Klenke (2013, pp. 180-185).

[10] That is, $v^A(\hat{x}, \theta^A) - v^A(x, \theta^A) \leq v^A(\hat{x}, \hat{\theta}^A) - v^A(x, \hat{\theta}^A)$ for all $x < \hat{x}$, $\theta^A < \hat{\theta}^A$.

As in most screening models, we assume that the range of transfers is not restricted (i.e. no liquidity constraints). This can be modified to only allowing for bounded transfers $t : \Theta \to [-K, K]$ provided that $K \in \mathbb{R}_+$ is sufficiently large. With a substantial constraint on the transfers (K small enough), our result in general will not hold, since in the extreme case of $K = 0$ the principal can only screen with the costly instruments.

Productive Component. Assumptions (1.1) and (1.2) are the classic assumptions of one-dimensional screening problems. For future reference, we say a one-dimensional screening problem is *standard* if it satisfies Assumptions (1.1) and (1.2). Assumption (1.3) is a sorting condition on the surplus function. It is weaker than the usual single-crossing differences condition for monotone comparative statics as in Milgrom and Shannon (1994).[11] It is satisfied in common one-dimensional screening problems. For example, any of the following conditions is sufficient: (i) s^A is strictly increasing in x; (ii) $s^A_{12} \geq 0$; or (iii) the principal's preferences are not interdependent (provided that Assumption (1.2) holds). This assumption ensures that there is a monotone efficient allocation rule. It is not satisfied when the principal's preference to trade with low types is so strong that any socially efficient allocation rule is not monotone. In Section 4, we show that some sorting condition on the surplus function is necessary for the result (see Remark 2 and Example 2).

Positive Correlation. Assumption (2.1) says that θ^B encodes the strength of the agent's preferences on the costly component such that higher θ^B represents higher willingness to pay for any y. Assumption (2.2) then defines the positive correlation structure between the agent's preferences for the two components. The condition is known as stochastic monotonicity (Müller and Stoyan, 2002). We say θ^B is *stochastically nondecreasing* in θ^A whenever Assumption (2.2) holds. This is an asymmetric condition. It says that observing a high θ^A conveys good news about θ^B in the sense of stochastic dominance. A sufficient condition for Assumption (2.2) is that $(\theta^0, \theta^1, \ldots, \theta^N)$ are affiliated in the sense of Milgrom and Weber (1982). Assumption (2.2) is weaker than affiliation. For example, when $N = 2$, (θ^1, θ^2) may be negatively correlated with each other, while $\theta^B = (\theta^1, \theta^2)$ is positively correlated with θ^A according to this notion.

[11] The usual single-crossing differences condition requires, in addition, that $s^A(\hat{x}, \theta^A) - s^A(x, \theta^A) \geq 0 \implies s^A(\hat{x}, \hat{\theta}^A) - s^A(x, \hat{\theta}^A) \geq 0$ for all $x < \hat{x}$, $\theta^A < \hat{\theta}^A$.

3 Main Result

Our main result says that if the agent has positively correlated preferences between the productive and costly components, then simply screening the one-dimensional productive component is optimal and essentially uniquely optimal:

Theorem 1. *Suppose Assumptions A1 and A2 hold. Then:*

(i) There exists an optimal mechanism that involves no costly screening.

(ii) If the instruments are strictly costly, then every optimal mechanism almost surely involves no costly screening.

In the case of negatively correlated preferences, we show a partial converse. We say the utility functions u^A, u^B, v^A, v^B are *admissible* if they satisfy all the assumptions in Section 2 including the strict version of (1). For a real-valued continuous random variable X, let $\beta(X) = \mathbb{1}_{X \geq \text{median}(X)}$ denote the "binarization" of X.

Proposition 1. *Suppose θ is absolutely continuous; $|\mathcal{X}| > 1$, $|\mathcal{Y}| > 1$; and there exists some $i \in \{1, \ldots, N\}$ such that θ^i is stochastically* nonincreasing *in θ^0 and $\beta(\theta^i)$, $\beta(\theta^0)$ are not independent. Then, there exist admissible utility functions such that any mechanism screening only the productive component is strictly dominated by a mechanism involving costly screening.*

We postpone the explanation and proof of Theorem 1 to Section 4. Proposition 1 can be shown by a simple construction that sets $v^A = v^B = 0$. The intuition is as follows. The principal can always create a menu of two nontrivial options for the agent: (i) getting the favorite allocation in $\mathcal{X}$ at a high price, and (ii) getting the same allocation at a low price but with some costly activity. The proof shows that if θ^i is negatively correlated with θ^0 as defined in the statement, then there exist some admissible utility functions for the agent such that this way of price discrimination is always more profitable for the principal than selling the elements in $\mathcal{X}$ alone. The appendix provides details.

4 Proof of the Main Result

The intuition behind the proof of Theorem 1 can be understood in two parts.

In the first part, we demonstrate that what costly instruments do is loosen one class of IC constraints while tightening another class of IC constraints. The argument runs as follows. Fix any mechanism that involves costly screening. We modify the mechanism by shifting the costly component payoffs to monetary transfers, *assuming* all types report truthfully. Of course, there is no reason for the modified mechanism to be incentive compatible; the modification can easily break a large set of IC constraints.[12]

A key observation is that it still maintains all the *downward* IC constraints on the costly component: no type (θ^A, θ^B) has an incentive to imitate $(\hat{\theta}^A, \hat{\theta}^B)$ if $\hat{\theta}^B < \theta^B$. The reason is simple. All types obtain the same payoffs as before if reporting truthfully. Any type deviating from truthful reporting will now receive transfers instead of costly allocations. Because a type with higher θ^B enjoys the costly instruments more, such change lowers the deviating payoff for any downward deviation.

So far, we have not used that the agent's preferences between the two components are positively correlated. The positive correlation of types converts the downward IC constraints on the costly component to the downward IC constraints on the productive component. Indeed, given the positive correlation, when receiving a report of low $\hat{\theta}^B$, the principal has statistical reasons to believe that $\hat{\theta}^A$ should also be low. This is precisely the case when the support of (θ^A, θ^B) falls on a monotonic path in $\mathbb{R}^{N+1}$. More generally, our notion of positive correlation allows us to represent the distribution of (θ^A, θ^B) as a mixture of such monotonic paths.

Therefore, for any IC mechanism involving costly screening, there exists a downward IC mechanism that involves no costly screening and provides each type the same payoff under truthful reporting. That is, using costly instruments instead of monetary transfers can only help in dealing with the upward not the downward IC constraints. Moreover, because costly instruments do not generate social surplus, the *only* purpose they serve is to loosen the upward IC constraints.

In the second part of the analysis, we show that in any standard one-dimensional screening problem, the set of downward IC constraints including the nonlocal ones is sufficient if the surplus function satisfies the weak single-crossing differences condition. To solve for the principal's problem, one may simply ignore all upward IC constraints. Because costly instruments only help loosen upward IC constraints when the agent's preferences are positively correlated, they are thus completely ineffective.

We prove the downward sufficiency result first for finite type spaces and then for general type spaces by approximation. We also show that the downward sufficiency result may not hold if no sorting condition is imposed on the surplus function, which can then lead to a reverse in the use of costly instruments (see Remark 1 and Remark 2). The difficulty of working only with the downward IC constraints arises from two factors: (i) there is no appropriate envelope condition because each type is not fully optimizing, and

[12] It is then imprecise to call it a mechanism; however, we will sometimes abuse the notion by referring to any measurable map as a mechanism and be clear about whether it satisfies the IC and IR constraints.

(ii) the set of (downward IC) implementable allocations is unknown (in the finite-type case we will show that *any* allocation is implementable by a downward IC mechanism).

With this outline in mind, we now dive into the details. For clarity, the proof proceeds in a slightly different order. We break it down into four steps: (i) decompose the type distribution into paths, (ii) "shift" the space of mechanisms, (iii) solve the downward IC screening problem for finite type spaces, and (iv) approximate.

Step 1: We start with path decomposition. The idea is that we reveal some information to the principal about θ^B that is orthogonal to θ^A. We let the principal design a mechanism conditional on the information. This provides an upper bound on what the principal can achieve. We then check that the resulting mechanism is actually implementable.

Suppose for illustration that $\theta = (\theta^0, \theta^1)$. Let ε be an independent uniform $[0,1]$ draw. Note that the random vector (θ^0, θ^1) can be simulated by sampling θ^0 and then setting

$$\theta^1 = F^{-1}(\varepsilon \mid \theta^0),$$

where $F^{-1}(\cdot|\theta^0)$ is the generalized inverse function of $F(\cdot|\theta^0)$. Our positive correlation condition states that $\theta^1|\theta^0$ shifts upward in the sense of stochastic dominance as θ^0 increases. This implies that $F^{-1}(\varepsilon|\cdot)$ is a nondecreasing function. Therefore, if we reveal the realization of ε to the principal, then the principal believes that (θ^0, θ^1) falls on a monotonic path. This then reduces stochastic monotonicity to deterministic monotonicity.

There is a canonical representation generalizing the above by inductively decomposing the joint distribution into the "initial" private information θ^0 and i.i.d. shocks ε^i.[13] If $(\theta^0, \theta^1, \ldots, \theta^N)$ are affiliated, then revealing the realization of the shocks ε would give a monotone decomposition similar to above. In fact, it is a classic result on stochastic dominance (dating back to Strassen 1965) that a monotone decomposition (coupling) always exists given stochastic monotonicity.[14] In particular, we use the following result:

Lemma 1 (Kamae and Krengel (1978)[15]). *If θ^B is stochastically nondecreasing in θ^A, then there exist a measurable space $\mathcal{E}$; an $\mathcal{E}$-valued random variable ε, independent of θ^A; and a measurable function $h : \Theta^A \times \mathcal{E} \to \Theta^B$ nondecreasing in the first argument such that*

$$\theta \overset{d}{=} (\theta^A, h(\theta^A; \varepsilon)).$$

[13] See Eso and Szentes (2007) and Pavan et al. (2014, Example 1).

[14] This type of decomposition was used earlier in Haghpanah and Hartline (2021).

[15] See Theorem 6 of Kamae and Krengel (1978). Joint measurability of h follows from that the sample path in their proof of Theorem 6 is left-continuous except possibly at a fixed countable set of points. Appendix B.2 provides details.

Let $\Theta_\varepsilon = \{(\theta^A, \theta^B) : \theta^B = h(\theta^A; \varepsilon), \theta^A \in \Theta^A\}$ be the decomposed monotonic path given a realization ε. For any type space Θ, recall $\mathcal{M}(\Theta)$ is the set of IC and IR mechanisms. Let $v(x, y, \theta) = v^A(x, \theta^A) + v^B(y, \theta^B)$. It then follows that

$$\begin{aligned}\sup_{(x,y,t)\in\mathcal{M}(\Theta)} \mathbb{E}[v(x(\theta), y(\theta), \theta) - t(\theta)] &\leq \mathbb{E}_\varepsilon\Big[\sup_{(x,y,t)\in\mathcal{M}(\Theta)} \mathbb{E}\Big[v(x(\theta), y(\theta), \theta) - t(\theta) \mid \varepsilon\Big]\Big] \\ &\leq \mathbb{E}_\varepsilon\Big[\sup_{(x,y,t)\in\mathcal{M}(\Theta_\varepsilon)} \mathbb{E}\Big[v(x(\theta), y(\theta), \theta) - t(\theta) \mid \varepsilon\Big]\Big]. \qquad (2)\end{aligned}$$

Because ε is independent of θ^A, the inner expectation integrates with respect to the same marginal distribution of θ^A regardless of the realization of ε.

We will establish that (i) for all realizations of ε,

$$\sup_{(x,y,t)\in\mathcal{M}(\Theta_\varepsilon)} \mathbb{E}\Big[v(x(\theta), y(\theta), \theta) - t(\theta) \mid \varepsilon\Big] \qquad (3)$$

can be attained by a single mechanism that involves no costly screening; and (ii) if the instruments are strictly costly, then all optimal solutions to (3) satisfy $\mathbb{P}(y(\theta) = y_0 \mid \varepsilon) = 1$. Any mechanism in $\mathcal{M}(\Theta_\varepsilon)$ that involves no costly screening is in $\mathcal{M}(\Theta)$. Thus, the first part of Theorem 1 follows. If a mechanism (x, y, t) in $\mathcal{M}(\Theta)$ has $y(\theta) \neq y_0$ for a positive measure of θ, then we have $\mathbb{P}(y(\theta) = y_0 \mid \varepsilon) < 1$ for a positive measure of ε. Hence, if the instruments are strictly costly, then (x, y, t) is strictly dominated by any optimal mechanism involving no costly screening, and thus the second part of Theorem 1 follows.

Step 2: We fix a realization of ε and suppress the dependency on ε whenever clear. Let us fix a mechanism $(x, y, t) \in \mathcal{M}(\Theta_\varepsilon)$. Because θ^B is now determined by θ^A, the relevant private information is summarized in θ^A; thus it is without loss to let (x, y, t) only depend on the report $\hat{\theta}^A$. We use a "shift" argument as follows. Consider the modification:

$$\tilde{x}(\theta^A) = x(\theta^A), \qquad \tilde{y}(\theta^A) = y_0, \qquad \tilde{t}(\theta^A) = t(\theta^A) - u^B(y(\theta^A), h(\theta^A)).$$

The modification $(\tilde{x}, \tilde{y}, \tilde{t})$ maintains the same allocations for the productive component, involves no costly screening, and uses transfers to keep all types at their previous utility levels, *assuming* they report truthfully.

Assuming truthful reporting, this increases the total surplus while giving the same surplus to the agent, and therefore increases the principal's payoff. Indeed, the change in principal's payoff is

$$\mathbb{E}\Big[v^B(y_0, h(\theta^A)) - v^B(y(\theta^A), h(\theta^A)) - u^B(y(\theta^A), h(\theta^A))\Big] = \mathbb{E}\Big[-s^B(y(\theta^A), h(\theta^A))\Big] \geq 0\,.$$

The last inequality is strict if $\mathbb{P}(y(\theta^A) \neq y_0) > 0$ and the instruments are strictly costly.

