Incentive separability*

Paweł Doligalski Piotr Dworczak Joanna Krysta Filip Tokarski†

September 14, 2023

Abstract

We consider a general mechanism-design environment in which the planner faces incentive constraints such as the ones resulting from agents' private information or ability to take hidden actions. We study the properties of optimal mechanisms when some decisions are *incentive-separable*: A set of decisions is incentive-separable if, starting at some initial allocation, perturbing these decisions along agents' indifference curves preserves incentive constraints. We show that, under regularity conditions, the optimal mechanism allows agents to make unrestricted choices over incentive-separable decisions, given some prices and budgets. Using this result, we extend and unify the Atkinson-Stiglitz theorem on the undesirability of differentiated commodity taxes and the Diamond-Mirrlees production efficiency result. We also demonstrate that the analysis of incentive separability provides a novel justification for in-kind redistribution programs similar to food stamps.

One of the central problems studied by public finance is the conflict between efficiency and redistribution. The trade-off arises whenever the conclusion of the second welfare theorem fails—primarily due to incentive constraints (Kaplow, 2011). Nevertheless, some of the most celebrated results in optimal taxation identify decisions for which the equity-efficiency trade-off can be avoided. Diamond and Mirrlees (1971) showed that production plans should remain undistorted—regardless of the planner's redistributive preferences—in economies in which only consumers possess private information. Atkinson and Stiglitz (1976) demonstrated that when consumers' preferences are weakly separable between consumption choices and labor supply, commodities should not be taxed in a distortionary

*The authors thank Árpád Ábrahám, Mohammad Akbarpour, Benjamin Brooks, Antoine Ferey, Stéphane Gauthier, Tai-Wei Hu, Ravi Jagadeesan, Scott Duke Kominers, Raghav Malhotra, Karl Schulz, Ilya Segal, Damián Vergara, Marek Weretka, Hakki Yazici, as well as several seminar audiences for helpful comments. The authors gratefully acknowledge funding from the European Union (ERC, IMD-101040122).

†*Doligalski:* University of Bristol; and Group for Research in Applied Economics; pawel.doligalski@bristol.ac.uk. *Dworczak:* Northwestern University; and Group for Research in Applied Economics; piotr.dworczak@northwestern.edu. *Krysta:* Stanford University; and Group for Research in Applied Economics; asia.krysta@gmail.com. *Tokarski:* Stanford GSB; and Group for Research in Applied Economics; filipstefantokarski@gmail.com.

manner. The Atkinson-Stiglitz result was subsequently shown to imply that certain distortions should be avoided in other settings such as dynamic capital allocation (Ordover and Phelps, 1979), public good provision (e.g., Christiansen, 1981), and economies with externalities (e.g., Cremer et al., 1998). Gauthier and Laroque (2009) further generalized the analysis of optimal taxation under weak separability by observing that the second welfare theorem can be applied to weakly separable goods alone, irrespective of distortions imposed on other decisions in the economy.

In this paper, we identify a common logic underlying these results, and explore its consequences using a mechanism-design framework. In our model, a planner maximizes a social welfare function subject to incentive constraints (which prohibit the implementation of the first-best allocation). Unlike previous work, we do not impose *a priori* restrictions on either the form of incentive constraints or agents' preferences. Instead, we study the notion of *incentive separability*: A set of decisions is called incentive-separable at some initial allocation if any modification of these decisions that keeps all agents indifferent to the initial allocation is guaranteed to preserve incentive constraints. In general, modifications that keep agents' utilities from their assigned allocations unchanged may nevertheless change the payoffs from deviations; thus, the restriction imposed by incentive separability is that no deviation becomes profitable. As a result, the set of incentive-separable decisions for a given application is determined jointly by what is assumed about agents' preferences and the nature of incentive constraints (available deviations). For example, all decisions are incentive-separable if the only incentive constraint is voluntary participation; weak separability implies incentive separability in settings with adverse selection; and decisions made conditional on a realized state are incentive-separable in moral-hazard environments.

Our main result (Theorem 1) states that it is always optimal to remove distortions between incentive-separable decisions. Moreover, under regularity conditions, these decisions can be decentralized: Agents are allocated type-specific budgets and make (otherwise unrestricted) choices over incentive-separable decisions, taking prices as given.

The main result is derived through the following thought exercise: Taking any mechanism satisfying incentive constraints, we can reoptimize over the allocation of incentive-separable decisions subject to delivering the same target utility to all agents. Crucially, we drop any incentive constraints from this auxiliary (re-)optimization problem. Nevertheless—by definition of incentive separability—the solution to the problem is guaranteed to satisfy them. Intuitively, considering the allocation of incentive-separable decisions in isolation allows us to ignore the incentive constraints. The two main predictions of Theorem 1 easily follow from this observation: By removing distortions between incentive-separable decisions, the planner can deliver the same utility profile while freeing up scarce resources in the economy; and once the allocation is undistorted, it can be implemented with budgets and prices, mimicking the logic of the second welfare theorem.

We consider three applications. First, we show that Theorem 1 implies an extension of the Atkinson-Stiglitz theorem in several directions, including incorporating moral-hazard constraints, relaxing any assumptions required for first-order analysis of the consumer problem, and generalizing the conclusion to suboptimality of any non-market mechanism (and not just differential taxation). While many of these extensions have been separately established in the literature following Atkinson and Stiglitz (1976), our approach leads to a particularly simple proof that is not tailored to any given extension. We also clarify that—once a general environment is considered—observability of earnings (and a non-linear income tax) is neither necessary nor sufficient for the Atkinson-Stiglitz result to hold.

Second, we show how the analysis of incentive separability naturally leads to a unification of the Atkinson-Stiglitz theorem with the production efficiency result of Diamond and Mirrlees (1971). We do this by replacing the simple linear production technology of Atkinson and Stiglitz (1976) with a complex production sector with many firms; we demonstrate that when consumers have incentive-separable preferences over consumption goods, they should face the same commodity prices that result from efficient production decisions of profit-maximizing firms. At its core, the result is a consequence of the observation that production decisions in the Diamond-Mirrlees setting are incentive-separable.

Third, we show that incentive separability provides a new justification for in-kind redistribution schemes, such as food stamps programs. To date, the literature has emphasized that in-kind redistribution of food—treated as a single category of consumption expenditures—can be optimal only when weak separability fails (Nichols and Zeckhauser, 1982, Currie and Gahvari, 2008). However, when consumption choices over food items are modeled at a more granular level, some of them may naturally become incentive-separable from other decisions, especially for consumers with low overall food consumption. Theorem 1 implies that it is then optimal to give such consumers budgets that they can spend on these food items at undistorted prices. This scheme closely resembles the design of the US food stamps program. Under this perspective, food stamps are beneficial because they isolate recipients' food consumption choices from potential tax distortions. Furthermore, relying on the flexibility of our framework in incorporating different incentive constraints, we argue that eligibility for food stamps should not be used to incentivize job search effort.

While we focus on applications to public finance, our model and definitions are framed in terms of an abstract mechanism-design problem; in particular, we study a general objective function and work in the space of direct mechanisms that assign decisions to types. The mechanism-design perspective lies at the very roots of public-finance theory (with Mirrlees, 1971, being the canonical example); however, most classical uses of mechanism design were attempts to characterize the optimal tax system overall—a task that is intractable in all but the most stylized cases.[1] Our mechanism-design approach differs from the classical one in that we do not attempt to fully characterize the optimal mechanism; instead, we identify ways in which an existing mechanism can be improved upon. In this sense, our approach is related to papers (such as Laroque, 2005, and Kaplow, 2006) that study welfare-improving tax reforms.

[1] These limitations led to the development and popularization of the "tax perturbation" approach (Piketty, 1997; Saez, 2001; Golosov et al., 2014).

1 General Framework

There is a unit mass of agents with types θ distributed on a (measurable) subset Θ of $\mathbb{R}^N$. A planner chooses a (measurable) allocation rule $x : \Theta \to \mathbb{R}_+^K$ that specifies the allocation $x(\theta)$ assigned to each type θ. We will refer to the respective dimensions of the allocation as "goods," with the understanding that they could also capture other types of decisions (monetary transfers, time, effort, labor supply etc). Agents' preferences over their assigned allocations are described by the utility function $U(x(\theta), \theta)$ which is continuous in the first argument, and measurable in the second.[2]

Our setting allows for multi-dimensional types. However, because we imposed no assumptions on how agents' utilities depend on their types (in particular, we do not impose any single-crossing conditions), we can simplify notation by normalizing θ so that it is distributed uniformly on the unit interval.[3]

Planner's problem. The planner chooses an allocation rule subject to *incentive constraints*:

$$x \in \mathcal{I}, \tag{I}$$

where $\mathcal{I} \subseteq \left(\mathbb{R}_+^K\right)^\Theta$ is an arbitrary subset of allocation rules. We impose no structure on $\mathcal{I}$; intuitively, it could capture individual rationality, incentive compatibility, or obedience constraints in settings with adverse selection or moral hazard. An allocation $x \in \mathcal{I}$ is called *feasible*.

For any allocation rule x, let $\mathcal{U}_x$ denote the utility profile associated with the allocation rule x, defined by $\mathcal{U}_x(\theta) = U(x(\theta), \theta)$ for any $\theta \in \Theta$, and let $\mathbf{x} := \int x(\theta)\, d\theta$ denote the corresponding *aggregate allocation*. (Throughout, we will use boldface font to denote aggregate allocations.) The planner maximizes

$$W(\mathcal{U}_x, \mathbf{x}), \tag{W}$$

over feasible allocation rules, with W assumed to be continuous and non-increasing in the second argument. The dependence on $\mathcal{U}_x$ is standard and captures any "welfarist" preferences (e.g., utilitarian preferences with social welfare weights). The dependence on the aggregate allocation $\mathbf{x}$ captures preferences over allocations beyond their consequences for agents' utilities, and is central to our analysis. It could directly represent the cost of providing resources, or the opportunity cost of using resources available to the planner.[4] In public-finance settings, dependence on $\mathbf{x}$ could capture the planner's preferences over tax revenue (see Section 3). Pure revenue maximization by a monopolist is a special case as well (W would then not depend on the utility profile). Finally, if the planner faces constraints that

[2]Whenever we introduce a function of θ, we assume that this function is measurable. To simplify exposition, we will use the convention that "for all θ" should be formally interpreted as "for almost all θ with respect to Lebesgue measure on Θ."

[3]By the Borel isomorphism theorem, we can map any Borel measure space Θ of cardinality at most continuum injectively into $[0, 1]$, and then redefine the type to be the quantile of the resulting distribution.

[4]If the planner could deliver a given utility profile for less, she could allocate the remaining resources to some socially valuable causes.

make certain aggregate allocations $\mathbf{x}$ infeasible, dependence on $\mathbf{x}$ could result from incorporating these constraints into the objective (see Subsection 2.1 and Section 5, where we model such constraints explicitly).

Incentive separability. To define our key notion of incentive separability, we introduce notation for partitioning the set of decisions $\{1, ..., K\}$ into two subsets (intuitively, the incentive-separable decisions and the remaining decisions). For any subset $\mathcal{S} \subseteq \{1, ..., K\}$ of goods, we let $x^{\mathcal{S}}(\theta) \coloneqq (x_i(\theta))_{i \in \mathcal{S}}$, and $x^{-\mathcal{S}}(\theta) \coloneqq (x_i(\theta))_{i \notin \mathcal{S}}$. We will also write $x = (x^{\mathcal{S}}, x^{-\mathcal{S}})$.

