Robustness and Separation in Multidimensional Screening

Gabriel Carroll, Stanford University
gdc@stanford.edu

November 4, 2016

Abstract

A principal wishes to screen an agent along several dimensions of private information simultaneously. The agent has quasilinear preferences that are additively separable across the various components. We consider a robust version of the principal's problem, in which she knows the marginal distribution of each component of the agent's type, but does not know the joint distribution. Any mechanism is evaluated by its worst-case expected profit, over all joint distributions consistent with the known marginals. We show that the optimum for the principal is simply to screen along each component separately. This result does not require any assumptions (such as single crossing) on the structure of preferences within each component. The proof technique involves a generalization of the concept of virtual values to arbitrary screening problems. Sample applications include monopoly pricing and a stylized dynamic taxation model.

Keywords: bundling; generalized virtual value; mechanism design; multidimensional screening; robustness; separation.

This paper has greatly benefited from comments from (in random order) Richard Holden, Dawen Meng, Peter Klibanoff, Andy Skrzypacz, Jean Tirole, Alex Wolitzky, Nima Haghpanah, Paul Milgrom, Takuo Sugaya, Drew Fudenberg, Alejandro Manelli, Eric Maskin, Ben Brooks, Iván Werning, Florian Scheuer, Yiqing Xing, Ilya Segal, Juuso Toikka, Itai Sher, Brad Larsen, and Rakesh Vohra; plus conference and seminar audiences at Jiangxi University of Finance and Economics, Becker-Friedman Institute, Simons Institute for Theoretical Computer Science, Northwestern, Western Ontario, Toronto, MIT, Penn State, UCL, Warwick, LSE, Princeton, NSF/CEME Decentralization, and UC San Diego; and three referees.

1 Introduction

Multidimensional screening stands out among current topics in economic theory as a source of problems that are simple to write down, yet analytically intractable. Whereas the canonical one-dimensional screening model is well understood, its multidimensional analogue, as traditionally formulated, is much more complex and unruly.

Consider, for example, the simplest (and most widely studied) version of the problem: A monopolist has two goods, 1 and 2, to sell to a buyer. Marginal costs are zero; the buyer's preferences are quasi-linear and additively separable in the two goods. The monopolist has a prior belief about the joint distribution of the buyer's values for the two goods, (θ^1, θ^2). She wants to find a mechanism for selling the goods so as to maximize expected revenue. She may simply post a price for each good separately; she may also wish to set a price for the bundle of both goods, which may be greater or less than the sum of the two separate prices. But she can also offer probabilistic combinations of goods, say, a 2/3 chance of getting good 1 and also a 3/4 chance of good 2, for yet another price.

The single-good version of the problem is simple: even if probabilistic mechanisms are allowed, the optimum is simply to post a single, take-it-or-leave-it price (Riley and Zeckhauser 1983). Not so in the two-good problem. In the natural case where the values for each good are independent uniform on $[0, 1]$, the optimal mechanism is relatively simple (prices for each good plus a price for the bundle), but known proofs of optimality are involved (Manelli and Vincent 2006). In other cases, the optimum may involve probabilistic bundling, and may even require a menu of infinitely many such bundles (Thanassoulis 2004, Manelli and Vincent 2007, Daskalakis, Deckelbaum, and Tzamos 2013). Moreover, revenue can be non-monotone — moving the distribution of buyer's types upwards (in the stochastic dominance sense) may decrease optimal revenue (Hart and Reny 2015). And these phenomena are not artifacts of the possibilities for correlation in the multidimensional problem, since there are examples even when the values for the two goods are independently distributed. Furthermore, if the values are allowed to be correlated, then it can happen that restricting the seller to any fixed number k of bundles can lead to an arbitrarily small fraction of the optimal revenue (Hart and Nisan 2013), and finding the optimal mechanism is computationally intractable (Conitzer and Sandholm 2002, Daskalakis, Deckelbaum, and Tzamos 2014). With these challenges arising even for the simple monopoly problem, the prospects for more complex screening problems seem even more daunting.

This paper puts forward an alternative and general framework for writing models that may offer a new and complementary perspective, by avoiding these complexities. We consider a principal screening an agent with quasi-linear preferences, who has several components of private information. For each component $g = 1, \dots, G$, the agent will be assigned a (possibly random) allocation x^g, from which he derives a value that depends on his corresponding type θ^g. His total payoff is additively separable across components, $\sum_g u^g(x^g, \theta^g)$. Unlike the traditional model, in which the principal maximizes her expected profit according to a prior distribution over the full type $\theta = (\theta^1, \dots, \theta^G)$, here we assume that the principal only has beliefs about the *marginal* distribution of each component θ^g, but does not know how the various components are correlated with each other. She wishes for a guarantee on her average profit that is robust to this uncertainty. More precisely, she evaluates any possible mechanism according to the worst-case expected profit, over all possible joint distributions that are consistent with the known marginals.

Our main result says that the optimal mechanism separates: for each component g, the principal simply offers the optimal mechanism for that marginal distribution. In the simple monopolist example above, this means that the optimal mechanism is to post a price for each good separately, without any bundling. But the result also allows, for example, that each component g represents a price discrimination problem in which different qualities of product may be offered at different prices, as in Mussa and Rosen (1978). In fact, our result is much more general: In the above examples, each component g represents a standard one-dimensional screening problem, satisfying (for example) single-crossing conditions, but we do not actually require this. Ours is a general separability result, in which the structure of preferences within each component can be totally arbitrary. Further applications will be discussed ahead.

Much of the literature on multidimensional mechanism design has emphasized the advantages of bundling (to use the monopolist terminology) — or, more generally, of creating interactions between the different dimensions of private information. In this context, our result expresses the following simple counterpoint: If you don't know enough to see how to bundle, then don't.

Indeed, a simple intuition behind the result is as follows: The solution to a maxmin problem (such as the principal's problem that we set up here) often involves equalizing payoffs across many possible environments. In this case, the separate mechanism screening does exactly that; since the principal knows the marginal distribution of each θ^g, she can calculate her expected profit, without needing to know anything about how the different components are correlated, whereas any bundling mechanism would be sensitive to the correlation structure. However, this intuition is incomplete. Hypothetically, there could be some bundling mechanism whose exact profit depends on the correlation structure yet is always better than separate screening.

Our proof here works instead by explicitly constructing a joint distribution for which no mechanism performs better than separate screening. The approach first develops a generalization of the concept of virtual values, applicable to any screening problem; as in traditional one-dimensional problems, a type's virtual value for an allocation represents the actual value, minus the shadow costs from the incentives of other types to imitate it, and the optimal mechanism assigns each type the allocation for which its virtual value is the highest. Our joint distribution in the multidimensional type space is then chosen so that the (generalized) virtual value of an allocation equals the sum of the virtual values of its components. This condition is expressed as a system of linear equations that the distribution should satisfy. We show that the system indeed has a solution, and we can give this solution at least a mathematical interpretation, if not an economic one; we derive it as the stationary distribution of a particular Markov process over types.

The main purpose of this of this paper is as a methodological contribution in an area where new models seem to be called for. Screening problems are now pervasive in many areas of economic theory, and in many applications, agents have private information that cannot be conveniently expressed along a single dimension: In a Mirrlees-style tax model, workers may have multiple dimensions of ability, e.g. ability in different kinds of jobs (Rothschild and Scheuer 2013, Rothschild and Scheuer 2014). In regulation of a monopolist with unknown cost structure (Baron and Myerson 1982), it is natural for the costs to be described by more than one parameter, e.g. fixed and marginal costs. In a model of insurance with adverse selection (Rothschild and Stiglitz 1976), consumers may have different probabilities of facing different kinds of adverse events. The survey of Rochet and Stole (2003) (which also describes more applications) puts the point more forcefully: "In most cases that we can think of, a multidimensional preference parameterization seems critical to capturing the basic economics of the environment" (p. 150). In addition, there is a recent growth of interest in dynamic mechanism design, in which information arrives in each period; even if each period's information is single-dimensional, this field presents some of the same analytical challenges as static multidimensional mechanism design (Battaglini and Lamba 2015).

This range of applications highlights the need for general and tractable theory in multidimensional screening. Now, typically, the ultimate interest is in interaction across the dimensions of private information, which is exactly what our model here rules out. But this paper offers a very general way of writing down a model where no interactions in the problem formulation lead to no interactions in the solution. Arguably, it is helpful for theory to work from such a baseline, as a way to understand why and when such interactions do arise in other models, and perhaps, as a blank canvas on which future research can systematically add interactions without losing tractability.

Aside from this methodological contribution, it is also possible to interpret our result more directly as a positive description of behavior. We defer this interpretation, and discussion of related issues, to the conclusion.

This work connects with several branches of literature. It is part of a growing body of work in robust mechanism design, seeking to explain intuitively simple mechanisms as providing guaranteed performance in uncertain environments, and formalizing this intuition by showing how such mechanisms can arise as solutions to worst-case optimization problems. This includes earlier work by this author on contracting problems in uncertain environments (Carroll 2015b, Carroll 2015a, Carroll and Meng 2016), as well as several others (Bergemann and Morris 2005, Brooks 2013, Chung and Ely 2007, Frankel 2014, Garrett 2014). It also relates to a recent spurt of interest in multidimensional screening, particularly in the algorithmic game theory world. Several recent papers in this area have given conditions under which simple mechanisms can be shown to be optimal (Haghpanah and Hartline 2015, Wang and Tang 2014); most relevantly among these, Haghpanah and Hartline (2015) introduce a virtual value maximization approach that is connected to ours. Others (Hart and Nisan 2012, Li and Yao 2013, Babaioff, Immorlica, Lucier, and Weinberg 2014, Rubinstein and Weinberg 2015, Yao 2015, Cai, Devanur, and Weinberg 2016) have argued that simple mechanisms for selling multiple goods are approximately optimal under broad conditions. Finally, there is a longstanding, more applied literature on bundling in industrial organization (Stigler 1963, Adams and Yellen 1976, McAfee, McMillan, and Whinston 1989, Chu, Leslie, and Sorensen 2011), to which we shall return periodically. In contrast to these approximate-optimality and applied strands of literature, the present paper focuses on exact optimality among all possible (deterministic or probabilistic) mechanisms. Still, all of these branches share a common theme, of offering models in which relatively simple mechanisms are advantageous.

In the concluding section, we give a more detailed comparison of our proof techniques with the closest previous approaches to multidimensional screening. In particular, the concept of generalized virtual values, a key analytical tool here, appears to be relatively new to the literature, even though it falls out straightforwardly from the Lagrangian of the principal's maximization problem. (Cai, Devanur, and Weinberg (2016) independently introduce this same concept.)

2 Model and result

We shall first lay out the model, and then describe a number of applications.

We will notate the metric on an arbitrary metric space by d; no confusion should result. For a compact metric space X, we write $\Delta(X)$ for the space of Borel probability distributions on X, with the weak topology.

2.1 Screening problems

We first formally define a *(Bayesian) screening problem*, the building block of our model. We assume throughout that there is a single agent with quasi-linear utility.

A *screening environment* (Θ, X, u) is defined by a *type space* Θ and an *allocation space* X, both assumed to be compact metric spaces endowed with the Borel σ-algebra; and a *utility function* $u : X \times \Theta \to \mathbb{R}$, which is assumed to be (jointly) Lipschitz continuous. We define $Eu : \Delta(X) \times \Theta \to \mathbb{R}$ as the extension of u by taking expectations over allocations. We will use the variable x to denote either an element of X or of $\Delta(X)$; sometimes we will also use a for elements of X to avoid ambiguity.

In such an environment, the agent may be screened by a *mechanism*. We allow for arbitrary probabilistic mechanisms; thus, a *mechanism* is defined as a pair $M = (x, t)$, with $x : \Theta \to \Delta(X)$ and $t : \Theta \to \mathbb{R}$, with t measurable, and satisfying the usual incentive compatibility (IC) and individual rationality (IR) conditions:[1]

$$Eu(x(\theta), \theta) - t(\theta) \geq Eu(x(\widehat{\theta}), \theta) - t(\widehat{\theta}) \qquad \text{for all } \theta, \widehat{\theta} \in \Theta; \tag{2.1}$$

$$Eu(x(\theta), \theta) - t(\theta) \geq 0 \qquad \text{for all } \theta \in \Theta. \tag{2.2}$$

As usual, we are using the revelation principle to justify this formulation of mechanisms as functions of the agent's type; we will generally stick to this formalism, although the verbal descriptions of mechanisms may use other, equivalent formulations.

Note that the formulation here of the IR constraints has assumed each type's outside option is zero. This is essentially without loss of generality: even if there is a type-dependent outside option, $\underline{u}(\theta)$, we can renormalize by replacing $u(x, \theta)$ with $\widetilde{u}(x, \theta) =$

[1] Thus we further abuse notation by letting x denote either an allocation or the allocation-rule component of a mechanism; this should not cause confusion in practice.

$u(x,\theta) - \underline{u}(\theta)$.

We write $\mathcal{M}$ for the space of all mechanisms.

A *Bayesian screening problem* — or *screening problem* for short[2] — consists of a screening environment as above, together with one more ingredient, a given probability distribution over types, $\pi \in \Delta(\Theta)$. Then, the principal evaluates any mechanism by the resulting expected payment, $E_\pi[t(\theta)]$. We will refer to this objective interchangeably as *revenue* or *profit*.

It is known from the literature (e.g. Balder 1996, Theorem 2.2) that there exists a mechanism maximizing expected payment. Indeed, if Θ is finite then this is easy to see by compactness arguments. The proof for general Θ is much more difficult and will not be presented here. However, for the sake of self-containedness, we may point out that we will not actually need to use the general existence result or its proof in what follows: the statement and proof of our main result, Theorem 2.1, will still be valid using the supremum over mechanisms (without needing to know whether it is attained).

Note that elsewhere in the screening literature, the principal often has preferences directly over the allocations (for example, costs of producing goods). Our model above does not incorporate any such preferences. However, we can accommodate them, again by renormalization: For example, if the agent's payoff is given by $u(x,\theta)$ and producing allocation x costs $c(x)$ for the principal, then we can redefine allocation x to be "receiving x and paying the production cost" instead of only "receiving x." Explicitly, we apply the original model with payoff function $\widetilde{u}(x,\theta) = u(x,\theta) - c(x)$.