Because the modification maintains the utility for each type under truthful reporting, $(\tilde{x}, \tilde{y}, \tilde{t})$ satisfies all IR constraints. However, this mechanism is not necessarily IC. Indeed, suppose for illustration that $u^B(y, \cdot)$ and $h(\cdot)$ are strictly increasing, and for some $\hat{\theta}^A > \theta^A$, IC$[\theta^A \to \hat{\theta}^A]$ binds under (x, y, t). Consider the same deviation under $(\tilde{x}, \tilde{y}, \tilde{t})$:

$$\begin{aligned} u^A(\tilde{x}(\theta^A), \theta^A) - \tilde{t}(\theta^A) &= u^A(x(\theta^A), \theta^A) + u^B(y(\theta^A), h(\theta^A)) - t(\theta^A) \\ &= u^A(x(\hat{\theta}^A), \theta^A) + u^B(y(\hat{\theta}^A), h(\theta^A)) - t(\hat{\theta}^A) \\ &< u^A(x(\hat{\theta}^A), \theta^A) + u^B(y(\hat{\theta}^A), h(\hat{\theta}^A)) - t(\hat{\theta}^A) \\ &= u^A(\tilde{x}(\hat{\theta}^A), \theta^A) - \tilde{t}(\hat{\theta}^A)\,, \end{aligned} \tag{4}$$

where the first and the last line follow by construction, the second line uses the binding IC constraint, and the third line uses that $u^B(y, \cdot)$ and $h(\cdot)$ are strictly increasing. Therefore, IC$[\theta^A \to \hat{\theta}^A]$ is not satisfied under $(\tilde{x}, \tilde{y}, \tilde{t})$.

This demonstrates that the modification does not work directly. However, the same reasoning also shows that all downward IC constraints are still satisfied after this modification. Indeed, consider a downward deviation $[\theta^A \to \hat{\theta}^A]$ for any $\hat{\theta}^A < \theta^A$:

$$\begin{aligned} u^A(\tilde{x}(\theta^A), \theta^A) - \tilde{t}(\theta^A) &= u^A(x(\theta^A), \theta^A) + u^B(y(\theta^A), h(\theta^A)) - t(\theta^A) \\ &\geq u^A(x(\hat{\theta}^A), \theta^A) + u^B(y(\hat{\theta}^A), h(\theta^A)) - t(\hat{\theta}^A) \\ &\geq u^A(x(\hat{\theta}^A), \theta^A) + u^B(y(\hat{\theta}^A), h(\hat{\theta}^A)) - t(\hat{\theta}^A) \\ &= u^A(\tilde{x}(\hat{\theta}^A), \theta^A) - \tilde{t}(\hat{\theta}^A)\,, \end{aligned} \tag{5}$$

where the first and the last line follow by construction, the second line follows from (x, y, t) being IC, and the third line follows from that $u^B(y, \cdot)$ and $h(\cdot)$ are nondecreasing. Therefore, $(\tilde{x}, \tilde{y}, \tilde{t})$ satisfies all downward IC constraints.

Let $\tilde{\mathcal{M}}(\Theta_\varepsilon)$ denote the set of mechanisms that are IR, involve no costly screening, and satisfy all downward IC constraints. We summarize the above discussion into a lemma:

Lemma 2. *Any mechanism $(x, y, t) \in \mathcal{M}(\Theta_\varepsilon)$ is dominated by some mechanism $(\tilde{x}, \tilde{y}, \tilde{t}) \in \tilde{\mathcal{M}}(\Theta_\varepsilon)$ (assuming truthful reporting). If the instruments are strictly costly and $\mathbb{P}(y(\theta) = y_0 \mid \varepsilon) < 1$, then (x, y, t) is strictly dominated by $(\tilde{x}, \tilde{y}, \tilde{t})$ (assuming truthful reporting).*

Step 3: By Lemma 2, it is, therefore, always an upper bound for the principal to optimize over the "shifted" space of mechanisms $\tilde{\mathcal{M}}(\Theta_\varepsilon)$. Because $v^B(y_0, \theta^B) = u^B(y_0, \theta^B) = 0$, the dependency of θ^B drops out in this problem. The principal then solves the following:

$$\sup_{(x,t)\colon \Theta^A \to \mathcal{X}\times\mathbb{R},\text{ measurable}} \quad \mathbb{E}[v^A(x(\theta^A), \theta^A) + t(\theta^A)] \tag{6}$$

$$\begin{aligned} \text{subject to} \quad & u^A(x(\theta^A), \theta^A) - t(\theta^A) \geq u^A(x(\hat{\theta}^A), \theta^A) - t(\hat{\theta}^A) \quad && \text{for all } \theta^A > \hat{\theta}^A, \\ & u^A(x(\theta^A), \theta^A) - t(\theta^A) \geq 0 && \text{for all } \theta^A. \end{aligned}$$

This problem does not depend on ε and is a standard one-dimensional screening problem *except* all upward IC constraints are ignored. For future references, we use $(6)^\dagger$ to denote the version of problem (6) with all IC constraints.

If we show that there exists (x^*, t^*) solving problem (6) and satisfying also all upward IC constraints, then Theorem 1 follows; that is, we want to show the set of downward IC constraints is sufficient for any standard one-dimensional screening problem, provided that the ex post surplus function satisfies our sorting condition. From now on, we drop the superscript A whenever clear, as we will focus only on the productive component.

Proposition 2 (Downward sufficiency)**.** *Consider any standard one-dimensional screening problem. Suppose the surplus function $s(x, \theta)$ satisfies the weak single-crossing differences condition. Then, there exists an optimal solution to (6) that satisfies all IC constraints.*

To proceed, we first prove Proposition 2 for the case of finite Θ and then for the general case in Step 4 using approximation. Let us suppose $|\Theta| = n < \infty$ and order types by $\theta_1 < \theta_2 < \cdots < \theta_n$. Let $\mu \in \Delta(\Theta)$ denote the distribution of θ. We assume μ has full support. Without loss of generality, suppose $0 \leq \theta_1$ and $\theta_n \leq 1$. A mechanism is then specified by $(x_1, x_2, \ldots, x_n)$ and $(t_1, t_2, \ldots, t_n)$. The principal's problem is given by

$$\max_{(x,t)\in\mathcal{X}^n\times\mathbb{R}^n} \sum_i \mu(\theta_i)(v(x_i, \theta_i) + t_i) \tag{7}$$

$$\begin{aligned} \text{subject to} \quad & u(x_i, \theta_i) - t_i \geq u(x_j, \theta_i) - t_j \quad && \text{for all } i > j, \\ & u(x_i, \theta_i) - t_i \geq 0 && \text{for all } i. \end{aligned}$$

We replace sup to max as the existence of the solution is easy to see by compactness arguments. We divide our analysis into two sub-steps. We first characterize the set of implementable x and the corresponding optimal transfers (Step 3.1), and then show by contradiction that any optimal solution to (7) must satisfy all IC constraints (Step 3.2).[16]

[16] This proves a stronger claim than Proposition 2 for finite Θ: relaxing upward IC constraints results in

Step 3.1: Fix any allocation rule $x \in \mathcal{X}^n$. We show that it is downward IC implementable and solve for the optimal transfer rule to implement it. To be precise, we want to solve

$$\max_{t \in \mathbb{R}^n} \sum_i \mu(\theta_i)(v(x_i, \theta_i) + t_i) \tag{8}$$

subject to the same IC and IR constraints as in (7).

We first identify regions where x is not monotone as follows. We start with x_1 and check if $x_{i+1} < x_i$ as we increase i. When this first happens, we denote $o_1 = i$, which marks the origin of our first U-shaped region. We then look for the next smallest index j such that $x_j > x_{o_1}$. We denote $d_1 = j$, which marks the destination of our first U-shaped region. We then start our index i at d_1 and repeat this process. Denote $r_l = \{o_l, \ldots, d_l\}$ as the lth U-shaped region.[17] Let L denote the number of such regions. Note that two U-shaped regions may share at most one point. Let

$$Q = \left\{1 \leq j \leq n : j \notin r_l \text{ for all } l, \text{ or } j = d_l \neq o_{l+1} \text{for some } l\right\}$$

be the set of all indices in the monotonic regions including the end points d_l but excluding the starting points o_l.

For notational convenience, we write IC$[i \to j]$ (or simply $[i \to j]$) and IR$[i]$ as a shorthand for IC$[\theta_i \to \theta_j]$ and IR$[\theta_i]$, respectively. We show the following claim:

Claim 1. *(i) There exists a unique optimal solution to* (8)*. (ii) The optimal solution to* (8) *is the unique solution to a system of equations defined by the following constraints with equality: IR$[1]$, IC$[(i+1) \to i]$ for all $i \in Q \backslash \{n\}$, and IC$[i \to o_l]$ for all $i \in r_l \backslash \{o_l\}$ and all l.*

Figure 1 illustrates how the U-shaped regions and the binding constraints in Claim 1 are identified. In short, the local IC constraints bind until one travels into a U-shaped region (beginning with, say, index o) where the binding constraints all point toward θ_o.

Proof of Claim 1. Relax all the constraints in (8) except the ones indicated in Claim 1. We will show the following: First, these constraints must bind in the relaxed problem. Second, these constraints binding imply all downward IC constraints and all IR constraints. Third, there is a unique solution to the system of equations defined by these binding constraints. Claim 1 then follows.

exactly the same set of solutions. For general Θ, our proof method will not extend this property. However, Proposition 2 suffices for our purpose.

[17] The end point may not be defined for the last U-shaped region. This does not pose any issue in the proof. Formally, let $d = n+1$ in that case and ignore any index $i > n$ in the proof.

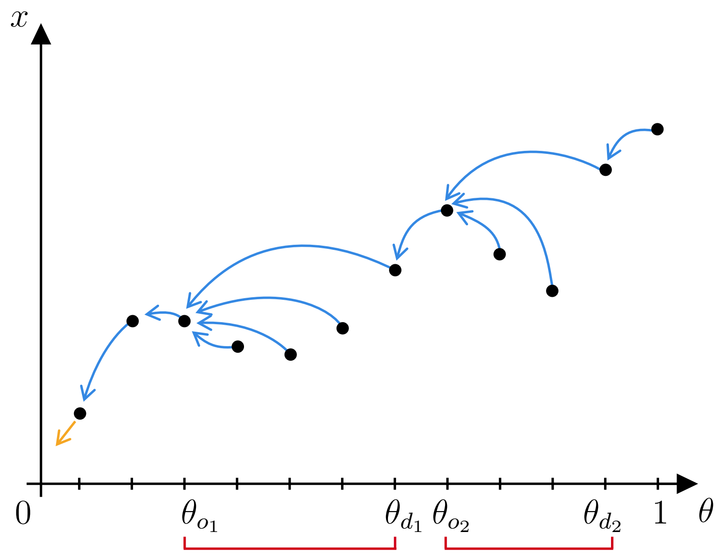

Figure 1: U-shaped regions and binding constraints for a fixed allocation rule

Note that for every $i > 1$, there is precisely one corresponding constraint $[i \to j]$ for some j. If this constraint does not bind at some mechanism (x, t), then simply set $\tilde{t}_i = t_i + \epsilon$ for some $\epsilon > 0$ small enough so that $[i \to j]$ still holds. This clearly increases the objective. It also does not distort other IC constraints. Indeed, the only other IC constraints this change affects are of the form $[k \to i]$ for some k, but

$$u(x_k, \theta_k) - t_k \geq u(x_i, \theta_k) - t_i \geq u(x_i, \theta_k) - \tilde{t}_i\,.$$

Therefore, all the IC constraints identified in Claim 1 must bind. Similarly, IR[1] binds.

Given that these constraints bind, we now show that they imply all the downward IC constraints in (8). We first collect two lemmas:

Lemma 3 (Local to global)**.** *Let $i > j > k$. If $[i \to j]$, $[j \to k]$ hold and $x_j \geq x_k$, then $[i \to k]$.*

Lemma 4 (Global to local)**.** *Let $i > j > k$. If $[i \to k]$, $[j \to k]$ bind and $x_j \leq x_k$, then $[i \to j]$.*

Lemma 3 is standard; it follows from a revealed-preference argument using the single-crossing property of u (we include a proof in the appendix for completeness). Lemma 4 appears to be new; it requires two binding IC constraints and follows from a revealed-preference argument that subtracts the two constraints. The appendix provides details.

We show all downward IC constraints are satisfied by induction on the number of U-shaped regions L. When $L = 0$, all downward IC constraints hold by successively applying Lemma 3 and building up from the adjacent local downward constraints. Suppose the claim holds for $L - 1$. Let us denote the last region as r with starting index o and end index d. By the inductive hypothesis, all downward IC constraints $[i \to j]$ are satisfied if $j < i \le o$. We divide the remaining pairs (j,i) with $j < i$ into two cases:

Case (1): $o \le j < i$. We make the following observations:

(a) if $o \le j < i \le d$, then $[i \to j]$ follows by the binding IC constraints $[i \to o]$, $[j \to o]$, $x_j \le x_o$, and Lemma 4;

(b) if $d \le j < i$, then $[i \to j]$ follows by successively applying Lemma 3;

(c) if $j < d < i$, then $[i \to j]$ follows by $[i \to d]$ from (b), $[d \to j]$ from (a), $x_d \ge x_j$, and Lemma 3.

Case (2): $j < o < i$. Note that $x_j \le x_o$ for all $j < o$. Then, $[i \to j]$ follows by $[i \to o]$ from Case (1), $[o \to j]$ from the inductive hypothesis, $x_o \ge x_j$, and Lemma 3.

Together, these cover all the downward IC constraints and prove the inductive step. The IR constraints follow easily from IR[1] and IC$[i \to 1]$ and that $u(x,\cdot)$ is nondecreasing.

The binding constraints define a system of n equations for t. It is not hard to see that these equations can be solved successively starting from the lowest one. By induction, the binding constraints uniquely define the following transfer formula:

$$t_i = u(x_i,\theta_i) - \underbrace{\sum_{j=1,2,\ldots,i-1:\ j\in Q} \Big(u(x_j,\theta_{j+1}) - u(x_j,\theta_j)\Big)}_{\text{local}} - \underbrace{\sum_{l=1,2,\ldots,L:\ o_l<i} \Big(u(x_{o_l},\theta_{d_l\wedge i}) - u(x_{o_l},\theta_{o_l})\Big)}_{\text{nonlocal}}, \tag{9}$$

where we use the notation $a \wedge b := \min(a,b)$. The first sum arises from the local downward IC constraints, and the second sum arises from the nonlocal ones. □

If x is monotonic, then the binding IC constraints are the local downward ones, and the transfer formula reduces to the standard solution. Note however that, in contrast to the standard setting with all IC constraints, *any* $x \in \mathcal{X}^n$ can be implemented by a downward IC mechanism and the corresponding optimal transfers depend on the *shape* of x.

Step 3.2: We now show the following claim:

Claim 2. *Any optimal solution to (7) must have a monotonic allocation rule:* $x_1 \le x_2 \le \cdots \le x_n$.

