Incentive separability*

Paweł Doligalski Piotr Dworczak Joanna Krysta Filip Tokarski†

August 3, 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 how the analysis of incentive separability can provide 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 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, Stéphane Gauthier, Tai-Wei Hu, Ravi Jagadeesan, Scott Duke Kominers, Raghav Malhotra, Ilya Segal, Hakki Yazici, as well as several seminar audiences for helpful comments. The authors gratefully acknowledge funding from the European Union (ERC, IMD-101040122): Views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or the European Research Council.

†*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 theorem proved portable to diverse settings such as dynamic capital allocation (Ordover and Phelps, 1979), public good provision (e.g., Christiansen, 1981), and taxation with externalities (e.g., Cremer et al., 1998). Gauthier and Laroque (2009) unified 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 examine the logic underlying all of these results using a mechanism-design framework. In our model, a planner maximizes a social welfare function subject to incentive constraints. 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 applications, the set of incentive-separable decisions is determined jointly by what is assumed about agents' preferences and incentive constraints. 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.

We prove two simple results in that general framework. First, it is always optimal to remove distortions between incentive-separable decisions. Second, under regularity assumptions, incentive-separable decisions can be *decentralized*: Agents are allocated type-specific budgets and make (otherwise unrestricted) choices over incentive-separable decisions, taking prices (derived endogenously from the constraints and the planner's objective function) as given. Both results follow easily from our main definition. The key observation is that considering incentive-separable decisions in isolation allows us to effectively ignore incentive constraints, regardless of their exact nature.

We consider three applications of our analysis. First, we show that our characterization 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 short 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.

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. Our results imply 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. These limitations led to the development and popularization of the "perturbation" approach (Piketty, 1997; Saez, 2001; Golosov et al., 2014). Our use of mechanism design 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 General Framework

There is a unit mass of agents with types θ, distributed uniformly on $\Theta \equiv [0, 1]$. 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.[1]

Assuming θ is uniformly distributed is without loss of generality since we impose no restrictions on how preferences depend on it. For example, our framework could be applied to a multidimensional-type environments by suitably mapping the type space to the unit

[1]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."

interval. We can also capture the case of finitely many types by assuming that preferences are piece-wise constant in θ.

Notation. Throughout, we use boldface font to denote aggregate allocations: for any allocation rule x, we let $\mathbf{x} \equiv \int x(\theta)\, d\theta$. For any subset $S \subset \{1, ..., K\}$ of goods, we let $x^S(\theta) \equiv (x_i(\theta))_{i \in S}$, and $x^{-S}(\theta) \equiv (x_i(\theta))_{i \notin S}$. With slight abuse of notation, we will write $x = (x^S, x^{-S})$. Analogously, for a function $\mathcal{S} : \Theta \to 2^{\{1,...,K\}}$ that maps types into subsets of $\{1, ..., K\}$, we let $x^{\mathcal{S}}(\theta) \equiv x^{\mathcal{S}(\theta)}(\theta)$ and $x^{-\mathcal{S}}(\theta) \equiv x^{-\mathcal{S}(\theta)}(\theta)$. 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 $k \in \{1, ..., K\}$. Note that $\mathbf{x} = \mathbf{x}^{\mathcal{S}} + \mathbf{x}^{-\mathcal{S}}$.

Planner's problem. The planner chooses an allocation rule subject to two types of constraints. First, the allocation must satisfy *aggregate constraints*:

$$x \in \mathcal{F} := \{x : \mathbf{x} \in \mathbf{F}\}, \tag{F}$$

where $\mathbf{F}$ is a non-empty subset of $\mathbb{R}^K$ satisfying free disposal. This captures constraints that only depend on aggregates, such as resource availability or technological constraints. Second, the allocation must satisfy *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. For the sake of generality, we impose no structure on $\mathcal{I}$. However, it could capture individual rationality, incentive compatibility, or obedience constraints in settings with adverse selection or moral hazard. An allocation $x \in \mathcal{F} \cap \mathcal{I}$ is called *feasible*.

Let $\mathcal{U}_x$ be the utility profile associated with the allocation rule x, defined by $\mathcal{U}_x(\theta) = U(x(\theta), \theta)$ for any $\theta \in \Theta$. The planner maximizes

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

over feasible allocation rules. 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 capture, in a reduced-form way, the opportunity cost of resources—if the planner could deliver a given utility profile for less, she could allocate the remaining resources to some socially valuable causes.[2] Alternatively, it could capture the planner's preferences over tax revenue.[3]

[2]For example, this formulation lets us capture the planner's preference for providing public goods under the assumption that agents' benefits from them enter additively into their utility functions.