2.2 Joint screening

In our main model, the agent is to be screened according to G pieces of private information, $g = 1,\ldots,G$, that enter separately into his utility function. Thus, we assume given a screening problem $(\Theta^g, X^g, u^g, \pi^g)$ for each g. We will refer to these as the *component screening problems*.

These G component screening problems give rise to a *joint screening environment* (Θ, X, u), where the type space is $\Theta = \times_{g=1}^G \Theta^g$ and the allocation space is $X = \times_{g=1}^G X^g$, with representative elements $\theta = (\theta^1,\ldots,\theta^G)$ and $x = (x^1,\ldots,x^G)$; and the utility func-

[2] This latter name is not perfect, since it appears to presume that any screening problem should come with a prior — which is contrary to the whole message of this paper! But we will stick with it for brevity.

tion $u : X \times \Theta \to \mathbb{R}$ is given by

$$u(x, \theta) = \sum_{g=1}^{G} u^g(x^g, \theta^g).$$

(Admittedly, there is potential for confusion with the use of these same variables (Θ, X, u) to refer to an arbitrary screening environment in the previous subsection. From now on, they will refer specifically to this product environment except when stated otherwise.)

We will use standard notation such as $\Theta^{-g} = \times_{h \neq g} \Theta^h$; θ^{-g} for a representative element of Θ^{-g}; and (θ^g, θ^{-g}) for an element of Θ with one component singled out.

The combined utility function u can be extended to probability distributions over X as before. Note however that if ρ is a probability distribution over X, then the expected utility of the agent of type θ is

$$E_\rho \left[\sum_g u^g(x^g, \theta^g) \right] = \sum_g E_{\mathrm{marg}_{X^g}(\rho)}[u^g(x^g, \theta^g)]$$

where $\mathrm{marg}_{X^g}(\rho)$ denotes the marginal distribution over X^g. That is, the agent's expected utility depends only on the marginal distribution over each X^g, and not the correlation among them. So we can simplify by thinking of a random allocation as an element of $\times_g \Delta(X^g)$, and define the expected utility $Eu : \times_g \Delta(X^g) \to \mathbb{R}$ accordingly. A *mechanism* in this environment then can be defined as a pair of functions (x, t), with $x : \Theta \to \times_g \Delta(X^g)$ and $t : \Theta \to \mathbb{R}$ measurable, satisfying conditions (2.1) and (2.2) as before.

In the joint problem, unlike in Subsection 2.1, we assume that the principal does not know the distribution of the agent's full type θ. Instead, she knows only the marginal distributions of each component, $\pi^1, \ldots, \pi^G$, but not how these components are correlated with each other. Our principal possesses non-Bayesian uncertainty about the correlation structure; she evaluates any mechanism (x, t) by its worst-case expected profit over all possible joint distributions. Formally, let Π be the set of all distributions $\pi \in \Delta(\Theta)$ that have marginal π^g on Θ^g for each g. Then, a mechanism (x, t) is evaluated by

$$\inf_{\pi \in \Pi} E_\pi[t(\theta)]. \tag{2.3}$$

This describes the joint screening problem. The class of mechanisms we have allowed is fully general: each component x^g of the allocation can depend on the agent's entire type θ. (For example, in selling goods as a bundle, whether the agent receives good 1 depends on his values for the other goods.)

But one thing the principal can always do is to screen each component separately. In particular, let us write R^{*g} for the optimal profit in screening problem $(\Theta^g, X^g, u^g, \pi^g)$, and (x^{*g}, t^{*g}) for the corresponding optimal mechanism. Then define $R^* = \sum_g R^{*g}$. The separate-screening mechanism (x^*, t^*) corresponding to these component mechanisms is given by

$$x^*(\theta) = (x^{*1}(\theta^1), \ldots, x^{*G}(\theta^G));$$

$$t^*(\theta) = t^{*1}(\theta^1) + \cdots + t^{*G}(\theta^G)$$

for each $\theta = (\theta^1, \ldots, \theta^G) \in \Theta$. It is immediate that this mechanism satisfies (2.1) and (2.2). And its expected profit is predictable despite the principal's uncertainty: for any possible $\pi \in \Pi$, we always have

$$E_\pi[t^*(\theta)] = \sum_g E_\pi[t^{*g}(\theta^g)] = \sum_g E_{\pi^g}[t^{*g}(\theta^g)] = \sum_g R^{*g} = R^*.$$

Our main result says that, with the robust objective, separate screening is optimal:

Theorem 2.1. *For any mechanism, the value of the objective (2.3) is at most R^*.*

We show this by constructing a specific distribution π on which the value is at most R^*. More precisely, in the text we give the proof for the finite setting (both Θ and X finite), and in that setting we construct a specific π; in the appendix, we show how the result then extends to the continuous case using an approximation argument. Some discussion about possible approaches to constructing π appears in Section 3, followed by the actual proof in Section 4. (A reader eager for the proof can skip ahead to Section 4 without serious loss of continuity.) In addition, a natural follow-up question is how sensitive the result is to the exact joint distribution; we explore this question in Section 5. The concluding Section 6 will compare our methods to previous multidimensional screening techniques, and discuss interpretive issues and future work.

2.3 Examples

Here we describe several applications.

Linear monopoly. In the monopoly sales setting described in the introduction, the principal is a seller with G goods available for sale, and the buyer's preferences are additively separable across goods. In this case, θ^g is the buyer's value for getting good

g, so Θ^g is an interval in $\mathbb{R}^+$, say $\Theta^g = [0, \overline{\theta}^g]$; and $X^g = \{0, 1\}$, with 1 corresponding to receiving the good g and 0 corresponding to not receiving the good. (Remember however that we allow for probabilistic mechanisms, which can give the good with probability between 0 and 1.) The utility function is $u^g(x^g, \theta^g) = \theta^g x^g$.

For each g, π^g is the prior distribution over the buyer's value for good g. In this model, it is well known (e.g. Riley and Zeckhauser 1983) that the optimal selling mechanism for a single good is a single posted price p^{*g}: that is $(x^g(\theta^g), t^g(\theta^g)) = (1, p^{*g})$ if $\theta^g \geq p^{*g}$ and $(0, 0)$ otherwise. Consequently, Theorem 2.1 says that, in our model where the seller knows the marginal distribution of values for each good but not the joint distribution of values, the worst-case-optimal mechanism simply consists of posting price p^{*g} for each good g separately.

We shall refer to this example as the *benchmark application*, and shall return to it periodically in discussion.

We have written this application with a single buyer, but note that it works equally well with a continuum population of buyers, in which π^g denotes the cross-sectional distribution of buyers' values for good g. (Indeed, one can think of other possible applications of the model that are amenable to a population-level interpretation; the taxation example below is an instance.) In this case, our restriction to direct mechanisms is not immediately without loss: since the seller does not know the joint distribution π, she could in principle do better by adaptively learning the distribution as in Segal (2003), or by asking each buyer for beliefs about the distribution of other buyers as in Brooks (2013). In a model where the seller has Bayesian uncertainty about the distribution of types in the population, such learning would typically be beneficial. However, our model uses a worst-case objective, and our proof actually constructs a specific joint distribution where no (direct) mechanism does better than R^*. If this turns out to indeed be the true distribution, then even with adaptive learning or elicitation of beliefs, the seller will not be able to do better than R^*. So our worst-case optimality result is unaffected by these issues.

Nonlinear monopoly. More generally, we can consider a multi-good sales model in which each good g can be allocated continuously: $X^g = [0, 1]$. Then we can regard each good g as divisible, and interpret x^g as a quantity; or, alternatively, x^g may be a measure of quality, as in Mussa and Rosen (1978). Now θ^g is some parameter describing preferences, and the payoff function $u^g(x^g, \theta^g)$ may be an arbitrary Lipschitz function in our model. Usually in the literature, one imposes structural assumptions on this function — most importantly, some form of single crossing condition — that make it possible to solve explicitly for the component-g optimal mechanism using standard techniques. But no such assumptions are needed for our separation result to hold.

In general, the optimal mechanism for component g will now be a menu of various bundles of (possibly probabilistic) quantities and prices (x^g, t^g), from which each type gets its favorite. Our result then says that, from the point of view of the principal who is uncertain about the joint distribution of the θ^g, the worst-case-optimal joint screening mechanism consists of a separate such menu for each component g.

Note also that, while we have insisted on allowing probabilistic mechanisms, there are some sufficient conditions under which the optimal mechanism is actually deterministic. For example, suppose preferences are linear, $u^g(x^g, \theta^g) = \theta^g x^g$, but there is a convex function $c^g(x^g)$ describing the cost of production. Then any probabilistic mechanism can be improved on by replacing each type θ^g's random allocation by its mean, since this reduces the principal's cost and has no effect on the IC and IR constraints. Strausz (2006) also gives sufficient conditions for deterministic mechanisms to be optimal, although his conditions fall under the "easy case" of our theorem (discussed near the end of Section 3 below), namely when the component problems are standard one-dimensional screening problems in which the monotonicity constraint does not bind.

Optimal taxation. A less obvious application of our model is to a Mirrlees-style taxation problem, in which preferences are quasi-linear and the planner has a Rawlsian objective, to maximize the payoff of the worst-off type. This setup is highly stylized and not meant to directly inform tax policy, but it illustrates that our results offer a possible framework for thinking about other applications besides multi-good monopoly.

To illustrate the connection, let us first ignore the joint screening apparatus and return to the general screening language of Subsection 2.1. The taxation problem would be formulated as follows. There is a population of heterogeneous agents, of unit mass, with types indexed by $\theta \in \Theta$ following a known population distribution π. There is a single consumption good. Each agent θ can produce any amount $x \in X = [0, \overline{x}]$ of the good, at a disutility cost $h(x, \theta)$, which we extend linearly to random x (and notate the extension by Eh). We need not make any structural assumptions (e.g. Θ single-dimensional, single-crossing preferences). A mechanism then consists of an allocation rule $x : \Theta \to \Delta(X)$, and consumption function $c : \Theta \to \mathbb{R}$, satisfying the incentive constraint

$$c(\theta) - Eh(x(\theta), \theta) \geq c(\widehat{\theta}) - Eh(x(\widehat{\theta}), \theta) \qquad \text{for all } \theta, \widehat{\theta}$$

and the resource constraint

$$\int c(\theta)\, d\pi \leq \int E[x(\theta)]\, d\pi.$$

There is no individual rationality constraint, since everyone can be forced to participate.

The planner's problem is to find a mechanism to maximize the payoff of the worst-off type, $\min_{\theta\in\Theta}(c(\theta) - Eh(x(\theta), \theta))$.

To see how this is equivalent to our formulation, we take the primitive $u(x, \theta)$ to represent the utility an agent would get from producing and consuming x, and the transfer as the net amount redistributed away. Thus, we write $u(x, \theta) = x - h(x, \theta)$. Notice that an allocation rule and consumption rule $(x(\theta), c(\theta))$ then satisfy IC in the taxation problem if and only if the allocation rule and transfer rule $(x(\theta), E[x(\theta)] - c(\theta))$ satisfy IC in the original screening-problem language. Moreover, for any mechanism $(x(\theta), c(\theta))$, adjusting every type's consumption to $c(\theta) + \Delta$ for constant Δ preserves IC, and changes the planner's objective by Δ. The optimal Δ is the one for which the resource constraint just binds in the adjusted mechanism, namely, $\Delta = \int(E[x(\theta)] - c(\theta))\, d\pi$. Thus, in the taxation model, the planner's problem is equivalent to maximizing

$$\min_{\theta}(c(\theta) - Eh(x(\theta), \theta)) + \int (E[x(\theta)] - c(\theta))\, d\pi$$

over all mechanisms satisfying the IC constraint only. Likewise, in the screening formulation, every type's transfer can be adjusted by a constant Δ, and the optimal Δ is the one that makes IR just bind, namely $\Delta = \min_\theta(c(\theta) - Eh(x(\theta), \theta))$; then, the principal's problem is equivalent to maximizing

$$\int (E[x(\theta)] - c(\theta))\, d\pi + \min_{\theta}(c(\theta) - Eh(x(\theta), \theta))$$

over all mechanisms satisfying IC only. So the two problems are equivalent.

Now we move to the joint screening model. Suppose there are multiple income-producing activities $g = 1, \ldots, G$; each agent is parameterized by a type for each activity, so the overall type and allocation spaces are $\Theta = \times_g \Theta^g$, $X = \times_g X^g$ with $X^g = [0, \overline{x}^g]$, and payoffs are given by $c - \sum_g h^g(x^g, \theta^g)$. The planner knows the marginal distribution π^g of each θ^g in the population, but not the joint distribution π.

How do we define mechanisms in this model? A mechanism should specify a (probabilistic) level of production in each activity g, and a consumption level, for each type

$\theta \in \Theta$, satisfying incentive compatibility. The resource constraint should specify that total consumption in the population must not exceed total production across all activities. However, writing down this constraint is not straightforward, since both sides of the inequality depend on the unknown population distribution π. So we need to make a modeling choice. For concreteness let us focus on the most permissive specification: we allow each agent's production and consumption to depend not only on his own θ but also on the entire realized distribution π, and we require the resource constraint

$$\int c(\theta|\pi)\, d\pi \leq \int \left(\sum_g E[x^g(\theta|\pi)] \right) d\pi$$

to hold for each π separately. (One might alternatively imagine more restrictive versions of the model: for example, requiring c and x to depend only on θ, and requiring the resource constraint to be satisfied for *every* possible π.)

In the joint problem, one natural mechanism is to separate across the activities g: If the optimal tax schedule for activity g (with marginal distribution π^g) is (x^{*g}, c^{*g}), then each agent θ is assigned to produce $x^{*g}(\theta^g)$ in each activity g, and receive consumption equal to $\sum_g c^{*g}(\theta^g)$, regardless of the realized joint distribution π. For any π with the correct marginals, this will ensure that total consumption does not exceed total production. So, any reasonable formulation of the aggregate resource constraint is indeed satisfied by this mechanism. Then, using the worst-case distribution constructed in proving Theorem 2.1, we can conclude that this mechanism is worst-case optimal: no other mechanism can guarantee a better value for the Rawlsian objective across all joint distributions. (And under alternative formulations of the resource constraint, this would remain true, since we have adopted the most permissive formulation here.)

