Optimal Redistribution via Income Taxation and Market Design

Paweł Doligalski ⓡ Piotr Dworczak
Mohammad Akbarpour ⓡ Scott Duke Kominers*

January 5, 2025

Abstract

Policymakers often intervene in goods markets to effect redistribution—for example, via price controls, differential taxation, or in-kind transfers. We investigate the optimality of such policies alongside the (optimally-designed) income tax. In our framework, agents possess private information about their ability to generate income and consumption preferences, and a planner maximizes a social welfare function subject to resource constraints. We uncover a generalization of the Atkinson-Stiglitz theorem by showing that goods markets should be undistorted if (i) individual utility functions feature no income effects, (ii) redistributive preferences depend only on agents' ability, and (iii) there is no statistical correlation between ability and taste for goods. We also show, however, that the conclusion of the Atkinson-Stiglitz theorem fails if any of the three assumptions is relaxed. In a special case of our model with linear utilities, binary ability, and continuous willingness to pay for a single good, we characterize the globally optimal mechanism and show that it may feature means-tested consumption subsidies, in-kind transfers, and differential commodity taxation.

*The order of the first and second as well as third and fourth co-author is randomized, as indicated by the ⓡ symbol. Dworczak acknowledges the support received under the ERC Starting grant IMD-101040122. Kominers acknowledges the support of the Ng Fund and the Mathematics in Economics Research Fund of the Harvard Center of Mathematical Sciences and Applications, and the Digital Data Design D^3 Institute at Harvard. Kominers is a Research Partner at a16z crypto (for general a16z disclosures, see https://www.a16z.com/disclosures/). Notwithstanding, the ideas and opinions expressed herein are those of the authors, rather than of a16z or its affiliates. *Doligalski:* University of Bristol. *Dworczak:* Northwestern University; and Group for Research in Applied Economics. *Akbarpour:* Graduate School of Business, Stanford University. *Kominers:* Entrepreneurial Management Unit, Harvard Business School; Department of Economics, Harvard University; and a16z crypto.

1 Introduction

Policymakers often distort market allocations as a way of addressing inequality. In developed countries, for example, it is common for local housing authorities to impose rent-control policies, provide affordable housing, or subsidize construction of rental housing for low-income tenants. Food assistance is also prevalent—for example, in the form of food stamps in the United States. And recently, many European countries imposed caps on electricity prices to shield vulnerable households from sharp increases in energy bills. In developing countries, likewise, in-kind provision of food items and subsidized energy consumption have long been an important part of the safety net.

Market-level redistributive policies defy conventional economic wisdom—rooted in the welfare theorems—that market interventions compromise efficiency and thus should be avoided (absent market failures). A recent literature on *inequality-aware market design* has pointed out, however, that policies such as price controls and rationing may be justified on welfare grounds if policymakers lack the information or instruments needed to effect redistribution through targeted lump-sum transfers.[1] The key intuition is that agents' behavior in the market may reveal information about their welfare weights: in settings where agents' redistribution-worthiness is not directly observed, willingness to pay for a good may be correlated with welfare-relevant characteristics (such as income level or wealth). By modifying market-clearing rules to induce appropriate self-selection, the designer can trade off efficiency in the market with equity—distorting the allocation in order to effect redistribution to agents with higher levels of need.

However, arguments in favor of market interventions remain incomplete without considering the role of broader policy instruments that address inequality. In particular, income taxation is often thought of as the primary—and ideal—tool for effecting redistribution in the presence of incentive constraints (see, e.g., Kaplow (2011) and the references therein). Thus, we ask: Can redistribution through markets be justified if the policymaker also controls income taxation? And if yes, how do market interventions interact with the income tax to strike the balance between equity and efficiency?

The public finance literature has provided the answer in a core benchmark case: By the Atkinson-Stiglitz theorem, if agents only differ in their ability to generate income (and preferences satisfy a weak-separability assumption), income taxes alone are sufficient to maximize social welfare for any set of welfare weights (Atkinson and Stiglitz (1976)). In other

[1] See, for example, the work of Condorelli (2013), Dworczak ⓡ al. (2021), and Akbarpour ⓡ al. (2024)—building on classical insights such as those of Weitzman (1977), Spence (1977) and Nichols and Zeckhauser (1982).

words, goods market interventions are redundant at best and harmful at worst. Under this perspective, redistribution through markets can be defended only if the policymaker lacks the ability to adjust income taxes to the socially optimal levels.

In this paper, we show that market interventions can be a valuable redistributive tool more broadly. Specifically, we examine the desirability of interventions in goods markets in the presence of *multidimensional* heterogeneity, where individuals differ both in their productive ability and in their tastes for goods. In line with the public finance literature studying the robustness of the Atkinson-Stiglitz theorem, we find that once heterogeneity in tastes is introduced, it may be beneficial to supplement the income tax with market interventions. In contrast to most of that literature, by employing a mechanism design framework under a few simplifying assumptions, we are able to characterize the optimal *combination* of income taxation and redistributive market design. Our characterization of the optimal multidimensional mechanism shows that it is typically optimal to use *both* income taxation and redistributive market design to resolve the equity-efficiency trade-off.

Our analysis has two parts. As a baseline, we first prove an extension of the Atkinson-Stiglitz theorem to our setting with multidimensional heterogeneity in ability and tastes; however, this is only possible under three strong assumptions. We show that income taxation alone is sufficient (and market interventions are redundant) when:

- A1. There are no income effects;[2]
- A2. Welfare weights depend only on agents' ability levels, and not on their tastes for goods;
- A3. Tastes for goods and ability are statistically independent.

Assumptions A1 and A2 together imply that the planner has no desire to redistribute between agents with different tastes; thus, the planner is concerned only with redistribution across ability levels. By Assumptions A1 and A3, meanwhile, willingness to pay for goods is uninformative about ability. Thus, under Assumptions A1–A3, distorting choices in the goods market neither serves a valuable redistributive role on its own, nor screens agents' ability levels—and hence it should be avoided.

In the second part of our analysis, we show that the assumptions used to derive the Atkinson-Stiglitz result in our setting are "tight" in the sense that if we relax any one of them, the conclusion of the Atkinson-Stiglitz theorem fails. We characterize the optimal mix of income

[2]Specifically, we require that there exists a consumption good which enters the utility of all individuals linearly (with an identical coefficient) and which can be consumed in any real (incl. negative) quantity.

tax and market intervention in a simplified specification that permits a tractable analysis of optimal mechanisms in the presence of multidimensional heterogeneity. The simplified framework features an intensive-margin choice of two commodities: a good x, which is singled out for potential redistributive market intervention, and a numeraire c, which can be thought of as an aggregate consumption good. Agents have utilities that are linear in consumption and labor supply. There is a continuum of taste types, but ability is a binary characteristic, with low-ability agents unable to generate any income.[3]

While stylized, our setting can be considered a best-case scenario for income taxation: The optimal income tax always leads to efficient labor market outcomes—which means that interventions in goods markets are never aimed at reducing distortions in the labor market; rather, they directly effect redistribution. Furthermore, rich heterogeneity in tastes for goods allows us to focus on market-design implications, which have received far less attention than the optimal income tax in the literature following Atkinson and Stiglitz (1976).

We first relax Assumption A1 (no income effects) by supposing that consumption of the numeraire cannot fall below some subsistence level $\underline{c}$. Absent any market intervention, agents with low ability (and, thus, low income) and high taste must limit their consumption of good x to maintain subsistence. When the designer has redistributive preferences, the optimal mechanism in this setup looks as follows: First, the income tax redistributes from rich (high-ability) to poor (low-ability) as much as possible subject to maintaining the high-ability agents' incentive to work. Second, the good x is subsidized at low consumption levels and taxed at high consumption levels. The subsidy at low levels of x creates a (first-order) welfare gain by relaxing low-ability, high-taste agents' subsistence constraint—allowing them to consume more of the good. This positive effect dominates the (second-order) negative distortion of making some low-ability agents over-consume good x. Meanwhile, the planner can be certain that agents at higher consumption levels of good x are of high ability, using the fact that their labor income relaxes their subsistence constraint. Thus, the planner can tax purchases of x on the margin to raise additional revenue (which in turn is used to subsidize the sale of lower amounts of the good at below-market prices). In particular, if the planner cares only about the low-ability types, the marginal after-tax price should be set to the price that would be chosen by the revenue-maximizing monopolist.