[3]If the planner ultimately only cares about agent's utilities, we make an implicit assumption that any improvement of W from the second argument (the aggregate allocation) can be translated into an improvement of W from the first argument (agents' utility profile) without breaking incentive constraints—a property related to "non-satiation" of the allocation in Gauthier and Laroque (2009).

Incentive separability. We now introduce our key notion of incentive separability.

Definition 1. *Decisions $\mathcal{S} : \Theta \to 2^{\{1,\dots,K\}}$ are incentive-separable (at a feasible allocation x_0) if*

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

To paraphrase, consider a feasible allocation x_0 where decisions $\mathcal{S}$ are incentive-separable. Suppose the allocation of goods $\mathcal{S}(\theta)$ is altered for any type θ in a way that keeps type θ's utility unchanged (and the aggregate constraints satisfied). 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 $\underline{\mathrm{U}}(\theta)$. Then,

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

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

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

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

Then, decisions $\mathcal{S}(\theta) \equiv S$ are incentive-separable. To see why, consider any altered allocation $\tilde{x}^S$ of incentive-separable goods that keeps all agents indifferent. This allocation must then preserve subutility levels given to every agent: $v(\tilde{x}^S(\theta)) = v(x^S(\theta))$ 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^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 $E(\theta) \subset \Theta$. 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.[4]

[4] 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.

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:

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

The consumption vector y is incentive-separable for each type: $\mathcal{S}(\theta) = \{1, ..., K-1\}$ for each $\theta = (\tau, \omega)$. That is, consumption choices are incentive-separable conditional on the realized state ω (but are not incentive-separable *across* the states).[5]

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

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

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 S are incentive-separable at an allocation x_0 for agents whose subutility under x_0 is below $\underline{v}$. That is, $\mathcal{S}(\theta) = S$ for θ such that $v(x_0^S(\theta)) \leq \underline{v}$, and $\mathcal{S}(\theta) = \varnothing$ otherwise.[6]

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

[6] Fix a feasible allocation x_0. Perturb decisions $\mathcal{S}$ so that the new allocation $(x^{\mathcal{S}}, x_0^{-\mathcal{S}})$ satisfies aggregate

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 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

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

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

where the dependence on the objects we keep fixed has been suppressed for brevity.

2 Results

In this section we assume that decisions $\mathcal{S} : \Theta \to 2^{\{1,\ldots,K\}}$ are incentive-separable. We begin with a key 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*

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

To simplify exposition, we assume that problem (7) has a solution, so that an $\mathcal{S}$-undistorted allocation exists (sufficient conditions for existence are provided in Appendix A).

Intuitively, fixing an initial $x_0^{-\mathcal{S}}$, the choice of $x^{\mathcal{S}}$ is undistorted if it maximizes the planner's objective subject to delivering the given utility profile. Importantly, $\mathcal{S}$-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); rather, the allocation of $\mathcal{S}$–goods is the first-best way to deliver the target utility from consuming them (which may itself be distorted).

---

constraints and all agents are indifferent. Indifference implies that $v(x^S(\theta')) = v(x_0^S(\theta'))$ for all θ' with $v(x_0^S(\theta')) \leq \underline{v}$; thus, mimicking these types by some θ will give payoff $U_L(v(x_0^S(\theta')), x_0^{-S}(\theta'), \theta)$, exactly the same as before the perturbation. As a result, incentive compatibility is preserved.

First, we observe that a mechanism that distorts incentive-separable decisions can be improved upon.

Lemma 1 (Optimality). *If x_0 is not $\mathcal{S}$-undistorted, then it can be improved upon in terms of the planner's objective* (W).

Proof. Fix a feasible x_0 that is not $\mathcal{S}$-undistorted. By definition, there exists $x_\star = (x_\star^{\mathcal{S}}, x_0^{-\mathcal{S}})$ that satisfies aggregate constraints, leaves all types' utilities unchanged, and yields $R(\mathbf{x}_\star^{\mathcal{S}}) > R(\mathbf{x_0}^{\mathcal{S}})$. By incentive separability, the first two properties imply that $x_\star \in \mathcal{I}$, and hence $x_\star$ is feasible. The third property implies that $x_\star$ achieves higher objective (W) than x_0. $\square$

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.