If we interpret each g as a time period, then this application of our model connects with a recent literature in dynamic public finance, in which agents' income-producing abilities evolve over time. A common prediction of such models is that, in the optimal mechanism, each agent's tax will typically depend on the entire history of his past income. This literature has tacitly acknowledged such history-dependent taxation schemes as unrealistically complicated, and responded by quantitatively comparing with the optima obtained using more restrictive tax instruments, such as ones depending only on age and current income (Farhi and Werning 2013, Stantcheva 2015, Findeisen and Sachs 2015). However, the theoretical foundations for this approach, or more generally for delineating which kinds of tax systems are or are not "simple," are yet to be established. Our model suggests one avenue for such foundations. In our model, if the planner knows the distribution of ability within each period but not the correlation structure across periods nor the information each agent has about his own future ability, then the optimally robust tax policy will tax and redistribute within each period separately.

3 Unsuccessful proof approaches

As indicated above, we will prove Theorem 2.1 by constructing a particular joint distribution on Θ for which no mechanism can generate profit greater than R^*. In order to better understand the content of the result, we consider some straightforward ways one might try to construct such a joint distribution.

3.1 Independent distributions

One natural first try would be to have the different components θ^g be independently distributed, $\pi = \times_g \pi^g$. A priori, an independent model seems to remove all interactions across components in the setup of the screening problem, so one might expect no interactions in the solution.

However, this approach is a nonstarter even in the benchmark monopoly problem, as is well known from the bundling literature. For example, with a large number of goods with i.i.d. values, the value for the bundle of all goods is approximately pinned down by the law of large numbers; hence, the seller can extract almost all surplus by bundling, which she could not do with separately posted prices (Armstrong 1999, Bakos and Brynjolfsson 1999). Hart and Nisan (2012) show that independence also fails in a minimal example: two goods, where the buyer's value for each good is either 1 or 2 with probability 1/2 each. The seller can then extract profit 1 for each good, by setting either price 1 or price 2; so her optimal profit from selling the two goods separately is 2. But if the values are independent, she can instead charge a price 3 for the bundle of both goods, which she then sells probability 3/4 and so earns expected profit $9/4 > 2$.

In fact, McAfee, McMillan, and Whinston (1989) show that with continuous distributions, separate pricing is essentially *never* optimal under independence. This follows from considering the first-order condition for charging a price for the bundle that is just slightly less than the sum of separate prices.

3.2 Maximal positive correlation

Another approach comes from considering the case where the G separate screening problems $(\Theta^g, X^g, u^g, \pi^g)$ are all identical. In this case, one possible joint distribution π is that all components of the agent's type are identical: π is distributed along the diagonal $\{\theta \in \Theta \mid \theta^1 = \cdots = \theta^G\}$. For this distribution, the multidimensional joint mechanism design problem is equivalent to the component problem, scaled up by a factor of G; it is not hard to see that indeed no mechanism can earn expected profit greater than $G \cdot R^{*1} = R^*$, as needed.

How might one generalize this construction when the component problems are not identical? One possibility to try is to have all G components of the type be "as positively correlated as possible," so as to again reduce the mechanism design problem to a single-dimensional type. For example, in the benchmark monopoly application where $\theta^g \in \Theta^g \subseteq \mathbb{R}$ is the value for good g, let $q^g : [0,1] \to \Theta^g$ be the inverse quantile function, defined by

$$q^g(z) = \min\{\theta^g \mid Pr_{\pi^g}(\theta \leq \theta^g) \geq z\},$$

and then define the joint distribution by randomly drawing $z \sim U[0,1]$ and taking $\theta^g = q^g(z)$ for each g. We refer to this as the *comonotonic* joint distribution.

A problem with this approach is that it is unclear how it would work in general, when each Θ^g is not necessarily single-dimensional. But even in the benchmark application it does not always succeed. Here is a counterexample. Consider two goods. Suppose the possible values for the first good are $1, 2$, with probability $1/2$ each; and the possible values for the second good are $2, 3, 4$, with probabilities $1/3, 1/6, 1/2$ respectively. The seller can earn an expected profit of 1 from the first good alone (either by setting price 1 or price 2), and 2 from the second good alone (either by price 2, 3 or 4), so the maximum profit from separate pricing is 3.

In the comonotonic joint distribution π, there are three possible types, $(1, 2)$, $(1, 3)$, $(2, 4)$, occurring with probabilities $1/3, 1/6, 1/2$ respectively (as shown in Figure 3.1(a)). If this is the joint distribution, then we propose the following mechanism: the buyer can either

- receive good 1 at a price of 1;
- receive good 2 at a price of 3;
- receive both goods at a price of 5; or

- receive nothing and pay nothing.

Figure 3.1(b) shows the regions of buyer space in which each option is chosen; in particular, it is incentive compatible for the $(1,2)$ buyer to buy only good 1, the $(1,3)$ buyer to buy only good 2, and the $(2,4)$ buyer to buy the bundle, leading to revenue

$$\frac{1}{3} \cdot 1 + \frac{1}{6} \cdot 3 + \frac{1}{2} \cdot 5 = \frac{10}{3} > 3.$$

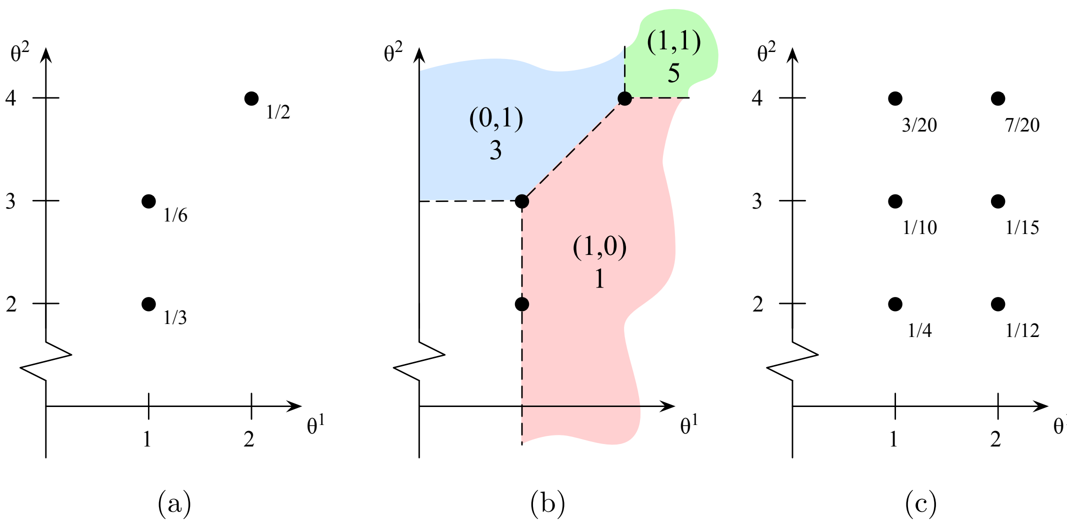

Figure 3.1: Counterexample with maximal positive correlation. (a) The candidate joint distribution. (b) A mechanism that outperforms separate sales. (c) A joint distribution for which no mechanism outperforms separate sales.

This shows that the comonotonic distribution cannot be used to prove Theorem 2.1 in general.

Why does the optimal mechanism design problem for this π not decompose into separate problems for each component? One way to think about what goes wrong, expressed in terms of the usual approach to one-dimensional screening problems, is that the monotonicity constraint that arises with multidimensional allocation is weaker than requiring monotonicity good-by-good. This can be seen in the example above: the low and high types both receive good 1, but the middle type does not. This allows the seller to charge a different marginal price for good 1 to the low type than the high type.

Indeed, when the standard solution to the one-dimensional problem has the monotonicity constraint not binding, the maximal-positive-correlation approach does succeed. This is formalized in the following proposition, which for convenience is expressed in terms of continuous distributions. (The counterexample above is discrete, but this difference is immaterial; it can be made continuous by perturbation.)

Proposition 3.1. *Consider the benchmark monopoly application. Suppose each marginal distribution π^g is represented by a continuous, positive density f^g, and write F^g for the cumulative distribution. Suppose that for each g, there is a type θ^{*g} such that the virtual value*

$$v^g = \theta^g - \frac{1 - F^g(\theta^g)}{f^g(\theta^g)}$$

is negative for $\theta^g < \theta^{*g}$ and positive for $\theta^g > \theta^{*g}$. Let π be the comonotonic joint distribution. For this π, no mechanism yields higher expected profit than R^*.

This can be shown by the usual method of ignoring the monotonicity constraint, using the infinitesimal downward IC constraints to rewrite profit in terms of virtual surplus, and then maximizing virtual surplus pointwise. The details are in the appendix. (In fact, Proposition 3.1 extends beyond the benchmark application, to any single-dimensional screening problems with single crossing in which the monotonicity constraint does not bind; we omit the details.)

In the light of Proposition 3.1, our Theorem 2.1 can be seen as a result about the relationship between multidimensional ironing and single-dimensional ironing. Indeed, the *generalized virtual value*, which will be a key object in our proof, extends the concept of an ironed virtual value in the benchmark case. After developing this concept, we will show how to use it to directly construct our joint distribution π. In the special case where the component problems involve no ironing, our construction actually recovers the comonotonic distribution (as discussed at the end of Subsection 4.2 below), but in general it is subtler. For the above example, the π we construct is as shown in Figure 3.1(c).

In this example, there are actually other joint distributions that also pin down profit to at most R^*. All of them place positive probability on at least five of the six possible types. However, the distribution shown is unique in the following respect: instead of imposing all of the IC and IR constraints, we can impose just IC against deviating one step downward for one good (and truthfully reporting the value for the other good), plus the IR constraint for the lowest type; this subset of constraints already ensures that no mechanism earns higher profit than R^*. The distribution in Figure 3.1(c) is the unique joint distribution for which this subset of incentive constraints is sufficient.

We turn now to the general proof.

4 The actual proof

We start with a verbal overview of the ideas. The substance of the proof consists of the case where Θ^g and X^g are finite, and we cover this case in the text. In the appendix, we complete the proof by extending to the general case via an approximation argument.

For the finite case, it helps to first develop some basics about the mathematics of screening problems in general. A finite screening problem can be thought of as a linear programming problem, with the components of the mechanism (the probability of each allocation at each type, and the payments) as variables. This allows us to use the language of LP duality, and talk about the dual variables on the IC and IR constraints. These dual variables (also known as Lagrange multipliers) will be labeled with the letters λ and κ.

We first show how, in any finite screening problem, the Lagrangian can be interpreted as a kind of generalized virtual surplus: For each type θ of the agent and each possible allocation x, one can define a generalized virtual value $\overline{u}(x, \theta)$, which represents the effect on revenue from a marginal increase in the probability assigned to allocation x; it equals the value of type θ for that allocation, minus the shadow cost from binding incentive constraints of other types that would like to imitate type θ. In the familiar case of one-dimensional screening problems, these generalized virtual values are just the usual ironed virtual values; but in fact they can be defined for any finite screening problem. An optimal mechanism then must maximize the virtual surplus for each type separately. Our strategy for proving Theorem 2.1 is to construct the distribution π for the joint problem so that, for each joint type θ, the virtual value for any outcome x is simply the sum of the virtual values $\overline{u}^g(x^g, \theta^g)$ from the component problems. If we can do this, it will immediately follow that the separate-screening mechanism maximizes the virtual surplus and so maximizes revenue.

One complication in this argument is that the generalized virtual values for a screening problem are functions not only of the probability of each type but also of the dual variables. In general, the dual LP may have many optimal solutions, and we need to choose one in order to define the generalized virtual values. (The traditional definition of ironed virtual values implicitly corresponds to one particular choice of dual solution.) Hence, rather than construct a joint distribution π by itself, we will simultaneously construct π and the dual variables for the joint screening problem, using dual variables from the component problems as inputs to the construction.

We can simplify the task of choosing dual variables by using only a subset of the IC constraints — namely, the ones for misreporting a single component of the type; that is, type $\theta = (\theta^g, \theta^{-g})$ imitating a type $(\widehat{\theta}^g, \theta^{-g})$, for some g and $\widehat{\theta}^g$. We are thus studying a relaxed problem, and showing that with this subset of constraints, already no mechanism earns more than R^*.

Once we have decided to focus on this set of constraints, it quickly becomes apparent that the generalized virtual values will separate in the desired way if the following relationship holds between the dual variables on the IC's in the joint problem and the corresponding dual variables from the component problems:

$$\lambda[(\theta^g, \theta^{-g}) \to (\widehat{\theta}^g, \theta^{-g})] = \frac{\pi(\widehat{\theta}^g, \theta^{-g})}{\pi^g(\widehat{\theta}^g)} \times \lambda^g[\theta^g \to \widehat{\theta}^g]. \tag{4.1}$$

This formula, in effect, tells us how we should construct the IC dual variables in terms of the joint distribution π (the dual variables on the IR's will be constructed separately). Of course, we cannot simply choose any old π and define the λ's by (4.1); we need to ensure that they are indeed a solution to the dual LP.

Here is an alternative perspective on what this requirement means. In any screening problem, the dual variables (λ, κ) need to be related to the type probabilities via a certain set of linear equations, called *flow conservation equations*. These equations are the constraints dual to the t variables. They also ensure that the principal's profit from any mechanism (the primal objective) is bounded by its generalized virtual surplus.

Our strategy in the joint screening problem, then, is to choose π and the multipliers (λ, κ) so that two conditions hold:

(i) the flow conservation equations are satisfied;

(ii) the virtual surplus, when maximized type-by-type, comes out to R^*.

If we can achieve both of these, then together they imply that the profit from any mechanism is at most R^*, which is what we need.

We will choose π, and then determine the multipliers via (4.1) above; this will ensure that virtual values in the joint problem separate into the component virtual values, which will give (ii). It remains to ensure (i). That is, we must show that $\pi \in \Pi$ can be chosen so that the resulting multipliers satisfy the flow conservation equations in the joint screening problem. We will use a connection between the flow conservation equations in the joint problem and the corresponding equations in the component screening problems to show that this can be done — and to obtain at least something of an interpretation for π in the process.

With this outline to guide us, it's time to begin the details.

4.1 Preliminary tools

We first gather several general-purpose facts about screening problems. For this subsection, we use the variables (Θ, X, u, π) to refer to an arbitrary screening problem as in Subsection 2.1, not the product environment of our main model. We also use R^* here to refer to the maximal profit in this problem, again setting aside its meaning in the main model.

We start with the discussion of generalized virtual values, introduced in the outline above. Suppose that (Θ, X, u, π) is a screening problem in which Θ and X are finite, and every type has positive probability: $\pi(\theta) > 0$ for each $\theta \in \Theta$. (This latter assumption is not without loss of generality: probability-zero types cannot simply be deleted from the type space, because their IR constraints may affect the set of possible mechanisms.)