Definition 1. *Decisions $\mathcal{S} \subseteq \{1, ..., K\}$ are incentive-separable (at a feasible allocation x_0) if*

$$\{(x^{\mathcal{S}}, x_0^{-\mathcal{S}}) : U(x^{\mathcal{S}}(\theta), x_0^{-\mathcal{S}}(\theta), \theta) = U(x_0(\theta), \theta),\ \forall \theta \in \Theta\} \subseteq \mathcal{I}.$$

To paraphrase, consider a feasible allocation x_0 where decisions $\mathcal{S}$ are incentive-separable. Suppose the allocation of goods $\mathcal{S}$ is altered for any type θ in a way that keeps type θ's utility unchanged. Then, this new allocation satisfies incentive constraints.

Incentive separability is a joint property of preferences and incentive constraints. Intuitively, the altered allocation only keeps unchanged the agents' utilities from *their assigned allocation*. The agents' utilities from *deviations* may change. Incentive separability ensures that they do not increase sufficiently to make any deviations profitable.

We illustrate the concept with a few examples.

Example 1 (Voluntary participation)**.** Suppose each θ has an outside option $\underline{U}(\theta)$. Then,

$$\mathcal{I} = \{\ x : U(x(\theta), \theta) \geq \underline{U}(\theta), \quad \forall \theta \in \Theta\ \}. \tag{1}$$

All decisions are incentive-separable: Any allocation that gives agents the same utility profile as the initial feasible allocation also satisfies participation constraints.

Example 2 (Private information and weak separability)**.** Suppose decisions $\mathcal{S} \subseteq \{1, ..., K\}$ are weakly separable for all types. That is, $U(x(\theta), \theta) = \tilde{U}(v(x^{\mathcal{S}}(\theta)), x^{-\mathcal{S}}(\theta), \theta)$ for some subutility function $v : \mathbb{R}_+^{|\mathcal{S}|} \to \mathbb{R}$ and $\tilde{U} : \mathbb{R} \times \mathbb{R}_+^{K-|\mathcal{S}|} \times \Theta \to \mathbb{R}$, where $\tilde{U}$ is strictly increasing in subutility level v. Let $\mathcal{I}$ represent incentive-compatibility constraints when types are private information:

$$\mathcal{I} = \left\{\ x : U(x(\theta), \theta) \geq \max_{\theta' \in \Theta} U(x(\theta'), \theta), \ \ \forall \theta \in \Theta\ \right\}. \tag{2}$$

Then, decisions $\mathcal{S}$ are incentive-separable (at any feasible x_0). To see why, consider some altered allocation $x^{\mathcal{S}}$ of incentive-separable goods that keeps all agents indifferent. This allocation must then preserve subutility levels given to every agent: $v(x^{\mathcal{S}}(\theta)) = v(x_0^{\mathcal{S}}(\theta))$ for all θ. Since the allocation of the remaining goods is unchanged, every agent's utility from every report is as before, and hence incentive compatibility is unaffected. Note the role played by weak separability with a common subutility function—it ensures that alterations to $x^{\mathcal{S}}$ that preserve an agent's utility from reporting her type θ also preserve *others'* utilities from reporting θ.

Incentive separability persists if each type θ can only mimic types in some subset $E(\theta) \subseteq \Theta$. This can arise, for instance, if some dimensions of the type space are observable to the planner, or agents need to provide some verifiable evidence to be eligible for certain allocations.[5]

Example 3 (Moral hazard with adverse selection)**.** An agent has a privately observed characteristic $\tau \in \mathcal{T}$ and takes an unobservable action $a \in \mathcal{A}$ at a utility cost $c(a, \tau)$. The action determines the distribution of an observable state $\omega \in \Omega$, which is given by $\mu(\omega | a, \tau)$. After the state is realized, the agent consumes a vector of goods $y \in \mathbb{R}_+^{K-1}$ and receives a state-contingent utility $u(y, \omega)$. A possible interpretation is that an agent with privately observed learning ability τ chooses an unobserved human capital investment a that affects her future earnings ω.

Formally, we treat $x \equiv (y, a)$ as the allocation, and $\theta \equiv (\tau, \omega)$ as the type. Define the type's payoff as

$$U(y, a, (\tau, \omega)) = u(y, \omega) - c(a, \tau).$$

The incentive constraints ensure that the agent prefers to report her private characteristic τ truthfully and take the recommended action $a(\tau)$ (taking "double deviations" into account):

$$\mathcal{I} = \Big\{ (y, a) : \int U(y(\tau, \omega), a(\tau), (\tau, \omega))\, d\mu(\omega | a, \tau) \geq \max_{\tau' \in \mathcal{T}, a' \in \mathcal{A}} \int U(y(\tau', \omega), a', (\tau, \omega))\, d\mu(\omega | a', \tau), \;\; \forall \tau \in \mathcal{T} \Big\}. \tag{3}$$

The consumption vector y is incentive-separable: $\mathcal{S} = \{1, ..., K-1\}$. That is, consumption choices are incentive-separable conditional on the realized state ω (but are not incentive-separable *across* the states).[6]

In some applications, it may be natural to let the set of incentive-separable decisions depend on the type. We can accommodate that more general case at the cost of complicating our notation.

Remark 1. Definition 1 can be extended *verbatim* to the case of type-dependent incentive-separable decisions in the following way. For a *function* $\mathcal{S} : \Theta \to 2^{\{1,...,K\}}$ that maps types into subsets of $\{1, ..., K\}$, we let $x^{\mathcal{S}}(\theta) := x^{\mathcal{S}(\theta)}(\theta)$ and $x^{-\mathcal{S}}(\theta) := x^{-\mathcal{S}(\theta)}(\theta)$. The corresponding aggregate allocations $\mathbf{x}^{\mathcal{S}}$ and $\mathbf{x}^{-\mathcal{S}}$ are defined by $\mathbf{x}_k^{\mathcal{S}} := \int x_k(\theta) \mathbb{1}_{k \in \mathcal{S}(\theta)}\, d\theta$ and $\mathbf{x}_k^{-\mathcal{S}} := \int x_k(\theta) \mathbb{1}_{k \notin \mathcal{S}(\theta)}\, d\theta$, respectively, for all $k \in \{1, ..., K\}$.

From now on, we abuse notation slightly by letting $\mathcal{S}$ denote a function from types to subsets of decisions (rather then a fixed subset of decisions), noting that when $\mathcal{S}$ is constant in type, the more general notation reduces to the one we considered so far (except for Remark 4 and Section 4, our results can be understood using the simpler baseline definition). We illustrate the generalized notion of incentive separability with an example.

[5] A richer verifiable-evidence model, such as the one considered by Ben-Porath et al. (2019), could also be captured at the cost of complicating notation.

[6] The "trick" of including the state in the type makes it possible to generate similar examples of incentive separability in more complex environments.

Example 4 ("Hierarchy of needs" preferences)**.** We extend Example 2 with privately observed types by allowing for additional preference heterogeneity; as a result, the set of incentive-separable decisions becomes type-dependent and endogenous to the initial allocation.

Take some subset of goods $E \subseteq \{1, ..., K\}$ and suppose that agents value goods E according to a common subutility function $v : \mathbb{R}_+^{|E|} \to \mathbb{R}$ as long as $v(x^E)$ is below a threshold $\underline{v}$. However, once $v(x^E)$ is above the threshold, agents can exhibit heterogeneous tastes over these goods. Thus, the utility function can be written as

$$U(x, \theta) = \begin{cases} U_L(v(x^E), x^{-E}, \theta) & \text{if } v(x^E) \leq \underline{v} \\ U_H(x, \theta), & \text{otherwise,} \end{cases} \tag{4}$$

where U_L is strictly increasing in its first argument. Intuitively, the threshold $\underline{v}$ marks the fulfillment of basic needs that are universal to all agents; once these are satisfied, agents can exhibit individual, idiosyncratic tastes. We provide an economic application of such preferences in Section 4. Incentive constraints are given by equation (2) from Example 2.

Decisions E are incentive-separable at an allocation x_0 for agents whose subutility under x_0 is below $\underline{v}$. That is, $\mathcal{S}(\theta) = E$ for θ such that $v(x_0^E(\theta)) \leq \underline{v}$, and $\mathcal{S}(\theta) = \varnothing$ otherwise.[7]

Our examples illustrate that the more potential deviations agents have, i.e., the tighter the incentive constraints $\mathcal{I}$, the more stringent the requirements incentive separability imposes on agents' preferences. With individual rationality alone, all decisions are incentive-separable regardless of agents' utility functions. In settings with private information, decisions are incentive-separable if they are weakly separable. In moral-hazard settings, incentive separability holds only conditionally on the realization of the observable state.

This list of examples is not exhaustive. Other potential applications include settings with dynamic private information where successive elements of θ are revealed to agents over time, economies with aggregate shocks and state-dependent resource constraints, or combinations of Examples 1-4.

Preliminaries. Our analysis focuses on the properties of optimal mechanisms with respect to incentive-separable goods. To derive these properties, we will take a feasible allocation rule x_0 and "reoptimize" over the allocation of incentive-separable goods $\mathcal{S}$, keeping fixed the allocation of all other goods $x_0^{-\mathcal{S}}$, as well as the original utility profile $\mathcal{U}_{x_0}$. It will thus be convenient to define the payoffs of the agents and the planner from adjusting the allocation of incentive-separable goods as

$$v_\theta(x^{\mathcal{S}}(\theta)) := U(x^{\mathcal{S}}(\theta), x_0^{-\mathcal{S}}(\theta), \theta), \quad \forall \theta \in \Theta, \tag{5}$$

$$R(\mathbf{x}^{\mathcal{S}}) := W(\mathcal{U}_{x_0}, \mathbf{x}^{\mathcal{S}} + \mathbf{x}_0^{-\mathcal{S}}), \tag{6}$$

[7] Fix a feasible allocation x_0. Perturb decisions E so that the new allocation (x^E, x_0^{-E}) leaves all agents indifferent. Indifference implies that $v(x^E(\theta')) = v(x_0^E(\theta'))$ for all θ' with $v(x_0^E(\theta')) \leq \underline{v}$; thus, by mimicking θ', θ will obtain a payoff of $U_L(v(x_0^E(\theta')), x_0^{-E}(\theta'), \theta)$, exactly the same as before the perturbation. As a result, incentive compatibility is preserved.

where the dependence on the objects we keep fixed has been suppressed for brevity, and $\mathbf{x}^{\mathcal{S}} + \mathbf{x}_0^{-\mathcal{S}}$ is the total aggregate allocation when the planner allocates incentive-separable goods according to $x^{\mathcal{S}}$ (using notation introduced in Remark 1).

Fixing some initial feasible allocation x_0, we make the following additional assumptions. Each v_θ is locally nonsatiated; moreover, there exists an integrable function $\bar{x}(\theta)$ such that $v_\theta(y) = v_\theta(x_0^{\mathcal{S}}(\theta))$ implies $y \leq \bar{x}(\theta)$. Finally, $v_\theta(x_0^{\mathcal{S}}(\theta)) \geq v_\theta(\mathbf{0})$. These assumptions are milder than typically imposed in the literature. The first assumption is standard. The second one is needed for technical reasons in a model with a continuum of agents: It ensures that agents' indifference curves over incentive-separable goods are properly bounded, which implies that the admissible aggregate consumption set is closed—a property we will need to prove that an optimal allocation exists.[8] The last assumption states that the initial allocation gives each agent more utility than consuming nothing, which will allow us to prove decentralization with positive prices even if some decisions are "bads" for some agents.

2 Results

In this section, we assume that decisions $\mathcal{S} : \Theta \to 2^{\{1,\ldots,K\}}$ are incentive-separable (see Definition 1 and Remark 1). We make two simple observations (Lemma 1 and 2) that lead to our main result—Theorem 1—predicting that it is optimal for the planner to allow agents to make undistorted choices among incentive-separable goods, by purchasing them at (endogenously derived) prices subject to budget constraints.