The structure of optimal policy—in particular, the involvement of some degree of market intervention—is robust to income effects generated by a strictly concave utility from numeraire. Under concave utility for the numeraire, agents with higher taste for good x

[3] To make the problem well-defined, we also assume that the planner attaches weakly higher welfare weights to low-ability agents.

consume less of numeraire and thus have higher marginal utility of income. By subsidizing the purchases of low amounts (or low quality) of good x, the planner is able to redistribute the numeraire towards high-taste, low-ability agents with high marginal utility of income. If the planner assigns no (or low) welfare weight to high-ability agents, she also taxes high levels of consumption of good x, which allows her to further redistribute from high-ability agents to low-ability ones.

Relaxing either Assumption A2 or Assumption A3 creates a correlation between welfare weights and tastes and, thus, leads to a motive to redistribute across the taste types. These two cases are technically similar and we discuss them together. The optimal mechanism takes one of two possible forms, depending on the relative strength of redistributive preferences across abilities and tastes. If the planner is mainly concerned with the "vertical" inequality between high-ability and low-ability agents, then the redistribution via the income tax is maximized (subject to preserving incentives of high-ability agents to work). This income redistribution is complemented by a simple market intervention: Purchases of the good are *taxed* when agents consuming it have lower welfare weights on average (e.g., when there is positive correlation between taste and ability) and *subsidized* otherwise (e.g., when there is negative correlation between taste and ability). If the planner is predominantly concerned with "horizontal" inequality between high-taste and low-taste agents, then redistribution via the income tax is scaled down to allow for a greater role of redistribution through markets.[4] Specifically, the income tax rate is lowered so that high-ability agents maintain a positive surplus from working. This leaves slack in the downward incentive constraints in ability, allowing the planner to use income-dependent prices for good x. The planner then offers a means-tested subsidy that is available only to low-income individuals, which benefits agents with low ability and high taste for the good.

In summary, our analysis suggests that there may be a number of markets for which it is optimal to complement income tax policies with redistributive market design. The key intuition for why this happens closely aligns with the motivation for redistributive market design described above: in many goods markets, it is possible to use agents' purchasing behavior to infer welfare-relevant information that an income tax policy has no way of conditioning on directly. Thus, distortions can be justified in markets where consumption choices are particularly informative for redistribution, e.g., because they induce significant income effects (related to A1), directly correlate with welfare weights (related to A2), or induce correlation with welfare weights through a statistical link to ability (related to A3). This lends support for some redistributive market interventions used in practice while at

[4] In fact, when there is no concern about vertical inequality, welfare weights depend only on tastes, and Assumptions A1 and A3 hold, then the income tax becomes redundant.

the same time informing the optimal interaction of such policies with the income tax.

The classic Atkinson-Stiglitz framework masks this potential role for redistributive market design by implicitly assuming that all of the welfare-relevant information revealed through market behavior is redundant (because it is already revealed by agents' labor supply choices). To illustrate the difference, it is instructive to contrast the reasoning just presented with the logic of the original Atkinson-Stiglitz theorem in the context of a specific example, for instance healthcare. In both frameworks, richer agents tend to consume more healthcare. However, because agents in the model of Atkinson and Stiglitz (1976) only differ in their earnings ability, high consumption of healthcare is not a signal of need but merely a direct manifestation of high income. It is thus most efficient to redistribute income directly by taxing it—any intervention in the healthcare market is an imperfect substitute for the optimal income tax policy. In practice, demand for healthcare stems from the combination of ability to pay (i.e., income) and need (i.e., taste). Heterogeneity in preferences makes need one of the factors shaping demand for healthcare in our framework. Thus, subsidies for low consumption (or low quality) of healthcare services can improve redistribution relative to income tax alone because they endogenously target the low-ability agents who have a particularly high marginal utility of income—i.e., the agents who are consuming low amounts of healthcare relative to need.

1.1 Structure of the paper

The remainder of the paper is organized as follows: The next subsection describes the related literature in market design and public finance. We introduce our framework in Section 2. Then, in Section 3, we extend the Atkinson-Stiglitz theorem to a setting with multidimensional heterogeneity under the assumptions A1–A3. In Section 4, we relax each of the three assumptions in turn and characterize the optimal interaction between the income tax and goods market interventions in the simplified framework. We discuss the general structure of the optimal mechanism in Section 5, along with some details of the characterization method. We present brief concluding remarks in Section 6.

1.2 Related literature

Previous work on inequality-aware market design (e.g., Condorelli (2013), Dworczak ⓡ al. (2021), Kang and Zheng (2022), Akbarpour ⓡ al. (2024), Kang (2023)) focused on the problem of designing a market for a single good. The underlying assumption of that approach has been that the designer does not control other redistributive tools such as income taxation—for example, because the designer is a local authority and taxes are set at national or regional level, or because political economy frictions block the government from adjusting taxes to the optimal redistributive target. We show that in fact redistribution through markets may be optimal *even when the designer does control income taxation*. At the same time, we show that the interaction between the two tools is non-trivial. For example, the market intervention may take the form of a means-tested subsidy for consumption or an income-independent in-kind transfer combined with a tax on top-up consumption; income taxes may sometimes be lowered in order to incentivize labor provision by high-ability agents when low-ability agents face lower prices in the goods market.

In a related approach, Kang (2024), Pai and Strack (2024), and Ahlvik et al. (2024) studied the problem of optimal regulation of consumption of goods generating an externality (e.g., pollution) when the designer has redistributive preferences and agents differ in their tastes for goods. Pai and Strack (2024) extended their results to the case when income taxes are present but are set exogenously, while Ahlvik et al. (2024) allowed for joint design of an income tax and a consumption tax. This work showed that Pigovian taxation—a classical solution to the externality problem—must be appropriately modified to account for redistributive concerns, and that the income tax alone is not sufficient to address those concerns in the presence of taste heterogeneity.

We contribute to the multidimensional screening literature by identifying a tractable class of models with the property that one dimension of the type space is continuous and the other binary. Models with a similar structure (although in different economic contexts) have been studied by Fiat et al. (2016) and Li (2021). Our approach to solving the model is different: Relying on the structure of the problem, we represent the incentive compatibility constraint across the two types as an outside option constraint, as in the work of Jullien (2000). We then adopt a recent solution technique developed by Dworczak and Muir (2024) to solve for the optimal allocation rule for one type, when the other type's allocation is fixed. Finally, we use the linearity of the problem to argue that the full solution features allocation rules that are step functions with a limited number of steps (see Section 5 for the technical details and additional comments on the related literature).

Our way of modeling the income effect as a subsistence constraint connects our framework to models with budget-constrained agents. Most closely related is the work of Che et al. (2013) who studied optimal allocation of resources when agents differ in values for the good and budgets. However, Che et al. (2013) study efficient allocations, without the redistributive concerns that are the core focus of our work.

Meanwhile, a rich public finance literature has studied the optimal design of income and consumption taxes under heterogeneity in abilities and preferences. Atkinson and Stiglitz

(1976) and Mirrlees (1976) set up the overall agenda and introduced the framework that is now standard in analyzing optimal taxation. Several papers have considered one-dimensional heterogeneity under which tastes and abilities are perfectly correlated (see, e.g., Cremer and Gahvari (1998); Golosov et al. (2013); Gerritsen et al. (2020); Scheuer and Slemrod (2020); Schulz (2021); Hellwig and Werquin (2024)). Within the work that considers multidimensional settings, Cremer et al. (2001, 2003), Diamond and Spinnewijn (2011) and Gauthier and Henriet (2018) characterized a nonlinear income tax and linear consumption (including capital and inheritance) taxes. Moser and Olea de Souza e Silva (2019) considered agents who are heterogeneous in abilities and in the strength of present-bias, and studied the optimal (paternalistic) savings policies. Doligalski et al. (2024) studied a setting in which agents have heterogeneous tastes over food items and showed that a food voucher program might be optimal. Other related studies, which are not directly concerned with a treatment of commodities, include those of Golosov and Krasikov (2023) and Spiritus et al. (2022) on taxation of couples and Boerma et al. (2022) on the optimal bunching patterns in the labor market with multidimensional sorting. We contribute to this literature by providing precise conditions under which the income tax is sufficient (and interventions in the goods markets are redundant) in the multidimensional setting. Furthermore, we provide a multidimensional framework where the optimal interaction between the income tax and market interventions is nontrivial—it can involve nonlinear consumption taxes—and can be characterized in closed form.

There are also related studies that examine desirability of goods market interventions by using sufficient statistics. Saez (2002) considered a setting where individuals differ in abilities and tastes, and derived conditions under which introducing a small tax on one of the goods cannot improve social welfare if the policymaker can use a nonlinear income tax. Ferey et al. (2021) extended the Saez (2002) approach by deriving the formulas for optimal commodity and income taxes and estimating the relevant sufficient statistics from the data. Our results are consistent with the conditions in Saez (2002) but provide a complementary perspective by fleshing out the precise assumptions on the model primitives—rather than endogenous sufficient statistics—that make commodity distortions redundant, and by characterizing the optimal interaction between income taxes and market interventions when these assumptions are not met. Furthermore, our results describe global optimum with respect to arbitrary mechanisms that could include rationing, quotas or public provision of goods. By contrast, conditions based on sufficient statistics are necessarily local and informative about the effects of small tax reforms only.