Let us refer to pure allocations (elements of X) by the variable a. A mechanism then consists of $|\Theta| \cdot (|X| + 1)$ numbers, namely the allocation probabilities $x_a(\theta)$ and payment $t(\theta)$ for each type θ and each $a \in X$, satisfying the IC and IR constraints, as well as the requirement that each $x(\theta)$ form a valid probability distribution:

$$\sum_{a \in X} u(a,\theta) x_a(\theta) - t(\theta) \geq \sum_{a \in X} u(a,\theta) x_a(\widehat{\theta}) - t(\widehat{\theta}) \qquad \text{for all distinct } \theta, \widehat{\theta} \in \Theta; \tag{4.2}$$

$$\sum_{a \in X} u(a,\theta) x_a(\theta) - t(\theta) \geq 0 \qquad \text{for all } \theta \in \Theta; \tag{4.3}$$

$$x_a(\theta) \geq 0 \qquad \text{for all } \theta \in \Theta, a \in X; \tag{4.4}$$

$$\sum_{a \in X} x_a(\theta) = 1 \qquad \text{for all } \theta \in \Theta. \tag{4.5}$$

The maximum expected profit over all such mechanisms is given by

$$\max_{(x,t)} \sum_{\theta} \pi(\theta) t(\theta).$$

Maximizing this profit, subject to constraints (4.2)–(4.5), is a linear programming problem. The maximum value is R^*, by definition.

Then, the value of the dual problem is also R^*. Consider an optimal solution to this dual problem. It consists of dual variables for each constraint (these dual variables are indexed by types, which we indicate in square brackets):

$$\begin{aligned} \lambda[\theta \to \widehat{\theta}] &\geq 0 && \text{for each } \theta \neq \widehat{\theta}; \\ \kappa[\theta] &\geq 0 && \text{for each } \theta; \\ \mu_a[\theta] &\geq 0 && \text{for each } \theta, a; \\ \nu[\theta] & && \text{for each } \theta. \end{aligned}$$

It will be useful for notational simplicity to also define $\lambda[\theta \to \theta] = 0$ for each θ.

Writing out explicitly the dual constraints, we have

$$\sum_{\widehat{\theta}} \lambda[\widehat{\theta} \to \theta] u(a, \widehat{\theta}) - \sum_{\widehat{\theta}} \lambda[\theta \to \widehat{\theta}] u(a, \theta) - \kappa[\theta] u(a, \theta) - \mu_a[\theta] - \nu[\theta] = 0 \tag{4.6}$$

for each θ and a, as the constraint corresponding to primal variable $x_a(\theta)$; and

$$\sum_{\widehat{\theta}} \lambda[\theta \to \widehat{\theta}] - \sum_{\widehat{\theta}} \lambda[\widehat{\theta} \to \theta] + \kappa[\theta] = \pi(\theta), \tag{4.7}$$

as the constraint corresponding to $t(\theta)$. We will refer to these latter as *flow conservation equations*, since they can be viewed as describing net flow in a network whose nodes are the types, where the flow along each edge $\theta \to \widehat{\theta}$ is given by $\lambda[\theta \to \widehat{\theta}]$. We also have

$$-\sum_{\theta} \nu[\theta] = R^*, \tag{4.8}$$

the optimal value of the objective in the dual.

Now, for each θ, define the *generalized virtual value* for each allocation $a \in X$:

$$\overline{u}(a, \theta) = u(a, \theta) - \sum_{\widehat{\theta} \in \Theta} \frac{\lambda[\widehat{\theta} \to \theta]}{\pi(\theta)} \left(u(a, \widehat{\theta}) - u(a, \theta) \right), \tag{4.9}$$

and extend this function to randomized allocations, $E\overline{u} : \Delta(X) \times \Theta \to \mathbb{R}$, by linearity. Also define the maximum generalized virtual value for each type:

$$\overline{u}_{\max}(\theta) = \max_{a \in X} \overline{u}(a, \theta).$$

As we shall momentarily show, these generalized virtual values extend two features of virtual values that are familiar from the traditional setting (e.g. Myerson 1981). First, we can use them to eliminate the payments from the principal's maximization problem: for any mechanism, its profit is at most the (expected) generalized virtual value generated by its allocation rule.[3] Consequently, we obtain an immediate upper bound on the profit of any mechanism by simply taking the maximum generalized virtual value for each type separately, $\overline{u}_{\max}(\theta)$. Second, any optimal mechanism attains this bound with equality. In particular, for each θ, the optimal mechanism chooses the allocation $a \in X$ with the highest generalized virtual value. (If the optimal mechanism is not deterministic, there must be multiple a's tied for highest virtual value, and only such a's can receive any probability weight. Note that this is not reversible: there may exist allocation rules x that put weight only on virtual-value-maximizing outcomes, yet are not part of any optimal mechanism (x, t).)

To be precise, we will show the following chain of inequalities: for any mechanism (x, t),

$$\sum_{\theta} \pi(\theta)t(\theta) \leq \sum_{\theta} \pi(\theta)E\overline{u}(x(\theta), \theta) \leq \sum_{\theta} \pi(\theta)\overline{u}_{\max}(\theta) \leq R^*. \tag{4.10}$$

It will then follow immediately that any optimal mechanism (x^*, t^*) satisfies these relations with equality. (In particular, the last inequality is always an equality since it does not depend on the mechanism (x, t).)

In fact, all of these inequalities are straightforward rearrangements of the duality conditions. For the first inequality in the chain: for any mechanism (x, t), the IC and IR constraints give

$$\begin{aligned} \sum_{\theta} \pi(\theta)t(\theta) \quad \leq \quad & \sum_{\theta} \Big(\pi(\theta)t(\theta) + \\ & \sum_{\widehat{\theta}} \lambda[\theta \to \widehat{\theta}] \left((Eu(x(\theta), \theta) - t(\theta)) - (Eu(x(\widehat{\theta}), \theta) - t(\widehat{\theta})) \right) + \\ & \kappa[\theta](Eu(x(\theta), \theta) - t(\theta)) \Big). \end{aligned} \tag{4.11}$$

The right-side terms can be reorganized as "x terms" and "t terms." For each θ, the terms containing $x(\theta)$ are

$$\sum_{\widehat{\theta}} \lambda[\theta \to \widehat{\theta}]Eu(x(\theta), \theta) - \sum_{\widehat{\theta}} \lambda[\widehat{\theta} \to \theta]Eu(x(\theta), \widehat{\theta}) + \kappa[\theta]Eu(x(\theta), \theta).$$

[3] With traditional (unironed) virtual values in the benchmark monopoly problem, and continuous types, this inequality actually is always an equality. This property does not extend to our generalized virtual values. Note that even in the benchmark problem, it does not hold for *ironed* virtual values, which are what our definition generalizes.

Using (4.7) to plug in for $\sum_{\widehat{\theta}} \lambda[\theta \to \widehat{\theta}] + \kappa[\theta]$, we see that this equals

$$\left(\sum_{\widehat{\theta}} \lambda[\widehat{\theta} \to \theta] + \pi(\theta) \right) Eu(x(\theta), \theta) - \sum_{\widehat{\theta}} \lambda[\widehat{\theta} \to \theta] Eu(x(\theta), \widehat{\theta}) \quad = \quad \pi(\theta) E\overline{u}(x(\theta), \theta).$$

And as for the t terms, we have $t(\theta)$ appearing on the right side of (4.11) with coefficient

$$\pi(\theta) - \sum_{\widehat{\theta}} \lambda[\theta \to \widehat{\theta}] + \sum_{\widehat{\theta}} \lambda[\widehat{\theta} \to \theta] - \kappa[\theta],$$

which is zero by (4.7). This verifies the first inequality in (4.10).

The second inequality is immediate, from the definition of $\overline{u}_{\max}$.

For the third inequality in (4.10), it suffices to show that *any* function $x : \Theta \to \Delta(X)$ (without imposing any IC or IR constraints) must satisfy $\sum_\theta \pi(\theta) E\overline{u}(x(\theta), \theta) \leq R^*$. For this, we reverse the "x term" calculations above to obtain

$$\begin{aligned}
&\sum_\theta \pi(\theta) E\overline{u}(x(\theta), \theta) \\
&\qquad = \sum_\theta \left(\left(\kappa[\theta] + \sum_{\widehat{\theta}} \lambda[\theta \to \widehat{\theta}] \right) Eu(x(\theta), \theta) - \sum_{\widehat{\theta}} \lambda[\widehat{\theta} \to \theta] Eu(x(\theta), \widehat{\theta}) \right) \\
&\qquad = \sum_\theta \sum_a \left(\left(\kappa[\theta] + \sum_{\widehat{\theta}} \lambda[\theta \to \widehat{\theta}] \right) u(a, \theta) - \sum_{\widehat{\theta}} \lambda[\widehat{\theta} \to \theta] u(a, \widehat{\theta}) \right) x_a(\theta) \\
&\qquad = \sum_\theta \sum_a -(\mu_a[\theta] + \nu[\theta]) x_a(\theta)
\end{aligned}$$

by (4.6)

$$\leq - \sum_\theta \sum_a \nu[\theta] x_a(\theta) = - \sum_\theta \nu[\theta] = R^*$$

by (4.8). This completes the argument for (4.10).

Notice, moreover, that the assumption that the variables $\lambda[\theta \to \widehat{\theta}]$, $\kappa[\theta]$ were the *optimal* dual variables was used only in proving the third inequality. For the first inequality, we just needed these variables to be nonnegative and satisfy the flow conservation equations (4.7).

We now summarize as a lemma the essential conclusions from the above discussion:

Lemma 4.1. *Consider a screening problem* (Θ, X, u, π) *in which* Θ *and* X *are finite, and* $\pi(\theta) > 0$ *for each* $\theta \in \Theta$*. Let* R^* *be the optimal profit. Then:*

(a) *There exist nonnegative numbers* $\lambda[\theta \to \widehat{\theta}]$, *for all* $\theta, \widehat{\theta} \in \Theta$ *(with* $\lambda[\theta \to \theta] = 0$*), and* $\kappa[\theta]$ *for all* $\theta \in \Theta$*, such that*

  - *the flow conservation equation (4.7) holds for each* θ*, and*
  - *with generalized virtual values* $\overline{u}(a, \theta)$ *defined as in (4.9), every possible mechanism* (x, t) *satisfies the chain of inequalities (4.10), and any optimal mechanism* (x^*, t^*) *satisfies them with equality.*

(b) *If* $\lambda[\theta \to \widehat{\theta}]$ *and* $\kappa[\theta]$ *are* any *nonnegative numbers satisfying (4.7) for each* θ*, and these values* $\lambda[\theta \to \widehat{\theta}]$ *are used to define* $\overline{u}(a, \theta)$ *via (4.9), then every mechanism satisfies the first inequality in (4.10).*

As indicated earlier, the generalized virtual values we have defined extend the traditional notion of ironed virtual values from the one-dimensional case. This is detailed in the supplementary appendix, which works through the correspondence in the case of the benchmark monopoly problem and shows how these two notions of virtual value coincide, so that our definition is a bona fide generalization.

We close this subsection by stating a couple of additional technical results about screening problems that are useful for the general proof. First is a simple continuity result, again for finite screening environments:

Lemma 4.2. *Consider a screening environment* (Θ, X, u)*, with* Θ, X *finite. Then, the maximum expected profit is continuous as a function of the distribution* π*.*

The proof is straightforward, so we leave it to the appendix.

Second is an approximation lemma due to Madarász and Prat (2012) that applies to continuous type spaces. It shows that when each type in a given screening problem is moved by a small amount, the principal's optimal profit does not degrade significantly, even though the optimal mechanism may be discontinuous. The idea is that any given mechanism can be made robust to slight changes in the type distribution by refunding a small fraction of the payment to the agent: doing so pads the incentive constraints in such a way that, if the agent is induced to change his chosen allocation, he does so in favor of more expensive allocations; and this effect outweighs any small change in the agent's location in type space.

Formally, say that two distributions $\pi, \pi' \in \Delta(\Theta)$ are *δ-close* if Θ can be partitioned into disjoint measurable sets $S_1, \ldots, S_r$ such that $d(\theta, \theta') < \delta$ for any θ, θ' in the same cell S_k, and $\pi(S_k) = \pi'(S_k)$ for each S_k.

Lemma 4.3. *(Madarász and Prat 2012) Take the environment (X, Θ, u) as fixed. For any $\epsilon > 0$, there exists a number $\delta > 0$ with the following property: For any mechanism (x, t), there exists a mechanism $(\widetilde{x}, \widetilde{t})$ such that*

(a) for any two types θ, θ' with $d(\theta, \theta') < \delta$, then $\widetilde{t}(\theta') > t(\theta) - \epsilon$;

(b) for any two distributions π, π' that are δ-close,

$$E_{\pi'}[\widetilde{t}(\theta)] > E_{\pi}[t(\theta)] - \epsilon.$$

For completeness, a proof is included in the supplementary appendix.

4.2 Main proof

Now we start the proof of the theorem proper. We return to the notation $(\Theta^g, X^g, u^g, \pi^g)$ for the component screening problems, and use (Θ, X, u) to refer to the joint environment; R^{*g} is the optimal profit in component problem g, and $R^* = \sum_g R^{*g}$. Again, we prove the theorem here under the assumption that the Θ^g and X^g are all finite, leaving the extension to the general case for the appendix.

Also assume that $\pi^g(\theta^g) > 0$ for each g and each type θ^g; this assumption will be dispensed with at the end.