We begin with the definition of an *$\mathcal{S}$-undistorted* allocation.

Definition 2. *A feasible allocation rule x_0 is $\mathcal{S}$-undistorted if $x_0^{\mathcal{S}}$ solves*

$$\max_{x^{\mathcal{S}}} R(\mathbf{x}^{\mathcal{S}}) \quad \textit{subject to} \quad v_\theta(x^{\mathcal{S}}(\theta)) = v_\theta(x_0^{\mathcal{S}}(\theta)), \ \forall \theta \in \Theta. \tag{7}$$

Conversely, a feasible allocation that does not solve problem (7) is called $\mathcal{S}$-distorted. We implicitly assumed that problem (7) has a solution, so that an $\mathcal{S}$-undistorted allocation exists—we later prove that this is indeed the case.

Intuitively, fixing $x_0^{-\mathcal{S}}$, the allocation $x^{\mathcal{S}}$ of incentive-separable goods $\mathcal{S}$ is undistorted if it maximizes the planner's objective subject to delivering the given utility profile. Importantly, problem (7) does not impose incentive constraints $\mathcal{I}$. Thus, an undistorted allocation of $\mathcal{S}$–goods is the first-best way (for the planner) to deliver the target utility from consuming them. Note, however, that $\mathcal{S}$-undistortedness does not imply that the allocation of $\mathcal{S}$–goods is the same as in the first-best solution (i.e., in the absence of incentive constraints). This is because the (fixed) target utility from consuming them may itself be distorted.

We first observe that a mechanism that distorts incentive-separable decisions can be improved upon.

[8] While the assumption may seem economically restrictive, it is not: Fixing an arbitrary compact set $B \subset \mathbb{R}_+^K$ of target consumption levels, any family $\{v_\theta\}_{\theta\in\Theta}$ of functions can be modified to satisfy this assumption while staying unchanged on B. In particular, the assumption would be superfluous if we imposed an exogenous bound on the amounts of goods that can be allocated to each agent.

Lemma 1 (Optimality)**.** *If x_0 is $\mathcal{S}$-distorted, then it can be improved upon in terms of the planner's objective* (W)*.*

Proof. Fix a feasible x_0 that is $\mathcal{S}$-distorted. By definition, there exists $x_\star = (x_\star^{\mathcal{S}}, x_0^{-\mathcal{S}})$ that leaves all types' utilities unchanged and yields $R(\mathbf{x}_\star^{\mathcal{S}}) > R(\mathbf{x_0}^{\mathcal{S}})$. By incentive separability, the first property implies that $x_\star \in \mathcal{I}$, and hence $x_\star$ is feasible. The second property implies that $x_\star$ achieves higher objective (W) than x_0. □

By Lemma 1, the planner can restrict attention to $\mathcal{S}$-undistorted allocations: If an allocation features any distortions between $\mathcal{S}$-goods, it can be replaced by a superior allocation that removes these distortions.

Even though $\mathcal{S}$-undistortedness is defined with respect to a fixed profile of agents' utilities, it is tightly linked to Pareto efficiency.

Remark 2 ($\mathcal{S}$-undistortedness implies conditional Pareto efficiency)**.** Suppose that R is strictly decreasing. If x_0 is $\mathcal{S}$-undistorted, then there does not exist an alternative allocation $x^{\mathcal{S}}$ of incentive-separable goods that results in the same aggregate allocation ($\mathbf{x}^{\mathcal{S}} = \mathbf{x}_0^{\mathcal{S}}$) and makes a positive mass of agents strictly better off without making any agent worse off.

The implication is intuitive. Suppose that, as in Remark 2, the planner benefits strictly from saving resources. If agents' utilities could be increased, the planner would be able to "reclaim" the resulting surplus by scaling agents' allocations down to guarantee their initial utilities, and pocketing the left-overs. This would contradict $\mathcal{S}$-undistortedness. (We formalize this argument in Claim 1 in Appendix A.) Hence, conditional on fixing the aggregate allocation of incentive-separable goods, the planner would never want to assign them to agents inefficiently.

This motivates our second observation, which is that, under certain conditions, $\mathcal{S}$–undistorted allocations can be decentralized. We first formalize the notion of decentralization and provide a necessary and sufficient condition under which decentralization is possible.

Definition 3. *An allocation $x^{\mathcal{S}}$ can be decentralized (with prices $\lambda \in \mathbb{R}^K_{++}$) if there exists a budget assignment $m : \Theta \to \mathbb{R}_+$ such that, for all $\theta \in \Theta$, $x^{\mathcal{S}}(\theta)$ solves*

$$\max_{y \in \mathbb{R}_+^{|\mathcal{S}(\theta)|}} \quad v_\theta(y) \quad \textit{subject to} \quad \lambda^{\mathcal{S}(\theta)} \cdot y \le m(\theta), \tag{8}$$

where we let $\lambda^{\mathcal{S}(\theta)} \equiv (\lambda_i)_{i \in \mathcal{S}(\theta)}$. When prices λ decentralize an $\mathcal{S}$–undistorted allocation, we refer to them as $\mathcal{S}$–undistorted.

Lemma 2 (Decentralization)**.** *Fix a feasible allocation x_0 and a price vector $\lambda \in \mathbb{R}^K_{++}$. Then, $x_0^{\mathcal{S}}$ can be decentralized with prices λ if and only if x_0 is regular with prices λ, that is, $x_0^{\mathcal{S}}$ solves*

$$\min_{x^{\mathcal{S}}} \quad \lambda \cdot \mathbf{x}^{\mathcal{S}} \quad \textit{subject to} \quad v_\theta(x^{\mathcal{S}}(\theta)) = v_\theta(x_0^{\mathcal{S}}(\theta)), \ \forall \theta \in \Theta. \tag{9}$$

Proof. The minimization problem (9) can be solved pointwise in θ, and thus is equivalent to solving, for all $\theta \in \Theta$,

$$\min_{x^{\mathcal{S}}(\theta)\in\mathbb{R}_+^{|\mathcal{S}(\theta)|}} \quad \lambda^{\mathcal{S}(\theta)} \cdot x^{\mathcal{S}}(\theta) \quad \text{subject to} \quad v_\theta(x^{\mathcal{S}}(\theta)) = v_\theta(x_0^{\mathcal{S}}(\theta)). \tag{10}$$

Note that, for each θ, problem (10) is an expenditure minimization problem. Therefore, by consumer duality (Proposition 3.E.1 in Mass-Colell et al., 1995), problem (10) is equivalent to a utility maximization problem given some budget $m(\theta)$, that is, problem (8).[9] Thus, x_0 is regular if and only if $x_0^{\mathcal{S}}$ can be decentralized. □

Lemma 2 shows that an allocation can be decentralized if and only if it minimizes some *linear* objective function (with strictly positive coefficients) subject to keeping the agents' utilities constant—a property we call "regularity." Regularity differs from $\mathcal{S}$-undistortedness in that the defining optimization problem for the former features a linear objective. Thus, if the planner's objective R is linear—as in our first two applications, including the original Atkinson-Stiglitz setting—regularity holds trivially, and $\mathcal{S}$-undistorted allocations can be decentralized.

We derive our main result by observing that a much weaker condition on R suffices for regularity. We say that a function $f : \mathbb{R}^K \to \mathbb{R}$ has *bounded marginals* if there exist constants $\bar{c} > \underline{c} > 0$ such that, for all $y \in \mathbb{R}^K$, $k \in \{1, ..., K\}$, and $\epsilon > 0$

$$\bar{c} \geq \frac{|f(y + \epsilon \mathbf{e}_k) - f(y)|}{\epsilon} \geq \underline{c}, \tag{11}$$

where $\mathbf{e}_k$ is a standard unit vector (equal to 1 at the k-th coordinate and equal to zero elsewhere). When applied to R, the definition of bounded marginals states that the planner's marginal benefit from freeing any resource is bounded away from zero (and infinity).

Theorem 1. *Let decisions $\mathcal{S}$ be incentive-separable and suppose that R has bounded marginals. Starting at any feasible allocation x_0, the planner's objective can be improved by allowing agents to purchase incentive-separable goods at $\mathcal{S}$–undistorted prices subject to type-dependent budgets (with the improvement being strict if x_0 is $\mathcal{S}$-distorted).*

The proof is relegated to Appendix A. Theorem 1 predicts that it is optimal to let agents trade incentive-separable goods freely given $\mathcal{S}$-undistorted prices and type-dependent budgets. In particular, the planner should always implement an allocation in which agents' choices among incentive-separable goods $\mathcal{S}$ are undistorted. We will show in applications that $\mathcal{S}$-undistorted prices must be proportional to marginal costs in settings with production.

In light of Remark 2, Theorem 1 can be seen as a second welfare theorem for the "incentive-separable" part of our economy. Our proof is standard in that it attempts to separate the solution to problem (7) (which we show exists) from a properly defined set of admissible allocations by a hyperplane. The separating hyperplane defines the prices that verify the

[9] Formally, Proposition 3.E.1 in Mass-Colell et al. (1995) does not cover the case when $x^{\mathcal{S}}(\theta) \equiv 0$ but in this case the equivalence holds trivially by assigning a zero budget.

regularity condition (9) in Lemma 2. While our setting with a continuum of agents automatically gives us convexity assumptions required for separation (see Aumann, 1966),[10] the main technical challenge is to ensure that the separating hyperplane is properly bounded, so that the corresponding prices are *strictly* positive. The requirement of strictly positive prices in Lemma 2 is vital. When only a weaker notion of regularity with non-negative prices is satisfied, the allocation of incentive-separable goods can be decentralized only for agents with strictly positive expenditures on these goods.[11]

The novel part of the proof of Theorem 1 is to argue that when the objective R has bounded marginals, the aggregate allocation selected by the planner can indeed be decentralized with strictly positive prices.[12] For intuition, suppose that, at some candidate allocation $x^{\mathcal{S}}$, all separating hyperplanes were horizontal in some direction. Then, there would exist some good k with a zero price, and thus a reduction in the aggregate consumption of good k would require only a slight compensation in the aggregate consumption of other goods to keep all agents indifferent. That, however, would contradict $x^{\mathcal{S}}$ being selected by the planner: When the planner has a bounded rate of substitution between any two goods (which is implied by bounded marginals of R), she would strictly prefer to reduce the aggregate consumption of good k.

Remark 3. An alternative strategy to proving Theorem 1 that avoids making any additional assumptions on R is to ensure that prices are strictly positive in *any* equilibrium, that is, regardless of which aggregate allocation is chosen by the planner. This, however, requires imposing additional restrictions on agents' preferences. For example, as we demonstrate in Claim 2 in Appendix A, it suffices to assume that all $v_\theta(x^{\mathcal{S}})$ have (uniformly) bounded marginals.

It is important to note that—in contrast to the classical second welfare theorem—Theorem 1 is a *partial* decentralization result, in that it only applies to incentive-separable goods. The planner may be using some complicated mechanism to allocate non-incentive-separable goods and type-dependent budgets (which our results are silent about). However, conditional on allocating these budgets to agents, the rest of the mechanism is pinned down by letting agents trade incentive-separable goods freely at $\mathcal{S}$-undistorted prices.

2.1 Additional constraints

To highlight the key intuitions, in Definition 1 of incentive separability we assumed that *any* perturbation of incentive-separable decisions is permitted. More generally, we can define

[10] With finitely many agents, our results would hold under additional convexity assumptions on preferences.