2 General Framework

Our framework features agents who are heterogeneous in ability and taste and make optimal labor supply and consumption choices, as well as a planner who aims to maximize social welfare subject to incentive-compatibility and resource constraints.

There is a unit mass of agents who differ in both their *taste for goods* $t \in \Theta_t$ and their *earning ability* $a \in \Theta_a$. Types $\theta = (t,a) \in \Theta_t \times \Theta_a = \Theta \subseteq \mathbb{R}^2$ are jointly distributed according to $F(\theta)$. Agents have preferences over a vector of *goods* $x \in \mathbb{R}_+^L$, a *numeraire* consumption good $c \in \mathbb{R}$, and *earnings* $z \in \mathbb{R}_+$, as given by the utility function

$$U((c,x,z),(t,a)) = u(c) + v(x,t) - w(z,a) \tag{1}$$

which is continuous in (c,x,z) and measurable in (t,a). The separation of consumption into a vector of goods x and a one-dimensional numeraire c is convenient for studying different cases of the model (e.g., with and without income effects). Conceptually, one can think of x as a set of goods that are singled out for potential intervention, and of c as aggregating the consumption of all remaining goods into a single composite commodity.[5] We assume that $u(c)$ is strictly increasing. While we do not (at this point) assume specific functional forms for the different components of utility, we do assume that utility is additively separable between the numeraire c, goods x, and earnings z.[6]

The planner chooses an *allocation rule* $Y = (c,x,z) : \Theta \to \mathbb{R} \times \mathbb{R}_+^{L+1}$ to maximize the expected utility of agents weighted with welfare weights $\lambda(\theta) \geq 0$. The average welfare weight is normalized to 1. That is, the social objective is

$$\int \lambda(\theta)\, U(Y(\theta), \theta)\, dF(\theta). \tag{2}$$

The planner faces the resource constraint

$$\int [z(\theta) - c(\theta) - k \cdot x(\theta)]\, dF(\theta) \geq G, \tag{3}$$

where $k \in \mathbb{R}_{++}^L$ is the marginal cost of producing goods x in terms of numeraire (or earnings) and G represents the minimum revenue requirement (which could be negative, representing exogenous revenue sources).

[5]Alternatively, c can be interpreted as leftover "money"; such an interpretation is often adopted in quasilinear models in which $u(c) = c$.

[6]The original Atkinson-Stiglitz theorem holds under a more permissive "weak separability" between commodities and earnings. We make a stronger assumption on preferences to focus solely on extending the analysis into a multidimensional setting.

The planner does not observe individual agents' types. Thus, the mechanism must satisfy the standard incentive compatibility (IC) constraints

$$U(Y(\theta), \theta) \geq U(Y(\theta'), \theta), \quad \forall \theta, \theta' \in \Theta. \tag{4}$$

The constraints (4) prevent agents from misreporting their taste, ability, or both simultaneously.

For an agent with consumption levels c (of numeraire) and x (of other goods), let $e = c + k \cdot x$ be the consumption expenditure, evaluated at the marginal costs.

A set of *efficient* (or *undistorted*) choices of goods x for an individual with taste type t and disposable income e is defined as[7]

$$X^*(t, e) := \arg\max_{x \in \mathbb{R}_+^L} u(e - k \cdot x) + v(x, t). \tag{5}$$

Intuitively, efficiency requires that, conditional on a given level of total expenditure (equal to disposable income e), the agent consumes an amount of x that she would have chosen if all goods were priced at their marginal costs. If $x \notin X^*(t, e)$, then x is called *inefficient* (or *distorted*). When $u(c) = c$, the set of efficient choices of x is independent of expenditure e, and we denote it by $X^*(t)$. We assume that $X^*(t, e)$ is not empty for any (t, e).

3 Baseline: When Interventions in Markets are Redundant

In this section, we establish a baseline result that organizes the rest of the analysis. We introduce three assumptions that jointly imply that the planner can achieve an optimal allocation with an income tax alone, without intervening in the goods markets.

Assumption A1. *There are no income effects:* $u(c) = c$, $\forall c \in \mathbb{R}$.

Assumption A2. *Welfare weights depend only on ability:* $\lambda(t, a) \equiv \bar{\lambda}(a)$.

Assumption A3. *Ability and tastes are statistically independent:* $F(t, a) \equiv F_t(t) F_a(a)$*, where* F_t *and* F_a *denote the marginal distributions.*

For the following result, we also assume existence of an optimal mechanism (which could be established under standard assumptions).

[7] Throughout the paper, our use of the term *efficiency* is always consistent with the usual definition of *Pareto efficiency*: There is no other feasible allocation that makes all agents weakly better off and a strictly positive mass of agents strictly better off.

Theorem 1. *Under Assumptions A1–A3, there is an optimal mechanism that induces an efficient choice of goods x, and can be decentralized with a competitive goods market and a (possibly nonlinear) income tax.*

3.1 Intuition

Theorem 1 can be understood as a generalization of the Atkinson-Stiglitz result to settings with multidimensional heterogeneity. The intuition is straightforward from the three assumptions: Assumptions A1 and A2 together imply that the planner wants to redistribute between agents with different abilities, but not between agents with identical abilities and different tastes. By Assumptions A1 and A3, meanwhile, agents' willingness to pay for goods is uninformative regarding (i.e., statistically independent of) ability. Thus, distorting goods choices does not help to distinguish ability types—and hence, has no value for the redistributive objective; hence, there is no reason to distort the goods market.

Note the Assumption A1 (no income effects) plays two separate roles: it is critical *both* for the lack of direct redistributive concerns along the taste dimension *and* for the independence of ability and willingness to pay. If income effects were present, then an agent's marginal social welfare weight (which is proportional to $\lambda(\theta)u'(c(\theta))$) would be endogenous to their spending on x (and, thus, to their taste for x) through its impact on the consumption of the numeraire c. Furthermore, the agent's willingness to pay would depend on their income, and hence be correlated with ability.

While Theorem 1 shows that we can recover the Atkinson-Stiglitz result in our multidimensional setting, the intuition just described already suggests what we prove in the sequel: all three of the assumptions are essentially necessary for the Atkinson-Stiglitz conclusion to hold. We thus interpret Theorem 1 as in effect revealing three channels through which intervention in goods markets may become optimal even when nonlinear income tax instruments are available.

3.2 Sketch of argument

Before turning to relaxations of the three assumptions, we briefly discuss the proof strategy we use for Theorem 1, and how it relates to the original Atkinson-Stiglitz theorem.

In the first step of the argument, we consider a relaxed problem in which the planner is able to directly *observe* the taste type t. Then, the problem can be solved for each t separately, ignoring agents' incentives to misreport their tastes. For each t, the relaxed problem becomes a one-dimensional optimal taxation problem in which agents have identical tastes over goods x. Thus, the original Atkinson-Stiglitz theorem implies that goods x should be undistorted, and that redistribution should be conducted via income taxes alone.

In the second step, we show that under our three assumptions, the income tax schedule that solves the relaxed problem in fact does *not* depend on the taste type t, and hence is feasible (and therefore optimal) in the original problem in which the planner does not observe tastes. This step highlights the key role played by Assumptions A1–A3. The social welfare function has the same shape for every t: The planner neither conditions welfare weights on t directly (Assumption A2), nor learns about the distribution of ability a by observing t (Assumption A3). Moreover, by Assumption A1 (no income effect) taste t does not interact with the agent's preferences over disposable income: Total expenditure on goods is the same for all t; all that changes with t is how disposable income is split between the goods x and the numeraire c. Consequently, the income tax that solves the relaxed subproblem does not condition on t even when t is freely observable. While individual consumption of c and x does depend on t in the relaxed problem, because the allocation is efficient, it can be implemented in an incentive compatible way by pricing goods at marginal costs and letting agents make unrestricted consumption choices.