Our second observation is that allocations of incentive-separable goods can, under certain conditions, be implemented in a decentralized manner, by letting agents choose incentive-separable goods freely given prices and some θ-dependent budgets.

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 $\lambda^{\mathcal{S}(\theta)} \equiv (\lambda_i)_{i \in \mathcal{S}(\theta)}$.

For the following result, and for the remainder of the paper, we assume that v_θ—the agents' payoff from incentive-separable goods—is locally nonsatiated, for all $\theta \in \Theta$.

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 $m(\theta)$, namely problem (8).[7] Thus, x_0 is regular if and only if $x_0^{\mathcal{S}}$ can be decentralized. $\square$

[7] 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.

Lemma 2 shows that an allocation can be decentralized if and only if it minimizes some *linear* objective function subject to keeping the agent's 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 and no aggregate constraints. Thus, if the planner's objective R is linear and aggregate constraints are slack—as in our first two applications, including the original Atkinson-Stiglitz setting—regularity holds trivially. However, regularity is satisfied also in more complex environments, such as the one we consider in Section 5. In Appendix A, we provide sufficient conditions on the primitives under which regularity holds. Intuitively, even if R is non-linear and the aggregate constraint $\mathcal{F}$ binds, appropriate convexity assumptions guarantee that the solution to problem (7) can be separated from the feasible set by a hyperplane that defines the prices appearing in the linear objective in (9).

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.[8]

Taken together, Lemmas 1 and 2 imply it is optimal to let agents trade incentive-separable goods freely given some prices and budgets. It is instructive to note an analogy to the second welfare theorem. The first step of our analysis (Lemma 1) states that the planner should implement an $\mathcal{S}$-undistorted allocation. Undistortedness is analogous to Pareto efficiency if the planner is viewed as one of the agents.[9] The second step (Lemma 2) states that $\mathcal{S}$-undistorted allocations can be decentralized, under a regularity condition.[10]

3 Atkinson-Stiglitz Theorem

In this section, we apply Lemmas 1 and 2 to prove a generalization of the Atkinson-Stiglitz theorem. We assume that the planner's payoff from providing incentive-separable goods, conditional on the utility profile, is linear: $R(\mathbf{x}^{\mathcal{S}}) = -\lambda \cdot \mathbf{x}^{\mathcal{S}} + \text{const}$, where $\lambda \in \mathbb{R}^K_{++}$. A natural interpretation is that λ represents constant marginal costs of producing incentive-separable goods and that the planner wants to minimize production cost. We also assume the aggregate constraints $x \in \mathcal{F}$ are slack at $\mathcal{S}$–undistorted allocations.

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 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

[8] It is well known that duality between utility maximization and expenditure minimization may fail when the consumer has zero total expenditure in the optimal solution; see Mass-Colell et al. (1995).

[9] $\mathcal{S}$-undistorted allocations are Pareto efficient in the constrained sense: Fixing the level of non-incentive-separable decisions and agents' utilities, it is not possible to increase the "utility" of the planner.

[10] 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. Because we work with a general objective function (W), we must construct the decentralizing prices explicitly.

$\lambda_k > 0$ for $k = 1, ..., K-1$. Thus, 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. The aggregate constraint, requiring that enough labor is supplied to cover the production costs, is slack.[11]

Theorem 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.*

Proof. By Lemma 1, the $\mathcal{S}$-undistorted allocation $(x_\star^{\mathcal{S}}, x_0^{-\mathcal{S}})$ improves upon the original one. Moreover, the $\mathcal{S}$-undistorted allocation is regular with prices proportional to marginal costs λ. Hence, by Lemma 2, $x_\star^{\mathcal{S}}$ can be decentralized with those prices. □

Let us describe Theorem 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, 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.

Theorem 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

[11] The aggregate constraint reads $\mathbf{L} - \sum_{k=1}^{K-1} \lambda_k \mathbf{y}_k \geq 0$. Since the planner's objective is to maximize the left-hand side of this inequality, which represents the tax revenue, the aggregate constraint can be dropped from the planner's problem without affecting the solution.

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 Theorem 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.[12] 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 Theorem 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*.[13]

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 $v(x^E)$ measure the nutritional value of bundle x^E, common to all agents. Idiosyncratic tastes affect food choices only once a nutritional

[12] 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.

[13] 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.