[11] It is well known that the second welfare theorem may fail when some prices are zero and there are consumers with zero total expenditure at the optimal solution; in such cases, it is only possible to support a quasi-equilibrium (see, for example, Proposition 16.D.3 in Mass-Colell et al., 1995). In the context of Lemma 2, this corresponds to the failure of consumer duality.

[12] In their analysis of weak separability, Gauthier and Laroque (2009) rely on the second welfare theorem directly to implement a Pareto efficient allocation of weakly-separable goods. While our proof strategy is similar, our argument for existence of strictly positive prices does not have a parallel in their analysis.

decisions $\mathcal{S}$ to be incentive-separable on some subset of allocations $\mathcal{F}$:

$$\{(x^{\mathcal{S}}, x_0^{-\mathcal{S}}) \in \mathcal{F} : U(x^{\mathcal{S}}(\theta), x_0^{-\mathcal{S}}(\theta), \theta) = U(x_0(\theta), \theta),\ \forall \theta \in \Theta\} \subseteq \mathcal{I}.$$

This generalized definition is useful when the planner faces a constraint of the form $x \in \mathcal{F}$ in the original problem of maximizing welfare (W). For example, in Section 5, we study a version of the model in which the planner faces technological constraints, and thus only a subset of possible aggregate allocations $\mathbf{x}^{\mathcal{S}}$ of incentive-separable goods is relevant.

To accommodate this extension, we can redefine $\mathcal{S}$-undistorted allocations by adding the constraint $(x^{\mathcal{S}}, x_0^{-\mathcal{S}}) \in \mathcal{F}$ to problem (7). Lemma 1 continues to hold (with the same proof), and implies that the planner should always use an allocation that is $\mathcal{S}$-undistorted (on $\mathcal{F}$). Lemma 2, on the other hand, continues to hold without any modification: Decentralization is possible if and *only if* the target allocation minimizes a linear objective subject to delivering the target utility profile.[13] Thus, an $\mathcal{S}$-undistorted allocation on $\mathcal{F}$ can be decentralized only if the constraint $(x^{\mathcal{S}}, x_0^{-\mathcal{S}}) \in \mathcal{F}$ can be subsumed by choosing appropriate prices λ. Intuitively, this is possible when $\mathcal{F}$ only includes constraints on the aggregate allocation. In Section 5, we formalize this observation by providing sufficient conditions for decentralizing an $\mathcal{S}$-undistorted allocation subject to aggregate constraints.

The extension can be useful even if the planner does not face an additional constraint in the original problem, as the following remark illustrates.

Remark 4. Consider a standard one-dimensional screening problem in which types and allocations are ordered by the single-crossing condition, and the planner can use monetary transfers. Take any feasible allocation x_0 that features under-provision of the non-monetary good (relative to first-best). Define $\mathcal{F}$ to be the set of allocations in which the non-monetary allocation is pointwise (weakly) higher than in x_0. Then, all decisions are incentive-separable for the highest type (by the single-crossing assumption).[14] Therefore, by Lemma 1, the highest type's allocation should be undistorted in the optimal mechanism—implying a version of the famous "no distortion at the top" result.

3 Atkinson-Stiglitz Theorem

In this section, we apply Theorem 1 to generalize the Atkinson-Stiglitz theorem. We assume that the planner's payoff from providing incentive-separable goods is linear, that is, $R(\mathbf{x}^{\mathcal{S}}) = -\lambda \cdot \mathbf{x}^{\mathcal{S}} + \text{const}$, $\lambda \in \mathbb{R}^K_{++}$. A natural interpretation is that λ represents constant marginal costs of producing incentive-separable goods and that the planner wants to minimize production costs (for a given utility profile).

We will show that this setup nests the original model of Atkinson and Stiglitz (1976) as a special case. To see that, let θ be the agent's privately observed ability which determines

[13] It would be tempting to modify Lemma 2 by adding the constraint $(x^{\mathcal{S}}, x_0^{-\mathcal{S}}) \in \mathcal{F}$ to problem (9). However, this modification would make Lemma 2 false.

[14] Here, we rely on the generalized definition of incentive separability from Remark 1.

her cost of supplying labor. Let $L \in \mathbb{R}_+$ represent labor in efficiency units. An agent consumes goods $y \in \mathbb{R}_+^{K-1}$ and receives payoff $U(v(y), L, \theta)$, that is, the consumption of goods is weakly separable from type and labor supply. Let $x = (y, L) : \Theta \to \mathbb{R}_+^K$ denote the allocation rule. All goods are produced using labor as the sole input with constant marginal costs $\lambda_k > 0$ for $k = 1, ..., K-1$. Hence, the tax revenue is equal to the aggregate labor supply net of the aggregate production cost. The planner's objective depends both on individual utilities and on the tax revenue, with $\alpha > 0$ representing the marginal value of public funds. Thus, the planner's objective can be written as

$$W(\mathcal{U}_x, \mathbf{x}) = V(\mathcal{U}_x) + \alpha(\mathbf{L} - \sum_{k=1}^{K-1} \lambda_k \mathbf{y}_k).$$

Note that in the model of Atkinson and Stiglitz (1976) all commodities are incentive-separable for all types, $\mathcal{S} \equiv \{1, ..., K-1\}$, by Example 2. Moreover, holding fixed the utility profile, the planner's objective is linear in commodities (up to a constant): $R(\mathbf{x}^{\mathcal{S}}) = -\alpha \sum_{k=1}^{K-1} \lambda_k \mathbf{y}_k$, as assumed.

Theorem 1 implies the following result.

Corollary 1. *(Atkinson-Stiglitz) Consider a feasible allocation x_0. The planner's objective can be (weakly) improved by allowing agents to purchase incentive-separable goods at prices proportional to marginal costs subject to type-dependent budgets.*

Note that, as long as an $\mathcal{S}$-undistorted allocation exists, Corollary 1 follows from Lemma 1 and 2 alone, giving a particularly short proof of the Atkinson-Stiglitz result.

Let us describe Corollary 1 in terms of tax systems, which was the original focus of Atkinson and Stiglitz (1976). Suppose that the relative consumption of incentive-separable commodities was initially distorted, e.g., by differentiated commodity taxes. The planner can then generate more revenue without affecting individual utilities by *(i)* removing taxes from the incentive-separable goods, which ensures that they are traded at undistorted prices equal to marginal costs, and *(ii)* implementing type-dependent budgets to be spent on incentive-separable goods. The second step can be done either by adjusting (potentially nonlinearly) taxes on non-separable goods or, equivalently, by introducing a tax on total expenditure on incentive-separable goods.

Corollary 1 is significantly more general than the original Atkinson-Stiglitz theorem and its subsequent extensions (Laroque, 2005; Kaplow, 2006; Gauthier and Laroque, 2009) along a few dimensions. First, we demonstrate that *any* distortion to the relative consumption of incentive-separable goods—including but not limited to indirect taxation—is suboptimal. For example, it is not optimal to have public provision of incentive-separable goods, or to use a rationing mechanism as in Dworczak ⓡ al. (2021).

Second, it is conventional wisdom that Atkinson-Stiglitz theorem is applicable only when the planner can use a nonlinear income tax, i.e., when individual earnings (or labor in efficiency units) are observable. In contrast, we show that observability of earnings is neither necessary nor sufficient for this result in a more general model. Instead, what is required is the observability of individuals' total expenditure on incentive-separable goods, since it allows the planner to implement type-dependent budgets for these goods. For instance, suppose that agents can conceal some of their earnings, either by engaging in informal employment or by underreporting business income. Even though earnings are not fully observable, by Corollary 1, the relative consumption of incentive-separable goods should be undistorted as long as the government observes individuals' total expenditures on these goods.

Finally, we show that the Atkinson-Stiglitz theorem is not specific to the standard taxation model with private types and weakly separable commodities. On the contrary: It holds in any environment (with constant marginal costs) with respect to decisions that are incentive-separable. It can be applied in models featuring, for instance, stochastic states (idiosyncratic or aggregate), hidden actions, verifiable information, or dynamic private information.[15] For a concrete example, consider a combination of Examples 2 and 3, where agents with private abilities choose *(i)* an unobserved effort (e.g., in education) that affects their subsequent productivity distribution and *(ii)* labor supply and consumption of various goods conditional on the realized (private) productivity. Such a model features both private information and moral hazard, and combines the redistributive (Mirrlees, 1971) and the social insurance (Varian, 1980) strands of income taxation literature. By Corollary 1, taxes on incentive-separable goods (here: goods that are weakly separable from labor, productivity, and effort) are superfluous.

4 Food Vouchers

In this section, we use our results to study the optimality of offering food vouchers. Setting aside reasons related to paternalism and consumption externalities, the literature (as surveyed in Currie and Gahvari, 2008) emphasized the role of in-kind transfers in relaxing incentive constraints (e.g., Nichols and Zeckhauser, 1982). For example, if the rich do not want to consume certain low-quality goods, providing them for free can work as an incentive-compatible way of making targeted transfers to the needy. This rationale requires that preferences over those goods *not* be weakly separable. In contrast, we emphasize the role of food voucher programs in removing distortions in consumption of individual food items when food purchases of the poor *are incentive-separable*.[16]

Jensen and Miller (2010) observe that poorer individuals tend to make food choices based on nutritional value, while richer individuals pay more attention to quality and taste. We use the "hierarchy of needs" preferences from Example 4 to model this observation. Let $E \subseteq \{1, ..., K\}$ be the set of food items and let $v(x^E)$ be a (common to all agents) function measuring the nutritional value of bundle x^E. Idiosyncratic tastes affect food choices only

[15] Applicability of the Atkinson-Stiglitz theorem to some of these environments has been already noted: see Da Costa and Werning (2002) for pure moral hazard and Golosov et al. (2003) for dynamic private information.

[16] These two perspectives are not mutually exclusive. Indeed, in our framework, the total food consumption is not assumed to be weakly separable from other decisions, so it may be distorted for the reasons Nichols and Zeckhauser (1982) and others identfied.

once a nutritional threshold $\underline{v}$ is met. Thus, an agent's utility is given by

$$U(x,\theta) = \begin{cases} U_L(v(x^E), x^{-E}, \theta) & \text{if } v(x^E) \le \underline{v} \\ U_H(x,\theta), & \text{otherwise.} \end{cases} \tag{12}$$

U_L is strictly increasing in its first argument; x^{-E} represents all other decisions (such as labor supply and other consumption choices). We assume type θ is private and consider $\mathcal{I}$ to be the set of incentive-compatible allocations. Finally, we assume that food items in E are produced with constant marginal costs $\lambda \in \mathbb{R}^{|E|}_{++}$, and that the planner, conditional on a given utility profile, minimizes the costs of production (which is equivalent to tax revenue maximization, as explained in Section 3).

By the "hierarchy of needs" preferences, food items E are incentive-separable at some allocation x_0 for types θ for which $v(x_0^E(\theta)) \le \underline{v}$ (see Example 4). Then, Theorem 1 implies the following:

Corollary 2 (Food vouchers)**.** *Consider a feasible allocation x_0. The planner's objective can be (weakly) improved by assigning budgets to all agents with θ such that $v(x_0^E(\theta)) \le \underline{v}$, and letting them spend these budgets on food items E (but not other goods) priced at marginal costs.*

Intuitively, under the initial allocation x_0, agents are separated into two groups, depending on whether the nutritional value of their food consumption is below or above $\underline{v}$. The first group (which we interpret as the poor) have weakly separable preferences between basic food items and other decisions. The second group (which we interpret as the rich) can have arbitrary, potentially heterogeneous preferences. Saez (2002) showed that preference heterogeneity can lead to optimal consumption distortions (e.g., via commodity taxes).[17] Thus, the planner can benefit from distorting food consumption of the rich, but not of the poor. She can achieve that by implementing distortionary taxes on food while isolating the poor from these distortions by offering budgets allowing for tax-free purchases, which we interpret as food vouchers. This rationale is in line with the structure of existing food vouchers programs. For example, in the US, SNAP offers food stamps to poor households (with, likely, poor nutrition), and exempts purchases made with food stamps from state and local consumption taxes.