It is instructive to compare our result to the canonical multidimensional screening framework used to study, e.g., nonlinear pricing by a multiproduct monopolist. Rochet and Choné (1998) showed that in such problems it is typically optimal to distort *all* dimensions of the allocation by bunching, i.e., assigning identical bundles of goods to different types. Perfect separation of types is suboptimal due to the tension between the individual rationality (or participation) constraints and the second-order incentive constraints. In contrast, we concluded that the allocation of goods x should never be distorted and, thus, bunching of different taste types can be easily ruled out.[8] This stark difference in conclusions is due to the absence of individual-rationality constraints in our framework. Indeed, Theorem 1 would fail if such constraints were included and were binding. In particular, if the planner in our framework wanted to maximize revenue, an individual-rationality constraint would be needed to make the problem well-defined, and the optimal mechanism would likely distort all decisions.[9]

3.3 Extensions

Theorem 1 allows for several noteworthy extensions. First, note that the assumption that earnings $z(\theta)$ are a scalar (say, total earnings) rather than a vector (say, earnings from differ-

[8]It is sufficient to suppose that $x^*(t)$ is injective, i.e., distinct taste types have distinct efficient choices.

[9]A Rawlsian planner would choose a mechanism similar to the revenue-maximizing planner (except for the choice of the lump-sum payment) but Rawlsian preferences additionally violate Assumption A2 because the worst-off agent has the lowest ability type *and* the lowest taste type.

ent activities) is immaterial, as the proof works in either case.[10] Moreover, we do not need to identify all entries of z as "earnings," per se: some of them could stand for consumption of commodities for which the agent's taste depends on (or is correlated with) their ability. We obtain the following result:

Corollary 1. *Suppose that utility of type (t, a) can be written as*

$$u(c) + v(x_t, t) + w(z, x_a, a), \tag{6}$$

where $x_t \in \mathbb{R}^{L_t}$ and $x_a \in \mathbb{R}^{L_a}$ are two vectors of goods, produced with marginal costs $k_t \in \mathbb{R}^{L_t}_{++}$ and $k_a \in \mathbb{R}^{L_a}_{++}$. Under Assumptions A1–A3, the optimal mechanism induces an efficient choice of goods x_t, and can be decentralized with a competitive market for goods x_t and a (possibly nonlinear) tax that depends on earnings z and consumption of goods x_a.

Corollary 1 shows that the logic behind Theorem 1 can be applied to subgroups of goods markets. If the taste for a subset of goods is uncorrelated with welfare weights and abilities, then there are no redistributive gains from intervening in markets for those goods.

For another extension, note that in our framework there is no fundamental reason why z (which, again, could be a vector) is labeled as "earnings" while x is labeled as "goods"; for the following result, we apply the proof of Theorem 1 while flipping these labels to examine what happens when welfare weights depend only on tastes and not on ability.

Corollary 2. *Suppose that welfare weights depend only on tastes: $\lambda(t, a) \equiv \tilde{\lambda}(t)$. Under Assumptions A1 and A3, the optimal mechanism induces an efficient choice of earnings: $z(t, a) \in \arg\max_{z'}\{z' - w(z', a)\}, \forall (t, a) \in \Theta$, and can be decentralized with a (possibly nonlinear) tax on goods x that does not depend on earnings.*

When the planner wants to redistribute across the taste dimension but not between agents with different abilities, tastes and abilities are uncorrelated, and there are no income effects, it is *the income tax* that becomes a redundant instrument. In that case, redistribution should be conducted *only* by intervening in the goods markets. In this way, we see precise assumptions under which focusing solely on redistributive market mechanisms—as in the setting of Dworczak ⓡ al. (2021)—is justified.

4 Main Analysis: Interaction of Income Taxation and Market Design

We now introduce a simplification of our full model under which we can fully characterize the optimal mechanism in cases when the assumptions of Theorem 1 fail. We assume that

[10] Solving a relaxed subproblem would be more difficult in the latter case, yet we actually do not need to explicitly find a solution in order to show that the solution is independent of t_0.

an agent with type $\theta = (t, a)$ has a utility function

$$u(c) + tx - \frac{z}{a},$$

with $c \in \mathbb{R}$, $x \in [0,1]$, and $z \in [0, \bar{z}]$. Unless stated otherwise, we assume that

$$u(c) = \begin{cases} c & c \geq \underline{c}, \\ -\infty & c < \underline{c}. \end{cases}$$

The interpretation is that c represents numeraire consumption whose level cannot fall below a "subsistence threshold" denoted $\underline{c}$. The introduction of a subsistence level for numeraire consumption provides a simple way of breaking the linear-utility assumption that we used in Theorem 1.[11] While this way of modeling the income effect is stylized, it allows us to derive tight predictions about the optimal mechanism; as we discuss later, the qualitative conclusions continue to hold whenever $u(c)$ is a smooth, strictly increasing, and concave function.

The variable x represents the level of consumption of a good or commodity; x can thus represent quantity, quality, or probability of allocation (in case the good is indivisible). For concreteness, we will refer to x as "quality." Note that we normalize the maximal quality of x to be 1 (which is convenient if we wanted to interpret x as probability). All individuals have the same preferences over goods that comprise the numeraire c, but they differ (through the taste parameter t) in their marginal rates of substitution between c and x. Our specification also assumes a linear disutility function for generating earnings z, and imposes a finite bound $\bar{z}$ on earnings.

For tractability, we assume that the ability type is binary: $a \in \{l, h\}$. We call agents with $a = h$ the high-ability types, and agents with $a = l$ the low-ability types. We let μ_a denote the mass of agents with ability type a. We assume that it is efficient for high-ability workers to work, $h > 1$, and that low-ability workers are effectively unable to work, $l = 0$ (our arguments apply as long as low-ability agents are sufficiently unproductive).

Taste types t are distributed according to a cumulative distribution function (cdf) F_a with strictly positive, absolutely continuous densities f_a, for $a \in \{l, h\}$, supported on the same interval $[0, \bar{t}]$. The assumption that the lowest type t is 0 is convenient because it implies

[11] The utility function can be seen as a limiting form (as $\beta \to \infty$) of a concave utility function for consumption:

$$u_\beta(c) = \begin{cases} c & c \geq \underline{c}, \\ (1-\beta)\underline{c} + \beta c & c < \underline{c}. \end{cases}$$

that—without loss of generality—the lowest taste type will only consume the numeraire, which allows us to interpret $c_a(0)$ as the lump-sum transfer to group a.

Letting $\lambda_a(t) \equiv \lambda(t,a)$, the objective function of the planner is to maximize

$$\sum_{a\in\{l,h\}} \mu_a \int_0^{\bar{t}} \lambda_a(t) \left(c_a(t) + t x_a(t) - \frac{z_a(t)}{a} \right) dF_a(t), \tag{7}$$

over $c_a(t) \geq \underline{c}$, $z_a(t) \in [0, \bar{z}]$, and $x_a(t) \in [0, 1]$, subject to the incentive-compatibility constraint,

$$c_a(t) + t x_a(t) - \frac{z_a(t)}{a} \geq c_{a'}(t) + t x_{a'}(t') - \frac{z_{a'}(t')}{a}, \forall t, t' \in [0, \bar{t}], \forall a, a' \in \{l, h\}. \tag{8}$$

We continue to assume that the average welfare weight $\bar{\lambda}_l$ on low-ability agents is weakly larger than the average welfare weight $\bar{\lambda}_h$ on high-ability agents. The resource constraint is

$$\sum_{a\in\{l,h\}} \mu_a \int_0^{\bar{t}} (z_a(t) - c_a(t) - k x_a(t))\, dF_a(t) \geq G. \tag{9}$$

To avoid trivial cases, we assume that $k \in (0, \bar{t})$, and that G is low enough that there exist feasible allocations at which all agents' utilities are finite. Under this assumption, it is without loss of generality to restrict attention to allocations such that $c_a(\bar{t}) \geq \underline{c}$—we refer to this constraint as the "subsistence constraint." (Note that in an incentive-compatible mechanism, if $c_a(\bar{t}) \geq \underline{c}$, then $c_a(t) \geq \underline{c}$ for all t).

4.1 Pareto efficiency

The following lemma pins down the key condition for an efficient allocation of goods in our specification with subsistence constraints.

Lemma 1. *The allocation $(z_a(t), c_a(t), x_a(t))$ is efficient if and only if the resource constraint (9) holds with equality and, for almost all t,*

$$z_h(t) = \bar{z} \text{ and } z_l(t) = 0,$$

$$x_a(t) = \begin{cases} 1 & t \geq k,\ c_a(t) > \underline{c} \\ \in [0, 1] & t \geq k,\ c_a(t) = \underline{c} \\ 0 & \text{otherwise} \end{cases}. \tag{10}$$

The properties of an efficient allocation in our setting are straightforward given the linear-utility model. For agents above the subsistence level, the taste type t is equal to their willingness to pay (WTP) for the good, and efficiency requires that these agents consume the good if and only if their WTP is above marginal cost. However, for agents whose numeraire consumption is at the subsistence level $\underline{c}$, WTP is not uniquely well-defined. Intuitively, agents at the subsistence constraint have a rate of substitution t for buying slightly less of the good, and a rate of substitution 0 for buying slighlty more of the good (since this would shift their numeraire consumption below subsistence). Consequently, for an agent at the subsistence constraint who has taste type t, any level of consumption of the good consistent with WTP being between 0 and t is Pareto efficient.[12]

In a direct mechanism, there is no notion of a "price" of the good; the price concept is nevertheless useful when thinking about distortions in the allocation.