Let $\lambda^g[\theta^g \to \widehat{\theta}^g]$ and $\kappa^g[\theta^g]$ be dual variables for the component-g problem, given by Lemma 4.1(a). Thus, they satisfy the flow conservation equations

$$\sum_{\widehat{\theta}^g} \lambda^g[\theta^g \to \widehat{\theta}^g] - \sum_{\widehat{\theta}^g} \lambda^g[\widehat{\theta}^g \to \theta^g] + \kappa^g[\theta^g] = \pi^g(\theta^g) \tag{4.12}$$

for each $\theta^g \in \Theta^g$, and with the virtual values defined by

$$\overline{u}^g(x^g, \theta^g) = u^g(x^g, \theta^g) - \sum_{\widehat{\theta}^g} \frac{\lambda^g[\widehat{\theta}^g \to \theta^g]}{\pi^g(\theta^g)} \left(u^g(x^g, \widehat{\theta}^g) - u^g(x^g, \theta^g) \right),$$

we have from the equality case of (4.10) that

$$\sum_{\theta^g} \pi^g(\theta^g) \overline{u}^g_{\max}(\theta^g) = R^{*g}, \qquad \text{where } \overline{u}^g_{\max}(\theta^g) = \max_{x^g} \overline{u}^g(x^g, \theta^g).$$

Also note for future reference that, for each fixed g, if we sum up (4.12) over all θ^g, the $\lambda^g[\cdots]$ terms cancel, and we are left with

$$\sum_{\theta^g} \kappa^g[\theta^g] = \sum_{\theta^g} \pi^g(\theta^g) = 1. \tag{4.13}$$

We now begin to construct our joint distribution π and dual variables $\lambda[\theta \to \widehat{\theta}], \kappa[\theta]$ for the joint problem. We first define the κ's, by putting

$$\kappa[\theta] = \prod_g \kappa^g[\theta^g].$$

If we multiply (4.13) across all g, we see that

$$\sum_{\theta \in \Theta} \kappa[\theta] = 1. \tag{4.14}$$

And if we fix any g and θ^g, and multiply (4.13) only for all other components $h \neq g$, we likewise get

$$\sum_{\theta^{-g} \in \Theta^{-g}} \kappa[\theta^g, \theta^{-g}] = \kappa^g[\theta^g] \tag{4.15}$$

which will also be useful later.

Next we will define π as the solution to a system of linear equations analogous to (4.12). This system is given by (4.19) below. To briefly motivate it, consider the formula for the generalized virtual value in the joint screening problem: for any $\theta = (\theta^1, \ldots, \theta^G)$ and $x = (x^1, \ldots, x^G)$,

$$\begin{aligned} \overline{u}(x,\theta) &= u(x,\theta) - \sum_{\widehat{\theta} \in \Theta} \frac{\lambda[\widehat{\theta} \to \theta]}{\pi(\theta)} \left(u(x,\widehat{\theta}) - u(x,\theta) \right) \\ &= \sum_g \left(u^g(x^g,\theta^g) - \sum_{\widehat{\theta} \in \Theta} \frac{\lambda[\widehat{\theta} \to \theta]}{\pi(\theta)} \left(u^g(x^g,\widehat{\theta}^g) - u^g(x^g,\theta^g) \right) \right) \end{aligned}$$

by additive separability of u. Our plan is to put positive weight only on IC constraints $\widehat{\theta} \to \theta$ for which $\widehat{\theta}, \theta$ differ in just one coordinate. Then, for each g, the difference terms $\left(u^g(x^g,\widehat{\theta}^g) - u^g(x^g,\theta^g) \right)$ for which $\widehat{\theta}$ and θ differ in a coordinate $h \neq g$ cancel, and we are only left with the terms where they differ in coordinate g; thus our virtual value will become

$$\sum_g \left(u^g(x^g, \theta^g) - \sum_{\widehat{\theta}^g \in \Theta^g} \frac{\lambda[(\widehat{\theta}^g, \theta^{-g}) \to (\theta^g, \theta^{-g})]}{\pi(\theta)} \left(u^g(x^g, \widehat{\theta}^g) - u^g(x^g, \theta^g) \right) \right). \tag{4.16}$$

We wish to choose the distribution π and the λ's so that the separate-screening mechanism maximizes generalized virtual value. This will be accomplished if the generalized virtual value of each type θ is the sum of the component virtual values; and for this it is sufficient that, given π, the λ's should be defined by

$$\lambda[(\widehat{\theta}^g, \theta^{-g}) \to (\theta^g, \theta^{-g})] = \frac{\pi(\theta)}{\pi^g(\theta^g)} \times \lambda^g[\widehat{\theta}^g \to \theta^g]. \tag{4.17}$$

We will also need the joint distribution π and the dual variables $\lambda[\theta \to \widehat{\theta}]$, $\kappa[\theta]$ to satisfy the flow conservation equations

$$\sum_g \sum_{\widehat{\theta}^g} \lambda[(\theta^g, \theta^{-g}) \to (\widehat{\theta}^g, \theta^{-g})] - \sum_g \sum_{\widehat{\theta}^g} \lambda[(\widehat{\theta}^g, \theta^{-g}) \to (\theta^g, \theta^{-g})] + \kappa[\theta] = \pi(\theta) \tag{4.18}$$

for each $\theta \in \Theta$. Substituting for the λ's from (4.17), this is equivalent to

$$\sum_g \sum_{\widehat{\theta}^g} \frac{\pi(\widehat{\theta}^g, \theta^{-g})}{\pi^g(\widehat{\theta}^g)} \times \lambda^g[\theta^g \to \widehat{\theta}^g] - \sum_g \sum_{\widehat{\theta}^g} \frac{\pi(\theta)}{\pi^g(\theta^g)} \times \lambda^g[\widehat{\theta}^g \to \theta^g] + \kappa[\theta] = \pi(\theta) \tag{4.19}$$

for each $\theta \in \Theta$.

System (4.19) consists of $|\Theta|$ linear equations in $|\Theta|$ unknowns (the probabilities $\pi(\theta)$), whose coefficients come from the component-g problems. The crucial step in our proof is showing that this system indeed has a solution π, which moreover is a probability distribution with the correct marginals.

Lemma 4.4. *Under the assumptions of this subsection (Θ^g finite and $\pi^g(\theta^g) > 0$), there exists a full-support distribution $\pi \in \Pi$ satisfying the flow conservation equations (4.19).*

Proof. First, for each g, we consider a continuous-time Markov process whose state space is Θ^g. (This should not be thought of as an agent's type changing over time; it is simply an abstract process that happens to have this state space.) The process works as follows. There are two types of state changes, both triggered by Poisson arrivals:

- To each type $\widehat{\theta}^g \in \Theta^g$ is associated a Poisson clock; when the clock ticks, the state changes to $\widehat{\theta}^g$. The intensity of this clock depends on the current state: it is

$\lambda^g[\widehat{\theta}^g \to \theta^g]/\pi^g(\theta^g)$ when the current state is θ^g. (The Poisson clocks corresponding to different $\widehat{\theta}^g$ run independently conditional on the current state θ^g.)

- In addition, there is an independent "reset" clock at constant Poisson rate 1. When a reset arrives, the type changes to each θ^g with probability $\kappa^g[\theta^g]$. (Equation (4.13) ensures this makes sense — the probabilities add up to 1.)

Let $\rho^g(\theta^g)$ denote the stationary distribution of this Markov process. Because there is a single recurrent set, the stationary distribution is uniquely determined as the solution to the system of linear equations

$$\sum_{\widehat{\theta}^g} \rho^g(\widehat{\theta}^g) \frac{\lambda^g[\theta^g \to \widehat{\theta}^g]}{\pi^g(\widehat{\theta}^g)} + \kappa^g[\theta^g] = \rho^g(\theta^g) \left(\sum_{\widehat{\theta}^g} \frac{\lambda^g[\widehat{\theta}^g \to \theta^g]}{\pi^g(\theta^g)} + 1 \right)$$

for all $\theta^g \in \Theta^g$. The left-hand side represents the steady-state frequency of changes into state θ^g, and the right-hand side the frequency of changes out of state θ^g. However, from (4.12), we see that distribution π^g is a solution to this system, so it must actually be the stationary distribution, $\rho^g(\theta^g) = \pi^g(\theta^g)$.

Now consider a new Markov process whose state space is the joint type space Θ, and which is governed by Poisson state changes as follows:

- For each component g, and each $\widehat{\theta}^g \in \Theta^g$, there is a Poisson clock. When it ticks, component g of the state changes to $\widehat{\theta}^g$, i.e. if the current state is θ then it changes to $(\widehat{\theta}^g, \theta^{-g})$. This clock has an intensity of $\lambda^g[\widehat{\theta}^g \to \theta^g]/\pi^g(\theta^g)$ when the current state is θ.

- In addition, there is a "reset" clock with rate 1; when a reset arrives, the entire state is redrawn, with probabilities $\kappa[\theta]$ for each θ. (This makes sense, by (4.14).)

These clocks are independent across g and $\widehat{\theta}^g$, conditional on the current state.

Let π denote the stationary distribution of this Markov process. This stationary distribution must satisfy the equations

$$\sum_g \sum_{\widehat{\theta}^g} \pi(\widehat{\theta}^g, \theta^{-g}) \frac{\lambda^g[\theta^g \to \widehat{\theta}^g]}{\pi^g(\widehat{\theta}^g)} + \kappa[\theta] = \pi(\theta) \left(\sum_g \sum_{\widehat{\theta}^g} \frac{\lambda^g[\widehat{\theta}^g \to \theta^g]}{\pi^g(\theta^g)} + 1 \right)$$

for all θ — which is exactly (4.19) above.

Also, because each θ^g is a recurrent state of the Markov process for component g, it is not hard to see that any state θ is reached with positive probability, so π has full support.

Finally, we can see that for each component g of the state, the arrival rate of each type of event that affects component g (either component changes or resets) depends only on the current component g and is otherwise independent of other components. Hence, the evolution of component g of the state is described exactly by our earlier Markov process with state space Θ^g. (This conclusion makes use of (4.15), which ensures that when a reset event occurs, the new θ^g is indeed drawn from distribution κ^g.) Consequently, the stationary distribution π has component-g marginal equal to π^g. Thus, $\pi \in \Pi$. $\square$

Now, armed with this joint distribution π, define the IC multipliers $\lambda[(\widehat{\theta}^g, \theta^{-g}) \to (\theta^g, \theta^{-g})]$ by (4.17) — and write $\lambda[\widehat{\theta} \to \theta] = 0$ whenever $\widehat{\theta}, \theta$ differ in more than one component. This ensures (4.18) is satisfied, since it follows from (4.19).

We can use these multipliers to show that no mechanism earns total profit more than R^* against distribution π. To start, define $\overline{u}(x, \theta)$, for each $x \in X$ and $\theta \in \Theta$, to be the virtual value in the joint screening problem with these multipliers and distribution. Then, the virtual values are additively separable, as we can confirm by retracing our steps above:

$$\begin{aligned}
\overline{u}(x,\theta) &= \sum_g \left(u^g(x^g,\theta^g) - \sum_{\widehat{\theta}^g} \frac{\lambda[(\widehat{\theta}^g,\theta^{-g}) \to (\theta^g,\theta^{-g})]}{\pi(\theta)} \Big(u^g(x^g,\widehat{\theta}^g) - u^g(x^g,\theta^g) \Big) \right) \\
&= \sum_g \left(u^g(x^g,\theta^g) - \sum_{\widehat{\theta}^g} \frac{\lambda^g[\widehat{\theta}^g \to \theta^g]}{\pi^g(\theta^g)} \Big(u^g(x^g,\widehat{\theta}^g) - u^g(x^g,\theta^g) \Big) \right) \\
&= \sum_g \overline{u}^g(x^g,\theta^g).
\end{aligned}$$

Here, the first line is the calculation we have already performed in (4.16), using the additive separability of u and the specification that $\lambda[\widehat{\theta} \to \theta]$ is zero unless $\widehat{\theta}$ and θ differ in just one coordinate. And in the second line, we have used (4.17).

Now, we have not established that the multipliers here are an optimal solution to the dual of the joint problem. Nonetheless, we know that they satisfy the flow conservation equations for the joint problem, (4.18). Therefore, we can apply part (b) of Lemma 4.1: For any mechanism (x, t) in the joint screening problem,

$$\begin{aligned}
\sum_\theta \pi(\theta)t(\theta) &\leq \sum_\theta \pi(\theta)E\overline{u}(x(\theta),\theta) \\
&= \sum_\theta \pi(\theta)\sum_g E\overline{u}^g(x^g(\theta),\theta^g).
\end{aligned}$$

The right side thus separates partially across components — but it does not separate completely, in that the argument to the component virtual utility is $x^g(\theta)$, which can potentially depend on every component of θ and not just on θ^g. However, we have $E\overline{u}^g(x^g(\theta),\theta^g) \leq \overline{u}^g_{\max}(\theta^g)$, by definition of $\overline{u}^g_{\max}$. Therefore, we can complete the chain:

$$\begin{aligned}
\sum_\theta \pi(\theta)t(\theta) &\leq \sum_\theta \sum_g \pi(\theta)E\overline{u}^g(x^g(\theta),\theta^g) \\
&= \sum_g \left(\sum_\theta \pi(\theta)E\overline{u}^g(x^g(\theta),\theta^g)\right) \\
&\leq \sum_g \left(\sum_\theta \pi(\theta)\overline{u}^g_{\max}(\theta^g)\right) \\
&= \sum_g \left(\sum_{\theta^g}\left(\sum_{\theta^{-g}} \pi(\theta^g,\theta^{-g})\right)\overline{u}^g_{\max}(\theta^g)\right) \\
&= \sum_g \left(\sum_{\theta^g} \pi^g(\theta^g)\overline{u}^g_{\max}(\theta^g)\right) \\
&= \sum_g R^{*g} \\
&= R^*.
\end{aligned}$$

So any joint screening mechanism gives expected profit at most R^* with respect to the type distribution π, which is what we wanted to show.

This finishes the case where each π^g has full support. To allow for probability-zero types, we use our earlier continuity result, Lemma 4.2.

For each g, suppose π^g is an arbitrary distribution on Θ^g, with R^{*g} the corresponding maximum expected profit. Let $\pi^g_1, \pi^g_2, \ldots$ be a sequence of full-support distributions on Θ^g that converges to π^g. Then, if we let $R^{*g}_1, R^{*g}_2, \ldots$ be the corresponding values of the maximum expected profit, Lemma 4.2 says that $R^{*g}_n \to R^{*g}$ as $n \to \infty$.

For each n, the proof we have just completed shows that there exists a joint distribution

π_n on Θ, with marginals π_n^g, such that no mechanism earns an expected profit of more than ${R^*}_n = \sum_g {R^*}_n^g$ with respect to π_n. By compactness, we can assume (taking a subsequence if necessary) that π_n converges to some distribution π on Θ. Then π has marginal π^g on each Θ^g. And for any mechanism (x, t), we have $E_{\pi_n}[t(\theta)] \leq {R^*}_n$; taking limits as $n \to \infty$, we have $E_\pi[t(\theta)] \leq R^*$. Thus, no mechanism earns expected profit more than R^* on distribution π.