Our analysis provides further prescriptions for the design of a food vouchers program. For instance, eligibility for SNAP is to a large extent contingent on employment, presumably to motivate job search effort.[18] It is natural to ask whether this is optimal. We can incorporate unobservable job search effort into our setting, similarly to Example 3. Each agent privately chooses job search intensity that determines a distribution over the state—her employment status. We assume that, conditional on the realized employment status, preferences over

[17] Gauthier and Henriet (2018) constructed a tractable model of taste heterogeneity in which optimal commodity taxes are partly shaped by a version of the many-person Ramsey rule.

[18] A non-disabled adult without children who is less than half-time employed can be eligible for food stamps for a maximum of 3 months every 3 years. However, work requirements are regularly waived by states with high unemployment (Ganong and Liebman, 2018) suggesting that their purpose is to motivate job search effort (since effort is less useful when the labor market is slack).

food consumption have the structure assumed in (12). By Example 3, Corollary 2 applies conditional on every state realization. Hence, distorting food consumption for agents below the nutritional threshold $\underline{v}$ is suboptimal regardless of whether they are employed or unemployed, and therefore should not be used as an incentive tool.[19]

5 Atkinson-Stiglitz meet Diamond-Mirrlees

We now apply our methods to production economies with a potentially complex input-output structure. We also show how to include additional constraints into the planner's problem on top of incentive constraints.

There are $J \in \mathbb{N}$ firms; $z^j \in \mathbb{R}^K$ denotes firm j's production vector. Negative entries stand for inputs and positive for outputs. Each firm is equipped with a production technology $Z_j \subseteq \mathbb{R}^K$ which allows for free disposal and inaction. A production vector z^j is available to firm j if $z^j \in Z_j$. We will denote a production plan for all firms by $z = (z^1, ..., z^J)$ and the corresponding aggregate production vector by $\mathbf{z} = \sum_{j=1}^J z^j$. Letting $\mathbf{Z}$ denote the Minkowski sum of Z_j for $j \in \{1, \dots, J\}$, the aggregate production $\mathbf{z}$ is technologically possible if $\mathbf{z} \in \mathbf{Z}$.[20]

Let (x, z) denote the overall allocation of consumption and production in this economy. An allocation (x, z) is feasible if it satisfies incentive constraints $x \in \mathcal{I}$, the production plan is technologically possible: $\mathbf{z} \in \mathbf{Z}$, and aggregate consumption does not exceed aggregate output: $\mathbf{z} \geq \mathbf{x}$. The planner chooses a feasible allocation to maximize the objective $\mathcal{W}(\mathcal{U}_x, \mathbf{z} - \mathbf{x})$, which is weakly increasing in $\mathbf{z} - \mathbf{x}$. The vector $\mathbf{z} - \mathbf{x}$, representing the "leftover" resources for the planner, can be interpreted as purchases of goods by the planner (financed, e.g., through tax revenue). Following Diamond and Mirrlees (1971), we will say that a feasible production plan z_0 is efficient if there does not exist an aggregate production vector $\mathbf{z} \in \mathbf{Z}$ such that $\mathbf{z} \geq \mathbf{z_0}$ and $\mathbf{z} \neq \mathbf{z_0}$.

By assumption, the production plan z affects neither individual utilities nor incentive constraints. Thus, it is trivially incentive-separable, implying the following result:[21]

Corollary 3. *For any feasible allocation, the planner's objective can be (weakly) improved by choosing an $\mathcal{S}$-undistorted allocation of incentive-separable goods and an efficient production plan.*

Corollary 3 merges the Atkinson-Stiglitz theorem with the Diamond-Mirrlees production efficiency result: It is welfare-improving to jointly remove distortions to consumption of incentive-separable goods and to production of all goods.

[19] Note, however, that this concerns *eligibility* for food stamps, i.e., the *extensive margin*. Policymakers could still provide incentives by varying the allocation of food vouchers on the *intensive margin*. Our framework remains silent about such adjustments—these should be made based on considerations related to relaxing incentive constraints, as in Nichols and Zeckhauser (1982).

[20] We can accommodate an aggregate endowment vector $\mathbf{q} \geq 0$ by including a fictitious firm with a production set containing only $\mathbf{q}$.

[21] Formally, to apply our results from Section 2, we can represent the production plan as a function of θ and include it in the allocation rule x. Because z enters neither agents' utilities nor incentive constraints, it can be included in the set of incentive-separable decisions.

Diamond and Mirrlees (1971) and Stiglitz and Dasgupta (1971) proved production efficiency at the optimum, assuming that agents with private types are taxed with linear consumption taxes. Hammond (2000) showed that there are welfare gains from implementing production efficiency also away from the full optimum. Gauthier and Laroque (2009) asserted (without proof) that a result analogous to Corollary 3 holds in their framework. Relative to these papers, we show that implementing production efficiency is desirable under general incentive constrains and when the planner can rely on arbitrary mechanisms (including nonlinear taxes or rationing systems).

Importantly, Corollary 3 requires that production decisions of firms do not affect incentive constraints. Otherwise, the planner could potentially relax incentive constraints by introducing production distortions. Naito (1999) studies such a motive in a model where labor supply of different types is not perfectly substitutable in production but cannot be distinguished by the planner.[22]

Our next result concerns decentralizing $\mathcal{S}$-undistorted allocations in production economies. The main technical difficulty is ensuring that $\mathcal{S}$-undistorted allocations exist and are regular in the presence of aggregate constraints. As we show in Lemma 3 in Appendix A, this can be guaranteed under the following assumption:

Assumption 1. *The aggregate production set* $\mathbf{Z}$ *is closed, bounded from above and convex. Furthermore, for any* $\mathbf{z}_0 \in \mathbf{Z}$ *and any nonempty proper subset* $A \subset \{1, ..., K\}$*, there exists* $\mathbf{z} \in \mathbf{Z}$ *such that* $\mathbf{z}^A \leq \mathbf{z}_0^A$, $\mathbf{z}^{-A} \geq \mathbf{z}_0^{-A}$ *and* $\mathbf{z}^{-A} \neq \mathbf{z}_0^{-A}$.

Once the existence of regular and $\mathcal{S}$-undistorted allocations is ensured, the main result of this section follows easily from Lemma 2 and Corollary 3 (see Appendix A for details).

Theorem 2 (Atkinson-Stiglitz meet Diamond-Mirrlees)**.** *Suppose Assumption 1 holds. For any feasible allocation, there exists a price vector* $\lambda \in \mathbb{R}^K_{++}$ *such that the planner's objective can be (weakly) improved by simultaneously: (i) allowing agents to purchase incentive-separable goods at prices* λ *subject to type-dependent budgets, and (ii) allowing firms to maximize profits and trade all goods taking prices* λ *as given, and taxing their profits lump-sum.*

Theorem 2 gives us a straightforward recipe for a welfare improving reform: Implement a competitive outcome in incentive-separable consumption goods and in the entire production sector by allowing agents and firms to trade goods while taking prices as given. Thus, it is beneficial to jointly remove any distortionary taxes levied on incentive-separable goods

[22] Naito (1999) considers an economy with two types of workers: high-skilled and low-skilled. The planner does not observe skill types and taxes earnings of all workers with the same tax schedule. The incentive constraint that prevents redistribution to low-skilled workers thus depends on their relative wage rate. The relative wage rate depends on the production decisions of firms because labor of different skill types is imperfectly substitutable. Hence, the planner may benefit from distorting the production decisions: By overhiring low-skilled workers in the public sector, she inflates their wage rate, which relaxes the incentive constraint and allows for more redistribution. In our mechanism-design framework, restrictions on tax instruments due to some decisions being unobservable correspond to obedience constraints in the set $\mathcal{I}$. In Naito's setup, however, such restrictions depend on wages, and through them on the production plan z. This in turn makes the set $\mathcal{I}$ depend on production z, breaking incentive separability of production decisions that is key for our results.

or on firm transactions. The resulting profits, if any, should be taxed lump-sum. Importantly, removing consumption distortions in incentive-separable goods without addressing production distortions (or vice versa) could fail to improve welfare. To see why, suppose there is a consumption tax on some incentive-separable commodity and an output subsidy on the same good. If the tax and the subsidy are of similar magnitude, they mostly offset each other, resulting in a small overall distortion. However, removing the consumption tax alone breaks that balance, increasing distortions. Intuitively, allowing consumers to buy incentive-separable goods at producer prices may fail to improve welfare if producer prices are distorted.

Recall that in Theorem 1 we proved existence of strictly positive prices under the assumption that rates of substitutions between any two goods are nondegenerate (as implied by the preferences of the planner, or the preferences of the agents, by Remark 3). By analogy, Theorem 2 guarantees strictly positive prices by assuming a nondegenerate rate of transformation between goods (which is the most restrictive part of Assumption 1). These alternative assumptions can be useful in different settings: For example, by assuming a nondegenerate rate of substitution we can accommodate the case of a fixed supply of goods (which violates Assumption 1).

Remark 5. Suppose that the aggregate supply of goods is fixed at $\bar{\mathbf{z}} \in \mathbb{R}^K_+$, meaning that the aggregate production set is given by $\mathbf{Z} = \{\mathbf{z} \in \mathbb{R}^K : \mathbf{z} \leq \bar{\mathbf{z}}\}$, and that $v_\theta(x^{\mathcal{S}})$ has bounded marginals for all $\theta \in \Theta$ (with bounds that are uniform across θ). Then, the conclusion of Theorem 2 holds (the proof can be found in Appendix A).

6 Concluding Remarks

In this paper, we relied on a mechanism-design approach to identify a simple but powerful principle underlying many classical results in public finance, including the seminal Diamond-Mirrlees and Atkinson-Stiglitz theorems. That principle is based on the observation that in many complex environments—indeed, often too complex to characterize the optimal mechanism in its entirety—some non-trivial set of decisions may be *incentive-separable*. There should be no distortions between incentive-separable decisions, and hence their choices can often be delegated to agents maximizing private utility given prices and budgets. Apart from extending and unifying the classical results, we presented a novel application to optimal design of food vouchers programs.

While we focused on public-finance applications throughout, neither our formal model nor our proofs are tailored towards them. There are other (notoriously difficult) problems with multidimensional types and allocations where the analysis of incentive-separable decisions could cast light on some features of the optimal mechanism. We leave this direction for future research.

References

ATKINSON, A. AND J. STIGLITZ (1976): "The design of tax structure: Direct versus indirect taxation," *Journal of Public Economics*, 6, 55–75.

AUMANN, R. J. (1965): "Integrals of set-valued functions," *Journal of Mathematical Analysis and Applications*, 12, 1–12.

——— (1966): "Existence of Competitive Equilibria in Markets with a Continuum of Traders," *Econometrica*, 34, 1–17.

BEN-PORATH, E., E. DEKEL, AND B. L. LIPMAN (2019): "Mechanisms With Evidence: Commitment and Robustness," *Econometrica*, 87, 529–566.

CHRISTIANSEN, V. (1981): "Evaluation of Public Projects under Optimal Taxation," *The Review of Economic Studies*, 48, 447–457.

CREMER, H., F. GAHVARI, AND N. LADOUX (1998): "Externalities and optimal taxation," *Journal of Public Economics*, 70, 343–364.

CURRIE, J. AND F. GAHVARI (2008): "Transfers in cash and in-kind: Theory meets the data," *Journal of Economic Literature*, 46, 333–383.