Proof of Claim 2. Suppose, for contradiction, that there is an optimal solution (x,t) such that x is not monotone. The proof idea is to perturb the allocation rule (as well as the optimal transfer rule implementing it) and show a strict improvement. Because x is not monotone, there exists a U-shaped region. Let r be the first U-shaped region, o its starting index, and d its end index. Moreover, let

$$g = \min\left\{j > o : x_j \geq x_o\right\}$$

denote the first index after o with associated allocation no less than x_o. Put $g = n+1$ if the above set is empty. Either $g = d$ or g is the first index in $r \backslash \{o\}$ such that $x_g = x_o$. Let

$$\hat{x} = \max\left\{x_j : o < j < g\right\}$$

denote the largest allocation for indices strictly between o and g. We have $\hat{x} < x_o$. Let $j^* \in \{o, o+1, \ldots, g\}$ be the first index achieving the maximum and let $\hat{\theta} = \theta_{j^*}$. Let

$$k = \min\left\{j : x_j > \hat{x}\right\}$$

denote the first index whose associated allocation is strictly higher than $\hat{x}$. Since $x_o > \hat{x}$, we have $k \leq o$. Because r is the first U-shaped region, we have

$$\hat{x} < x_k \leq x_{k+1} \leq \cdots \leq x_o .$$

Consider the following perturbation: Let $\tilde{x}$ be the same as the original allocation rule except $\tilde{x}_j = \hat{x}$ for all $j = k, k+1, \ldots, o$. Let $\tilde{t}$ be the optimal transfer rule implementing $\tilde{x}$. Figure 2 illustrates.

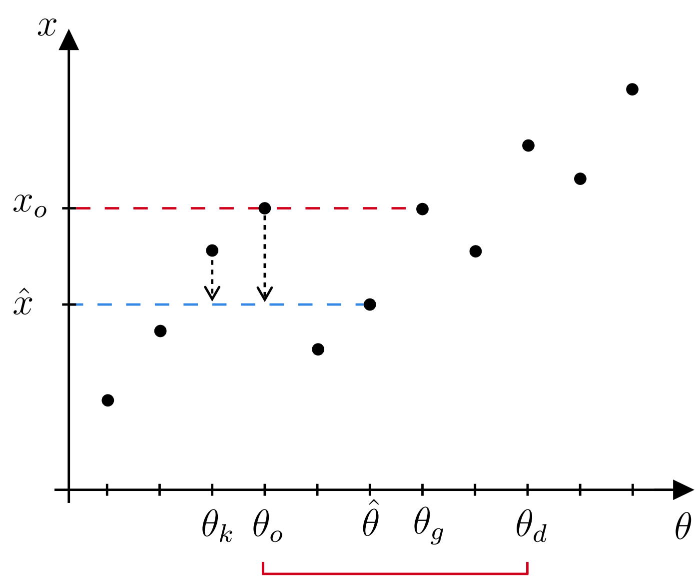

Figure 2: Perturbation of the allocation rule

This perturbation creates minimal changes to the shape of x. In particular, we show that the set of binding constraints for the transfers (identified in Claim 1) is preserved under the perturbation. This is easy to see for the ones pointing from $\{k, k+1, \ldots, o\}$: by Claim 1, these are the local downward IC constraints both before and after the perturbation. For the ones in the U-shaped region r, by Claim 1, they are of the form $[j \to o]$ before the perturbation. If $g = d$, then after perturbation the binding constraints are the same since the U-shaped region stays essentially the same.[18] If $g < d$, then $x_g = x_o$ and we can split the region r into $\{o, o+1, \ldots, g\}$ and $\{g, g+1, \ldots, d\}$ by replacing $[j \to o]$ to $[j \to g]$ for $g < j \le d$. Because the perturbation moves $x_k, x_{k+1}, \ldots, x_o$ downward, by Claim 1, the binding constraints afterward are still of the form $[j \to o]$ for $o < j \le g$ and $[j \to g]$ for $g < j \le d$. Therefore, the set of binding constraints is preserved under the perturbation (after splitting if necessary).

Because the set of binding constraints is preserved, the objective takes the same *form* before and after perturbation. We show that the objective weakly increases on the parts involving $x_k, x_{k+1}, \ldots, x_{o-1}$ (which may be an empty set) and strictly increases on the parts involving x_o (which always exist). To start, fix some $j \in \{k, k+1, \ldots, o-1\}$. Plugging (9) into the objective of (7) and collecting terms involving x_j gives

$$s(x_j, \theta_j)\mu(\theta_j) - \Big(u(x_j, \theta_{j+1}) - u(x_j, \theta_j)\Big) \sum_{i>j} \mu(\theta_i). \tag{10}$$

This is the discrete analog of the virtual surplus function multiplied by $\mu(\theta_j)$. Now consider the terms involving x_{j^*}. Because $o < j^* < g$, there is no IC constraint pointing toward j^*. Therefore, there is only one such term:

$$s(x_{j^*}, \theta_{j^*})\mu(\theta_{j^*}).$$

Note that $x_j \in \mathcal{X}$ is feasible to assign to θ_{j^*}. Moreover, doing so does not change the shape of x since $x_j \le x_o$, thus generating a payoff also according to the above formula. The fact that x is optimal then implies

$$s(\hat{x}, \hat{\theta}) \ge s(x_j, \hat{\theta});$$

that is,

$$s(x_j, \hat{\theta}) - s(\hat{x}, \hat{\theta}) \le 0.$$

[18] If $j^* > o+1$, then the U-shaped region stays exactly the same. Otherwise, $j^* = o+1$ and the U-shaped region starts with $o+1$ afterward but we can simply replace all constraints $[j \to o+1]$ by $[j \to o]$ as $\tilde{x}_o = \tilde{x}_{o+1}$.

Because $x_j > \hat{x}$ and $\theta_j < \hat{\theta}$, by the weak single-crossing differences property of s,

$$s(x_j, \theta_j) - s(\hat{x}, \theta_j) \leq 0. \tag{11}$$

Moreover, because $x_j > \hat{x}$, by the strict increasing differences property of u,

$$u(x_j, \theta_{j+1}) - u(x_j, \theta_j) > u(\hat{x}, \theta_{j+1}) - u(\hat{x}, \theta_j). \tag{12}$$

Combining (11) and (12) gives

$$s(x_j,\theta_j)\mu(\theta_j) - \Big(u(x_j,\theta_{j+1}) - u(x_j,\theta_j)\Big)\sum_{i>j}\mu(\theta_i) \leq s(\hat{x},\theta_j)\mu(\theta_j) - \Big(u(\hat{x},\theta_{j+1}) - u(\hat{x},\theta_j)\Big)\sum_{i>j}\mu(\theta_i),$$

proving that the part of the objective involving x_j increases. Because this holds for all $j \in \{k, k+1, \ldots, o-1\}$, to conclude our proof, it remains to show that the part of the objective involving x_o strictly increases. Recall that we split the region r into $\{o, o+1, \ldots, g\}$ and $\{g, g+1, \ldots, d\}$ if $g < d$. Plugging (9) into (7) and collecting terms involving x_o gives

$$s(x_o,\theta_o)\mu(\theta_o) - \sum_{i=o+1}^{g}\mu(\theta_i)\Big(u(x_o,\theta_i) - u(x_o,\theta_o)\Big) - \Big(u(x_o,\theta_g) - u(x_o,\theta_o)\Big)\sum_{i>g}\mu(\theta_i).$$

By the same argument as the previous case, we have

$$s(x_o, \theta_o) \leq s(\hat{x}, \theta_o).$$

For any $i > o$, by the strict increasing differences property of u,

$$u(x_o, \theta_i) - u(x_o, \theta_o) > u(\hat{x}, \theta_i) - u(\hat{x}, \theta_o).$$

Together they imply

$$\begin{aligned} & s(x_o,\theta_o)\mu(\theta_o) - \sum_{i=o+1}^{g}\mu(\theta_i)\Big(u(x_o,\theta_i) - u(x_o,\theta_o)\Big) - \Big(u(x_o,\theta_g) - u(x_o,\theta_o)\Big)\sum_{i>g}\mu(\theta_i) \\ & < s(\hat{x},\theta_o)\mu(\theta_o) - \sum_{i=o+1}^{g}\mu(\theta_i)\Big(u(\hat{x},\theta_i) - u(\hat{x},\theta_o)\Big) - \Big(u(\hat{x},\theta_g) - u(\hat{x},\theta_o)\Big)\sum_{i>g}\mu(\theta_i), \end{aligned}$$

where the strict inequality also uses that μ has full support. □

To finish the proof of Proposition 2 for finite type spaces, we recall the following:

Lemma 5. *If x is monotone and $[(i+1) \to i]$ binds for all i, then all upward IC constraints are satisfied.*

This result is standard (we include a proof in the appendix for completeness). It follows from that binding local downward IC constraints with monotonicity of x imply local upward IC constraints and that local upward IC constraints with monotonicity of x imply all upward IC constraints.

Now fix any optimal solution (x, t) to (7). By Claim 2, we have x is monotone. Then, by Claim 1, all local downward IC constraints bind. Hence, by Lemma 5, we have Proposition 2 hold for finite type spaces.

Step 4: We prove Proposition 2 for general type space Θ by approximation.[19] Because this step is mostly technical, we give a sketch of the argument and leave the details to the appendix. Let $\mu \in \Delta(\Theta)$ denote the distribution on Θ. Recall that $(6)^{\dagger}$ denotes the version of program (6) with all IC constraints (both downward and upward). Let $V(\Theta, \mu)$ denote the optimal value of $(6)^{\dagger}$ given (Θ, μ). We show that $V(\Theta, \mu)$ equals to the optimal value of (6). Suppose, for contradiction, there exists some $(\hat{x}, \hat{t})$ feasible for (6) such that

$$V(\Theta, \mu) < \mathbb{E}^{\mu}[v(\hat{x}(\theta), \theta) + \hat{t}(\theta)]. \tag{13}$$

We first construct an appropriate sequence $\{(\Theta^{(n)}, \mu^{(n)})\}$ approximating (Θ, μ).

Lemma 6. *Suppose $v(x, \theta)$ is Lipschitz continuous on $\mathcal{X} \times \Theta$. Then, there exists a sequence $\{(\Theta^{(n)}, \mu^{(n)})\}$ with $\Theta^{(n)} \subseteq \Theta$ finite and $\mu^{(n)} \in \Delta(\Theta^{(n)})$ full support such that*

(i) $\mu^{(n)} \to_w \mu$;

(ii) $\limsup_{n \to \infty} V(\Theta^{(n)}, \mu^{(n)}) \le V(\Theta, \mu)$.

Because atomic measures are dense in $\Delta(\Theta)$ (with the weak-* topology), one can easily find a sequence satisfying (i). However, the upper semicontinuity property (ii) does not necessarily hold for an arbitrary approximation sequence. We construct a particular approximation sequence such that for each $(\Theta^{(n)}, \mu^{(n)})$, we can convert the optimal mechanism there into a mechanism for the problem with (Θ, μ), with small loss in the objective. The argument relies on the problem being one-dimensional.

[19] The results so far imply a version of Theorem 1 for finite type spaces. One may hope to approximate directly in the original problem as in Carroll (2017). However, Carroll (2017)'s approximation argument is based on Madarász and Prat (2017)'s result, which relies crucially on there being no interdependent values. Our problem allows for interdependent values, and so requires a different approximation argument.

Now note that $(\hat{x}, \hat{t})$ restricted to $\Theta^{(n)}$ is a feasible solution to the finite-type version of (6) with $(\Theta^{(n)}, \mu^{(n)})$. By Step 3, we have

$$V(\Theta^{(n)}, \mu^{(n)}) \geq \mathbb{E}^{\mu^{(n)}}[v(\hat{x}(\theta), \theta) + \hat{t}(\theta)].$$

Suppose for a moment that $\hat{x}, \hat{t}$ are continuous on Θ and v is Lipschitz continuous. Then $v(\hat{x}(\theta), \theta) + \hat{t}(\theta)$ is a bounded continuous function on Θ. Using Lemma 6 and taking limits on both sides of the above, we have

$$V(\Theta, \mu) \geq \limsup_{n\to\infty} V(\Theta^{(n)}, \mu^{(n)}) \geq \limsup_{n\to\infty} \mathbb{E}^{\mu^{(n)}}[v(\hat{x}(\theta), \theta) + \hat{t}(\theta)] = \mathbb{E}^{\mu}[v(\hat{x}(\theta), \theta) + \hat{t}(\theta)],$$

contradicting (13).

In general, the situation is more delicate. We prove it first for Lipschitz continuous v and then for all continuous v by extension (via the Stone–Weierstrass theorem). For any measurable $\hat{x}, \hat{t}$, we invoke Lusin's theorem to identify a compact set $\tilde{\Theta} \subset \Theta$ such that $\hat{x}, \hat{t}$ are continuous on $\tilde{\Theta}$ with $\mu(\Theta \backslash \tilde{\Theta})$ sufficiently small. Let $\tilde{\mu}$ be the conditional measure of μ on $\tilde{\Theta}$. Because $\hat{x}, \hat{t}$ are continuous on $\tilde{\Theta}$, the above argument applies to $(\tilde{\Theta}, \tilde{\mu})$. The proof shows that the difference of objective values between $(\tilde{\Theta}, \tilde{\mu})$ and (Θ, μ) can be made arbitrarily small. This then gives a contradiction.

Finally, to conclude Proposition 2, it suffices to show the existence of an optimal solution to the full IC program $(6)^{\dagger}$. Even though this is a standard one-dimensional problem, the existence result appears to be new at this generality. As it may be of independent interest, we record it in the following lemma:

Lemma 7. *Any standard one-dimensional screening problem has a solution.*

The proof (in the appendix) proceeds by showing the space of IC and IR mechanisms is sequentially compact in the product topology. The argument uses a generalized version of Helly's selection theorem from Fuchino and Plewik (1999). *Q.E.D.*

Remark 1. The following example shows that some sorting condition on the surplus function is needed for Proposition 2:

Example 1. Suppose $\Theta = \mathcal{X} = \{0, 1\}$; types are uniformly distributed; $u(x, \theta) = \theta x$, $v(x, \theta) = \kappa(\frac{1}{2} - \theta)x$ for some $\kappa > 2$. The efficient allocation is to assign $x(0) = 1$, $x(1) = 0$. The optimal downward IC mechanism is to set $x(0) = 1$, $t(0) = 0$ and $x(1) = 0$, $t(1) = -1$. This does not satisfy all IC constraints: IC$[0 \to 1]$ is violated.

Remark 2. Building on the above example, the next example shows that some sorting condition on the surplus function s^A is needed for Theorem 1:

Example 2. In the setting of Example 1, we set $\kappa = 2.5$ and add another component, $\Theta^B = \{-1, 0\}$, $\mathcal{Y} = \{0, 1\}$; $u^B(y, \theta^B) = \theta^B y$, $v^B(y, \theta^B) = 0$; types are uniformly distributed and comonotonic. Screening only with x yields at most a payoff of 0. However, with the costly instrument, the menu $\{(1, 0, 0), (0, 1, -1)\}$ yields a payoff of $\frac{1}{8} > 0$.