This completes the proof of the theorem when the type and allocation spaces are finite. The extension to the general case is basically technical and consists of two steps. First, we keep type spaces finite but extend to general allocation spaces; this part is completely straightforward. Second, we remove the restriction to finite type spaces, by approaching the distributions π^g using arbitrarily fine discrete approximations; here is where we use Lemma 4.3 to show that the approximation has only a small effect on optimal profit, both in the component problems and in the joint problem. The details of these arguments are left to the appendix. $\square$

At this point, we can briefly remark on the construction of π embedded in Lemma 4.4, and informally relate it back to the attempts in Section 3. Consider the benchmark monopoly application, and suppose the marginal distributions π^g have increasing virtual values. Take a fine discretization of the type space. For each good g separately, the standard solution to the dual problem puts positive weight only on the adjacent downward incentive constraints and the IR constraint of the lowest type. (This dual solution is constructed explicitly in Section C of the supplementary appendix.) Thus, the component-g Markov process in the proof of Lemma 4.4 runs as follows: the state θ^g gradually increases from one type to the next higher type according to a Poisson clock, interrupted by resets at which the state moves back to the lowest type. In the continuous-type limit, the state drifts deterministically upward in the type space, until a reset occurs, when the process returns to the lowest type. Accordingly, the corresponding joint type process runs as follows: in between resets, the state θ drifts deterministically upward in every component; when a reset occurs, all components jump back to their lowest possible value. Consequently the movements of all components are perfectly synchronized, so the stationary distribution must be the comonotonic distribution that we saw in Section 3. (However, for any finite approximation of the type space, the components are not perfectly synchronized, and the resulting joint distribution π has full support.)

5 Sensitivity analysis

A natural response to Theorem 2.1 is to ask how sensitive the result is to the assumption of extreme uncertainty. We have assumed that the principal solves a maxmin problem over the entire space of joint distributions Π. On the other hand, if we were to assume that the principal knows π exactly, typically separate screening would not be optimal. But what happens if we shrink the space of uncertainty just slightly? That is, does separate screening remain optimally robust if the principal has a little bit of information about the joint distribution? That is the question we explore in this section. For concreteness, we focus on the benchmark monopoly application.

One immediate difficulty is that it is not clear exactly what it should mean to have "a little information" about π. For example, if we assume that the principal knows that some alternative mechanism (x', t') earns higher profit than separate screening, then this knowledge might be considered "a little information," but it trivially overturns the result.

To express this more formally: One might describe the principal's additional information by imposing some one-dimensional moment restriction on π, of the form

$$E_{\pi}[z(\theta)] = 0 \tag{5.1}$$

or

$$E_{\pi}[z(\theta)] \geq 0 \tag{5.2}$$

for some function $z : \Theta \to \mathbb{R}$, and let Π' be the set of all $\pi \in \Pi$ satisfying the restriction. We would then evaluate a mechanism (x, t) according to $\inf_{\pi \in \Pi'} E_{\pi}[t(\theta)]$ instead of the original objective (2.3), and ask whether separate sales is still optimal. The example in the preceding paragraph shows how we can easily cook up a moment restriction to make the answer negative. The question then seems to be, what are the interesting restrictions to consider?

We explore two approaches here. First, we consider one particular moment restriction suggested by the literature, namely negative correlation in values between two goods, and explore its consequences numerically. Second, we show that in the finite-type version of the model, in general there is an open set of distributions π for which separate pricing is optimal. In the continuous-type version, this is not the case, but at least there is an open set of moment restrictions $z(\theta)$ one can impose on the maxmin problem such that separate pricing remains optimal. These findings give some assurance that Theorem 2.1 is not a knife-edge result, without needing to take a stand on specific moment restrictions.

Proofs of all results from this section can be found in the supplementary appendix.

5.1 Negative correlation

One of the main intuitions from the early bundling literature is that bundling is profitable when values for goods are negatively correlated. This argument has been passed down as received wisdom (see e.g. the undergraduate text of Church and Ware (2000, pp. 169–170)), although it comes chiefly from examples rather than any general theorem (Stigler 1963, Adams and Yellen 1976): for example, in the extreme case where the total value for all goods is deterministic, the seller can extract the full surplus by selling the bundle of all goods. Negative correlation is also in some sense opposite to the positively-correlated case of Subsection 3.2. For both these reasons, negative correlation seems like a natural place to look for restrictions that might overturn the separation result.

Accordingly, let us take the restriction of negative correlation literally, and test it by considering $G = 2$ goods and imposing the moment restriction

$$E_{\pi}[\theta^1\theta^2] \leq E_{\pi^1}[\theta^1] \times E_{\pi^2}[\theta^2]$$

on the possible joint distributions $\pi \in \Pi'$. It turns out that with this restriction, it may or may not happen that the worst-case-optimal revenue is still the R^* from selling separately.

To get some sense of whether one case or the other is common, random numerical experiments were run in Matlab, using finite type spaces Θ. These were computationally implemented via the following more general observation: For any one-dimensional moment restriction of the form $E_{\pi}[z(\theta)] \geq 0$, the worst-case optimization problem

$$\max_{(x,t)\in\mathcal{M}} \left(\min_{\pi\in\Pi'} E_{\pi}[t(\theta)] \right) \tag{5.3}$$

can be rewritten as a standard linear program. To see this, observe that for any given mechanism (x, t), the inner minimization in (5.3) is a linear program: choose probabilities $\pi(\theta) \geq 0$ to minimize $\sum_{\theta} t(\theta)\pi(\theta)$ subject to constraints $\sum_{\theta^{-g}} \pi(\theta^g, \theta^{-g}) = \pi^g(\theta^g)$ for each $g = 1, \ldots, G$ and each $\theta^g \in \Theta^g$, and $\sum_{\theta} z(\theta)\pi(\theta) \geq 0$. By duality, this program's value can also be obtained by a maximization problem, namely

$$\max_{\alpha^g[\theta^g],\beta} \sum_{g} \sum_{\theta^g} \pi^g(\theta^g)\alpha^g[\theta^g]$$

over all choices of real numbers $\alpha^g[\theta^g]$ (for each g and θ^g) and $\beta \geq 0$ satisfying

$$\sum_g \alpha^g[\theta^g] + \beta z(\theta) \leq t(\theta) \qquad \text{for all } \theta = (\theta^1, \ldots, \theta^G) \in \Theta. \tag{5.4}$$

Therefore, the maxmin problem in (5.3) can be expressed as a single maximization — over all choices of the mechanism (x, t), and all choices of $\alpha^g[\theta^g]$ and β satisfying (5.4) — that can be solved by standard LP methods.

In the simulations, we randomly generated 1000 choices for the marginal distributions (π^1, π^2), calculated the mechanism that maximizes the worst-case revenue over negatively correlated distributions by the method described above, and checked whether the worst-case revenue was strictly higher than obtained by selling the two goods separately. The set of possible values for each good was $\Theta^1 = \Theta^2 = \{1, 2, 3, 4, 5\}$, and the marginals π^1, π^2 were generated by drawing a probability uniformly from $[0, 1]$ for each value, then rescaling to make the probabilities sum to 1.

The result was that in 975 of the 1000 trials, the worst-case revenue was still the R^* from selling separately. That is, based on the simulation results, the optimality of separate sales is usually robust to assuming that the values are known to be negatively correlated.

One might protest that this sensitivity test is too weak, because it is inappropriate to take negative correlation so literally; it is no surprise that one misspecified inequality restriction often fails to rule out some worst-case distributions. It is not clear what an appropriate sharper test would be. One possible test would be to impose negative *affiliation* on π — that is, negative correlation conditional on $\theta \in \widetilde{\Theta}^1 \times \widetilde{\Theta}^2$, for all nonempty measurable subsets $\widetilde{\Theta}^1 \subseteq \Theta^1, \widetilde{\Theta}^2 \subseteq \Theta^2$. This limits the possible π to a much smaller set, so one would expect separate sales to be worst-case optimal much less often.

It is not immediately clear how to repeat the above computational exercise with the restriction of negative affiliation, since it is not a linear constraint on π. However, if we move to the continuous-type setting, we can obtain a negative result: separate sales is (essentially) *always* dominated by bundling. The dominance is weak for a given bundle price, but can be made strict by randomizing the bundle price.

Proposition 5.1. *Consider the benchmark monopoly application, with $G = 2$. Suppose each set of values Θ^g is an interval in $\mathbb{R}^{++}$, and each marginal distribution π^g is represented by a continuous, positive density f^g. Assume that for each g, the optimal separate price p^{*g} is strictly in the interior of Θ^g. Then there exist $\underline{\epsilon}, \overline{\epsilon} > 0$, $\underline{q} \in (0, 1)$, and $R' > R^*$*

such that the following hold:

(a) *The mechanism that lets the buyer choose either good g with price p^{*g}, or the bundle of both goods at price $p^{*1}+p^{*2}-\underline{\epsilon}$, earns expected profit at least R^* for any negatively affiliated joint distribution $\pi \in \Pi$.*

(b) *Consider the following mechanism: first randomly choose $\epsilon = \underline{\epsilon}$ with probability $\underline{q}$ or $\epsilon = \overline{\epsilon}$ with probability $\overline{q} = 1 - \underline{q}$; then offer each good g with price p^{*g}, or the bundle of both goods at price $p^{*1} + p^{*2} - \epsilon$. This mechanism earns expected profit at least R', for any negatively affiliated joint distribution $\pi \in \Pi$.*

The argument (in the supplementary appendix) is a straightforward extension of that given by McAfee, McMillan, and Whinston (1989) for the independent case, obtained by looking at the first-order condition for ϵ.

5.2 Open sets of distributions

Another approach to sensitivity analysis, which avoids making any particular choices of moment restrictions, is to look at the set of distributions π for which separate pricing is optimal, and ask how small this set is. If it contained only the one specific distribution constructed in Section 4, this might be cause to view the conclusion of Theorem 2.1 as a knife-edge result. However, we will show here that the situation is not so extreme: in particular, when the type space is *finite*, there typically exists an open set of such worst-case π's (in the natural topology on Π).

We could also simply try to show that there is an open set of possible moment restrictions z in (5.1) (or (5.2)) for which the worst-case optimal revenue is R^*. Note however that this is a much weaker statement than existence of an open set of π's as above. In fact, as long as there is more than one π — say π_1 and π_2 — for which optimal revenue is R^*, then we can allow any z in the open set satisfying $E_{\pi_1}[z(\theta)] < 0$ and $E_{\pi_2}[z(\theta)] > 0$. There is then some convex combination π' of π_1 and π_2 for which $E_{\pi'}[z(\theta)] = 0$, and for this distribution, no mechanism can earn more than R^*. So for any such z, the worst-case objective (5.3) equals R^*.

To get to our open-set result, the important step is the following:

Lemma 5.2. *Consider the benchmark monopoly application. Suppose each set of values Θ^g is finite, that each π^g has full support, and also suppose that for each single-good problem π^g, the optimal price is unique. Then, there exists a joint distribution $\pi \in \Pi$ for which selling each good separately constitutes the* unique *optimal mechanism.*

The proof (in the supplementary appendix) involves a more careful retracing of the generalized-virtual-value maximization arguments in Section 4.

From here we can infer the result:

Corollary 5.3. *In the setting of Lemma 5.2, there exists a set of joint distributions $\widehat{\Pi} \subseteq \Pi$, which is open (in the relative topology on Π, as a subset of $\mathbb{R}^{|\Theta|}$), and such that for any $\pi \in \widehat{\Pi}$, no joint screening mechanism earns an expected profit higher than R^*.*

This quickly follows because the set of possible mechanisms is effectively a convex polytope, so there are only finitely many choices for the optimum (namely, the polytope's vertices). Lemma 5.2 shows that there exists π for which the separate sales mechanism is strictly better than all the other candidate optima, so by continuity it remains optimal for nearby distributions. The formal proof is, again, in the supplementary appendix.

We note that Corollary 5.3 does not extend to the continuous-type case. A brief argument for why can be seen following McAfee, McMillan, and Whinston (1989): Assume two goods; assume that each Θ^g is an interval in $\mathbb{R}^+$, with a unique optimal price p^{*g} that is in the interior of the interval, and assume that each π^g has a continuous density. Restrict attention to joint distributions π that also have continuous densities.[4] Consider the mechanism that offers good 1 for price p^{*1}, good 2 for price p^{*2}, or both goods for the price $p^{*1} + p^{*2} - \epsilon$, where ϵ may be positive, negative, or zero. The profit of this mechanism, under any given π, is differentiable as a function of ϵ. Separate pricing is optimal only if the first-order condition is satisfied: the derivative of profit with respect to ϵ should be zero at $\epsilon = 0$. If one writes out this condition explicitly, it depends on the joint distribution, and is easily seen to be violated for generic choices of π. Note that this argument does not apply in the discrete-type situation of Corollary 5.3 because then the profit is not differentiable as a function of ϵ.

Even though the optimal mechanism is more sensitive to changing π in the continuous-type case, there is still typically an open set of moment restrictions z under which the worst-case optimum remains R^*. As described at the beginning of this subsection, this holds as long as there is more than one choice of joint distribution π that pins profit down to R^* (which will be true in general).

In the continuous-type setting, it should be no surprise that slight perturbations of the joint distribution lead to changes in the optimal mechanism: The principal optimizes

[4] This restriction is not innocuous here, because Theorem 2.1 had no such restriction, and in fact our proof technique can lead to distributions that do not have a density. However, one can formulate a proposed counterpart of Corollary 5.3 for continuous types, without restricting π to have a density, and show that such a counterpart is false by building on the intuition here. Details are omitted for brevity.

over a large space of mechanisms, and we cannot expect to get strong functional form conclusions about the solution to such an optimization problem without similarly strong functional form assumptions on the setup of the problem. Still, continuity arguments (such as Lemma 4.3) can be used to argue that if the distribution π is perturbed slightly, separate screening remains approximately optimal.

6 Concluding comments

We wrap up by discussing past, present, and future: comparing our analytical techniques to previous methods in multidimensional screening; discussing interpretation of the present model; and aspirations for subsequent work.