Definition 1. *For an incentive-compatible mechanism $(z_a(t), x_a(t), c_a(t))$, we define the per-unit price for a good with (strictly positive) quality $q \in Im(x_a)$ faced by ability type a as*

$$p_a(q) := \frac{c_a(0) - c_a(x_a^{-1}(q))}{q}. \tag{11}$$

Intuitively, the numerator of (11) is equal to the total payment that any type t consuming quality q (i.e., $q = x_a(t)$) must be making, compared to a type 0 who does not consume the good at all. Dividing by q turns the total payment into the per-unit price. The following simple lemma confirms that our definition of a per-unit price coincides with its intuitive meaning.

Lemma 2. *If an incentive-compatible mechanism $(z_a(t), x_a(t), c_a(t))$ is efficient, then for any $a \in \{l, h\}$ and any strictly positive $q \in Im(x_a)$,*

$$p_a(q) = k.$$

Unsurprisingly, Pareto efficiency in our setting requires that all agents face a per-unit price for the good equal to its marginal cost. This result is convenient because we will often be able to characterize distortions in the goods market in terms of prices diverging from marginal cost.

[12]For further intuition, we may imagine approximating the utility function for numeraire consumption with a smooth concave function $u(c)$, with $u'(c) = 1$ for all $c \geq \underline{c}$; under this approximation, WTP would be uniquely defined by $t/u'(c)$ and efficiency would require that $x_a(t) = 1$ if and only if $t/u'(c) \geq k$. In the limit, $t/u'(c)$ would be discontinuous at $c = \underline{c}$, with a right limit of t and a left limit of 0.

4.2 Preliminary results

We first show a simple lemma stating that redistribution via income tax does not conflict with efficiency.

Lemma 3. *In any optimal mechanism, labor supply is efficient:* $z_h(t) = \bar{z}$, $z_l(t) = 0$, $\forall t \in [0, \bar{t}]$.

Lemma 3 confirms our earlier assertion that we study the best-case scenario for income taxation: labor markets remain efficient even under strong redistributive motives.

To provide intuition for further analysis, it is instructive to consider the benchmark case in which ability is observed, and the subsistence constraint is not binding—then, it is optimal to offer quality 1 to agents with ability a at a price $p_a^\star$ that (assuming an interior solution) must satisfy

$$p_a^\star = k + (\bar{\lambda}_a - \Lambda_a(p_a))\gamma_a(p^\star), a \in \{l, h\}, \tag{12}$$

where $\Lambda_a(p)$ is the average welfare weight on agents with ability a and taste type above p, and $\gamma_a(p) = \frac{1-F_a(p)}{f_a(p)}$ is the inverse hazard rate.[13] Intuitively, agents with taste type $t \geq p_a^\star$ are buying the good; thus, if the designer attaches a higher-than-average welfare weight to buyers of the good, then it is optimal to subsidize the price of the good below marginal cost. In the opposite case, it is optimal to tax purchases of the good. Only when welfare weights do not depend on taste—that is, under Assumption A2—is the efficient allocation optimal.

The logic just described makes it intuitive that relaxing Assumption A2 would result in optimal mechanisms incorporating distortions in the goods market. However, we show that in fact—with unobserved ability and income effects—relaxing of *any* of the Assumptions A1-A3 leads to a violation of the conclusion of Theorem 1 (i.e., violations of the Atkinson-Stiglitz conclusion). We focus on economic intuitions when presenting our results in this section—we explain the underlying proof technique and its broader relation to multidimensional screening in Section 5.

4.3 Optimality of market distortions under income effects

We first consider the case in which Assumptions A2 and A3 of Theorem 1 hold but the utility for numeraire consumption is no longer linear. We investigate the properties of optimal mechanisms when the subsistence constraint binds for the low-ability agents. In the

[13] This follows, for example, from the analysis of Akbarpour ⓡ al. (2024) by replacing their fixed-supply assumption with a constant marginal cost.

following subsection, we show that the main results are robust to introducing strictly concave utility from numeraire. Whenever Assumption A3 holds, we write just "F" (without a subscript) to denote the (common across the two ability types) cdf of the distribution of taste types.

Theorem 2. *Suppose that Assumptions A2 and A3 hold. Furthermore, assume that* $\frac{1-F(t)}{f(t)}$ *is non-increasing, that* $\frac{F(t)}{f(t)}$ *is non-decreasing, and that*

$$\mu_h \bar{z}\left(1-\frac{1}{h}\right)-G<k+\underline{c} \text{ and } \frac{\bar{z}}{h} \geq \bar{t}. \tag{13}$$

There exists an optimal mechanism in which:

1. *Every agent chooses between purchasing*
   - *quality* q_l *of the good at a per-unit price* $p_l < k$ *or*
   - *(maximal) quality* 1 *of the good at a price* $p_l q_l + p_h(1-q_l)$*, where* $p_h > k$*;*

   if $\lambda_h = 0$*, then* p_h *coincides with the revenue-maximizing price.*

2. *Every agent chooses to work or not, with each unit of earnings taxed at the rate* $1-1/h$*.*

3. *Every agent receives a lump-sum transfer equal to* $\underline{c}+p_l q_l$*.*

Under the optimal mechanism, all agents face the same price schedule in the goods market. However, the allocation of goods is not efficient: The mechanism offers a low-quality good at an average price strictly below marginal cost. Low-ability agents of sufficiently high taste type consume the subsidized option. The mechanism allows high-ability agents—who work and therefore have higher disposable income—to "top up" their consumption of the good; the price charged for the additional units is strictly above marginal cost, and thus can be understood as extracting revenue from wealthier agents (which then subsidizes the lump-sum payment). In fact, when the welfare weight on high-ability agents is 0, the top-up price p_h is equal to the revenue-maximizing (monopoly) price.

Because the lump-sum payment is equal to subsistence consumption plus $p_l q_l$—which is exactly the price of the subsidized good—we can also interpret the mechanism as giving the agents the choice between receiving the good with quality q_l in-kind, or opting for a higher cash transfer. Note that high-ability agents are also allowed to use the subsidy just like low-ability agents are. This is because income taxes are set at a maximal level that extracts all surplus from working (i.e., the net wage is just enough to cover the cost of labor supply). If high-ability agents were excluded from the subsidy, given the tax regime, some of them would choose not to work (which is not optimal by Lemma 3). At the same time, though, the subsidy is phased out by charging agents a higher marginal price for topping up. This combination of prices achieves the screening effect that underlies the main results of Dworczak ⓡ al. (2021): high-ability agents with strong enough preferences for the good choose to top up; because low-ability agents never do so, market selection in effect identifies the higher-ability agents with particularly strong preferences for the good, and extracts surplus from them to redistribute via the lump-sum transfer.

The following intuition helps explain why it is optimal to lower the price of the low-quality good below marginal cost. Suppose instead that the good is offered at marginal cost k. This means that low-ability agents with $t \geq k$ spend all their disposable income (equal to the lump-sum payment) on the good with quality $q < 1$ which puts their consumption c at subsistence level. Consider now a slight perturbation of the price to $k - \epsilon$. This perturbation has a negative effect due to allocative inefficiency (some agents with $t < k$ consume the good); however, this effect is of second-order in ϵ (it is an order-ϵ distortion for an order-ϵ mass of agents). The perturbation also has a positive effect, which is that all low-ability agents with types $t \geq k$ now consume $q\epsilon$ more units of the good; this is a first-order effect since even inframarginal taste types enjoy an increase in utility at the order of ϵ. Thus, for small ϵ, the positive effect dominates the negative effect, and it is optimal to lower the price below marginal cost.