5 Monopoly Pricing with Costly Signals

Before discussing other applications of Theorem 1 in Section 6, we specialize the main model to a setting of monopoly pricing with costly signals. In this setting, a monopolist sells a spectrum of quality-differentiated goods and can make the menu of offers contingent on the costly actions that a buyer may take.

In Section 5.1, we show that if the buyer's utility functions are *multiplicatively separable* within each component, then the positive correlation of preferences condition can be weakened to the positive correlation between the preferences for the productive component and the *marginal rates of substitution* between the productive and costly components.

In Section 5.2, we consider a multiple-good monopolist selling different qualities of bundles (with no costly signals). This environment generalizes the classic multiple-good monopoly problem by allowing for *both* probabilistic bundling and quality discrimination. We show that (a relaxation of) this problem can be mapped to a monopolist selling a spectrum of quality-differentiated goods without bundling but with costly signals as in Section 5.1. A key insight is that one can view selling the grand bundle as the productive component, and selling a smaller bundle *instead of* the grand bundle as a costly instrument for screening a consumer's value for the grand bundle. Using this perspective, we generalize a result of Haghpanah and Hartline (2021). In particular, we show that under their stochastic ratio monotonicity condition, the general feature of the optimal mechanism is to post a menu of different qualities of the grand bundle — the monopolist screens only the productive component and does not use any of the "costly signals."

Setup. A monopolist sells a quality-differentiated spectrum of goods. A buyer of type $\theta^A \in \Theta^A$ receives utility $u^A(x, \theta^A)$ from consuming the good of quality $x \in \mathcal{X}$. The seller incurs a cost $C(x, \theta^A)$ to produce the good of quality x for type θ^A. Suppose $u^A(x, \theta^A)$ is nondecreasing in θ^A and has strict increasing differences, and that the surplus function $u^A(x, \theta^A) - C(x, \theta^A)$ has weak single-crossing differences. (The continuity and compactness assumptions in Section 2 are also maintained.)

Besides offering a menu of products of different qualities and prices, the monopolist can make the offers contingent on various costly signals (e.g. waiting in line, collecting coupons, walking up stairs). A costly signal is represented by $y \in \mathcal{Y}$. To obtain a signal y, a buyer of type $\theta^B \in \Theta^B$ incurs a cost $c(y, \theta^B)$ that is nonincreasing in θ^B (so θ^B represents the willingness to endure various costly activities).

Theorem 1 then says that if θ^B is positively correlated with θ^A according to our notion, then the monopolist never makes more profits by using these costly signals. Therefore, if the monopolist in fact uses these instruments, then we should expect that the consumers with higher willingness to pay tend to incur higher costs to obtain the signals (both measured with respect to the constant marginal value for money). In fact, sometimes we can say more when the buyer's utility functions are multiplicatively separable within each component, which we turn to next.

5.1 Marginal Rates of Substitution Between Two Components

We follow the notation in the above setup, and let $\mathcal{X}$ be $[0,1]$, $\mathcal{Y}$ any measurable space, Θ^A any compact subset of $\mathbb{R}_{++}$, and Θ^B any compact subset of $\mathbb{R}^N_-$. We say that the buyer has *multiplicatively separable* utilities within each additive component if for any quality x, signal y, and price t, the buyer's payoff can be written as

$$\theta^A u(x) + \theta^B \cdot c(y) - t$$

where $u : \mathcal{X} \to \mathbb{R}$ is a continuous and strictly increasing function satisfying $u(0) = 0$, and $c : \mathcal{Y} \to \mathbb{R}^N_+$ is a bounded measurable function satisfying $c(y_0) = 0$ for some $y_0 \in \mathcal{Y}$.[20]

We say that the monopolist's cost function is *not interdependent* if $C(x, \theta^A)$ does not depend on θ^A, in which case without loss of generality we let $C(0) = 0$.

Recall the notation $\theta^B = (\theta^1, \ldots, \theta^N)$. Let $r^i = \frac{\theta^i}{\theta^A}$ and $r^B = (r^1, \ldots, r^N)$. Note that $r^B \leq 0$. We interpret r^B as the (negative) *marginal rates of substitution* between the productive and costly components.[21] In this setting, we show that our assumption of positive correlation between θ^A and θ^B can be weakened to that between θ^A and r^B.[22]

Proposition 3. *Suppose the seller's cost function C is continuous, nondecreasing in x, and not interdependent; the buyer's utilities are multiplicatively separable; and r^B is stochastically nondecreasing in θ^A. Then, there exists an optimal mechanism that involves no costly screening.*

The intuition behind this result can be understood in the same way as in the proof of the main result (Section 4). We show that when the marginal rates of substitution

[20] Note that a utility of the form $f^A(\theta^A)u(x) + f^B(\theta^B) \cdot c(y)$ provides (essentially) no additional generality.

[21] The substitution here is between the *utility* $u(x)$ from the productive component and the *disutility* $c(y)$ from the costly component, and hence has negative marginal rates.

[22] This is in general not necessarily a weaker condition; it is so in this case because $\theta^B \leq 0$.

are increasing in the values, instead of using the costly signals, the principal can simply adjust the allocations of the productive component while maintaining the downward IC constraints. Because downward IC constraints are sufficient, the result follows. Unlike in Section 4, in this case we substitute the costly signals with a decrease in the productive allocations holding the monetary transfers fixed, which is why the marginal rates of substitution between the two components play an important role here.

Proof of Proposition 3. By Lemma 1, as in Step 1 of Section 4, it suffices to show the case where $r^B = h(\theta^A)$ for some nondecreasing function $h : \Theta^A \to \mathbb{R}^N$. Thus, we may assume for all i, r^i is deterministic conditional on θ^A and nondecreasing in θ^A. Fix any (x, y, t) that is IC and IR. We may assume $t \geq 0$, because the monopolist can simply replace all options with negative profits in the menu with $(0, y_0, 0)$ and weakly increase the total profit (since the monopolist's cost function does not depend on the buyer's type). Now we apply a "shift" argument as follows. Consider the modification: $\tilde{t} = t, \tilde{y} = y_0$, and

$$\tilde{x}(\theta) = u^{-1}\Big(u(x(\theta)) + \frac{1}{\theta^A}[\theta^B \cdot c(y(\theta))]\Big).$$

Because $u(\cdot)$ is continuous and strictly increasing with $u(0) = 0$, u^{-1} is defined on $[0, u(1)]$. Moreover, because (x, y, t) is IR and $t \geq 0$, we have $0 \leq u(x(\theta)) + \frac{1}{\theta^A}[\theta^B \cdot c(y(\theta))] \leq u(x(\theta))$ for all θ. So the modification is well-defined and $0 \leq \tilde{x} \leq x$ pointwise. In other words, the modified mechanism decreases the productive allocation to substitute the costly screening so that all types have the same utilities as before, *assuming* truthful reporting.

Because $C(\cdot)$ is nondecreasing, this modification increases the objective, assuming truthful reporting. It is IR by construction. Moreover, it is downward IC: for any $\hat{\theta}^A < \theta^A$,

$$\begin{aligned}
\theta^A u(\tilde{x}(\theta)) - \tilde{t}(\theta) &= \theta^A u(x(\theta)) + \theta^B \cdot c(y(\theta)) - t(\theta) \\
&\geq \theta^A u(x(\hat{\theta})) + \theta^B \cdot c(y(\hat{\theta})) - t(\hat{\theta}) \\
&= \theta^A\Big(u(x(\hat{\theta})) + r^B \cdot c(y(\hat{\theta}))\Big) - t(\hat{\theta}) \\
&\geq \theta^A\Big(u(x(\hat{\theta})) + \hat{r}^B \cdot c(y(\hat{\theta}))\Big) - t(\hat{\theta}) = \theta^A u(\tilde{x}(\hat{\theta})) - \tilde{t}(\hat{\theta}).
\end{aligned}$$

The first inequality holds because (x, y, t) is IC. The second inequality holds because $\hat{r}^B \leq r^B$ and $c \geq 0$. Invoking Proposition 2 concludes the proof. □

5.2 Bundling and Quality Discrimination

We now show an application to a multiple-good monopoly problem allowing for both probabilistic bundling and quality discrimination.

A monopolist sells G many goods to a unit mass of consumers. For each bundle b, a random consumer has value v^b for getting the highest quality version of the bundle with probability one. We assume that $v^b \le v^{b'}$ for all $b \subset b'$ and $v^{\varnothing} = 0$. The monopolist can use probabilistic bundling, captured by a bundling allocation rule $v \mapsto \alpha(v) \in \Delta(2^G)$. In addition, the monopolist can adjust the quality of each bundle, captured by a quality allocation rule $v \mapsto q(v) \in [0,1]^{2^G}$. A type-$v$ consumer's payoff is given by

$$\sum_b \alpha^b q^b v^b - t.$$

The monopolist incurs a cost to improve the quality of a bundle, with a payoff given by

$$-\sum_b \alpha^b C(q^b) + t,$$

where $C(\cdot)$ is a continuous, nondecreasing, and convex function on $[0,1]$ with $C(0) = 0$. This cost structure assumes that the cost of producing a bundle of some quality does not depend on the size of the bundle, which is perhaps more suitable for digital goods.

Let v^* be the value of a random consumer for the grand bundle and $\tau = (\frac{v^b}{v^*})_{b=1,\dots,2^G}$ be the profile of values for each bundle relative to the grand bundle.

Proposition 4. *If τ is stochastically nondecreasing in v^*, then an optimal mechanism exists and can be implemented by a menu of prices for different qualities of the grand bundle.*

This result is a natural consequence of Proposition 3 once one views selling the grand bundle as the productive component, and selling a smaller bundle *instead of* the grand bundle as a costly instrument for screening a consumer's value for the grand bundle.

Proof of Proposition 4. By convexity of $C(\cdot)$ and Jensen's inequality, we have

$$\sum_b \alpha^b(v) C(q^b(v)) \ge C\Big(\sum_b \alpha^b(v) q^b(v)\Big).$$

Therefore, it is an upper bound on the monopolist's revenue to maximize the objective

$$\mathbb{E}\Big[-C\Big(\sum_b \alpha^b(v) q^b(v)\Big) + t(v)\Big]. \tag{14}$$

For this auxiliary problem, let us also relax the constraint $\sum_b \alpha^b = 1$ to $\sum_b \alpha^b \le 1$. Then, because α, q enter both the consumer's utility and the objective in the same way, it is without loss of generality to let $q^b = 1$ for all b.

We now reformulate this problem as a problem of monopoly pricing with costly signals. Let $\theta^A = v^*$ be the value of the grand bundle. For any proper bundle b, let

$$\theta^b = v^b - v^*$$

be the difference of values for bundle b and the grand bundle b^*. In words, θ^b is the negative value for getting bundle b instead of b^*. Let $N = 2^G - 1$, and let $\theta^B = (\theta^1, \ldots, \theta^N)$ be the profile of the differences.

We use $x : \Theta \to [0,1]$ to denote the *initial* allocation of the grand bundle, and $y : \Theta \to [0,1]^N$ to denote the allocation of the "costly signals" as follows. An assignment $y^b \in [0,1]$ represents assigning bundle b with probability y^b while *decreasing* the probability of the grand bundle b^* also by y^b. The consumer's payoff can be rewritten as

$$\theta^A x + \theta^B \cdot y - t.$$

For any substochastic allocation α (i.e. $\sum_b \alpha_b \leq 1$), we can replicate it by setting

$$x = \sum_b \alpha^b, \qquad y^b = \alpha^b \text{ for all } b \neq b^*$$

Therefore, the auxiliary problem (14) can be further relaxed to

$$\sup_{(x,y,t)\in\mathcal{M}(\Theta)} \mathbb{E}[-C(x(\theta)) + t(\theta)]. \tag{15}$$

For any $b = 1, \ldots, N$, we have $\frac{v^b}{v^*} = \frac{\theta^A + \theta^b}{\theta^A} = 1 + \frac{\theta^b}{\theta^A}$. Since τ is stochastically nondecreasing in v^*, we have $r^B := \frac{1}{\theta^A}\theta^B$ is stochastically nondecreasing in θ^A. So Proposition 3 applies to (15). Let $(x^*, 0, t^*)$ be the optimal solution to (15) that involves no costly screening.

We construct an allocation rule in the original problem as follows:

$$\alpha^{b^*} = 1,\ \alpha^b = 0 \text{ for all } b \neq b^*; \quad q^{b^*} = x^*,\ q^b = 0 \text{ for all } b \neq b^*.$$

Because probabilities and qualities enter the consumer's utility in the same way and $(x^*, 0, t^*)$ is IC and IR, (α, q, t^*) is also IC and IR. The revenue of the monopolist under (α, q, t^*) is

$$\mathbb{E}[-C(q^{b^*}(\theta)) + t^*(\theta)] = \mathbb{E}[-C(x^*(\theta)) + t^*(\theta)],$$

the optimal value of (15). Hence, (α, q, t^*) is optimal for the monopolist in the original problem; moreover, (α, q, t^*) screens using only the qualities of the grand bundle. □

Remark 3. Proposition 4 says that under the stochastic ratio monotonicity condition, the monopolist can restrict attention to selling only the grand bundle at various qualities. Because that is a one-dimensional problem à la Mussa and Rosen (1978), the solution can be explicitly characterized. When there is no cost for quality improvement ($C = 0$), the optimal mechanism is to sell the grand bundle at the highest quality with a posted price. This special case is due to Haghpanah and Hartline (2021). In general, however, the optimal mechanism involves price discrimination. But Proposition 4 shows that such price discrimination is only done by creating different qualities of the grand bundle.

Remark 4. In independent work, Bergemann et al. (2021) also use the term "marginal rates of substitution" but for a different condition. (They study the optimality of a menu of nested bundles in a multiple-good monopoly problem without quality choices.)

6 Additional Applications

6.1 Optimal Regulation

Consider the monopoly regulation problem first analyzed by Baron and Myerson (1982). For simplicity, we follow the formulation of Laffont and Tirole (1993). Let x be the quantity of production; $p(x)$ the inverse demand curve; and $\psi(x, \theta^A)$ the cost function for a firm of type θ^A, decreasing in θ^A. Suppose the marginal cost $\psi_x(x, \theta^A)$ is decreasing in θ^A so that a higher efficiency parameter θ^A represents a firm with lower marginal cost. Let

$$S(x) = \int_0^x (p(q) - p(x))dq$$

denote the consumer surplus at quantity x. The regulator wants to maximize the sum of consumer surplus and firm profits. Moreover, for any (lump-sum) tax t collected, the regulator can reduce distortionary taxes elsewhere and generate benefits $(1+\lambda)t$ for some $\lambda > 0$.[23] The regulator's payoff is thus given by $S(x) + p(x)x - \psi(x, \theta^A) + \lambda t$. After scaling, we may let the regulator's utility function be

$$v^A(x, \theta^A) = \frac{1}{\lambda}\Big(S(x) + p(x)x - \psi(x, \theta^A)\Big).$$

The firm's utility function is given by

$$u^A(x, \theta^A) = p(x)x - \psi(x, \theta^A).$$

[23]This is equivalent to assuming that the regulator places some weights on both surplus and revenue.