DA COSTA, C. E. AND I. WERNING (2002): "Commodity taxation and social insurance," *MIT Working Paper*.

DIAMOND, P. A. AND J. A. MIRRLEES (1971): "Optimal taxation and public production I: Production efficiency," *The American Economic Review*, 61, 8–27.

DWORCZAK, P. ⓡ S. D. KOMINERS ⓡ M. AKBARPOUR (2021): "Redistribution through Markets," *Econometrica*, 89(4), 1665–1698.

GANONG, P. AND J. B. LIEBMAN (2018): "The decline, rebound, and further rise in SNAP enrollment: Disentangling business cycle fluctuations and policy changes," *American Economic Journal: Economic Policy*, 10, 153–76.

GAUTHIER, S. AND F. HENRIET (2018): "Commodity taxes and taste heterogeneity," *European Economic Review*, 101, 284–296.

GAUTHIER, S. AND G. LAROQUE (2009): "Separability and public finance," *Journal of Public Economics*, 93, 1168–1174.

GOLOSOV, M., N. KOCHERLAKOTA, AND A. TSYVINSKI (2003): "Optimal indirect and capital taxation," *The Review of Economic Studies*, 70, 569–587.

GOLOSOV, M., A. TSYVINSKI, AND N. WERQUIN (2014): "A variational approach to the analysis of tax systems," *NBER Working Paper*.

HAMMOND, P. J. (2000): "Reassessing the Diamond-Mirrlees Efficiency Theorem," *Incentives, organization, and public economics: Papers in honour of Sir James Mirrlees*, 193.

JENSEN, R. T. AND N. H. MILLER (2010): "A Revealed Preference Approach to Measuring Hunger and Undernutrition," *NBER Working Paper*.

KAPLOW, L. (2006): "On the undesirability of commodity taxation even when income taxation is not optimal," *Journal of Public Economics*, 90, 1235–1250.

——— (2011): *The theory of taxation and public economics*, Princeton University Press.

LAROQUE, G. R. (2005): "Indirect taxation is superfluous under separability and taste homogeneity: A simple proof," *Economics Letters*, 87, 141–144.

MASS-COLELL, A., M. WHINSTON, AND J. GREEN (1995): *Microeconomic theory*, Oxford University Press.

MIRRLEES, J. A. (1971): "An exploration in the theory of optimum income taxation," *The review of economic studies*, 175–208.

NAITO, H. (1999): "Re-examination of uniform commodity taxes under a non-linear income tax system and its implication for production efficiency," *Journal of Public Economics*, 71, 165–188.

NICHOLS, A. L. AND R. J. ZECKHAUSER (1982): "Targeting transfers through restrictions on recipients," *The American Economic Review*, 72, 372–377.

ORDOVER, J. AND E. PHELPS (1979): "The concept of optimal taxation in the overlapping-generations model of capital and wealth," *Journal of Public Economics*, 12, 1–26.

PIKETTY, T. (1997): "La Redistribution fiscale contre le chômage," *Revue française d'économie*, 12, 157–203.

SAEZ, E. (2001): "Using Elasticities to Derive Optimal Income Tax Rates," *The Review of Economic Studies*, 68, 205–229.

——— (2002): "The desirability of commodity taxation under non-linear income taxation and heterogeneous tastes," *Journal of Public Economics*, 83, 217–230.

STIGLITZ, J. E. AND P. DASGUPTA (1971): "Differential taxation, public goods, and economic efficiency," *The Review of Economic Studies*, 38, 151–174.

VARIAN, H. R. (1980): "Redistributive taxation as social insurance," *Journal of Public Economics*, 14, 49–68.

A Proofs and additional results

A.1 Proof of Remark 2

Definition 4. *A feasible allocation x_0 is conditionally Pareto efficient if there is no allocation of incentive-separable goods $x^{\mathcal{S}}$ that uses the same aggregate resources: $\mathbf{x}^{\mathcal{S}} = \mathbf{x}_0^{\mathcal{S}}$, leaves no agent worse off, and strictly improves the utilities of a positive measure of agents (while keeping the allocation of non-incentive-separable goods fixed at $x_0^{-\mathcal{S}}$).*

Claim 1. *Suppose the planner's objective R is strictly decreasing. If an allocation is $\mathcal{S}$-undistorted, then it is conditionally Pareto efficient.*

Proof. First, note the following fact. Suppose an agent's utility at $x^{\mathcal{S}}(\theta)$ is strictly greater than at $x_0^{\mathcal{S}}(\theta)$: $v_\theta(x^{\mathcal{S}}(\theta)) > v_\theta(x_0^{\mathcal{S}}(\theta)) \geq v_\theta(\mathbf{0})$. Then, starting from $x^{\mathcal{S}}(\theta)$, we can reduce the agent's payoff to that at $x_0^{\mathcal{S}}(\theta)$ by reducing her consumption of all goods. To that end, consider an allocation $\tilde{x}^{\mathcal{S}}(\theta) = a x^{\mathcal{S}}(\theta)$, where a is a scalar. By continuity of v_θ, there is $a \in [0, 1)$ such that $v(\tilde{x}^{\mathcal{S}}(\theta)) = v(x_0^{\mathcal{S}}(\theta))$.

Take an allocation x_0 and suppose that it is not conditionally Pareto efficient. Then, there exists x that uses the same aggregate resources where a positive mass of agents is strictly better off and no agent is worse off. By the fact shown above, we can scale agents utilities down to the initial levels while reducing the aggregate allocation of at least some goods. As a result, the planner's objective R strictly increases. Therefore, x_0 is not $\mathcal{S}$-undistorted. □

A.2 Proof of Theorem 1

We begin by constructing the set $\mathcal{X}^{\mathcal{S}}$ which will be useful throughout. For every $\theta \in \Theta$, define

$$\mathcal{X}_\theta^{\mathcal{S}} := \left\{ y \in \mathbb{R}_+^{|\mathcal{S}(\theta)|} : v_\theta(y) = v_\theta(x_0^{\mathcal{S}}(\theta)) \right\},$$

and

$$\mathcal{X}^{\mathcal{S}} := \left\{ \mathbf{x}^{\mathcal{S}} \in \mathbb{R}_+^K : x^{\mathcal{S}}(\theta) \in \mathcal{X}_\theta^{\mathcal{S}}, \ \forall \theta \in \Theta \right\}.$$

Set $\mathcal{X}^{\mathcal{S}}$ contains aggregate vectors of incentive-separable goods which can be distributed among agents to keep their utility fixed at the initial level, without any resources to spare.

By Theorem 1 of Aumann (1965), $\mathcal{X}^{\mathcal{S}}$ is convex. Because each v_θ is continuous, $\mathcal{X}_\theta^{\mathcal{S}}$ is closed. We assumed that there exists an integrable function $\bar{x}(\theta)$ such that $v_\theta(y) = v_\theta(x_0^{\mathcal{S}}(\theta))$ implies $y \leq \bar{x}(\theta)$. As a result, Theorem 4 of Aumann (1965) implies that $\mathcal{X}^{\mathcal{S}}$ is compact.

We now show that an $\mathcal{S}$-undistorted allocation exists. An allocation rule x_0 is $\mathcal{S}$-undistorted if $x_0^{\mathcal{S}}$ solves

$$\max_{x^{\mathcal{S}}} R(\mathbf{x}^{\mathcal{S}}) \quad \text{subject to} \quad v_\theta(x^{\mathcal{S}}(\theta)) = v_\theta(x_0^{\mathcal{S}}(\theta)), \ \forall \theta \in \Theta. \tag{13}$$

Using the above construction, this problem can be rewritten as:

$$\max_{\mathbf{x}^{\mathcal{S}} \in \mathcal{X}^{\mathcal{S}}} R(\mathbf{x}^{\mathcal{S}}). \tag{14}$$

Set $\mathcal{X}^{\mathcal{S}}$ is compact and non-empty (it contains $\mathbf{x}_0^{\mathcal{S}}$), and R is continuous and non-increasing. Therefore, there exists a solution to (14), denoted by $\mathbf{x}_\star^{\mathcal{S}}$, that lies on the lower boundary of $\mathcal{X}^{\mathcal{S}}$, meaning that there is no $\mathbf{x}^{\mathcal{S}} \in \mathcal{X}^{\mathcal{S}} \setminus \{\mathbf{x}_\star^{\mathcal{S}}\}$ such that $\mathbf{x}^{\mathcal{S}} \leq \mathbf{x}_\star^{\mathcal{S}}$. Notice that, despite changing the choice variable, the aggregate vector $\mathbf{x}_\star^{\mathcal{S}}$ has a corresponding allocation $x_\star^{\mathcal{S}}$ that solves (13). Hence, $x_\star = (x_\star^{\mathcal{S}}, x_0^{-\mathcal{S}})$ is $\mathcal{S}$-undistorted.

We now show that this $x_\star$ is regular. To this end, we construct the set $\mathcal{F}_\epsilon$ and separate it from $\mathcal{X}^{\mathcal{S}}$ with a hyperplane, which will define the coefficients λ from (9).

Let $\mathbf{e}_k \in \mathbb{R}^K$ be a vector equal to one at coordinate k and equal to zero at all other coordinates. Similarly, let $\mathbf{e}_{-k} = \mathbf{1} - \mathbf{e}_k$ be a vector equal to zero at coordinate k and equal to one at all other coordinates. For any $\epsilon \geq 0$, define

$$\mathcal{F}_\epsilon := \{\mathbf{y} \in \mathbb{R}^K : \text{there exists } k \in \{1, ..., K\} \text{ and } \alpha > 0 \text{ such that } \mathbf{y} \ll \mathbf{x}_\star^{\mathcal{S}} - \alpha \mathbf{e}_k + \alpha\epsilon \mathbf{e}_{-k}\}.$$

In particular, $\mathcal{F}_0$ is the set of all aggregate allocations strictly lower than $\mathbf{x}_\star^{\mathcal{S}}$: $\mathcal{F}_0 = \{\mathbf{y} \in \mathbb{R}^K : \mathbf{y} \ll \mathbf{x}_\star^{\mathcal{S}}\}$. For $\epsilon > 0$, $\mathcal{F}_\epsilon$ additionally contains aggregate allocations that are obtained from $\mathbf{x}_\star^{\mathcal{S}}$ by decreasing one of its coordinates by α, and increasing all other coordinates by at most $\alpha\epsilon$.

We claim that if $\epsilon > 0$ is small enough, then $\mathcal{F}_\epsilon \cap \mathcal{X}^{\mathcal{S}} = \varnothing$. Since R has bounded marginals, there exist bounds $\bar{c} \geq \underline{c} > 0$ satisfying definition (11). Fix any $\epsilon > 0$ satisfying $\epsilon < \underline{c}/(\bar{c}(K-1))$. Take any $\mathbf{x}^{\mathcal{S}} \in \mathcal{F}_\epsilon$; by definition of $\mathcal{F}_\epsilon$ there exist $k \in \{1, ..., K\}$ and $\alpha > 0$ such that $\mathbf{x}^{\mathcal{S}} \ll \mathbf{x}_\star^{\mathcal{S}} - \alpha \mathbf{e}_k + \alpha\epsilon \mathbf{e}_{-k}$. Then, using the fact that R is nonincreasing and has bounded marginals, we obtain

$$\begin{aligned} R(\mathbf{x}^{\mathcal{S}}) - R(\mathbf{x}_\star^{\mathcal{S}}) &\geq R(\mathbf{x}_\star^{\mathcal{S}} - \alpha \mathbf{e}_k + \alpha\epsilon \mathbf{e}_{-k}) - R(\mathbf{x}_\star^{\mathcal{S}}) \\ &= \underbrace{[R(\mathbf{x}_\star^{\mathcal{S}} - \alpha \mathbf{e}_k) - R(\mathbf{x}_\star^{\mathcal{S}})]}_{\geq \underline{c}\alpha} - \underbrace{[R(\mathbf{x}_\star^{\mathcal{S}} - \alpha \mathbf{e}_k) - R(\mathbf{x}_\star^{\mathcal{S}} - \alpha \mathbf{e}_k + \alpha\epsilon \mathbf{e}_{-k})]}_{\leq \bar{c}(K-1)\alpha\epsilon} \\ &\geq \alpha(\underline{c} - (K-1)\bar{c}\epsilon) > 0. \end{aligned}$$

Thus, $R(\mathbf{x}^{\mathcal{S}}) > R(\mathbf{x}_\star^{\mathcal{S}})$. Since $\mathbf{x}_\star^{\mathcal{S}}$ solves (14), it must be the case that $\mathbf{x}^{\mathcal{S}} \notin \mathcal{X}^{\mathcal{S}}$. Therefore, $\mathcal{F}_\epsilon \cap \mathcal{X}^{\mathcal{S}} = \varnothing$.