6.1 Related multidimensional screening work

There is a substantial previous literature in economics and computer science on multidimensional screening, too diverse to survey exhaustively here. Much of this work studies versions of the (Bayesian) multi-good monopolist problem, usually using continuous type distributions. In some (e.g. Armstrong 1996), special cases are solved by partitioning the type space into one-dimensional paths, solving on each path, and then checking that the resulting mechanism is actually incentive compatible. A more widespread, and more broadly applicable, approach generalizes the standard method of integration by parts from single-dimensional screening problems. In principle, the method extends naturally to multiple dimensions: one constructs a virtual value that is related to the original type distribution by a differential equation, so that profit can be expressed as the integral of virtual value via the divergence theorem, and one then maximizes this virtual value pointwise to find the optimal mechanism. However, as explained in Rochet and Choné (1998), there is a large multiplicity of choices of virtual value, due to freedom in choosing boundary conditions, and most such choices do not lead to the correct solution. A second difficulty is that even if the boundary is identified correctly, typically the pointwise virtual value maximizer still is not incentive compatible because ironing is necessary; giving reasonable assumptions on primitives to rule out ironing is apparently more difficult in the multidimensional model than the one-dimensional model.

Several significant contributions (Rochet and Choné 1998, Daskalakis, Deckelbaum, and Tzamos 2013, Daskalakis, Deckelbaum, and Tzamos 2015) directly address ironing by offering "guess-and-check" methods that can be sometimes used to find the ironing regions, to thereby identify the optimal mechanism, and to find a dual solution to certify optimality. The arguments in these papers optimize over indirect utility functions (specifying the utility that each type of agent obtains in the mechanism), rather than working directly in the space of mechanisms, as in the present paper. Another work, closer to the present paper, is Haghpanah and Hartline (2015). That paper works in the space of mechanisms, and solves a class of problems by using the optimality of a particular candidate mechanism to pin down the correct choice of virtual value, then verifies that the mechanism is indeed optimal by showing that it is a virtual value maximizer. (The main results concern specifications where no ironing is needed, although ironing does receive some attention.)

The differential equation characterizing virtual values (appearing, for example, in equation (4.8) of Rochet and Choné (1998), equation (3.4) of Rochet and Stole (2003), or the "divergence density equality" in Haghpanah and Hartline (2015)) is a continuous-type analogue of our flow conservation equations (4.7). Again, though, it is based only on local incentive constraints and does not take account of ironing. In this paper, an important feature is that we make no assumptions about the structure of preferences within each component, and so there is no presumption that only local incentive constraints are needed. Indeed, we could have restricted the model to ensure that only local incentive constraints are needed, but as the discussion in Subsection 3.2 shows, doing so would miss out on much of the subtlety of the result.

The concept of generalized virtual values, central to this paper, does have antecedents in the literature. Most closely related is the independent and simultaneous work of Cai, Devanur, and Weinberg (2016). That paper gives essentially the same definition (4.9) for generalized virtual values, and proves Lemma 4.1 (both parts). However, the main focus of that paper is not on identifying the optimal mechanism as here, but rather on showing that some simple mechanisms are approximately optimal; to do so, it uses a simple but non-optimal choice of dual variables to obtain bounds on the optimal revenue. Other, earlier incarnations of our formula (4.9) appear in Sher and Vohra (2015), which gives a similar formula in a variant of a single-good monopoly problem; and Myerson (1991, Section 10.5), which studies a problem of welfare maximization, with multiple agents and no transfers.[5] The present paper seems to be the first to use these generalized virtual values to characterize the exact optimum in a model of this generality.

[5]The work Cai, Daskalakis, and Weinberg (2012) also defines a different object called "virtual types" in the context of a multiple-good, multi-buyer auction. The virtual types there play a role in the algorithmic description of the optimal auction, but do not have an economic interpretation corresponding to ours.

The papers by Sher and Vohra (2015) and Cai, Devanur, and Weinberg (2016) also walk through the connection with traditional single-good virtual values, just as we describe in the supplementary appendix. And these two papers introduce the "flow conservation" terminology, which we have followed here.

6.2 Interpretation

Our model can be viewed in various ways. One interpretation is positive: a problem faced by an actual decision maker, who is uncertain about the joint distribution governing the agent's type and has maxmin preferences (which could be interpreted as extreme ambiguity aversion à la Gilboa and Schmeidler (1989), with set of priors given by our Π). Indeed the prediction of the model squares with some common experience: when a buyer walks into a store that sells a thousand different items, she sees separate posted prices for each item (and perhaps special deals for a few combinations of items that are naturally grouped together). Why has the storekeeper not solved the high-dimensional problem of optimally selling all items simultaneously? One perspective suggested by our model invokes a particular form of bounded rationality: perhaps the manager can easily estimate the distribution of values for each item separately, but thinking about joint distributions becomes too unwieldy, or unreliable due to the curse of dimensionality, so the storekeeper avoids it. Then, selling separately provides the best guarantee that can be reached using only this limited information.

Such an interpretation is open to criticism: is it reasonable to make the extreme combination of assumptions, that the principal has perfect knowledge of the marginals yet no knowledge of the joint? One response is that, once we set out to avoid the extreme assumptions of either perfect knowledge of π or no knowledge at all, almost any specific modeling assumption in between would be open to question. The assumption made here at least is easy to state and gives a simple solution. But in fact, one can also give versions of the result with other informational assumptions. For example, suppose the principal is not even sure about the marginals: for each component g, she knows only some set of distributions Π^g. (Bergemann and Schlag (2011), Carrasco, Luz, Monteiro, and Moreira (2015) and Kos and Messner (2015), among others, study maxmin screening problems of this form.) Suppose her overall uncertainty set Π consists of all distributions whose marginal for each g lies in Π^g; she seeks a mechanism to maximize worst-case profit over this set. If each Π^g is sufficiently well-behaved to contain a single worst-case distribution, then as a corollary of our theorem, this principal will again optimize by solving the maxmin screening problem for each g separately.

Still, one might be uncomfortable giving such a literal interpretation to any model that juxtaposes the extremes of Bayesian and worst-case behavior. This is why our preferred view of Theorem 2.1 is as a methodological contribution: Previous literature on multidimensional screening does not offer a general model that makes the "null" prediction of no bundling. Arguably, it is helpful to have such a model, and this paper gives one.

This model helps to formalize a particular robustness property of selling separately. It also provides a theoretical benchmark, from which to better understand why various forms of bundling do arise in other models. For example, as discussed already, the elegant argument in McAfee, McMillan, and Whinston (1989) shows that, for generic continuous distributions, the first-order condition for separate pricing to be optimal fails: some form of bundling does better. That paper had left open the question of what happens in the non-generic case where the condition does hold. If bundling remained optimal, then arguably the reason for bundling would really not come from the distribution at all but from something even more basic. Theorem 2.1 responds to this, by showing that we cannot conclude that bundling is desirable without at least some knowledge of the distribution.

6.3 Future directions

As described in the introduction, the main purpose of this paper has been to advance a possible new perspective on multidimensional screening problems, to contrast with the traditional approach, which has often given intractable, or at least complicated, models. Specifically, we have formulated a model in which assuming no interactions between the components leads to no interactions in the solution.

This is a first step. There are two natural goals for further work to build on this model and proof:

- First, to connect formally with the simple intuition sketched back in the introduction, that the separate screening mechanism has the following property: knowing the marginal type distributions is enough to determine the expected profit, without knowing anything further about the joint distribution. Call this property *joint-independence*. This property makes the mechanism an intuitively natural candidate for the maxmin optimum. Can our proof (or perhaps some completely different proof technique) be expressed in a way that centers on this property?

  This might seem an obvious approach to try, but there is subtlety involved. If we give up the assumption that the agent's preference be additively separable across components g, and allow for more general preferences, then it is not hard to write down examples of similar-looking problems — maxmin profit with known marginal type distributions but unknown joint distribution — where the unique optimal mechanism does not satisfy joint-independence. (We omit the details here.) To put it another way, the connection from joint-independence to optimality is not automatic, but rather depends on some deeper properties of the problem.

- Second, to have fruitful applications of the method to new economic questions. More specifically, this would mean identifying some applied theory question involving multidimensional screening, where a traditional Bayesian model would be intractable, and where using the robust approach to obtain a solution turns out to be helpful in thinking through the economics of the problem. As noted before, the main applied value of studying multidimensional screening models is in understanding interactions between the dimensions, which is exactly what the model here has ruled out. However, this paper may be useful as a clean starting point, from which interactions can be systematically introduced.

  Earlier work by this author followed this course of development in a moral hazard setting, taking a simple robust approach to principal-agent contracting (Carroll 2015b) and applying it to gain traction on the more complex problem of incentivizing information acquisition (Carroll 2015a). To do the same in a multidimensional screening context is still a task that lies ahead.

A Omitted proofs

Note: the separate supplementary appendix contains further material (Sections B and C).

Proof of Proposition 3.1. For notational convenience, let us parameterize types directly by z, and write $\theta^g(z)$ for the corresponding value for good g. Thus, for any mechanism (x,t), $x(z) \in [0,1]^G$ indicates type z's probability of receiving each good, and $t(z) \in \mathbb{R}$ indicates z's payment; incentive compatibility (2.1) becomes

$$\sum_g \theta^g(z)x^g(z) - t(z) \geq \sum_g \theta^g(z)x^g(\widehat{z}) - t(\widehat{z}) \qquad \text{for all } z, \widehat{z} \in [0,1], \tag{A.1}$$

and individual rationality (2.2) becomes

$$\sum_g \theta^g x^g(z) - t(z) \geq 0. \tag{A.2}$$

The claim is that for any such mechanism, expected revenue satisfies

$$\int_0^1 t(z)\, dz \leq R^*.$$

Following the standard method, let $U(z) = \sum_g \theta^g(z)x^g(z) - t(z)$ denote the payoff of type z. So for $z' \geq z$, (A.1) gives

$$\sum_g (\theta^g(z') - \theta^g(z))x^g(z') \geq U(z') - U(z) \geq \sum_g (\theta^g(z') - \theta^g(z))x^g(z). \tag{A.3}$$

Thus, U is Lipschitz, hence absolutely continuous, and so equal to the integral of its derivative. Moreover, at each point of differentiability, the envelope theorem applied to (A.3) gives us

$$\frac{dU}{dz} = \sum_g \frac{d\theta^g(z)}{dz} x^g(z) = \sum_g \frac{1}{f^g(\theta^g(z))} x^g(z).$$

Therefore,

$$U(z) = U(0) + \int_0^z \frac{dU(\widetilde{z})}{d\widetilde{z}}\, d\widetilde{z} = U(0) + \sum_g \int_0^z \frac{x^g(\widetilde{z})}{f^g(\theta^g(\widetilde{z}))}\, d\widetilde{z}.$$

Consequently, profit is

$$\begin{aligned}
\int_0^1 t(z)\, dz &= \int_0^1 \left(\sum_g \theta^g(z)x^g(z) - U(z) \right) dz \\
&= \int_0^1 \left(\sum_g \theta^g(z)x^g(z) - U(0) - \sum_g \int_0^z \frac{x^g(\widetilde{z})}{f^g(\theta^g(\widetilde{z}))}\, d\widetilde{z} \right) dz \\
&\leq \sum_g \left[\int_0^1 \theta^g(z)x^g(z)\, dz - \int_0^1 \left(\int_0^z \frac{x^g(\widetilde{z})}{f^g(\theta^g(\widetilde{z}))}\, d\widetilde{z} \right) dz \right]
\end{aligned}$$

where the last inequality holds because $U(0) \geq 0$ by (A.2). Switching the variables in the second integral and changing the order of integration gives

$$= \sum_g \left[\int_0^1 \left(\theta^g(z) - \frac{1-z}{f^g(\theta^g(z))} \right) x^g(z)\, dz \right]. \tag{A.4}$$

Now, for each g, we can see that the expression in parentheses is exactly the virtual value v^g corresponding to marginal type $\theta^g(z)$. Hence, by assumption, this quantity is negative for $\theta^g(z) < \theta^{*g}$ and positive for $\theta^g(z) > \theta^{*g}$ — which correspond to $z < F^g(\theta^{*g})$ and $z > F^g(\theta^{*g})$, respectively. In particular, an upper bound for the value of (A.4) is found by taking $x^g(z)$ to be as small as possible, namely 0, when $z < F^g(\theta^{*g})$, and as large as possible, namely 1, when $z > F^g(\theta^{*g})$. Thus, the principal's profit is at most

$$\begin{aligned} \sum_g \left[\int_{F^g(\theta^{*g})}^1 \left(\theta^g(z) - \frac{1-z}{f^g(\theta^g(z))} \right) dz \right] &= \sum_g \left[\int_{F^g(\theta^{*g})}^1 \frac{d}{dz} \left(-\theta^g(z)(1-z)\right) dz \right] \\ &= \sum_g \theta^{*g}(1 - F^g(\theta^{*g})). \end{aligned}$$

This is exactly the profit attained by the mechanism that sells the goods separately, with a price of θ^{*g} for good g; therefore it cannot exceed R^*. This establishes that the profit from any mechanism is at most R^*, as claimed.

□

Proof of Lemma 4.2. Let $\Delta = \max_{x,\theta} u(x,\theta) - \min_{x,\theta} u(x,\theta)$, and notice that for any mechanism satisfying IC, any two types' payments can differ by at most Δ. Consequently, when looking for optimal mechanisms, we can restrict attention to mechanisms whose payments are in $[-\bar{t}, \bar{t}]$, for a sufficiently high value of $\bar{t}$ (which is independent of the distribution π): each type's payment is bounded below, because a sufficiently low payment for one type would imply low payments for all types, and then the mechanism would not be optimal (it could be improved by increasing all types' payments by a small constant without violating IR); and each type's payment is bounded above, simply by IR.

Then, the maximum expected profit, as a function of π, is the upper envelope of a uniformly bounded family of linear functions (one such function for each possible mechanism $M \in \mathcal{M}$ satisfying the bounds on payments). It readily follows that this upper envelope is Lipschitz in π, and in particular is continuous.

□

Completion of proof of Theorem 2.1. The argument in the text proved the theorem in the case of finite type and allocation spaces; here we extend this to prove the full theorem. We maintain the notation of Subsection 2.2.

First, we extend to general allocation spaces, but keep type spaces finite. So suppose each Θ^g is finite. Let R^{*g} be the optimal revenue in component g, and $R^* = \sum_g R^{*g}$. Suppose there exists a mechanism (x', t') that achieves revenue at least $R^* + \epsilon$ with respect to every joint distribution $\pi \in \Pi$, where $\epsilon > 0$.