Suppose the regulator can observe some inspection outcomes correlated with the firm's production efficiency.[24] Let $y \in \{0, 1, \dots, k\}$ represent a certificate level. Without exerting any effort, the firm can achieve a performance level θ^B, or any level less than that if the firm chooses to do so. The firm may exert effort to increase from its baseline θ^B; there is a cost $C > 0$ for each level above θ^B. So achieving level y for a firm of type θ^B costs

$$c(y, \theta^B) = C \max\{y - \theta^B, 0\}.$$

Such effort does not affect the real cost of production. So for a quantity x, tax t, and certificate level y, a type-θ firm gets a payoff

$$p(x)x - \psi(x, \theta^A) - c(y, \theta^B) - t.$$

Suppose the baseline level θ^B is positively correlated with the efficiency parameter θ^A according to our notion. Our result then says that it is optimal for the regulator *not* to make the tax and subsidy schedule contingent on the inspection outcomes in any way.

Remark 5. If the regulator were able to directly observe θ^B as in Laffont and Tirole (1986), then the data would always be helpful as they alleviate the information asymmetry. Besides that the outcomes are manipulable, another important feature of this setup is that no effort is needed to *decrease* the performance on an inspection.

6.2 Labor Market Screening

A monopsonistic firm wants to hire a worker. The firm gets a profit $v(x, \theta^0) - w$ for hiring a worker of ability θ^0 to produce x amount of work at wage w. Suppose the marginal productivity v_x is nondecreasing in the ability θ^0. A worker of ability θ^0 incurs a cost $\psi(x, \theta^0)$ decreasing in θ^0 to produce x amount of work. Suppose the marginal cost ψ_x is decreasing in the ability θ^0.

The firm has many costly instruments at its disposal; for example, it can ask the applicant to participate in various interviews or pass some tests.[25] Suppose there are N such activities. For $i = 1, \cdots, N$, let $c^i(y^i, \theta^i)$ denote the cost of obtaining a level-y^i signal in activity i that is nonincreasing in the cost type θ^i. For an offer (x, w) contingent on the

[24] For example, the regulator may observe the outcomes from energy performance inspections.

[25] We assume away many realistic features that may be important in labor markets, such as verification aspects of interviews and potential moral hazard problems after hiring.

costly signals $(y^1, \ldots, y^N)$, the agent gets a payoff

$$w - \psi(x, \theta^0) - \sum_{i=1}^{N} c^i(y^i, \theta^i).$$

Suppose $\theta^B = (\theta^1, \ldots, \theta^N)$ is positively correlated with θ^0 according to our notion so that a higher ability worker tends to find the costly activities easier. Our result then says that the firm should *not* make the menu of job offers contingent on any costly activity.

Remark 6. This implication contrasts with the common perception of costly signals in competitive labor markets. To clarify the difference, in Appendix B.1 we consider a competitive screening model in which multiple firms compete and are allowed to screen with both work allocations and costly instruments. We show that costly screening can appear in equilibrium. The intuition can be understood from the direction of binding incentive constraints, just as we have seen in Section 4. Suppose, for contradiction, there is no costly screening involved. Then the game reduces to a standard one-signal game. Because the firms compete, each type captures all surplus generated. But then it is the *upward* IC constraint that binds as the firms want to trade with the high type; the work assignment to the high type is distorted upward. A firm can then lower the work assignment and profit from the additional surplus generated. To deter the low type from imitating the high type, the firm can ask for a small amount of costly signals. If a small y costs much less for the high type, then it can deter upward deviation without destroying much surplus.

6.3 Monopoly Pricing with Costly Production

Consider a monopolist selling a quality-differentiated spectrum of goods as in the setting of Section 5 but with no costly signals. That is, a buyer of type $\theta^A \in \mathbb{R}$ receives utility $u^A(x, \theta^A)$ from consuming the good of quality x. The seller incurs a cost $C(x, \theta^A)$ to produce the good of quality x for type θ^A. Suppose that $u^A(x, \theta^A)$ is nondecreasing in θ^A and has strict increasing differences, and that the surplus function $u^A(x, \theta^A) - C(x, \theta^A)$ has weak single-crossing differences.

The monopolist also produces a different kind of goods; it costs $C^B(y, \theta^B)$ to produce the second good of quality y for type θ^B, and the consumer has a utility function $u^B(y, \theta^B)$ nondecreasing in θ^B. For the second good, however, the costs exceed the potential values: $C^B(y, \theta^B) \geq u^B(y, \theta^B)$ for all y and θ^B. Suppose θ^A, θ^B are positively correlated as in our definition. Our result then says that it is optimal for the monopolist *not* to sell the second good of any quality. This is generally not true when the preferences are negatively correlated. A related intuition is that negative correlation of values makes bundling profitable (Adams and Yellen, 1976). For a concrete example, consider the following:

Example 3. Suppose $\Theta^A = \Theta^B = \{1, 2\}$; the types are uniformly distributed on $\{(1, 2), (2, 1)\}$. The consumer has linear utility functions. Item A costs 0 to produce, and item B costs 2 to produce. Selling the bundle at price 3 and item A at price 2 gives a profit of 1.5, strictly higher than the monopoly profit of 1 from selling item A.

Remark 7. The intuition here still differs from that for bundling versus separate sales. In particular, it is known from Carroll (2017) that separate sales can be suboptimal even when the values are maximally positively correlated (i.e. on a monotonic path).

7 Concluding Remarks

This paper studies the effectiveness of costly instruments in a general multidimensional screening model. The model consists of two components: a one-dimensional productive component and a multidimensional costly component. Our main result says that if the agent's preferences are positively correlated between the two components in a suitably defined sense, then the costly instruments are ineffective — the optimal mechanism simply screens the one-dimensional productive component.

Our proof also provides clear insights into *why* this result holds. First, we show that costly instruments can loosen upward but not downward IC constraints on the costly component. Next, we show that positive correlation of preferences then converts the IC constraints on the costly component to those on the productive component without changing the direction. Finally, we show that the set of downward IC constraints is sufficient for any standard one-dimensional screening problem satisfying a single-crossing condition. Therefore, costly instruments cannot help the principal when the agent's preferences are positively correlated between the two components.

Armed with this understanding, we have also shown how additional results follow naturally. With negatively correlated preferences, we show a partial converse. With multiplicatively separable preferences within each component, we show a stronger result in terms of the marginal rates of substitution between the two components. Using the perspective of screening with costly instruments, we also show new results even in multidimensional screening models *without* any costly instruments.

Our model assumes that the principal has full commitment power. However, this assumption can be somewhat relaxed as a consequence of our result. In particular, consider a game in which the agent takes a costly action $y \in \mathcal{Y}$ first, and then the principal contracts with the agent on the productive allocations and transfers with commitment after observing the agent's costly action. Suppose the actions are costly to the agent, i.e. $u^B \leq 0$. Then note that the optimal contract screening only the productive component can be supported in a perfect Bayesian equilibrium of this game by letting the principal's belief be the prior γ at every history $y \in \mathcal{Y}$.

We view this paper as a starting point toward a systematic understanding of how costly instruments interact with the traditional design of optimal contracts. As we have demonstrated, the direction of binding incentive constraints, when no costly instrument is used, plays a key role in determining whether costly instruments should be used at all. In models different from the monopolistic screening model we consider, the binding incentive constraints may very well be different. However, our insights may still be helpful, just as we have seen with competition in labor market screening.

A Omitted Proofs

Proof of Proposition 1. Without loss of generality, we may assume $i = 1$. Let $v^A = v^B = 0$. Because θ^0 has a continuous distribution, there exists some constant m^0 such that

$$\mathbb{P}(\theta^0 > m^0) = \mathbb{P}(\theta^0 \leq m^0) = \frac{1}{2}.$$

Similarly define m^1 for θ^1. Since θ^1 is stochastically nonincreasing in θ^0, we have θ^1 and $-\theta^0$ are positively upper orthant dependent (see e.g. Müller and Stoyan (2002), pp. 121-125), and hence

$$\mathbb{P}(-\theta^0 > -m^0, \theta^1 > m^1) \geq \mathbb{P}(-\theta^0 > -m^0)\mathbb{P}(\theta^1 > m^1) = \frac{1}{4}.$$

Because $\beta(\theta^0)$, $\beta(\theta^1)$ are not independent, we have

$$\mathbb{P}(\theta^0 < m^0, \theta^1 > m^1) > \frac{1}{4}$$

and thus

$$\mathbb{P}(\theta^0 > m^0, \theta^1 < m^1) > \frac{1}{4}.$$

Define

$$f(\theta^0) = \begin{cases} 1 & \text{if } \theta^0 \leq m^0 \\ 2 & \text{if } \theta^0 > m^0 \end{cases}, \qquad g(\theta^1) = \begin{cases} -1 & \text{if } \theta^1 \leq m^1 \\ -\epsilon & \text{if } \theta^1 > m^1 \end{cases},$$

where $\epsilon > 0$ will be determined shortly. Let $\tilde{f}$ be a continuous approximation of f such that $\tilde{f}(\theta^0) = f(\theta^0)$ for all $\theta^0 \notin (m^0 - \epsilon, m^0 + \epsilon)$. It is clear that we may select $\tilde{f}$ to be nondecreasing. Let $x_0 = \min \mathcal{X}$ and $\hat{x} = \max \mathcal{X}$. Since $|\mathcal{X}| > 1$, $\hat{x} \neq x_0$. Since $|\mathcal{Y}| > 1$, there exists some $\hat{y} \neq y_0 \in \mathcal{Y}$. Now let

$$u^A(x, \theta^A) = \tilde{f}(\theta^0)\frac{x - x_0}{\hat{x} - x_0}, \qquad u^B(y, \theta^B) = g(\theta^1)\mathbb{1}_{y \neq y_0}.$$

This construction gives admissible utility functions. Consider offering the following menu of three options:

$$\Big\{(\hat{x}, y_0, 2 - \epsilon), (\hat{x}, \hat{y}, 1 - \epsilon), (x_0, y_0, 0)\Big\}.$$

Let the agent choose among these, breaking tie in favor of the principal. This yields a payoff of at least

$$r(\epsilon) := (1 - \epsilon)\mathbb{P}(\theta^1 > m^1) + (2 - \epsilon)\mathbb{P}(\theta^0 \geq m^0 + \epsilon, \theta^1 \leq m^1)$$

for the principal. Screening the productive component alone yields a payoff of at most

$$q(\epsilon) := 2\mathbb{P}(m^0 - \epsilon \leq \theta^0 \leq m^0 + \epsilon) + 1$$

for the principal. Note that $r(\epsilon), q(\epsilon)$ are both continuous on $(0, \frac{1}{2})$, and

$$\lim_{\epsilon \downarrow 0} r(\epsilon) = \frac{1}{2} + 2\mathbb{P}(\theta^0 > m^0, \theta^1 < m^1) > 1 = \lim_{\epsilon \downarrow 0} q(\epsilon).$$

Thus, there exists some $\epsilon^* > 0$ such that $r(\epsilon^*) > q(\epsilon^*)$. With this choice of ϵ^*, the above construction then gives admissible utility functions such that the menu of three options strictly dominates any mechanism screening only the productive component. □

Proof of Lemma 3. Write out $[i \to j]$ and $[j \to k]$:

$$u(x_i, \theta_i) - t_i \geq u(x_j, \theta_i) - t_j;$$

$$u(x_j, \theta_j) - t_j \geq u(x_k, \theta_j) - t_k.$$

Adding these two yields

$$u(x_i, \theta_i) - t_i + u(x_j, \theta_j) - t_j \geq u(x_j, \theta_i) - t_j + u(x_k, \theta_j) - t_k.$$

Hence,

$$u(x_i,\theta_i)-t_i \ge (u(x_j,\theta_i)+u(x_k,\theta_j)-u(x_j,\theta_j))-t_k.$$

Using $x_j \ge x_k$, $\theta_i > \theta_j$, and the strict increasing differences property of u, we have

$$u(x_j,\theta_i)+u(x_k,\theta_j)-u(x_j,\theta_j) \ge u(x_k,\theta_i).$$

Thus $[i \to k]$ follows. □

Proof of Lemma 4. Write out the binding constraints $[i \to k]$ and $[j \to k]$:

$$u(x_i,\theta_i)-t_i = u(x_k,\theta_i)-t_k;$$

$$u(x_j,\theta_j)-t_j = u(x_k,\theta_j)-t_k.$$

Subtracting these two yields

$$u(x_i,\theta_i)-u(x_j,\theta_j)-t_i = u(x_k,\theta_i)-u(x_k,\theta_j)-t_j.$$

Hence,

$$u(x_i,\theta_i)-t_i = (u(x_j,\theta_j)+u(x_k,\theta_i)-u(x_k,\theta_j))-t_j.$$

Using $x_k \ge x_j$, $\theta_i > \theta_j$, and the strict increasing differences property of u, we have

$$u(x_j,\theta_j)+u(x_k,\theta_i)-u(x_k,\theta_j) \ge u(x_j,\theta_i).$$

Thus $[i \to j]$ follows. □

Proof of Lemma 5. Write out the binding constraint $[(i+1) \to i]$:

$$u(x_{i+1},\theta_{i+1})-t_{i+1} = u(x_i,\theta_{i+1})-t_i.$$

By monotonicity of x and the strict increasing differences property of u, we have

$$u(x_{i+1},\theta_{i+1})-u(x_i,\theta_{i+1}) \ge u(x_{i+1},\theta_i)-u(x_i,\theta_i).$$

Therefore,

$$t_{i+1}-t_i \ge u(x_{i+1},\theta_i)-u(x_i,\theta_i),$$

which is equivalent to $[i \to (i+1)]$. Now suppose $[i \to j]$ and $[j \to k]$ hold, with $i < j < k$. As in the proof of lemma 3, adding these two gives

$$u(x_i, \theta_i) - t_i + u(x_j, \theta_j) - t_j \geq u(x_j, \theta_i) - t_j + u(x_k, \theta_j) - t_k .$$

Hence,

$$u(x_i, \theta_i) - t_i \geq (u(x_j, \theta_i) + u(x_k, \theta_j) - u(x_j, \theta_j)) - t_k .$$

Using $x_k \geq x_j$, $\theta_j > \theta_i$, and the strict increasing differences property of u, we have

$$u(x_j, \theta_i) + u(x_k, \theta_j) - u(x_j, \theta_j) \geq u(x_k, \theta_i).$$

Thus $[i \to k]$ follows. It is now immediate that all upward incentive constraints hold. □