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

$$U(x,\theta) = \begin{cases} U_L(v(x^E), x^{-E}, \theta) & \text{if } v(x^E) \leq \underline{v} \\ U_H(x,\theta), & \text{otherwise.} \end{cases} \tag{11}$$

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 the set of incentive-compatible allocations $\mathcal{I}$ as in Example 2. 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).

Theorem 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 $v(x_0^E(\theta)) \leq \underline{v}$, and letting them spend these budgets on food items E (but not other goods) priced at marginal costs.*

Proof. By Example 4, decisions E are incentive-separable for types θ for which $v(x_0^E(\theta)) \leq \underline{v}$ at the allocation x_0. Furthermore, constant marginal costs of food production imply regularity. Thus, by Lemmas 1 and 2, the planner can improve the objective by allocating type-dependent budgets to these types that can be spent on food items at prices λ. □

Intuitively, under the initial allocation x_0, agents are separated into two groups. 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).[14] 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 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.[15] 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

[14] 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.

[15] 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).

status. We assume that, conditional on the realized employment status, preferences over food consumption have the structure assumed in (11). By Example 3, Theorem 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.[16]

5 Atkinson-Stiglitz meet Diamond-Mirrlees

We now apply our methods to production economies with a potentially complex input-output structure. 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 feasible for 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 $j \in \{1, \ldots, J\}$, we can express feasibility of the aggregate production simply as $\mathbf{z} \in \mathbf{Z}$.[17]

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 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 $\mathbf{z}_1 \in \mathbf{Z}$ such that $\mathbf{z}_1 \geq \mathbf{z_0}$ and $\mathbf{z}_1 \neq \mathbf{z_0}$.

Note that the production plan z affects neither individual utilities nor incentive constraints. Thus, it is trivially incentive-separable, implying the following result:[18]

Corollary 1. *Consider a 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 1 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.

[16] 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).

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

[18] 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 1 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 1 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 motive in a model where labor supply of different types is not perfectly substitutable in production but cannot be observed by the planner.[19]

Our next result concerns decentralizing $\mathcal{S}$-undistorted allocations in production economies. The main difficulty is ensuring that such allocations exist and are regular. Lemma 3 in Appendix A guarantees exactly that, provided the following assumptions hold:

Assumption 1. *For every allocation $x_0 : \Theta \to \mathbb{R}^K_+$, there exists an integrable function $\theta \mapsto \bar{x}(\theta) \in \mathbb{R}^{|\mathcal{S}(\theta)|}_+$ such that $v_\theta(y) = v_\theta(x_0^{\mathcal{S}}(\theta))$ implies $y \leq \bar{x}(\theta)$.*

Assumption 2. *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 $\mathbf{z}_1 \in \mathbf{Z}$ such that $\mathbf{z}_1^A \leq \mathbf{z}_0^A$, $\mathbf{z}_1^{-A} \geq \mathbf{z}_0^{-A}$ and $\mathbf{z}_1^{-A} \neq \mathbf{z}_0^{-A}$.*

By Assumption 1, the agents' indifference curves over incentive-separable goods are bounded, and the bound is an integrable function of θ. As a result, the set of admissible aggregate consumption allocations is compact (see Appendix A for an additional discussion). Assumption 2 implies that the set of admissible production plans is compact as well. This guarantees that an $\mathcal{S}$-undistorted allocation exists. In addition, Assumption 2 states that $\mathbf{Z}$ is convex and that it is always technologically possible to produce less of any subset of goods in exchange for producing more of some of the other goods. These two properties ensure regularity: The former lets us use the separating hyperplane theorem to find the decentralizing prices, while the latter implies that these prices are strictly positive.

Theorem 3 (Atkinson-Stiglitz meet Diamond Mirrlees)**.** *Suppose Assumptions 1 and 2 hold. 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*

[19] Naito (1999) considers an economy with two unobserved types of workers: high-skilled and low-skilled. The payoff of the high-skilled from mimicking the low-skilled depends on their relative wage rates. When the incentive constraint of the high-skilled is binding, the planner finds it optimal to overhire low-skilled workers (relative to the production-efficient benchmark) to inflate their wage rate and, thus, relax the binding incentive constraint. In our mechanism-design approach, unobservability of different types of labor means that the set $\mathcal{I}$ must include obedience constraints. The agents' payoffs from deviating from the recommended labor supply decisions necessarily depend on market wages. Hence, the set $\mathcal{I}$ varies with aggregate variables determining these wages, which breaks incentive-separability.