We have established that $\mathcal{F}_\epsilon$ and $\mathcal{X}^{\mathcal{S}}$ are disjoint convex sets. Therefore, there exists a hyperplane with coefficients $\lambda \in \mathbb{R}^K \setminus \{\mathbf{0}\}$ that separates them at $\mathbf{x}_\star^{\mathcal{S}}$:

$$\forall \mathbf{x}^{\mathcal{S}} \in \mathcal{X}^{\mathcal{S}}, \quad \lambda \cdot \mathbf{x}_\star^{\mathcal{S}} \leq \lambda \cdot \mathbf{x}^{\mathcal{S}} \tag{15}$$
$$\forall \mathbf{x}^{\mathcal{S}} \in \mathcal{F}_\epsilon, \quad \lambda \cdot \mathbf{x}_\star^{\mathcal{S}} \geq \lambda \cdot \mathbf{x}^{\mathcal{S}}. \tag{16}$$

Since $\mathcal{F}_\epsilon$ extends downwards in every dimension, it must be that $\lambda \geq \mathbf{0}$. Moreover, we will show that $\lambda \gg \mathbf{0}$. Suppose $\lambda_k = 0$ for some $k \in \{1, ..., K\}$. Since $\lambda \neq \mathbf{0}$, there is some good $l \neq k$ such that $\lambda_l > 0$. Take any scalars $a, b > 0$ defining a point $\mathbf{x}_\star^{\mathcal{S}} - a\mathbf{e}_k + b\mathbf{e}_{-k} \in \mathcal{F}_\epsilon$. But then,

$$\lambda \cdot (\mathbf{x}_\star^{\mathcal{S}} - a\mathbf{e}_k + b\mathbf{e}_{-k}) = \lambda \cdot (\mathbf{x}_\star^{\mathcal{S}} + b\mathbf{e}_{-k}) > \lambda \cdot \mathbf{x}_\star^{\mathcal{S}},$$

which contradicts (16).

Recall that the aggregate vector $\mathbf{x}_\star^{\mathcal{S}}$ had a corresponding allocation $x_\star^{\mathcal{S}}$. By (15), it solves:

$$\min_{x^{\mathcal{S}}} \quad \lambda \cdot \mathbf{x}^{\mathcal{S}} \quad \text{subject to} \quad v_\theta(x^{\mathcal{S}}(\theta)) = v_\theta(x_0^{\mathcal{S}}(\theta)), \ \forall \theta \in \Theta, \tag{17}$$

where $\lambda \gg \mathbf{0}$. Hence, $x_\star = (x_\star^{\mathcal{S}}, x_0^{-\mathcal{S}})$ is regular. By Lemma 2, $x_\star^{\mathcal{S}}$ can be decentralized with prices λ. Since $(x_\star^{\mathcal{S}}, x_0^{-\mathcal{S}})$ is $\mathcal{S}$-undistorted, by Lemma 1, it improves upon the original allocation x_0 (with the improvement being strict when x_0 is $\mathcal{S}$-distorted).

A.3 Proof of Remark 3

Claim 2. *The conclusion of Theorem 1 remains true if we replace the assumption that R has bounded marginals with the assumption that v_θ has bounded marginals for all $\theta \in \Theta$ (with bounds that are uniform across θ).*

Proof. The assumption of R having bounded marginals is used in the proof of Theorem 1 only to show that sets $\mathcal{X}^{\mathcal{S}}$ and $\mathcal{F}_\epsilon$ (defined therein) are disjoint for some $\epsilon > 0$. We show that this remains true if v_θ has uniformly bounded marginals.

Denote the bounds from Definition (11) by $\bar{c}_v \geq \underline{c}_v > 0$. Take an agent with type θ and an allocation of incentive separable decisions $x^{\mathcal{S}} \in \mathbb{R}_+^{|\mathcal{S}(\theta)|}$. Notice that, for any $k = \{1, ..., |\mathcal{S}(\theta)|\}$, her utility is either strictly increasing in $x_k^{\mathcal{S}}$, or strictly decreasing in $x_k^{\mathcal{S}}$ (keeping her allocation of other decisions fixed).[23] In the former case, we will say that k is a *good* for this agent, in the latter, we will say that it is a *bad* (with the understanding that whether k is a good or a bad may depend on the agent's allocation of other goods).

Recall that we consider an aggregate vector $\mathbf{x}_\star^{\mathcal{S}}$ on the lower boundary of $\mathcal{X}^{\mathcal{S}}$, meaning that there is no $\mathbf{x}^{\mathcal{S}} \in \mathcal{X}^{\mathcal{S}} \backslash \{\mathbf{x}_\star^{\mathcal{S}}\}$ such that $\mathbf{x}^{\mathcal{S}} \leq \mathbf{x}_\star^{\mathcal{S}}$. Further recall that there exists an allocation $x_\star^{\mathcal{S}}$ (corresponding to $\mathbf{x}_\star^{\mathcal{S}}$) that specifies the assignment of incentive-separable decisions to types. We first show that the mass of agents receiving bads under $x_\star^{\mathcal{S}}$ is zero.

Suppose otherwise, and denote by $\bar{x}^{\mathcal{S}}$ an alternative allocation of incentive-separable decisions which differs from $x_\star^{\mathcal{S}}$ only in the following way: for each agent who receives any bads under $x_\star^{\mathcal{S}}$, her allocation of one of these bads is now reduced to zero. Then $v_\theta(\bar{x}^{\mathcal{S}}(\theta)) \geq v_\theta(x_\star^{\mathcal{S}}(\theta)) \geq v_\theta(\mathbf{0})$ for all θ, with the left-most inequality strict for a positive mass of agents. By continuity of v_θ, for any type θ, we can find a scalar $a_\theta \in [0,1]$ such that $\tilde{x}^{\mathcal{S}}(\theta) = \alpha_\theta \bar{x}^{\mathcal{S}}(\theta)$ and $v_\theta(\tilde{x}^{\mathcal{S}}(\theta)) = v_\theta(x_\star^{\mathcal{S}}(\theta))$. Thus, $\tilde{\mathbf{x}}^{\mathcal{S}} \in \mathcal{X}^{\mathcal{S}}$. Note that $\alpha_\theta < 1$ for agents who previously received bads, and that there is a positive mass of such agents. Hence, $\tilde{\mathbf{x}}^{\mathcal{S}} \leq \mathbf{x}_\star^{\mathcal{S}}$ and $\tilde{\mathbf{x}}^{\mathcal{S}} \neq \mathbf{x}_\star^{\mathcal{S}}$, contradicting $\mathbf{x}_\star^{\mathcal{S}}$ being at the lower boundary of $\mathcal{X}^{\mathcal{S}}$.

We now show that $\mathcal{F}_\epsilon$ and $\mathcal{X}^{\mathcal{S}}$ are disjoint for small enough ϵ. Fix a positive ϵ such that $\epsilon < \underline{c}_v / (\bar{c}_v (K-1))$. Consider a type θ for whom decision $k \in \{1, ..., |\mathcal{S}(\theta)|\}$ is a good, $x_{\star k}^{\mathcal{S}}(\theta) > 0$, and who receives no bads under $x_\star^{\mathcal{S}}$. Consider an alternative allocation for this type, denoted by $\bar{x}^{\mathcal{S}}(\theta)$, where her assignment of k is decreased by $\alpha > 0$ and her assignment of

[23] Suppose otherwise; then by the continuity of v_θ there exists $\epsilon > 0$ such that $v_\theta(x^{\mathcal{S}} + \epsilon \mathbf{e}_k) = v_\theta(x^{\mathcal{S}})$. But then $|v_\theta(x^{\mathcal{S}} + \epsilon \mathbf{e}_k) - v_\theta(x^{\mathcal{S}})| = 0 < \epsilon \underline{c}_v$, violating bounded marginals.

other incentive-separable decisions is increased *in total* by at most $\alpha\epsilon(K-1)$.[24] By bounded marginals of v_θ, the agent is strictly worse off under $\bar{x}^{\mathcal{S}}(\theta)$:

$$v_\theta(\bar{x}^{\mathcal{S}}(\theta)) - v_\theta(x_*(\theta)) = \underbrace{[v_\theta(\bar{x}^{\mathcal{S}}(\theta)) - v_\theta(x_\star^{\mathcal{S}}(\theta) - \alpha\mathbf{e}_k)]}_{\leq \bar{c}_v(K-1)\alpha\epsilon} - \underbrace{[v_\theta(x_\star^{\mathcal{S}}(\theta)) - v_\theta(x_\star^{\mathcal{S}}(\theta) - \alpha\mathbf{e}_k)]}_{\geq \underline{c}_v\alpha}$$
$$\leq \alpha(\bar{c}_v(K-1)\epsilon - \underline{c}_v) < 0.$$

Thus, if we take α of good k from type θ but want to keep her indifferent, we need to increase her other incentive-separable decisions by more than $(K-1)\alpha\epsilon$ *in total*. This holds for every agent whose allocation of that good is positive and who receives no bads. As noted above, the mass of agents who receive bads under $x_\star^{\mathcal{S}}$ is zero. Hence, if for any $k \in \{1, ..., K\}$ we decrease $\mathbf{x}_{\star k}^{\mathcal{S}}$ by α, we must increase the sum of the aggregate allocations of other incentive-separable decisions by more than $(K-1)\alpha\epsilon$ for the allocation to remain in $\mathcal{X}^{\mathcal{S}}$. In particular, the aggregate allocation of some decision $l \neq k$ needs to increase by more than $\alpha\epsilon$.

Suppose there is some $\mathbf{x}^{\mathcal{S}} \in \mathcal{F}_\epsilon \cap \mathcal{X}^{\mathcal{S}}$. By definition of $\mathcal{F}_\epsilon$, there exist $k \in \{1, ..., K\}$ and $\alpha > 0$ such that $\mathbf{x}^{\mathcal{S}} \ll \mathbf{x}_\star^{\mathcal{S}} - \alpha\mathbf{e}_k + \alpha\epsilon\mathbf{e}_{-k}$. However, since $\mathbf{x}^{\mathcal{S}} \in \mathcal{X}^{\mathcal{S}}$, it follows from the argument above that there is a good $l \neq k$ such that $\mathbf{x}_l^{\mathcal{S}} > \mathbf{x}_{\star l}^{\mathcal{S}} + \alpha\epsilon$, which leads to a contradiction. □

A.4 Proof of Theorem 2

First, we state and prove a useful technical lemma.