Proof of Lemma 6. We maintain the notation of Step 4 in Section 4. Without loss, let $\Theta \subseteq [0,1)$ and $0 \in \Theta$. The construction works as follows. Fix any $n \in \mathbb{N}$. Partition $[0,1)$ into intervals $\{[\frac{i-1}{n}, \frac{i}{n})\}_{i=1,\ldots,n}$. Let

$$I = \left\{ i : \mu([\frac{i-1}{n}, \frac{i}{n})) > 0 \right\}.$$

For any $i \in I$, let

$$\theta_i^{(n)} = \min \left\{ [\frac{i-1}{n}, \frac{i}{n}) \cap \Theta \right\}.$$

(The minimum is attained since Θ is compact.) For notational convenience, we reindex i so that it runs over from 1 to $|I|$. Let

$$\Theta^{(n)} = \{\theta_i^{(n)}\}_{i \in I} ;$$

$$\mu^{(n)}(\theta_i^{(n)}) = \mu([\theta_i^{(n)}, \theta_{i+1}^{(n)})).$$

We have $\Theta^{(n)} \subseteq \Theta$ finite and $\mu^{(n)} \in \Delta(\Theta^{(n)})$ full support. Note that

$$\mu(\{\theta \in \Theta : \theta \in [\theta_i^{(n)}, \theta_{i+1}^{(n)}) \text{ and } |\theta - \theta_i^{(n)}| > \frac{1}{n}\}) = 0. \tag{A.1}$$

We first show property (ii) in the statement. Recall that for this lemma we assume v is Lipschitz continuous on $\mathcal{X} \times \Theta$. Then, there exists some constant $K > 0$ such that for any $\theta, \theta' \in \Theta$,

$$\max_{x \in \mathcal{X}} |v(x, \theta') - v(x, \theta)| \leq K |\theta' - \theta|. \tag{A.2}$$

Let $(x^{(n)}, t^{(n)})$ be any optimal solution to the full IC program (6)† with $(\Theta^{(n)}, \mu^{(n)})$. Let $\bar{x}^{(n)}$ be the extension of $x^{(n)}$ to the right:

$$\bar{x}^{(n)}(\theta) = x^{(n)}(\theta_i^{(n)}) \quad \text{for all } \theta \in [\theta_i^{(n)}, \theta_{i+1}^{(n)}).$$

Note that $\bar{x}^{(n)}$ is a monotonic function on $[0,1)$. Define $\bar{t}^{(n)}$ in the same way. We claim $(\bar{x}^{(n)}, \bar{t}^{(n)})$, when restricted to Θ, is a feasible solution to (6)† with (Θ, μ). To see this, offer the menu $\{(x_i^{(n)}, t_i^{(n)})\}_{i \in I}$ to all types in Θ. Type $\theta_{i+1}^{(n)}$ is indifferent between $(x_{i+1}^{(n)}, t_{i+1}^{(n)})$ and $(x_i^{(n)}, t_i^{(n)})$. Type $\theta_i^{(n)}$ finds $(x_i^{(n)}, t_i^{(n)})$ optimal. Therefore, any type θ between $\theta_i^{(n)}$ and $\theta_{i+1}^{(n)}$ finds $(x_i^{(n)}, t_i^{(n)})$ optimal since u has strict increasing differences. By construction,

$$V(\Theta^{(n)}, \mu^{(n)}) = \mathbb{E}^{\mu^{(n)}}[v(\bar{x}^{(n)}(\theta), \theta) + \bar{t}^{(n)}(\theta)].$$

Since $(\bar{x}, \bar{t})$ is feasible to (6)† with (Θ, μ), we have

$$V(\Theta, \mu) \geq \mathbb{E}^{\mu}[v(\bar{x}^{(n)}(\theta), \theta) + \bar{t}^{(n)}(\theta)].$$

Because $\bar{x}^{(n)}, \bar{t}^{(n)}$ are constant over each interval $[\theta_i^{(n)}, \theta_{i+1}^{(n)})$, by (A.1) and (A.2), we have

$$\begin{aligned}
&\left| \mathbb{E}^{\mu}[v(\bar{x}^{(n)}(\theta), \theta) + \bar{t}^{(n)}(\theta)] - \mathbb{E}^{\mu^{(n)}}[v(\bar{x}^{(n)}(\theta), \theta) + \bar{t}^{(n)}(\theta)] \right| \\
&= \left| \int v(\bar{x}^{(n)}(\theta), \theta) d\mu - \int v(\bar{x}^{(n)}(\theta), \theta) d\mu^{(n)} \right| \\
&\leq \sum_{i \in I} \mu^{(n)}(\theta_i^{(n)}) \sup_{\theta \in [\theta_i^{(n)}, \theta_i^{(n)} + \frac{1}{n}] \cap \Theta} \left\{ \max_{x \in \mathcal{X}} \left| v(x, \theta) - v(x, \theta_i^{(n)}) \right| \right\} \\
&\leq \sum_{i \in I} \mu^{(n)}(\theta_i^{(n)}) \frac{K}{n} = \frac{K}{n}.
\end{aligned} \tag{A.3}$$

Then, it follows that

$$V(\Theta, \mu) \geq V(\Theta^{(n)}, \mu^{(n)}) - \frac{K}{n}.$$

Taking lim sup on both sides gives property (ii) in the statement.

We now show property (i) in the statement. It suffices to prove the weak convergence in $\Delta([0,1])$. Let F, $F^{(n)}$ be the CDFs of μ, $\mu^{(n)}$. We have $F^{(n)}(1) = F(1) = 1$. Fix any $\theta \in [0,1)$. Note that $\mu^{(n)} \preceq \mu$ in the stochastic dominance order, and hence

$$F^{(n)}(\theta) \geq F(\theta). \tag{A.4}$$

Let i be such that $[\theta_i^{(n)}, \theta_{i+1}^{(n)}) \ni \theta$. Note that

$$F^{(n)}(\theta) = \mu^{(n)}([0,\theta]) \le \mu^{(n)}([0,\theta_{i+1}^{(n)})) = \mu([0,\theta_{i+1}^{(n)})).$$

If $\theta + \frac{1}{n} \ge \theta_{i+1}^{(n)}$, then we have

$$\mu([0,\theta_{i+1}^{(n)})) \le F(\theta + \frac{1}{n}).$$

Otherwise, since $\theta + \frac{1}{n} \ge \theta_i^{(n)} + \frac{1}{n}$, we have $\mu([\theta + \frac{1}{n}, \theta_{i+1}^{(n)})) = 0$. Thus,

$$\mu([0,\theta_{i+1}^{(n)})) = \mu([0,\theta + \frac{1}{n})) \le F(\theta + \frac{1}{n}).$$

Hence, in either case, we have

$$F^{(n)}(\theta) \le F(\theta + \frac{1}{n}). \tag{A.5}$$

Using (A.4), (A.5), and that F is right-continuous, we have

$$F(\theta) \le \lim_{n\to\infty} F^{(n)}(\theta) \le \lim_{n\to\infty} F(\theta + \frac{1}{n}) = F(\theta).$$

Therefore, $F^{(n)}$ converges to F pointwise, and hence $\mu^{(n)} \to_w \mu$. □

Proof of Lemma 7. Recall $\mathcal{M}(\Theta)$ is the set of IC and IR mechanisms for the one-dimensional type space Θ. We want to show the following program has a solution:

$$\sup_{(x,t)\in\mathcal{M}(\Theta)} \mathbb{E}[v(x(\theta),\theta) + t(\theta)].$$

We first show that it is without loss to restrict the range of t to some interval $[-K, K]$ for K large enough. By the IR constraints, we have $t(\theta) \le \max_{x,\theta} |u(x,\theta)|$. By the IC constraints, for any θ, θ', we have

$$|t(\theta) - t(\theta')| \le 2\max_{x,\theta} |u(x,\theta)|.$$

Hence, for all θ,

$$t(\theta) \ge -3\max_{x,\theta} |u(x,\theta)| - 2\max_{x,\theta} |v(x,\theta)|, \tag{A.6}$$

because if the above is violated at any type θ, the principal gets strictly less than

$$-\max_{x,\theta} |u(x,\theta)| - \max_{x,\theta} |v(x,\theta)|$$

but that can be easily obtained by offering a single option. Thus, the claim holds for $K = 3\max_{x,\theta}|u(x,\theta)| + 2\max_{x,\theta}|v(x,\theta)|$.

Then, $\mathcal{M}(\Theta) \subseteq \mathcal{X}^\Theta \times [-K,K]^\Theta$ (with the product topology); we use the notation $\mathcal{X}^\Theta := \times_{\theta\in\Theta}\mathcal{X}$. By the dominated convergence theorem, the objective is sequentially continuous on $\mathcal{M}(\Theta)$. It is clear that $\mathcal{M}(\Theta)$ is nonempty. The existence result follows once we show $\mathcal{M}(\Theta)$ is sequentially compact. Fix any sequence $\{(x^{(n)}, t^{(n)})\}_n$ in $\mathcal{M}(\Theta)$. Let

$$U^{(n)}(\theta) = u(x^{(n)}(\theta), \theta) - t^{(n)}(\theta)$$

be the equilibrium payoff of type θ. For any $\hat{\theta} < \theta$, by IC$[\theta \to \hat{\theta}]$, we have

$$U^{(n)}(\hat{\theta}) = u(x^{(n)}(\hat{\theta}), \hat{\theta}) - t^{(n)}(\hat{\theta}) \le u(x^{(n)}(\hat{\theta}), \theta) - t^{(n)}(\hat{\theta}) \le U^{(n)}(\theta).$$

Therefore, $U^{(n)} \in [-K,K]^\Theta$ is a monotone function (increase K if necessary). Since u has strict increasing differences, $x^{(n)} \in \mathcal{X}^\Theta$ is also a monotone function. Note that $\Theta, \mathcal{X} \subset \mathbb{R}$ are linearly ordered and sequentially compact sets. By the Helly's selection theorem for monotone functions on linearly ordered sets (Fuchino and Plewik 1999, Theorem 7), there exists a subsequence $\{x^{(n_k)}\}$ that converges pointwise. Applying the same theorem again on $\{U^{(n_k)}\}$, we obtain a subsubsequence $\{U^{(n_{k_l})}\}$ that converges pointwise. Therefore,

$$t^{(n_{k_l})}(\theta) = u(x^{(n_{k_l})}(\theta), \theta) - U^{(n_{k_l})}(\theta)$$

also converges pointwise by continuity of u. Thus, there exists some $(x^*, t^*) \in \mathcal{X}^\Theta \times [-K,K]^\Theta$ such that

$$(x^{(n_{k_l})}, t^{(n_{k_l})}) \to (x^*, t^*)$$

in the product topology. Being the pointwise limit of measurable real-valued functions, x^* is measurable; so is t^*. Moreover, for any $\theta, \hat{\theta} \in \Theta$,

$$\begin{aligned} u(x^*(\theta), \theta) - t^*(\theta) &= \lim_{l\to\infty}\left(u(x^{(n_{k_l})}(\theta), \theta) - t^{(n_{k_l})}(\theta)\right) \\ &\ge \lim_{l\to\infty}\left(u(x^{(n_{k_l})}(\hat{\theta}), \theta) - t^{(n_{k_l})}(\hat{\theta})\right) = u(x^*(\hat{\theta}), \theta) - t^*(\hat{\theta}) \end{aligned}$$

by continuity of u and that $(x^{(n)}, t^{(n)}) \in \mathcal{M}(\Theta)$ for all n. Therefore, (x^*, t^*) satisfies all IC constraints. Similarly, (x^*, t^*) satisfies all IR constraints. So $(x^*, t^*) \in \mathcal{M}(\Theta)$, and hence $\mathcal{M}(\Theta)$ is sequentially compact. □

Completion of Proof of Proposition 2. We complete the proof of Proposition 2 by filling in the details of Step 4 in Section 4.

Recall that we want to show the optimal value of (6) equals to $V(\Theta,\mu)$. We first show it for Lipschitz continuous v, and then extend it to all continuous v. Without loss, we assume $0 \in \Theta \subseteq [0,1]$. Suppose for contradiction that there exist some $(\hat{x},\hat{t})$ feasible for (6) and some $\epsilon > 0$ such that

$$V(\Theta,\mu) + \epsilon \le \mathbb{E}^{\mu}[v(\hat{x}(\theta),\theta) + \hat{t}(\theta)]. \tag{A.7}$$

Let $\bar{S} = 3\max_{x,\theta}|u(x,\theta)| + 3\max_{x,\theta}|v(x,\theta)|$. By Lusin's theorem (see e.g. Aliprantis and Border 2006, Theorem 12.8), there exists a compact set $\tilde{\Theta} \subseteq \Theta$ such that $\hat{x},\hat{t}$ are continuous on $\tilde{\Theta}$ and $\alpha := \mu(\Theta\backslash\tilde{\Theta}) < \epsilon/(3\bar{S})$. Since $\tilde{\Theta}$ is compact, $\underline{\tilde{\Theta}} := \min\{\tilde{\Theta}\}$ is attained. If $\underline{\tilde{\Theta}} > 0$, we augment $\tilde{\Theta}$ by adding $\theta = 0$. Since $\{0\}$ is a singleton disjoint from the compact set $\tilde{\Theta}$, we have $\hat{x},\hat{t}$ continuous on the augmented set as well. Since $(\hat{x},\hat{t})$ is IR, $\hat{t}(\theta) \le \max_{x,\theta}|u(x,\theta)|$, and hence

$$\mathbb{E}^{\mu}[v(\hat{x}(\theta),\theta) + \hat{t}(\theta)] \le (1-\alpha)\,\mathbb{E}^{\tilde{\mu}}[v(\hat{x}(\theta),\theta) + \hat{t}(\theta)] + \alpha\bar{S}, \tag{A.8}$$

where $\tilde{\mu}$ is the distribution of θ conditional on $\theta \in \tilde{\Theta}$. We pick an approximation sequence $\{(\Theta^{(n)},\mu^{(n)})\}$ for $(\tilde{\Theta},\tilde{\mu})$ according to Lemma 6. By (A.3), for all n large enough, we have

$$\mathbb{E}^{\mu^{(n)}}[v(\bar{x}^{(n)}(\theta) + \bar{t}^{(n)}(\theta)] - \frac{\epsilon}{3(1-\alpha)} \le \mathbb{E}^{\tilde{\mu}}[v(\bar{x}^{(n)}(\theta) + \bar{t}^{(n)}(\theta)], \tag{A.9}$$

where $(x^{(n)},t^{(n)})$ is an optimal solution to the full IC problem $(6)^{\dagger}$ with $(\Theta^{(n)},\mu^{(n)})$, and $(\bar{x}^{(n)},\bar{t}^{(n)})$ is the extension of $(x^{(n)},t^{(n)})$ to the right, as defined in the proof of Lemma 6. As in the proof of Lemma 6, $(\bar{x}^{(n)},\bar{t}^{(n)})$ satisfies all IC and IR constraints for type space Θ. As in the proof of Lemma 7, (A.6) then holds for $\bar{t}^{(n)}$. By feasibility, (A.6), and (A.9), we have

$$\begin{aligned}
V(\Theta,\mu) &\ge \mathbb{E}^{\mu}[v(\bar{x}^{(n)}(\theta),\theta) + \bar{t}^{(n)}(\theta)] \\
&\ge (1-\alpha)\,\mathbb{E}^{\tilde{\mu}}[v(\bar{x}^{(n)}(\theta),\theta) + \bar{t}^{(n)}(\theta)] - \alpha\bar{S} \\
&\ge (1-\alpha)\,\mathbb{E}^{\mu^{(n)}}[v(\bar{x}^{(n)}(\theta),\theta) + \bar{t}^{(n)}(\theta)] - \frac{2}{3}\epsilon \\
&= (1-\alpha)V(\Theta^{(n)},\mu^{(n)}) - \frac{2}{3}\epsilon \\
&\ge (1-\alpha)\,\mathbb{E}^{\mu^{(n)}}[v(\hat{x}(\theta),\theta) + \hat{t}(\theta)] - \frac{2}{3}\epsilon.
\end{aligned}$$

In the last inequality, we have used that $(\hat{x},\hat{t})$ is a downward IC and IR mechanism for $(\Theta^{(n)},\mu^{(n)})$ and that Proposition 2 holds for finite type spaces (see Step 3 in Section 4).