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.

Proof of Theorem 3. By Lemma 3 (and Remark 1) in Appendix A, there exists an $\mathcal{S}$-undistorted allocation $x_* = (x_*^{\mathcal{S}}, x_0^{-\mathcal{S}})$ that is regular with some prices $\lambda \in \mathbb{R}_{++}^K$ and an associated production plan z_* that is efficient.[20] Then, Corollary 1 implies that welfare can be (weakly) improved by implementing (x_*, z_*).

Statement *(i)* follows from Lemma 2. It remains to show statement *(ii)*. Remark 1 to Lemma 3 states that the aforementioned production plan z_* satisfies

$$\mathbf{z}_* \in \arg\max_{\mathbf{z} \in \mathbf{Z}} \lambda \cdot \mathbf{z}. \tag{12}$$

Recall that $\mathbf{z}_* = \sum_{j=1}^J z_*^j$, where $z_*^j \in Z^j, \forall j$. Since the objective in (12) is linear, we can rewrite this problem as a collection of J maximization problems of individual firms:

$$z_*^j \in \arg\max_{z^j \in Z^j} \lambda \cdot z^j, \quad \forall j = 1, \ldots, J. \tag{13}$$

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 (13). $\square$

Theorem 3 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 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 an extra consumption tax on some incentive-separable commodity and an extra output subsidy on the same good. If the tax and 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.

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

[20] Assumption A0 of Lemma 3 is satisfied trivially in the production setting by identifying the set $\mathbf{F}$ with the aggregate production set $\mathbf{Z}$.

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—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.[21] 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.

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.

[21] Note that the objective function (W) nests revenue-maximization as a special case: It suffices to include money as one of the decisions, and drop the dependence of W on the utility profile.

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 Existence and regularity of $\mathcal{S}$-undistorted allocations

In this appendix, we provide sufficient conditions for existence and regularity of $\mathcal{S}$-undistorted allocations. Apart from completing the analysis of Section 2, this is a key step in the proof of Theorem 3 in Section 5. We first assume that the general problem admits an interpretation similar to that of the production setting we consider in Section 5—this allows us to streamline exposition.

Assumption A0. *The planner's objective function takes the form: $W(\mathcal{U}_x, \mathbf{x}) = max_{\mathbf{z}\geq\mathbf{x},\, \mathbf{z}\in\mathbf{F}} \mathcal{W}(\mathcal{U}_x, \mathbf{z}-\mathbf{x})$, where $\mathcal{W}$ is continuous and non-decreasing in the second argument.*

Assumption A0 makes explicit our intuitive description of the second argument in the planner's objective W: The planner wants to maximize the amount of resources that remain available to her after allocating them to agents to deliver the given utility profile. Under this assumption, the auxiliary variable $\mathbf{z}$ represents supply decisions, and the planner's problem admits a convenient separation under which $\mathbf{z}$ is chosen subject to aggregate constraints $\mathbf{F}$, x is chosen subject to incentive constraints $\mathcal{I}$, and the two parts of the problem are connected via the market-clearing constraint $\mathbf{x} \leq \mathbf{z}$:

$$\max_{x,\mathbf{z}} \mathcal{W}(\mathcal{U}_x, \mathbf{z}-\mathbf{x}) \text{ subject to } \mathbf{z} \in \mathbf{F},\ x \in \mathcal{I},\ \mathbf{x} \leq \mathbf{z}.$$

A problem of this form is introduced explicitly in Section 5.

We can now state the remaining assumptions separately for the "demand" and "supply" sides of the problem.

Assumption A1. *For every allocation $x_0 : \Theta \to \mathbb{R}^K_+$, there exists an integrable function $\theta \mapsto \bar{x}(\theta) \in \mathbb{R}^{|\mathcal{S}(\theta)|}_+$ such that $v_\theta(y) = v_\theta(x_0^{\mathcal{S}}(\theta))$ implies $y \leq \bar{x}(\theta)$.*

Assumption A1 ensures the agents' indifference curves over incentive-separable goods are bounded, and the bound is an integrable function of θ. This is needed to prove that the feasible aggregate consumption set is closed in a setting with a continuum of agents. While the assumption may seem 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.