For additional illustration, consider an example in which the good is treatment for a medical condition. The taste type captures whether (and to what extent) treatment is needed. When the cost of the treatment is high enough, agents with low income must substantially decrease their consumption of other goods to afford it—pushing them below the subsistence constraint in the language of our model. Thus, even if the designer has no inherent preference for redistribution to agents who are sick (i.e., welfare weights do not depend on the taste type) and the likelihood of getting sick is unrelated to ability (taste type and ability types are independent), she still wants to redistribute towards agents who consume treatment because of the associated income effects; the planner achieves this by subsidizing the price of treatment below its marginal cost.[14]

[14]The same story could be told by replacing treatment for a disease with consumption of a luxury good (perhaps a yacht), in which case the appeal of the intuition is lost. The easiest way to distinguish between these two cases (essential healthcare versus yachts) is to think of the planner maximizing a strictly concave transformation of agents' utilities. In such a framework, the *level* of utility matters for the marginal social welfare weight. An agent consuming yachts could have high marginal utility from them but since their overall utility is presumably already very high, their contribution to social welfare would be small, turning off the effect described here. In contrast, an agent in need of medical treatment who chooses low-quality healthcare is likely to have low consumption of other goods as well, and hence both high marginal utility and

Interpreted through the lens of the income effect, *the mechanism subsidizes the consumption of the low-quality good because consuming the good is a signal of high marginal value of consumption*. Low-ability agents who consume the good must be at subsistence level, which means that the effective social welfare weight on their consumption is high. It is thus optimal to give more resources to them—and if high-ability agents have enough demand to top up their consumption, taxing that consumption helps finance the subsidy.

Finally, we comment on the regularity assumptions in Theorem 2: First, the monotonicity of hazard rates is a mathematically restrictive assumption but its only role in the proof is to rule out ironing and ensure that first-order conditions are sufficient.[15] With ironing, as we explain in Section 5, the optimal mechanism may need to offer additional options to high-ability agents, which complicates the exposition without adding new economic insight. The condition $\mu_h \bar{z}\left(1 - \frac{1}{h}\right) - G < k + \underline{c}$, by contrast, is economically important: it states that the economy does not have enough resources to put all agents strictly above the subsistence level without raising at least some revenue (the left side of (13) is production minus government expenditure, and the right side is the minimal lump-sum transfer that makes low-ability agents capable of consuming at subsistence level while still affording the good priced at marginal cost). If (13) fails, then the subsistence constraint is moot, and optimality of efficient provision of the good (pricing at marginal cost) follows from our Atkinson-Stiglitz theorem (Theorem 1) because no agent's consumption induces income effects in equilibrium (effectively, restoring Assumption A1). Finally, the assumption $\frac{\bar{z}}{h} \geq \bar{t}$ ensures that the subsistence constraint is slack for high-ability agents—it says that high-ability agents are sufficiently productive relative to the strongest possible taste for buying the good. This assumption is not crucial; if the subsistence constraint is binding also for some high types, the solution is qualitatively very similar, with the main difference being that high-ability agents choose between not buying, consuming q_l, or consuming some $q_h < 1$.[16]

4.3.1 Robustness to curvature in utility functions

A potential concern with Theorem 2 is that its conclusion could be driven by the stylized form of the income effect, modeled for tractability as a subsistence constraint. Suppose instead that preferences take the general form from equation (1) with the functions u, v,

low overall utility, resulting in high redistribution-worthiness. In our current framework, we can replicate that logic by making welfare weights a function of tastes (i.e., by relaxing Assumption A2).

[15]The condition is stronger than usual because the analog of virtual surplus in our analysis is endogenous to the Lagrange multiplier on the subsistence constraint—see the proof in Appendix A for details.

[16]We omit this solution for brevity, but note that it could be easily reproduced following a similar proof strategy; see Proposition 2 that covers the structure of the optimal mechanism in the general case.

and $-w$ assumed smooth and strictly concave. While a full characterization of the optimal mechanism is no longer attainable in that case, we find that the main features of the optimal mechanism from Theorem 2 are maintained. Here, we summarize and discuss the findings, while the formal results are relegated to Appendix B.

The optimal mechanism under strictly concave utility functions has the following features. First, consumption of good x is distorted for at least some types—except for degenerate cases, the conclusion of the Atkinson-Stiglitz theorem fails. Second, if the consumption of the low-ability agents is distorted, then it must be distorted upwards, meaning that it is optimal to subsidize good x for them.[17] Third, if the planner does not value the utility of high-ability types, then it is optimal to charge a revenue-maximizing price for any additional unit of x on top of what the low-ability agents consume. Thus, optimal mechanism under curvature in the utility function preserves the main features of the optimum under the subsistence constraints.

To build intuition, notice that agents with high taste type choose relatively higher x and lower c, resulting in higher marginal utility for numeraire $u'(c)$. The income effect thus creates a positive correlation between purchases of x and the marginal utility $u'(c)$. The optimal mechanism exploits this correlation by subsidizing purchases of good x to support individuals with high marginal utility. Note that preference heterogeneity is crucial for this effect to be present. Agents with different tastes endogenously differ in their marginal value for the numeraire, which creates a motive for redistribution that cannot be addressed through the income tax alone.

The proof strategy closely follows the intuition. If both the low- and high-ability agents of the same (interior) taste consume an undistorted amount of x, then the planner can always improve the social objective by distorting their allocations of x upwards. Such change is beneficial because it allows to redistribute numeraire from lower taste (and lower $u'(c)$) agents to higher taste (and higher $u'(c)$) agents. When the welfare weight on the high-ability types is zero and they consume more x than the low-ability types, the planner can further improve by distorting the allocation of the high-ability types downwards, which allows to extract the maximal revenue. In that case only the downward incentive constraint in taste binds for high-ability agents and, thus, the solution coincides with the solution to the one-dimensional monopolistic screening problem.

[17] To prove this property, we assume that ironing is not required in the optimal mechanism.

4.4 Optimality of distortions with no income effect

Next, we assume that utilities are linear in numeraire consumption (taking $\underline{c} = -\infty$ in our specification), restoring Assumption A1, and examine what happens when we relax Assumptions A2 and/or A3 of Theorem 2. Because both relaxations have a similar effect, we exposit them together.

Even though we assumed that the designer cares more about the low-ability agents than high-ability agents on average, we cannot conclude that the price for the high-ability agents is weakly higher than for low-ability agents. For example, it may be optimal to post a lower price for high-ability agents if their distribution of taste types is lower (in an appropriate sense) than that for the low-ability agents (because low-ability agents are not able to mimic high-ability agents, that type of solution is feasible). We consider this case to be less economically interesting,[18] so we rule it out by assuming

$$\forall t \in [0, \bar{t}], (1 - \Lambda_h(t))\gamma_h(t) \geq (1 - \Lambda_l(t))\gamma_l(t), \tag{14}$$

where again $\Lambda_a(t)$ is the average welfare weight on agents with ability a and taste type above t, and $\gamma_a(t) = \frac{1-F_a(t)}{f_a(t}$ is the inverse hazard rate.

To interpret condition (14), note that if the welfare weights on low- and high-ability agents are the same, then the assumption states that the distribution of high-ability agents' tastes is higher—in the hazard-rate order—than that of low-ability agents. Analogously, if the taste distribution in the two different groups is the same, then the assumption states that the welfare weights on low-ability agents are weakly higher than on high-ability agents, conditional on the taste type exceeding any threshold.

Theorem 3. *Suppose that Assumption A1 and inequality (14) hold, and that*

$$(t - k - (1 - \Lambda_h(t))\gamma_h(t))\, f_h(t) \tag{15}$$

is non-decreasing whenever it is negative. Then, there exists an optimal mechanism that takes one of two forms:

1. *Labor income is taxed at a rate $1 - 1/h$ and all agents face the same price p for quality 1 of the good satisfying*

$$p = k + (1 - \Lambda(p))\gamma(p),$$

[18] Note that this class of solutions features inefficient allocation of the good, by definition, since the two groups face a different price for the good, and at most one of those two prices can be equal to marginal cost.

where $\Lambda(p)$ is the average welfare weight on all agents with taste type t above p, and γ is the inverse hazard rate of the unconditional distribution of taste; or

2. *Labor income is taxed at a rate strictly lower than $1 - 1/h$, all agents choose whether to purchase quality 1 of the good, agents who work face a price p_h, and agents who do not work face a price p_l, where*

$$k + (1 - \Lambda_h(p_h))\gamma_h(p_h) \geq p_h > p_l \geq k + (1 - \Lambda_l(p_l))\gamma_l(p_l).$$

Without income effects, the optimal mechanism in our framework becomes simple. In particular, (after ironing is ruled out by the regularity conditions), it is always optimal to sell the good to each group at a single price. If income taxes are set to the maximal level, then the prices are in fact the same for both groups; when income taxes leave high-ability agents with a strictly positive surplus from working, the price for the good faced by high-ability agents is higher than the price faced by low-ability agents.