For each component g, define an auxiliary screening problem $(\widetilde{\Theta}^g, \widetilde{X}^g, \widetilde{u}^g, \widetilde{\pi}^g)$ as follows: $\widetilde{\Theta}^g = \Theta^g$ and $\widetilde{\pi}^g = \pi^g$; $\widetilde{X}^g = \{x'^g(\theta) \mid \theta \in \Theta\}$; $\widetilde{u}^g(x^g, \theta^g) = Eu^g(x^g, \theta^g)$. That is, we consider only the component-g allocations which were actually assigned to some type in the mechanism (x', t'); such an allocation may have been a lottery, but we treat it as a pure allocation in the new screening problem. It is evident that any mechanism for this new screening problem translates to a mechanism for the original component-g screening problem, with the same expected profit; hence, the optimal profit for the new component-g screening problem is $\widetilde{R}^{*g} \leq R^{*g}$. Because $\widetilde{\Theta}^g$ and $\widetilde{X}^g$ are finite, we can apply the result for the finite case to see that there is a joint distribution $\widetilde{\pi}$ for which no joint screening mechanism can give expected profit more than R^*. But (x', t') evidently gives us a joint screening mechanism for the new problem, which produces expected profit $\geq R^* + \epsilon$ for every possible joint distribution — a contradiction.

Finally, we can prove the theorem in general, without the restriction to finite type spaces. Write $R^* = \sum_g R^{*g}$ as usual, and suppose, for contradiction, that there exists a mechanism (x', t') that achieves profit at least $R^* + \epsilon$ against every possible joint distribution π, where $\epsilon > 0$.

For each component g, let δ^g be as given by Lemma 4.3 with $\epsilon/(2G+1)$ as the allowable profit loss. By making δ^g smaller if necessary, we may also assume that any two types at distance at most δ^g have their utility for every $x^g \in X^g$ differ by at most $\epsilon/(2G+1)$. Also, consider the joint screening environment, with the metric on Θ given by the sum of componentwise distances, and let δ be as given by Lemma 4.3 with $\epsilon/(2G+1)$ as allowable loss again. Now let $\overline{\delta} = \min\{\delta^1, \ldots, \delta^G, \delta/G\}$.

For each g, we form an approximate distribution on Θ^g with finite support, as follows: Let $\widetilde{\Theta}^g$ be a finite subset of Θ^g, with the property that every element of Θ^g is within distance $\overline{\delta}$ of some element of $\widetilde{\Theta}^g$ (this can be done by compactness). Arbitrarily partition Θ^g into disjoint measurable subsets $S^g_{\theta^g}$, for $\theta^g \in \widetilde{\Theta}^g$, such that each element of any $S^g_{\theta^g}$ is within distance $\overline{\delta}$ of θ^g, and θ^g itself is in $S^g_{\theta^g}$. Then define a distribution $\widetilde{\pi}^g \in \Delta(\Theta^g)$, supported on $\widetilde{\Theta}^g$, by $\widetilde{\pi}^g(\theta^g) = \pi^g(S^g_{\theta^g})$.

Evidently, $\widetilde{\pi}^g$ is $\overline{\delta}$-close to π^g. Therefore, the maximum profit attainable in the screening problem $(\Theta^g, X^g, u^g, \widetilde{\pi}^g)$ is at most $R^{*g} + \epsilon/(2G+1)$: otherwise, Lemma 4.3 would be violated in going from $\widetilde{\pi}^g$ from π^g.

In turn, we can view $\widetilde{\pi}^g$ as a distribution just on $\widetilde{\Theta}^g$, and any mechanism for screening problem $(\widetilde{\Theta}^g, X^g, u^g, \widetilde{\pi}^g)$ can be converted to a mechanism on the whole type space Θ^g, by assigning each type its preferred outcome from the set $\{(x^g(\theta^g), t^g(\theta^g)) \mid \theta^g \in \widetilde{\Theta}^g\}$, and subtracting $\epsilon/(2G+1)$ from all payments (in order to make sure that IR is still satisfied for types outside of $\widetilde{\Theta}^g$). This conversion causes a profit loss of $\epsilon/(2G+1)$. We conclude that the maximum profit attainable in screening problem $(\widetilde{\Theta}^g, X^g, u^g, \widetilde{\pi}^g)$ is at most $R^{*g} + 2\epsilon/(2G+1)$.

Since this is true for each g, we can apply the finite-type-space version of our result to conclude the following: there is a distribution $\widetilde{\pi}$ on $\widetilde{\Theta} = \times_g \widetilde{\Theta}^g$, with marginals $\widetilde{\pi}^g$, for which any mechanism earns expected profit at most $R^* + 2G\epsilon/(2G+1)$.

Now we will construct a measure on Θ based on our discretization, but with marginals given by the original π^g. For each $\theta = (\theta^1, \dots, \theta^G) \in \widetilde{\Theta}$, let $S_\theta \subseteq \Theta$ be the product set $\times_g S^g_{\theta^g}$; notice that as θ varies over $\widetilde{\Theta}$, the sets S_θ form a partition of Θ.

For each such θ, we define a measure π_θ on S_θ as follows. If $\widetilde{\pi}^g(\theta^g) = 0$ for some g, let π_θ be the zero measure. Otherwise, consider the conditional probability measure $\pi^g | S^g_{\theta^g}$ for each component g (which is well-defined since π^g assigns positive probability to $S^g_{\theta^g}$). Define π_θ to be the product of these conditional measures, multiplied by the scalar $\widetilde{\pi}(\theta)$.

We can extend π_θ to a measure on all of Θ (by taking it to be zero outside S_θ). Now simply define a measure π on Θ as the sum of π_θ, over all $\theta \in \widetilde{\Theta}$.

Note that π is indeed a probability measure; this follows from the fact that the measure assigned to any S_θ equals $\widetilde{\pi}(\theta)$, and hence the total measure of Θ is

$$\pi(\Theta) = \sum_{\theta \in \widetilde{\Theta}} \pi(S_\theta) = \sum_{\theta \in \widetilde{\Theta}} \widetilde{\pi}(\theta) = 1.$$

In fact, the marginal of π along component g must equal π^g, for each g. This follows from two facts: first, the probability assigned to any cell $S^g_{\theta^g}$ under this marginal is equal to

$$\sum_{\theta^{-g} \in \widetilde{\Theta}^{-g}} \pi(S_{(\theta^g, \theta^{-g})}) = \sum_{\theta^{-g} \in \widetilde{\Theta}^{-g}} \widetilde{\pi}(\theta^g, \theta^{-g}) = \widetilde{\pi}^g(\theta^g) = \pi^g(S^g_{\theta^g});$$

and second, conditional on cell $S^g_{\theta^g}$, the distribution along the g-component follows $\pi^g | S^g_{\theta^g}$ (because this is true for each of the cells $S_{(\theta^g, \theta^{-g})}$).

Consequently, our mechanism (x', t') satisfies $E_\pi[t'(\theta)] \geq R^* + \epsilon$, by assumption.

In addition, the fact that $\pi(S_\theta) = \widetilde{\pi}(\theta)$ for every $\theta \in \widetilde{\Theta}$ implies that $\widetilde{\pi}$ is δ-close to π, since every element of $S^g_{\theta^g}$ is within distance $\overline{\delta} \leq \delta/G$ of θ^g, for each component g. Consequently, Lemma 4.3 assures us the existence of a mechanism whose expected profit with respect to $\widetilde{\pi}$ is greater than $(R^* + \epsilon) - \epsilon/(2G+1) = R^* + 2G\epsilon/(2G+1)$.

But we already showed it is impossible to earn profit greater than $R^* + 2G\epsilon/(2G+1)$ against $\widetilde{\pi}$. Contradiction.

□

References

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

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

——— (1999): "Price Discrimination by a Many-Product Firm," *Review of Economic Studies*, 66(1), 151–168.

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, pp. 21–30.

Bakos, Y., and E. Brynjolfsson (1999): "Bundling Information Goods: Pricing, Profits, and Efficiency," *Management Science*, 45(12), 1613–1630.

Balder, E. J. (1996): "On the Existence of Optimal Contract Mechanisms for Incomplete Information Principal-Agent Models," *Journal of Economic Theory*, 68(1), 133–148.

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

Battaglini, M., and R. Lamba (2015): "Optimal Dynamic Contracting: the First-Order Approach and Beyond," working paper, Pennsylvania State University.

Bergemann, D., and S. Morris (2005): "Robust Mechanism Design," *Econometrica*, 73(6), 1771–1813.

Bergemann, D., and K. Schlag (2011): "Robust Monopoly Pricing," *Journal of Economic Theory*, 146(6), 2527–2543.

Brooks, B. (2013): "Surveying and Selling: Belief and Surplus Extraction in Auctions," working paper, University of Chicago.

Cai, Y., C. Daskalakis, and S. M. Weinberg (2012): "Optimal Multi-Dimensional Mechanism Design: Reducing Revenue to Welfare Maximization," in *Proceedings of the IEEE 53rd Annual Symposium on Foundations of Computer Science*, FOCS '12, pp. 130–139.

Cai, Y., N. 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, pp. 926–939.

Carrasco, V., V. F. Luz, P. Monteiro, and H. Moreira (2015): "Robust Selling Mechanisms," working paper, University of British Columbia.

Carroll, G. (2015a): "Robust Incentives for Information Acquisition," working paper, Stanford University.

——— (2015b): "Robustness and Linear Contracts," *American Economic Review*, 105(2), 536–563.

Carroll, G., and D. Meng (2016): "Robust Contracting with Additive Noise," working paper, Stanford University.

Chu, C. S., P. Leslie, and A. Sorensen (2011): "Bundle-Size Pricing as an Approximation to Mixed Bundling," *American Economic Review*, 101(1), 263–303.

Chung, K.-S., and J. C. Ely (2007): "Foundations of Dominant-Strategy Mechanisms," *Review of Economic Studies*, 74(2), 447–476.

Church, J., and R. Ware (2000): *Industrial Organization: A Strategic Approach*. McGraw-Hill, Boston.

Conitzer, V., and T. Sandholm (2002): "Complexity of Mechanism Design," in *Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence*, UAI '02, pp. 103–110.

Daskalakis, C., A. Deckelbaum, and C. Tzamos (2013): "Mechanism Design via Optimal Transport," in *Proceedings of the 14th ACM Conference on Electronic Commerce*, EC '13, pp. 269–286.

——— (2014): "Complexity of Optimal Mechanism Design," in *Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '14, pp. 1302–1318.

——— (2015): "Strong Duality for a Multiple-Good Monopolist," working paper, Massachusetts Institute of Technology.

Farhi, E., and I. Werning (2013): "Insurance and Taxation over the Life Cycle," *Review of Economic Studies*, 80(2), 596–635.

Findeisen, S., and D. Sachs (2015): "Redistribution and Insurance with Simple Tax Instruments," working paper, University of Mannheim.

Frankel, A. (2014): "Aligned Delegation," *American Economic Review*, 104(1), 66–83.

Garrett, D. (2014): "Robustness of Simple Menus of Contracts in Cost-Based Procurement," *Games and Economic Behavior*, 87, 631–641.

Gilboa, I., and D. Schmeidler (1989): "Maxmin Expected Utility with Non-Unique Prior," *Journal of Mathematical Economics*, 18(2), 141–153.

Haghpanah, N., and J. Hartline (2015): "Multi-dimensional Virtual Values and Second-degree Price Discrimination," working paper, Northwestern University.

Hart, S., and N. Nisan (2012): "Approximate Revenue Maximization with Multiple Items," working paper, Hebrew University of Jerusalem.

——— (2013): "The Menu-Size Complexity of Auctions," working paper, Hebrew University of Jerusalem.

Hart, S., and P. J. Reny (2015): "Maximal Revenue with Multiple Goods: Nonmonotonicity and Other Observations," *Theoretical Economics*, 10(3), 893–922.

Kos, N., and M. Messner (2015): "Selling to the Mean," working paper, Bocconi University.

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 (2012): "Sellers with Misspecified Models," working paper, London School of Economics.

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

——— (2007): "Multidimensional Mechanism Design: Revenue Maximization and the Multiple-Good Monopoly," *Journal of Economic Theory*, 137(1), 153–185.

McAfee, R. P., J. McMillan, and M. D. Whinston (1989): "Multiproduct Monopoly, Commodity Bundling, and Correlation of Values," *Quarterly Journal of Economics*, 104(2), 371–383.

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.

——— (1991): *Game Theory: Analysis of Conflict*. Harvard University Press, Cambridge, Massachusetts.

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. Choné (1998): "Ironing, Sweeping, and Multidimensional Screening," *Econometrica*, 66(4), 783–826.

Rochet, J.-C., and L. Stole (2003): "The Economics of Multidimensional Screening," in *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress, Vol. I*, pp. 150–197. Cambridge University Press, New York.

Rothschild, C., and F. Scheuer (2013): "Redistributive Taxation in the Roy Model," *Quarterly Journal of Economics*, 128(2), 623–668.

——— (2014): "A Theory of Income Taxation under Multidimensional Skill Heterogeneity," working paper, Stanford University.

Rothschild, M., and J. Stiglitz (1976): "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information," *Quarterly Journal of Economics*, 90(4), 629–649.

Rubinstein, A., and S. M. Weinberg (2015): "Simple Mechanisms for a Subadditive Buyer and Applications to Revenue Monotonicity," in *Proceedings of the 16th ACM Conference on Electronic Commerce*, EC '15, pp. 377–394.

Segal, I. (2003): "Optimal Pricing Mechanisms with Unknown Demand," *American Economic Review*, 93(3), 509–529.

Sher, I., and R. Vohra (2015): "Price Discrimination through Communication," *Theoretical Economics*, 10(2), 597–648.

Stantcheva, S. (2015): "Optimal Taxation and Human Capital Policies over the Life Cycle," *Journal of Political Economy*, forthcoming.

Stigler, G. J. (1963): "United States v. Loew's Inc.: A Note on Block-Booking," *Supreme Court Review*, pp. 152–157.

Strausz, R. (2006): "Deterministic Versus Stochastic Mechanisms in Principal-Agent Models," *Journal of Economic Theory*, 128(1), 306–314.

Thanassoulis, J. (2004): "Haggling over Substitutes," *Journal of Economic Theory*, 117(2), 217–245.

Wang, Z., and P. Tang (2014): "Optimal Mechanisms with Simple Menus," in *Proceedings of the Fifteenth ACM Conference on Economics and Computation*, EC '14, pp. 227–240.

Yao, A. C.-C. (2015): "An n-to-1 Bidder Reduction for Multi-Item Auctions and its Applications," in *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '15, pp. 92–109.