Assumption A2. *The aggregate production set* $\mathbf{F}$ *is closed, bounded from above and convex. Furthermore, for any* $\mathbf{z}_0 \in \mathbf{F}$ *and any nonempty proper subset* $A \subset \{1, ..., K\}$, *there exists* $\mathbf{z}_1 \in \mathbf{F}$ *such that* $\mathbf{z}_1^A \leq \mathbf{z}_0^A$, $\mathbf{z}_1^{-A} \geq \mathbf{z}_0^{-A}$ *and* $\mathbf{z}_1^{-A} \neq \mathbf{z}_0^{-A}$.

Assumption A2 states the set of feasible supply plans is convex and compact, and that it is always technologically possible to produce less of any subset of goods in exchange for producing more of some of the other goods. While compactness is needed for existence, the remaining assumptions ensure that we will be able to separate the set of feasible supply plans by a hyperplane. The last part of Assumption A2 is restrictive; in particular, it rules out the case of exogenously fixed supply of some resource. Intuitively, it ensures that the boundary of $\mathbf{F}$ never becomes vertical or horizontal, which allows us to prove that prices pinned down by the separating hyperplane are strictly positive, as required in the definition of regularity in Lemma 2.[22]

Lemma 3. *Under Assumptions A0, A1, and A2, for any feasible allocation* x_0, *there exists a feasible allocation* $x_\star = (x_\star^{\mathcal{S}}, x_0^{-\mathcal{S}})$ *that is* $\mathcal{S}$*–undistorted, regular with some strictly positive prices, and satisfies* $v_\theta(x_\star^{\mathcal{S}}(\theta)) = v_\theta(x_0^{\mathcal{S}}(\theta))$ *for all* $\theta \in \Theta$.

When the above assumption holds, Lemma 1, 2 and 3 together guarantee that, starting at any feasible allocation x_0, the planner can increase her objective (W) by swapping to an $\mathcal{S}$-undistorted allocation and decentralizing it.

Proof of Lemma 3. Consider the following maximization problem:

$$\max_{x^{\mathcal{S}}, \mathbf{z}} \mathcal{W}(\mathcal{U}_{x_0}, \mathbf{z} - (\mathbf{x}^{\mathcal{S}} + \mathbf{x}_0^{-\mathcal{S}})) \text{ subject to} \tag{14}$$
$$\mathbf{z} \in \mathbf{F}, \quad \mathbf{z} \geq \mathbf{x}^{\mathcal{S}} + \mathbf{x}_0^{-\mathcal{S}}, \quad v_\theta(x^{\mathcal{S}}(\theta)) = v_\theta(x_0^{\mathcal{S}}(\theta)), \ \forall \theta \in \Theta.$$

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.

Let us reformulate this problem. For every $\theta \in \Theta$, define

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

and

$$\mathcal{X} = \left\{ \mathbf{x}^{\mathcal{S}} + \mathbf{x}_0^{-\mathcal{S}} : \ x^{\mathcal{S}}(\theta) \in \mathcal{X}_\theta^{\mathcal{S}}, \ \forall \theta \in \Theta \right\}.$$

Set $\mathcal{X}$ contains aggregate vectors of goods which can be distributed among agents to keep their allocation of non-incentive-separable goods fixed and their utility from incentive-separable goods unchanged, without any resources to spare.

By Theorem 1 of Aumann (1965), $\mathcal{X}$ is convex. Because each v_θ is continuous, $\mathcal{X}_\theta^{\mathcal{S}}$ is closed; by Assumption A1 the set-valued function $\mathcal{X}_\theta^{\mathcal{S}}$ is uniformly bounded, as defined in Aumann

[22] It is easy to construct examples with a fixed supply of some good in which it is not possible to decentralize the allocation with strictly positive prices; intuitively, the planner may have a (locally) infinite rate of substitution between two goods, and if the desired good is in fixed supply, the price of the other good must be zero. We could relax the assumption if we instead appropriately bounded the marginal rates of substitution in the planner's preferences.

(1965). Thus, by Theorem 4 of Aumann (1965), $\mathcal{X}$ is compact. It follows that our problem can be written simply as

$$\max_{(\mathbf{x},\mathbf{z})\in\mathcal{X}\times\mathbf{F}} \mathcal{W}(\mathcal{U}_{x_0}, \mathbf{z}-\mathbf{x}) \text{ subject to } \mathbf{z}\geq\mathbf{x}. \tag{15}$$

Notice that, despite changing the choice variable, every solution $(\mathbf{x}_\star, \mathbf{z}_\star)$ to (15) has a corresponding $(x_\star^{\mathcal{S}}, \mathbf{z}_\star)$ that solves (14). While the planner optimizes with respect to the entire aggregate consumption allocation $\mathbf{x}$, the choice set $\mathcal{X}$ implicitly keeps the allocation of non-incentive-separable goods fixed at $x_0^{-\mathcal{S}}$.