The assumption that (15) is non-decreasing whenever it is negative is precisely what rules out ironing. Like with our anti-ironing condition from Theorem 2, this regularity condition is mathematically restrictive, but not essential for our results. When it is relaxed, we cannot rule out the optimality of offering an additional option of a low-quality good to high-ability agents at a low per-unit price in case 1. Intuitively, in case 1, in order to maintain incentive-compatibility, the designer must ensure that high-ability agents face weakly better terms of trade in the goods market than low-ability agents. At the same time, our assumption (14) states that it is not optimal to offer a strictly lower price for the good to high-ability agents. If it is optimal to set a single price for high-ability agents, then it follows that the price offered to low- and high-ability agents is the same. However, it may sometimes be optimal to offer two prices to high-ability agents: a low-quality good with a per-unit price lower than p_l and a high-quality good with a per-unit price higher than p_l. Even in this case, though, allocation in the goods market is inefficient.[19]

In case 1, efficiency in the goods market is generically suboptimal when either Assumption A2 or A3 fails. Indeed, efficiency in the goods market is only optimal if the average welfare weight on agents buying the good, $\Lambda(p)$, is equal to the unconditional average of 1—and when welfare weights depend directly on the taste type (i.e., when Assumption A2 is relaxed), $\Lambda_a(t)$ (and hence its average over a) will typically diverge from 1. Moreover, even if welfare weights do not depend on the taste type directly, $\Lambda(p)$ might deviate from 1 if taste types are correlated with ability types (as occurs when we relax Assumption A3).

[19] Analytical expressions for these prices are not generally available, which is why we focus on the case in which ironing can be ruled out.

For intuition, suppose that $\bar{\lambda}_h = 0$, so that the designer only cares about low-ability agents. It is then straightforward to show that

$$\Lambda(p) < 1 \quad \iff \quad \frac{1 - F_l(p)}{\mu_l(1 - F_l(p)) + \mu_h(1 - F_h(p))} < 1 \quad \iff \quad F_l(p) > F_h(p).$$

That is, the good is taxed if low-ability agents have a lower distribution of the taste type (in first-order stochastic dominance sense), and is subsidized if low-ability agents have a higher distribution of the taste type. In both cases, the designer uses the market for the good to transfer more resources from high- to low-ability agents, relying on the statistical dependence between taste and ability. Note the role played by taste heterogeneity: Even though the designer cannot take away more resources from all high-ability agents directly (their income is already taxed as much as is possible while still satisfying the incentive constraints), she can transfer more resources away from high-ability agents with high taste by taxing the good (when high-ability agents have higher taste on average) or from high-ability agents with low taste by lowering the lump-sum transfer and subsidizing the good instead (in the case that low-ability agents have higher taste on average).

In case 2, the allocation of the good is never efficient because high-ability agents face a strictly higher price than low-ability agents (hence, it cannot be that both prices are equal to marginal cost). The mechanism can be implemented as a means-tested subsidy for the good (in the sense that the subsidy is only available to agents who have no labor income). For this mechanism to be incentive-compatible, labor income cannot be taxed maximally, which is why the tax rate is strictly below $1 - 1/h$. For intuition, recall that if the ability type were observed, the designer would like to implement ability-dependent prices given by (12). However, to implement $p_h > p_l$ when ability is not observed, the designer must provide enough rents to high-ability types to maintain incentive compatibility—and the larger the gap $p_h - p_l$, the larger the rent that high-ability agents must receive in the labor market. This creates a trade-off and implies that the optimal prices p_h and p_l will be closer together than the benchmark prices $p_h^\star$ and $p_l^\star$ given by (12).

It remains to discuss the determinants of which of the two candidate optimal mechanisms is used. The result below shows that both mechanisms are sometimes optimal, and the choice between them depends on whether the designer has a strong motive to redistribute across ability types.

Proposition 1. *Under the assumptions of Theorem 3, suppose additionally that*

$$(t - k - (1 - \Lambda_l(t))\gamma_l(t))\, f_l(t)$$

is non-decreasing whenever it is positive. Parametrize $\lambda_h(t) \equiv \bar{\lambda}_h \bar{\lambda}(t)$ for some $\bar{\lambda}(t)$ with mean 1. Then, fixing all parameters other than $\bar{\lambda}_h$, there exists a cutoff value $\lambda_h^0 \in [0, 1]$ such that mechanism 1 from Theorem 3 is optimal if $\bar{\lambda}_h < \lambda_h^0$ and mechanism 2 is optimal if $\bar{\lambda}_h > \lambda_h^0$.

Moreover, for any $\bar{\lambda}_h < 1$, mechanism 1 is optimal when μ_l is sufficiently small; and for $\bar{\lambda}_h = 1$, mechanism 2 is optimal whenever inequality (14) is strict for interior t.

Intuitively, when welfare weights depend on both ability and taste—directly, or indirectly through statistical correlation—the designer wants to condition redistribution on both dimensions. The gap between $\bar{\lambda}_l$ and $\bar{\lambda}_h$ can be seen as a measure of the strength of vertical redistributive preferences. When these preferences for redistribution are very strong, the designer uses mechanism 1 from Theorem 3, which achieves the maximal level of redistribution across ability types. The goods market may still be distorted but the redistribution through the goods market merely "complements" the redistribution through the income tax. On the other hand, when vertical redistributive preferences are not that strong, it becomes optimal to let high-ability workers retain some surplus from working in order to redistribute more effectively through the goods market.

5 General Structure of the Optimal Mechanism

In this section, we explore the general structure of optimal mechanisms in the simplified setting introduced in Section 4. In particular, we uncover the reasons for the relatively simple mechanism form identified in Theorems 2 and 3, and sketch the proofs of these results. The main goal here is to explain several key technical ideas behind our construction that could be useful in studying other similar multidimensional screening problems.

We center our discussion in this section around the following result that predicts the form of the optimal mechanism in the absence of any regularity conditions:

Proposition 2. *In the framework of Section 4, there always exists an optimal mechanism in which:*

1. *Low-ability agents consume one quality of the good: $x_l(t) \in \{0, q_l\}$, where $q_l \in (0, 1]$;*
2. *High-ability agents consume at most three distinct qualities of the good: $x_h(t) \in \{0, q_i, q_l, q_h\}$, where $0 \leq q_i \leq q_l \leq q_h$;*
3. *If the subsistence constraint does not bind, then the highest quality consumed is 1: If $c_l(\bar{t}) > \underline{c}$, then $q_l = q_h = 1$, and if $c_h(\bar{t}) > \underline{c}$, then $q_h = 1$;*

4. *High-ability agents either (i) receive a post-tax wage equal to the cost of labor provision but face a weakly better average price for the good consumed by low-ability agents, or (ii) receive a post-tax wage strictly higher than the cost of labor provision but face a weakly higher average price for the good consumed by low-ability agents: i.e., whenever $q_l \in Im(x_h)$, either*
   (a) $c_h(0) - \frac{\bar{z}}{h} = c_l(0)$ *and* $p_h(q_l) \leq p_l(q_l)$ *, or*
   (b) $c_h(0) - \frac{\bar{z}}{h} > c_l(0)$ *and* $p_h(q_l) > p_l(q_l)$.

In an optimal mechanism, low-ability agents face a simple choice (recall that they do not work, by Lemma 3). They receive a lump-sum transfer, equal to $c_l(0)$, and decide whether to buy the good with quality q_l at a per-unit price of p_l, or spend their entire disposable income on the numeraire (agents with taste type above p_l will decide to do the former). Whether $q_l = 1$ or $q_l < 1$ depends on whether the subsistence constraint binds: $q_l < 1$ can only be optimal if high-taste low-ability agents consume at the subsistence level.

High-ability agents face a potentially more complicated choice. They choose from up to three distinct quality levels. Proposition 2 asserts that there are two possible cases: Either *(i)* Labor income is taxed maximally, so that high-ability agents are indifferent between working or not, in which case they face (weakly) lower goods market prices than low-ability agents face; or *(ii)* labor taxation leaves a strictly positive surplus from working, in which case high-ability workers face strictly higher prices in the goods market than low-ability agents face.

For intuition, recall that the planner cares more (on average) about the welfare of low-ability agents, and that low-ability agents cannot mimic high-ability agents. In case *(i)*, because labor income is taxed maximally, it is not possible to offer better prices in the market to low-ability agents without violating incentive compatibility. Optimal prices often turn out to be independent of income under this fully extractive income tax. In case *(ii)*, high-ability agents receive strictly positive surplus from working, which they would lose if they pretended to be of low ability; thus, in this case, it is feasible (and optimal) to offer strictly better prices to low-ability agents.