Because $\hat{x}, \hat{t}$ are bounded and continuous on $\tilde{\Theta}$, and v is continuous on the compact space $\mathcal{X} \times \Theta$, we have $v(\hat{x}(\theta), \theta) + \hat{t}(\theta)$ is bounded and continuous on $\tilde{\Theta}$. But, since $\mu^{(n)} \to_w \tilde{\mu}$ in $\Delta(\tilde{\Theta})$, taking limits on both sides of the above and using (A.8), we see that

$$\begin{aligned} V(\Theta, \mu) &\geq (1-\alpha)\,\mathbb{E}^{\tilde{\mu}}[v(\hat{x}(\theta), \theta) + \hat{t}(\theta)] - \frac{2}{3}\epsilon \\ &\geq \mathbb{E}^{\mu}[v(\hat{x}(\theta), \theta) + \hat{t}(\theta)] - \alpha \bar{S} - \frac{2}{3}\epsilon \\ &> \mathbb{E}^{\mu}[v(\hat{x}(\theta), \theta) + \hat{t}(\theta)] - \epsilon, \end{aligned}$$

which is a direct contradiction to (A.7).

Now we let v be any continuous function on $\mathcal{X} \times \Theta$. Since $\mathcal{X} \times \Theta$ is compact, as a consequence of the Stone–Weierstrass theorem (see e.g. Aliprantis and Border (2006), Theorem 9.13), the set of Lipschitz continuous real-valued functions on $\mathcal{X} \times \Theta$ is dense in the space of continuous functions on $\mathcal{X} \times \Theta$ (with the sup norm). Therefore, there exists a sequence of Lipschitz continuous functions $\{v_k\}$ converging uniformly to v. Passing to a subsequence if necessary, we may assume that for all k,

$$\sup_{x \in \mathcal{X}, \theta \in \Theta} |v_k(x, \theta) - v(x, \theta)| < \frac{1}{k}.$$

Using the above and the earlier result applied to v_k, we have for all k,

$$\begin{aligned} \sup_{(x,t) \in \tilde{\mathcal{M}}(\Theta)} \mathbb{E}\Big[v(x(\theta), \theta) + t(\theta)\Big] - \frac{1}{k} &\leq \sup_{(x,t) \in \tilde{\mathcal{M}}(\Theta)} \mathbb{E}\Big[v_k(x(\theta), \theta) + t(\theta)\Big] \\ &\leq \sup_{(x,t) \in \mathcal{M}(\Theta)} \mathbb{E}\Big[v_k(x(\theta), \theta) + t(\theta)\Big] \leq \sup_{(x,t) \in \mathcal{M}(\Theta)} \mathbb{E}\Big[v(x(\theta), \theta) + t(\theta)\Big] + \frac{1}{k}. \end{aligned}$$

Taking $k \to \infty$ then gives the desired inequality.

Invoking the existence result of Lemma 7, we conclude the proof of Proposition 2. □

References

Acemoglu, D. and A. Wolitzky (2011): "The Economics of Labor Coercion," *Econometrica*, 79(2), 555–600.

Adams, W. J. and J. L. Yellen (1976): "Commodity Bundling and the Burden of Monopoly," *Quarterly Journal of Economics*, 90(3), 475–498.

Akbarpour, M., P. Dworczak, and S. D. Kominers (2020): "Redistributive Allocation Mechanisms," Available at SSRN: http://dx.doi.org/10.2139/ssrn.3609182.

Aliprantis, C. D. and K. C. Border (2006): *Infinite Dimensional Analysis: A Hitchhiker's Guide*, Berlin: Springer.

Armstrong, M. (1996): "Multiproduct Nonlinear Pricing," *Econometrica*, 64(1), 51–75.

Azar, J., I. Marinescu, M. Steinbaum, and B. Taska (2020): "Concentration in US Labor Markets: Evidence from Online Vacancy Data," *Labour Economics*, 66, 101886.

Babaioff, M., N. Immorlica, B. Lucier, and S. M. Weinberg (2014): "A Simple and Approximately Optimal Mechanism for an Additive Buyer," in *Proceedings of the IEEE 55th Annual Symposium on Foundations of Computer Science*, FOCS '14, 21–30.

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

Bergemann, D., A. Bonatti, A. Haupt, and A. Smolin (2021): "The Optimality of Upgrade Pricing," arXiv:2107.10323 [econ.TH].

Cai, Y., N. R. Devanur, and S. M. Weinberg (2016): "A Duality-based Unified Approach to Bayesian Mechanism Design," in *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing*, STOC '16, 926–939.

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

Chakravarty, S. and T. R. Kaplan (2013): "Optimal Allocation without Transfer Payments," *Games and Economic Behavior*, 77(1), 1–20.

Che, Y.-K. and W. Zhong (2021): "Robustly-Optimal Mechanism for Selling Multiple Goods," arXiv:2105.02828 [econ.TH].

Condorelli, D. (2012): "What Money Can't Buy: Efficient Mechanism Design with Costly Signals," *Games and Economic Behavior*, 75(2), 613–624.

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

Deb, R. and A.-K. Roesler (2021): "Multi-Dimensional Screening: Buyer-Optimal Learning and Informational Robustness," arXiv:2105.12304 [econ.TH].

Dube, A., J. Jacobs, S. Naidu, and S. Suri (2020): "Monopsony in Online Labor Markets," *American Economic Review: Insights*, 2(1), 33–46.

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

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

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

Hart, S. and N. Nisan (2017): "Approximate Revenue Maximization with Multiple Items," *Journal of Economic Theory*, 172, 313–347.

Hartline, J. D. and T. Roughgarden (2008): "Optimal Mechanism Design and Money Burning," in *Proceedings of the 40th Annual ACM SIGACT Symposium on Theory of Computing*, STOC '2008, 75–84.

Kamae, T. and U. Krengel (1978): "Stochastic Partial Ordering," *Annals of Probability*, 6(6), 1044–1049.

Klenke, A. (2013): *Probability Theory: A Comprehensive Course* (Second Ed.), Germany: Springer London.

Koçyiğit, Ç., f. N. Rujeerapaiboon, and D. Kuhn (2021): "Robust Multidimensional Pricing: Separation without Regret," *Mathematical Programming*, 1–34.

Laffont, J.-J. and J. Tirole (1986): "Using Cost Observation to Regulate Firms," *Journal of Political Economy*, 94(3), 614–641.

——— (1993): *A Theory of Incentives in Procurement and Regulation*, Cambridge, MA: MIT Press.

Li, X. and A. C.-C. Yao (2013): "On Revenue Maximization for Selling Multiple Independently Distributed Items," *Proceedings of the National Academy of Sciences*, 110(28), 11232–11237.

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

Malladi, S. (2020): "Delegated Screening and Robustness," Available at SSRN: https://ssrn.com/abstract=3740206.

Milgrom, P. and C. Shannon (1994): "Monotone Comparative Statics," *Econometrica*, 62(1), 157–180.

Milgrom, P. R. and R. J. Weber (1982): "A Theory of Auctions and Competitive Bidding," *Econometrica*, 50(5), 1089–1122.

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

Mussa, M. and S. Rosen (1978): "Monopoly and Product Quality," *Journal of Economic Theory*, 18(2), 301–317.

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

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

Prager, E. and M. Schmitt (2021): "Employer Consolidation and Wages: Evidence from Hospitals," *American Economic Review*, 111(2), 397–427.

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.

Spence, M. (1973a): "Job Market Signaling," *Quarterly Journal of Economics*, 87(3), 355–374.

——— (1973b): "Time and Communication in Economic and Social Interaction," *Quarterly Journal of Economics*, 87(4), 651–660.

——— (1974): "Competitive and Optimal Responses to Signals: An Analysis of Efficiency and Distribution," *Journal of Economic Theory*, 7(3), 296–332.

Stiglitz, J. E. (2002): "Information and the Change in the Paradigm in Economics," *American Economic Review*, 92(3), 460–501.

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

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

Zeckhauser, R. (2021): "Strategic Sorting: The Role of Ordeals in Health Care," *Economics & Philosophy*, 37(1), 64–81.

B Online Appendix

B.1 Competitive Screening

Our main model assumes monopolistic screening. It delivers a prediction different from the usual perception of costly screening in competitive labor markets (see Remark 6). In this appendix, we formulate a stylized competitive screening model consisting of two screening devices and show how competition can reverse the use of costly instruments.

There are two types of workers $\theta_H > \theta_L \geq 0$ in a perfectly competitive labor market.[26] A type-θ_i worker incurs a cost $\psi_i(x)$ for producing $x \in [0,1]$ units of work where ψ_i is a strictly increasing, continuously differentiable, and strictly convex function on $[0,1]$ with $\psi_i(0) = 0$. A firm gets a payoff $\theta_i x$ from x units of work by a type-θ_i worker.

Suppose the marginal cost is lower for the higher type: $\psi_H'(x) < \psi_L'(x)$ for all $x \in [0,1]$. The efficient amount of production for type θ_i is $x_i^e := (\psi_i')^{-1}(\theta_i)$, assumed to be in the interior of $[0,1]$. Suppose that

$$\theta_L x_L^e - \psi_L(x_L^e) < \theta_H x_H^e - \psi_L(x_H^e)$$

so the low type wants to imitate the high type when given the menu of the efficient allocations with competitive prices. Without this assumption, there is no adverse selection problem. Suppose also there exists some $x \geq x_L^e$ such that $\theta_L x_L^e - \psi_L(x_L^e) \geq \theta_H x - \psi_L(x)$ so it is possible to separate the types using only the work allocations.

There is one costly instrument. For a level $y \in [0,1]$ of the costly activity, a type-θ_i worker incurs a cost $c_i(y)$ where c_i is a strictly increasing, continuously differentiable function on $[0,1]$ with $c_i(0) = 0$. Suppose $c_L'(0) > \psi_L'(1)$ and $c_H'(0) = 0$. This says that a small amount of y costs nothing for the high type but a lot for the low type.

The firms commit to a set of offers. Each offer specifies an amount of work x, a level of costly activity y, and a wage w. The literature has not reached a consensus on the choice of solution concept for competitive screening models. We say a set of offers $\{(x, y, w)\}$ is a *separating set* if (i) the types separate and (ii) the firms earn zero payoff on each offer. A set of offers is a *Pareto-optimal separating set* if it is (constrained) Pareto-optimal among all separating sets. This solution concept is weaker than the Pareto-dominant separating set, which is known to be equivalent to the reactive equilibrium of Riley (1979) in settings with one screening device (Engers and Fernandez, 1987).

This competitive screening model is analogous to the labor market application in Section 6.2. However, costly screening now emerges in equilibrium:

[26]This part of the setup is standard; see e.g. Spence (1978) and Stantcheva (2014).

Proposition 5. *A Pareto-optimal separating set exists and any Pareto-optimal separating set involves costly screening.*

Proof of Proposition 5. We first prove the second part. Suppose for contradiction that there exists a Pareto-optimal separating set $\{(x, y, w)\}$ that does not involve costly screening ($y = 0$). By the definition of a separating set, x_H, x_L must differ. By the single-crossing property of ψ, we then have $x_H > x_L$. Note that x_H cannot be x_H^e because if so IC$[\theta_L \to \theta_H]$ will be violated:

$$\theta_L x_L - \psi_L(x_L) \le \theta_L x_L^e - \psi_L(x_L^e) < \theta_H x_H^e - \psi_L(x_H^e),$$

where the first inequality holds by definition of x_L^e and the second inequality holds by assumption. Therefore, IC$[\theta_L \to \theta_H]$ must be binding. To see this, note that if the upward IC constraint is not binding, then one can move x_H by small enough δ toward x_H^e without breaking the upward IC constraint. Since the surplus function $\theta_H x - \psi_H(x)$ is strictly concave, the modification increases the payoff of the high type and hence also preserves the downward IC constraint. But this means that the original set of offers is dominated by a separating set and hence impossible.

Since $x_H > x_L$ and the upward IC constraint is binding, the downward IC constraint must be slack by the single-crossing property of ψ. This implies that $x_L = x_L^e$ because otherwise moving x_L slightly toward x_L^e gives a contradiction by the same argument as above. We claim that $x_H > x_H^e$. To see this, let

$$f(x) = (\theta_H x - \psi_L(x)) - (\theta_L x_L^e - \psi_L(x_L^e)).$$

Note that it is concave on $[x_L^e, x_H^e]$. Moreover, $f(x_L^e) = (\theta_H - \theta_L)x_L^e > 0$, and $f(x_H^e) = (\theta_H x_H^e - \psi_L(x_H^e)) - (\theta_L x_L^e - \psi_L(x_L^e)) > 0$ by assumption. Thus, $f(x) > 0$ for all $x \in [x_L^e, x_H^e]$ and hence x_H cannot be in that region. Therefore, $x_H > x_H^e$.

Now consider the menu $\{(x_L, 0, \theta_L x_L), (x_H - \epsilon, \epsilon, \theta_H(x_H - \epsilon))\}$ for $\epsilon > 0$. We claim that for ϵ small enough, the offer $(x_H - \epsilon, \epsilon, \theta_H(x_H - \epsilon))$ increases the payoff of the high type. Let

$$u_H(\epsilon) = \theta_H(x_H - \epsilon) - \psi_H(x_H - \epsilon) - c_H(\epsilon).$$

It is a continuously differentiable function of ϵ. The right derivative of this function at 0 is strictly positive because

$$\partial_+ u_H(0) = -(\theta_H - \partial_- \psi_H(x_H)) - \partial_+ c_H(0) = -(\theta_H - \partial_- \psi_H(x_H)) > 0,$$

where the second equality holds by assumption, and the last inequality holds by strict concavity of ψ_H and that $x_H > x_H^e$. Therefore, there exists some $\epsilon > 0$ such that $u_H'(s) > 0$ for all $s \in [0, \epsilon]$; the claim follows immediately.