Recall that $\mathbf{F}$ is closed and bounded from above by assumption. We can also put a lower bound on admissible elements of $\mathbf{F}$—the resource constraint requires that $\mathbf{z}\geq\mathbf{x}$. Since aggregate consumption cannot be negative, we can bound the admissible subset of $\mathbf{F}$ by 0 in each dimension. Thus, the choice set is compact. Since, in addition, $\mathcal{W}$ is continuous and the choice set nonempty—it contains $(\mathbf{x}_0, \mathbf{z}_0)$—the solution to the problem (15) exists.

Denote a solution to problem (15) by $(\mathbf{x}_*, \mathbf{z}_*)$. Since $\mathcal{W}$ is non-decreasing in $\mathbf{z}-\mathbf{x}$, without loss of generality there is no admissible $\mathbf{z}'-\mathbf{x}'$ such that $\mathbf{z}'-\mathbf{x}' \gg \mathbf{z}_*-\mathbf{x}_*$. This implies that $\mathrm{int}\mathbf{F}\cap\mathrm{int}(\mathcal{X}+\mathbf{z}_*-\mathbf{x}_*)=\varnothing$. Graphically, set $\mathcal{X}$ translated by $\mathbf{z}_*-\mathbf{x}_*$ is to the north-east of $\mathbf{F}$, with only the boundaries of the two sets touching. By the separating hyperplane theorem, there exists a hyperplane with coefficients $\lambda\in\mathbb{R}^K\neq\mathbf{0}$ that separates $\mathrm{int}(\mathbf{F})$ and $\mathrm{int}(\mathcal{X}+\mathbf{z}_*-\mathbf{x}_*)$ at $\mathbf{z}_*$:

$$\mathbf{z}_*\in\arg\max_{\mathbf{z}\in\mathbf{F}}\lambda\cdot\mathbf{z}, \tag{16}$$

$$\mathbf{z}_*\in\arg\min_{\mathbf{z}\in\mathcal{X}+\mathbf{z}_*-\mathbf{x}_*}\lambda\cdot\mathbf{z}. \tag{17}$$

Since $\mathbf{F}$ allows for free disposal, it must be that $\lambda\geq\mathbf{0}$—the hyperplane is (weakly) downwards sloping. Moreover, we will show that $\lambda\gg 0$. Suppose that λ is zero at coordinates from a nonempty subset $A\subset\{1,\dots,K\}$. It follows that the set of remaining coordinates where it is strictly positive, denoted by $-A$, is non-empty. By Assumption A2, there exists $\mathbf{z}_1\in\mathbf{F}$ such that $\mathbf{z}_1^A\leq\mathbf{z}_*^A$, $\mathbf{z}_1^{-A}\geq\mathbf{z}_*^{-A}$ and $\mathbf{z}_1^{-A}\neq\mathbf{z}_*^{-A}$. Then, $\lambda\cdot\mathbf{z}_1>\lambda\cdot\mathbf{z}_*$, which contradicts (16). Thus, $\lambda\gg 0$.

By a change of variable, (17) can be rewritten as

$$\mathbf{z}_*-(\mathbf{z}_*-\mathbf{x}_*)\in\arg\min_{\mathbf{x}\in\mathcal{X}}\lambda\cdot\mathbf{x}. \tag{18}$$

Note that $\mathbf{z}_*-(\mathbf{z}_*-\mathbf{x}_*)=\mathbf{x}_*$. Given that $\lambda\gg 0$, this implies that x_* is regular. □

Remark 1. Note that the proof of Lemma 3 also establishes the existence of the supply vector $\mathbf{z}_*$ associated with the $\mathcal{S}$-undistorted allocation x_*. Furthermore, $\mathbf{z}_*\in\arg\max_{\mathbf{z}\in\mathbf{F}}\lambda\cdot\mathbf{z}$, which implies that $\mathbf{z}_*\geq\mathbf{z}'$ for all $\mathbf{z}'\in\mathbf{F}$. Thus, interpreting $\mathbf{z}_*$ as an aggregate production vector from Section 5, the corresponding production plan z_* is efficient.