Lemma 3. *Consider the setting of Section 5 and suppose Assumption 1 holds. For any feasible allocation (x_0, z_0) there exists a feasible allocation $(x_\star, z_\star)$ where $x_\star = (x_\star^{\mathcal{S}}, x_0^{-\mathcal{S}})$ is $\mathcal{S}$–undistorted, regular and satisfies $v_\theta(x_\star^{\mathcal{S}}(\theta)) = v_\theta(x_0^{\mathcal{S}}(\theta))$ for all $\theta \in \Theta$, and $z_\star$ is an efficient production plan.*

Proof. In the proof of Theorem 1 we defined the set $\mathcal{X}^{\mathcal{S}}$ that contains the aggregate allocations of incentive-separable goods which can be distributed among agents to keep their utilities fixed at the initial level. Here, it will be convenient to work with this set translated by the aggregate allocation of non-incentive-separable goods:

$$\mathcal{X} := \mathcal{X}^{\mathcal{S}} + \mathbf{x}_0^{-\mathcal{S}}. \tag{18}$$

Since $\mathcal{X}^{\mathcal{S}}$ is compact and convex (see the proof of Theorem 1), so is $\mathcal{X}$.

Consider the following maximization problem:

$$\max_{x^{\mathcal{S}}, \mathbf{z}} \mathcal{W}(\mathcal{U}_{x_0}, \mathbf{z} - (\mathbf{x}^{\mathcal{S}} + \mathbf{x}_0^{-\mathcal{S}})) \text{ subject to} \tag{19}$$
$$\mathbf{z} \in \mathbf{Z}, \quad \mathbf{z} \geq \mathbf{x}^{\mathcal{S}} + \mathbf{x_0}^{-\mathcal{S}}, \quad v_\theta(x^{\mathcal{S}}(\theta)) = v_\theta(x_0^{\mathcal{S}}(\theta)), \forall\theta \in \Theta.$$

[24] Formally, $\bar{x}^{\mathcal{S}}(\theta) = x_\star^{\mathcal{S}}(\theta) - \alpha\mathbf{e}_k + \alpha\epsilon\tilde{\mathbf{e}}_{-k}$, where $\tilde{\mathbf{e}}_{-k} \in \mathbb{R}_+^{|S(\theta)|}$ is some non-negative vector that is equal to zero at coordinate k and sums up to (at most) $K-1$: $\sum_l \tilde{\mathbf{e}}_{-k,l} \leq K-1$.

Note that if some $(x_\star^{\mathcal{S}}, \mathbf{z}_\star)$ solves this problem, then $(x_\star^{\mathcal{S}}, x_0^{-\mathcal{S}})$ will be $\mathcal{S}$-undistorted. Using the set $\mathcal{X}$, we can express the above problem in terms of aggregates as follows:

$$\max_{(\mathbf{x},\mathbf{z})\in\mathcal{X}\times\mathbf{Z}} \mathcal{W}(\mathcal{U}_{x_0}, \mathbf{z}-\mathbf{x}) \text{ subject to } \mathbf{z} \geq \mathbf{x}. \tag{20}$$

Recall that $\mathbf{Z}$ is closed and bounded from above by assumption. We can also put a lower bound on admissible elements of $\mathbf{Z}$—the resource constraint requires that $\mathbf{z} \geq \mathbf{x}$. Since aggregate consumption cannot be negative, we can bound the admissible subset of $\mathbf{Z}$ by 0 in each dimension. Thus, the choice set is compact. Since $\mathcal{W}$ is continuous and the choice set nonempty—it contains $(\mathbf{x}_0, \mathbf{z}_0)$—a solution to problem (20) exists. Moreover, $\mathcal{W}$ is nondecreasing in $\mathbf{z}-\mathbf{x}$, so there exists a solution $(\mathbf{x}_\star, \mathbf{z}_\star)$ such that there is no $(\mathbf{x}', \mathbf{z}') \in \mathcal{X}\times\mathbf{Z}$ satisfying $\mathbf{z}'-\mathbf{x}' \gg \mathbf{z}_\star - \mathbf{x}_\star$. This implies that $\text{int}\mathbf{Z} \cap \text{int}(\mathcal{X} + \mathbf{z}_\star - \mathbf{x}_\star) = \varnothing$.[25] Graphically, set $\mathcal{X}$, translated by $\mathbf{z}_\star - \mathbf{x}_\star$, is to the north-east of $\mathbf{Z}$, with only the boundaries of the two sets touching.

By the separating hyperplane theorem, there exists a hyperplane with coefficients $\lambda \in \mathbb{R}^K\backslash\{\mathbf{0}\}$ that separates $\text{int}\mathbf{Z}$ and $\text{int}(\mathcal{X} + \mathbf{z}_\star - \mathbf{x}_\star)$ at $\mathbf{z}_\star$:

$$\mathbf{z}_\star \in \arg\max_{\mathbf{z}\in\mathbf{Z}} \quad \lambda\cdot\mathbf{z}, \tag{21}$$

$$\mathbf{z}_\star \in \arg\min_{\mathbf{z}\in\mathcal{X}+\mathbf{z}_\star-\mathbf{x}_\star} \lambda\cdot\mathbf{z}. \tag{22}$$

Since $\mathbf{Z}$ allows for free disposal, it must be that $\lambda \geq \mathbf{0}$. Moreover, we will show that $\lambda \gg 0$. Suppose that λ is zero at coordinates from a nonempty subset $A \subset \{1,...,K\}$ and is strictly positive elsewhere. Since $\lambda \neq \mathbf{0}$, the set of remaining coordinates, denoted by $-A$, is nonempty. By Assumption 1, there exists $\mathbf{z} \in \mathbf{Z}$ such that $\mathbf{z}^A \leq \mathbf{z}_\star^A$, $\mathbf{z}^{-A} \geq \mathbf{z}_\star^{-A}$ and $\mathbf{z}^{-A} \neq \mathbf{z}_\star^{-A}$. Then, $\lambda\cdot\mathbf{z} > \lambda\cdot\mathbf{z}_\star$, which contradicts (21). Thus, $\lambda \gg 0$. Then, (21) implies that there is no $\mathbf{z}\in\mathbf{Z}\backslash\{\mathbf{z}_\star\}$ such that $\mathbf{z} \geq \mathbf{z}_\star$. Thus, the production plan $z_\star$ that corresponds to $\mathbf{z}_\star$ is efficient.

By a change of variable, (22) can be rewritten as

$$\mathbf{z}_\star - (\mathbf{z}_\star - \mathbf{x}_\star) \in \arg\min_{\mathbf{x}\in\mathcal{X}} \lambda\cdot\mathbf{x}. \tag{23}$$

Note that $\mathbf{z}_\star - (\mathbf{z}_\star - \mathbf{x}_\star) = \mathbf{x}_\star$. Given that $\lambda \gg 0$, this implies that $x_\star$—the allocation corresponding to $\mathbf{x}_\star$—is regular. □

We now proceed to the proof of Theorem 2. By Lemma 3, there exist an $\mathcal{S}$-undistorted allocation $x_\star = (x_\star^{\mathcal{S}}, x_0^{-\mathcal{S}})$ that is regular with some prices $\lambda \in \mathbb{R}^K_{++}$ and an associated production plan $z_\star$ that is efficient. Then, Corollary 3 implies that welfare can be (weakly) improved by implementing $(x_\star, z_\star)$.

Statement *(i)* follows from Lemma 2. It remains to show statement *(ii)*. In the proof of

[25] Denote $\mathbf{g}_\star = \mathbf{z}_\star - \mathbf{x}_\star$; suppose $\text{int}\mathbf{Z} \cap \text{int}(\mathcal{X} + \mathbf{g}_\star)$ is nonempty and contains some $\mathbf{z}$. Then we will construct $(\mathbf{z}', \mathbf{x}') \in \mathcal{X}\times\mathbf{Z}$ such that $\mathbf{z}'-\mathbf{x}' \gg \mathbf{z}_\star - \mathbf{x}_\star$ (note this also implies $\mathbf{z}' \geq \mathbf{x}'$). Notice $\mathbf{z} \in \mathcal{X} + \mathbf{g}_\star$, so $\mathbf{x}' := \mathbf{z} - \mathbf{g}_\star \in \mathcal{X}$. Since $\text{int}\mathbf{Z}$ is open, there exists $\mathbf{z}' \in \mathbf{Z}$ such that $\mathbf{z}' \gg \mathbf{z}$. But then $\mathbf{z}'-\mathbf{x}' = \mathbf{z}' - \mathbf{z} + \mathbf{g}_\star \gg \mathbf{g}_\star$.

Lemma 3 we showed that the aforementioned production plan z_* satisfies

$$\mathbf{z}_* \in \arg\max_{\mathbf{z}\in\mathbf{Z}} \lambda \cdot \mathbf{z}. \tag{24}$$

Recall that $\mathbf{z}_* = \sum_{j=1}^{J} z_*^j$, where $z_\star^j \in Z^j, \forall j$. Since the objective in (24) is linear, we can rewrite this problem as a collection of J maximization problems of individual firms:

$$z_\star^j \in \arg\max_{z^j\in Z^j} \lambda \cdot z^j, \quad \forall j = 1, ..., J. \tag{25}$$

Each of these problems is the profit maximization problem of an individual firm facing the production set Z^j and prices λ. Firm's profits are taxed lump-sum, which does not affect the solution to problem (25).

A.5 Proof of Remark 5

We will show that Lemma 3 (and, thus, Theorem 2) continues to hold in the fixed supply case when we replace Assumption 1 with the assumption that $v_\theta(x^{\mathcal{S}})$ has uniformly bounded marginals.

By the proof of Lemma 3, there exists a solution $(\mathbf{z}_*, \mathbf{x}_*) = (\bar{\mathbf{z}}, \mathbf{x}_*^{\mathcal{S}} + \mathbf{x}_0^{-\mathcal{S}})$ to problem (20) that satisfies $\text{int}\mathbf{Z} \cap \text{int}(\mathcal{X} + \bar{\mathbf{z}} - \mathbf{x}_*) = \varnothing$. By the proof of Claim 2, the set

$$\mathcal{F}_\epsilon := \{\mathbf{y} \in \mathbb{R}^K : \text{there exists } k \in \{1, ..., K\} \text{ and } \alpha > 0 \text{ such that } \mathbf{y} \ll \mathbf{x}_\star^{\mathcal{S}} - \alpha\mathbf{e}_k + \alpha\epsilon\mathbf{e}_{-k}\}$$

is disjoint from $\mathcal{X}^{\mathcal{S}} = \mathcal{X} - \mathbf{x}_0^{-\mathcal{S}}$ for sufficiently small $\epsilon > 0$, and these sets can be separated by a hyperplane with coefficients $\lambda \gg 0$. Translating both sets by a vector $\bar{\mathbf{z}} - \mathbf{x}_\star^{\mathcal{S}}$ does not change this conclusion: There is a hyperplane with coefficients $\lambda \gg 0$ that separates $\mathcal{F}_\epsilon + \bar{\mathbf{z}} - \mathbf{x}_\star^{\mathcal{S}}$ and $\mathcal{X}^{\mathcal{S}} + \bar{\mathbf{z}} - \mathbf{x}_*^{\mathcal{S}} = \mathcal{X} + \bar{\mathbf{z}} - \mathbf{x}_*$. Notice that $\mathbf{Z} \subset \mathcal{F}_\epsilon + \bar{\mathbf{z}} - \mathbf{x}_\star^{\mathcal{S}}$; thus, the same hyperplane also separates $\mathbf{Z}$ and $\mathcal{X} + \bar{\mathbf{z}} - \mathbf{x}_*$. The rest of the proof of Lemma 3 applies without any modifications.