We also claim that for $\epsilon > 0$ small enough, the modification still deters the low type from imitating the high type. To see this, let

$$\hat{u}_L(\epsilon) = \theta_H(x_H - \epsilon) - \psi_L(x_H - \epsilon) - c_L(\epsilon).$$

It is a continuously differentiable function of ϵ. The right derivative of this function at 0 is strictly negative because

$$\partial_+ \hat{u}_L(0) = -\theta_H + \partial_- \psi_L(x_H) - \partial_+ c_L(0) \leq \partial_- \psi_L(1) - \partial_+ c_L(0) < 0,$$

where the first inequality uses convexity of ψ_L and the second inequality holds by assumption. Therefore, there exists some $\epsilon > 0$ such that $\hat{u}_L'(s) < 0$ for all $s \in [0, \epsilon]$; the claim follows immediately.

Hence, for $\epsilon > 0$ sufficiently small, the proposed menu is a separating set that Pareto-improves on the original one. Contradiction.

For the first part of the statement, consider the following optimization problem:

$$\max_{(x,y) \in [0,1]^2} \theta_H x - \psi_H(x) - c_H(y)$$

$$\text{subject to} \quad \theta_L x_L^e - \psi_L(x_L^e) \geq \theta_H x - \psi_L(x) - c_L(y).$$

An optimizer (x^*, y^*) exists by standard compactness arguments. Moreover, $y^* \neq 0$ because otherwise it can be strictly improved by the argument above.

We claim that $\{(x_L^e, 0, \theta_L x_L^e), (x^*, y^*, \theta_H x^*)\}$ is a separating set. The low type chooses the first offer by construction. To see that the high type chooses the second offer, recall that by assumption there exists some $x \geq x_L^e$ such that $\theta_L x_L^e - \psi_L(x_L^e) \geq \theta_H x - \psi_L(x)$. Since this inequality is violated at x_L^e, by continuity, there exists some $x_H > x_L^e$ such that $\theta_L x_L^e - \psi_L(x_L^e) = \theta_H x_H - \psi_L(x_H)$. Then, by the single-crossing property of ψ, $\theta_H x_H - \psi_H(x_H) > \theta_L x_L^e - \psi_H(x_L^e)$. Thus,

$$\theta_H x^* - \psi_H(x^*) - c_H(y^*) \geq \theta_H x_H - \psi_H(x_H) > \theta_L x_L^e - \psi_H(x_L^e),$$

where the first inequality uses that $(x_H, 0)$ is a feasible solution. So the high type chooses the second offer.

Observe that $\{(x_L^e, 0, \theta_L x_L^e), (x^*, y^*, \theta_H x^*)\}$ must be Pareto-optimal among all separating sets. Suppose for contradiction that there is a separating set that Pareto-dominates it. Then the separating set must provide strictly higher payoff for the high type and maintain the same payoff for the low type because the low type already gets the maximal payoff. But that is impossible subject to IC$[\theta_L \to \theta_H]$ by the construction of (x^*, y^*). □

B.2 Additional Proofs

In this appendix, we adapt the proof of Theorem 6 in Kamae and Krengel (1978) to prove Lemma 1. We start with a technical lemma.

Lemma 8. *Suppose X, Y are two Θ-valued random variables where Θ is a compact subset of $\mathbb{R}^N$. Let $\preceq_{st}$ and $\preceq'_{st}$ denote the stochastic dominance partial orders on $\Delta(\mathbb{R}^N)$ and $\Delta(\Theta)$ (i.e. $X \preceq'_{st} Y$ if $\mathbb{E}[f(X)] \leq \mathbb{E}[f(Y)]$ for all bounded nondecreasing measurable $f : \Theta \to \mathbb{R}$). Then $X \preceq_{st} Y$ if and only if $X \preceq'_{st} Y$.*

Proof of Lemma 8. ($\Longleftarrow$) Suppose $X \preceq'_{st} Y$. Note that for any bounded monotone measurable $f : \mathbb{R}^N \to \mathbb{R}$, the restriction $f|_\Theta : \Theta \to \mathbb{R}$ is also a bounded monotone measurable function, and moreover $\mathbb{E}[f(X)] = \mathbb{E}[f|_\Theta(X)] \leq \mathbb{E}[f|_\Theta(Y)] = \mathbb{E}[f(Y)]$ since X, Y are Θ-valued and $X \preceq'_{st} Y$. So $X \preceq_{st} Y$.

($\Longrightarrow$) Suppose $X \preceq_{st} Y$. To show $X \preceq'_{st} Y$, by Theorem 1 of Kamae et al. (1977), it suffices to show that for any increasing set $B \subseteq \Theta$ closed in Θ, we have $\mathbb{E}[\mathbb{1}_{X \in B}] \leq \mathbb{E}[\mathbb{1}_{Y \in B}]$ (we say a set B is *increasing* if $\mathbb{1}_B$ is a nondecreasing function).

Fix any such B. Let $B^\uparrow := \{y \in \mathbb{R}^N : y \geq x, x \in B\}$ be the increasing hull of B in $\mathbb{R}^N$. We claim that $B^\uparrow$ is closed in $\mathbb{R}^N$. To see this, fix any $y^n \to y$ in $\mathbb{R}^N$ where $y^n \in B^\uparrow$. Since $y^n \in B^\uparrow$, there exists $x^n \in B$ such that $y^n \geq x^n$. Since B is a closed subset of a compact set Θ, B is compact. Therefore, there exists a subsequence x^{n_l} converging to some $x \in B$. Passing to this subsequence, we have $y = \lim_{l \to \infty} y^{n_l} \geq \lim_{l \to \infty} x^{n_l} = x \in B$ and hence $y \in B^\uparrow$. This proves that $B^\uparrow$ is closed in $\mathbb{R}^N$, and hence measurable.

Because X is Θ-valued, we have $\mathbb{E}[\mathbb{1}_{X \in B^\uparrow}] = \mathbb{E}[\mathbb{1}_{X \in B^\uparrow \cap \Theta}]$. We claim that $B^\uparrow \cap \Theta = B$. Since $B \subseteq \Theta$ and $B \subseteq B^\uparrow$, we have $B \subseteq B^\uparrow \cap \Theta$. Now take any $y \in B^\uparrow \cap \Theta$. Then $y \in \Theta$ and there exists some $x \in B$ such that $y \geq x$. But because B itself is an increasing set in Θ, we must have $y \in B$. Thus $B^\uparrow \cap \Theta \subseteq B$. Therefore, $B^\uparrow \cap \Theta = B$. Now, we have

$$\mathbb{E}[\mathbb{1}_{X \in B}] = \mathbb{E}[\mathbb{1}_{X \in B^\uparrow \cap \Theta}] = \mathbb{E}[\mathbb{1}_{X \in B^\uparrow}] \leq \mathbb{E}[\mathbb{1}_{Y \in B^\uparrow}] = \mathbb{E}[\mathbb{1}_{Y \in B^\uparrow \cap \Theta}] = \mathbb{E}[\mathbb{1}_{Y \in B}]$$

where the inequality follows from that $X \preceq_{st} Y$ and $B^\uparrow$ is a measurable increasing set in $\mathbb{R}^N$. Since this holds for all closed increasing sets B in Θ, the claim follows. □

Proof of Lemma 1. Let $\mathcal{B}_{\Theta^B}$ be the Borel σ-algebra of Θ^B. Let $\kappa : \Theta^A \times \mathcal{B}_{\Theta^B} \to [0,1]$ be the regular conditional distribution of θ^B given θ^A. For any $t \in \Theta^A$, define measure $P_t(\cdot) = \kappa(t,\cdot)$. Let S be the support of θ^A. By assumption, $\{P_t\}_{t\in S}$ is a stochastically non-decreasing family of probability measures on $\mathbb{R}^N$. By Lemma 8, it is also a stochastically nondecreasing family of probability measures on Θ^B.

Let

$$\varphi(s) = \begin{cases} \max\{t : t \le s, t \in S\} & \text{if } s \ge \min(S) \\ \min(S) & \text{otherwise} \end{cases}.$$

Because $(-\infty, s] \cap S$ is compact, we have $\varphi(s) \in S$. For all $s \notin S$, define $P_s = P_{\varphi(s)}$. Because $\varphi(\cdot)$ is nondecreasing and $\varphi(s) = s$ for all $s \in S$, $\{P_t\}_{t\in\mathbb{R}}$ is a stochastically nondecreasing family of probability measures on Θ^B. Invoking Theorem 6 of Kamae and Krengel (1978), we have a Θ^B-valued stochastic process $\{X_t\}_{t\in\mathbb{R}}$ on a probability space (Ω, ν) such that (i) $X_s(\omega) \le X_t(\omega)$ for all $s < t$ and all ω and (ii) P_t is the distribution of X_t for all t.

By the proof of Theorem 6 in Kamae and Krengel (1978), there exists a dense countable set $D \subset \mathbb{R}$ such that for all ω and all $s \notin D$

$$X_s(\omega) = \lim_{t\to s;\, t\in D,\, t\le s} X_t(\omega).$$

Let X_t^i denote the i-th coordinate of X_t. We claim that for all i and all ω, the sample path $X_t^i(\omega)$ is left-continuous at all $t \notin D$. To see this, fix any i, $\omega \in \Omega$, $t \notin D$, and $\epsilon > 0$. By construction, $X_t^i(\omega) = \lim_{k\to\infty} X_{t_k}^i(\omega)$ for some sequence $t_k \uparrow t$, with $t_k \in D$. So there exists some $K \in \mathbb{N}$ such that $X_t^i(\omega) - X_{t_K}^i(\omega) < \epsilon$. But then for any $s \in (t-\delta, t)$ where $\delta := t - t_K$, we have $|X_t^i(\omega) - X_s^i(\omega)| \le X_t^i(\omega) - X_{t_K}^i(\omega) < \epsilon$ by monotonicity of $X_t^i(\omega)$.

Because D is countable, D^c is dense in $\mathbb{R}$. Pick any dense countable set $Q \subset D^c$. For all ω, define $\bar{X}_t(\omega) = X_t(\omega)$ for all $t \in D^c$ and

$$\bar{X}_t(\omega) = \lim_{s\to t;\, s\in Q,\, s\le t} \bar{X}_s(\omega)$$

for all $t \in D$. Note that $\{\bar{X}_t\}_{t\in\mathbb{R}}$ is also a nondecreasing stochastic process. By a similar argument as above, $\bar{X}_t^i(\omega)$ is left-continuous at all $t \in D$. Moreover, for any $t \in D^c$, and any sequence $t_k \uparrow t$, with $t_k \in Q$, we have $\bar{X}_t^i(\omega) = X_t^i(\omega) = \lim_{k\to\infty} X_{t_k}^i(\omega) = \lim_{k\to\infty} \bar{X}_{t_k}^i(\omega)$ by left continuity of $X_t^i(\omega)$ at $t \in D^c$. Therefore, by a similar argument as above, $\bar{X}_t^i(\omega)$ is also left-continuous at all $t \in D^c$. So $\{\bar{X}_t\}_{t\in\mathbb{R}}$ is a left-continuous stochastic process, and thus the map $(t,\omega) \mapsto \bar{X}_t^i(\omega)$ is jointly measurable (see e.g. Karatzas and Shreve 1998, p. 5). Since D is countable, $(t,\omega) \mapsto X_t^i(\omega)\mathbb{1}_{t\in D}$ is also jointly measurable. Then, because $X_t^i(\omega) =$

$\bar{X}_t^i(\omega)\mathbb{1}_{t\notin D} + X_t^i(\omega)\mathbb{1}_{t\in D}$, $(t,\omega)\mapsto X_t^i(\omega)$ is jointly measurable. Therefore, $(t,\omega)\mapsto X_t(\omega)$ is jointly measurable.

Let $\mathcal{E}=\Omega$ and $h(\theta^A;\varepsilon)=X_{\theta^A}(\varepsilon)$ for all $\theta^A\in\Theta^A$ and $\varepsilon\in\mathcal{E}$. Then $h:\Theta^A\times\mathcal{E}\to\Theta^B$ is jointly measurable and nondecreasing in the first argument. Let μ denote the marginal distribution of θ^A. By the construction of h, for any $(a,b)\in\mathbb{R}\times\mathbb{R}^N$, we have

$$\begin{aligned}(\mu\times\nu)(\{(\theta^A,\varepsilon):\theta^A\le a,\, h(\theta^A;\varepsilon)\le b\}) &= \int \mathbb{1}_{\theta^A\le a}\mathbb{1}_{\theta^A\in S}\Big(\int \mathbb{1}_{h(\theta^A;\varepsilon)\le b}\, d\nu(\varepsilon)\Big)d\mu(\theta^A)\\ &= \int \mathbb{1}_{\theta^A\le a}\mathbb{1}_{\theta^A\in S}\,\kappa(\theta^A,\{\theta^B:\theta^B\le b\})\,d\mu(\theta^A)\\ &= \int \mathbb{1}_{\theta^A\le a}\,\kappa(\theta^A,\{\theta^B:\theta^B\le b\})\,d\mu(\theta^A)\\ &= \mathbb{P}(\theta^A\le a,\,\theta^B\le b),\end{aligned}$$

where we have also used $\mu(S)=1$ and Fubini's theorem. Thus, $\theta\overset{d}{=}(\theta^A,h(\theta^A,\varepsilon))$ when θ^A, ε are independently drawn from μ, ν respectively. □

References

Engers, M. and L. Fernandez (1987): "Market Equilibrium with Hidden Knowledge and Self-selection," *Econometrica*, 55(2), 425–439.

Kamae, T. and U. Krengel (1978): "Stochastic Partial Ordering," *Annals of Probability*, 6(6), 1044–1049.

Kamae, T., U. Krengel, and G. L. O'Brien (1977): "Stochastic Inequalities on Partially Ordered Spaces," *Annals of Probability*, 5(6), 899–912.

Karatzas, I. and S. Shreve (1998): *Brownian Motion and Stochastic Calculus* (Second Ed.), New York: Springer-Verlag.

Riley, J. G. (1979): "Informational Equilibrium," *Econometrica*, 47(2), 331–359.

Spence, M. (1978): "Product Differentiation and Performance in Insurance Markets," *Journal of Public Economics*, 10(3), 427–447.

Stantcheva, S. (2014): "Optimal Income Taxation with Adverse Selection in the Labour Market," *Review of Economic Studies*, 81(3), 1296–1329.