The preceding intuition also helps explain why up to three quality levels are needed for high-ability agents. When $q_l < 1$, some intermediate-taste high-ability agents may consume the same quality as low-ability agents when the downward IC constraint (in ability) is binding. Additionally, the higher quality $q_h > q_l$ may be needed because high-ability agents have more disposable income, so their subsistence constraint is relaxed compared to same–taste-type low-ability agents. Finally, the lowest quality level q_i may be needed if the optimal solution requires ironing—roughly, if the planner's objective function is non-monotone in the taste type, it may be preferable to satisfy the downward IC constraint (in ability) by offering a low-quality good at a low price to high-ability agents (this possibility is ruled out by the regularity conditions imposed in Theorems 2 and 3).

5.1 Sketch of argument

The proof of Proposition 2—which is then specialized to prove Theorems 2 and 3—is relatively involved but can be decomposed into several steps, most of which use familiar ideas. First, we reduce the problem to maximizing the welfare function over allocation rules $x_l(t)$ and $x_h(t)$ that are monotone in t, using the envelope formula to express consumption of the numeraire in terms of the allocation rules and ability-specific lump-sum transfers. We use the resource constraint to pin down the lump-sum transfer to the low-ability agents.

Second, we argue that the subsistence constraints can only bind for the highest-taste type within each ability level. This is intuitive, as agents with the same ability share the same disposable income, by Lemma 3. This observation allows us to incorporate all subsistence constraints into the objective function via a pair of Lagrange multipliers, after we parametrize the highest quality level consumed by each ability type. In the final stage of the construction, we optimize over these highest quality levels; intuitively, the highest quality level is 1 (the maximal consumption of good x) if the subsistence constraints for the given ability type turn out to be slack, but it could be interior otherwise.

Third, we argue that the incentive constraint preventing low-ability agents from pretending to have high ability is slack (as a consequence of Lemma 3). The opposite constraint could bind but—conditional on misreporting ability—high-ability agents find it optimal to report their taste truthfully. This key step reduces the incentive constraints in our multidimensional model to two standard one-dimensional constraints (within each ability level) and an outside-option–like constraint for the high-ability agents: High-ability agents must receive a minimal utility level pinned down by the allocation to low-ability agents with the corresponding taste type. Fixing the allocation rule for low-ability agents, our problem thus becomes a standard one-dimensional screening problem with a type-dependent outside option (as in Jullien (2000)).

Fourth, we fix the allocation rule for low-ability agents and solve for the optimal allocation rule for high-ability agents. We rely on an ironing technique that extends the analysis of Myerson (1981) to problems with type-dependent outside options. The extension we use was recently introduced by Dworczak and Muir (2024): the key take-away is that the allocation rule for high-ability agents is *linear* in their outside option, i.e., in the allocation rule for low-ability agents.

Fifth, we maximize the Lagrangian over the allocation rule for low-ability agents, accounting for how that choice affects the optimal allocation rule for high-ability agents. As a consequence of the previous step, this problem is *linear*, with no constraints. By a standard argument, the optimal allocation rule for low-ability agents is therefore a posted price for a single quality level (pinned down by the subsistence constraint).

Sixth, we show that the allocation rule for high-ability agents derives its simple structure from the cutoff allocation rule for low-ability agents. Up to two additional quality (and price) levels may be needed for high-ability agents: a lower quality may be introduced by the ironing procedure, while a higher quality level may be required due to a more permissive subsistence constraint—consistent with the informal discussion of Proposition 2. We also establish the validity of the Lagrangian approach by showing that the subsistence constraints can always be satisfied by choosing appropriate highest quality levels consumed by the two ability types.

Finally, part 4 of Proposition 2 follows from the optimal choice of a lump-sum payment to high-ability agents. Intuitively, the planner faces a trade-off: She can satisfy the endogenous outside-option constraint for high-ability agents (created by the allocation to low-ability agents) by either *(i)* increasing the allocation to high-ability agents, or *(ii)* giving those agents a higher lump-sum payment (implemented in an incentive-compatible way as a reduction in income taxes). The generalized ironing procedure that we deploy to solve for the optimal allocation rule for high-ability agents determines how to optimally use options *(i)* and *(ii)*, leading to the two cases in part 4 of Proposition 2.

5.2 Literature notes

As the proof sketch makes clear, the tractability of our model relies crucially on the simplifying assumption of a binary ability type (with an incentive constraint that can only bind in one direction). A mathematically similar structure arises in the so-called "FedEx problem" in which agents differ in their (continuous) value for receiving a package, as well as a discrete (possibly binary) deadline by which they need to receive it: Relying on duality techniques, Fiat et al. (2016) derived the structure of the revenue-maximizing mechanism in such an environment; Saxena et al. (2018) showed that the number of prices required for full optimality grows exponentially with the number of "deadline" types (which are analogous to ability types in our framework).[20]

Ahlvik et al. (2024) solved a multidimensional screening problem (assuming no bunching in the solution) in a model with a binary (and deterministic) choice in the goods market.

[20]In unreported simulation results, we have also found that our optimal mechanism grows increasingly complex with more ability types, although the economic insights from the binary case carry through.

We instead assume that productivity is binary but allow for continuous types and choices in the goods market. This makes the two papers complementary: we obtain a richer design of the goods market (e.g., with several levels of consumption, rationing etc.), while Ahlvik et al. (2024) obtain richer predictions about the optimal income tax schedule. Additionally, the techniques used to solve the respective problems are different (Ahlvik et al. (2024) rely on first-order conditions using a perturbation approach and optimal-control methods).

A natural benchmark for our analysis is that of Dworczak ⓡ al. (2021), who solved an analogous market-level redistribution problem without integrating income taxation. The optimality of offering a single quality to low-ability agents (and at most three quality levels to high-ability agents) is a consequence of the assumption of constant marginal cost in our framework; if, instead, we had assumed a fixed supply of goods as Dworczak ⓡ al. (2021) did, then an additional quality level might be needed in the optimal mechanism.[21] (Avoiding this additional complexity in the optimal mechanism is why we decided to work with a fixed–marginal cost model, which also happens to be closer to the original work of Atkinson and Stiglitz (1976).) With a fixed supply and linear utilities as in the setting of Dworczak ⓡ al. (2021), rationing (interior quality level) is always inefficient. In our model, rationing may be efficient for agents whose numeraire consumption is at subsistence—and instead, inefficiency is manifested by the price diverging from marginal cost.

6 Concluding Remarks

We investigated the joint problem of income taxation and goods market design in a mechanism design framework, and showed that goods market design plays an important role in balancing equity and efficiency under redistributive preferences. While our paper is far from the first to point out the limitations of the celebrated Atkinson-Stiglitz theorem, we believe that—by characterizing the optimal mechanism in closed form—-it shows the value of goods market interventions in a particularly salient way. Several decades after the original work of Atkinson and Stiglitz (1976), conventional economic wisdom seems to have embraced the idea of using income taxes rather than goods market interventions to redistribute; it would appear that this intuition needs to be revisited, and more research is needed to understand whether it constitutes good policy advice under realistic scenarios.

[21] Mathematically, the Dworczak ⓡ al. (2021) way of modeling scarcity is closely related to ours because a fixed supply constraint results in a Lagrange multiplier that enters the problem in the same way as a constant marginal cost. However, the need to satisfy the supply constraint may result in the optimal mechanism being a convex combination of different maximizers of the Lagrangian (see, e.g., Doval and Skreta (2024)).

References

AHLVIK, L., M. LISKI, AND M. MÄKIMATTILA (2024): "Pigouvian Income Taxation," *CEPR discussion paper*.

AKBARPOUR, M. ⓡ P. DWORCZAK ⓡ S. D. KOMINERS (2024): "Redistributive Allocation Mechanisms," *Journal of Political Economy*, 132, 1831–1875.

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

BOERMA, J., A. TSYVINSKI, AND A. P. ZIMIN (2022): "Bunching and taxing multidimensional skills," Tech. rep., National Bureau of Economic Research.

CHE, Y.-K., I. GALE, AND J. KIM (2013): "Assigning resources to budget-constrained agents," *Review of Economic Studies*, 80, 73–107.

CONDORELLI, D. (2013): "Market and non-market mechanisms for the optimal allocation of scarce resources," *Games and Economic Behavior*, 82, 582–591.

CREMER, H. AND F. GAHVARI (1998): "On optimal taxation of housing," *Journal of Urban Economics*, 43, 315–335.

CREMER, H., P. PESTIEAU, AND J.-C. ROCHET (2001): "Direct versus indirect taxation: the design of the tax structure revisited," *International Economic Review*, 42, 781–800.

——— (2003): "Capital income taxation when inherited wealth is not observable," *Journal of Public Economics*, 87, 2475–2490.

DIAMOND, P. AND J. SPINNEWIJN (2011): "Capital income taxes with heterogeneous discount rates," *American Economic Journal: Economic Policy*, 3, 52–76.

DOLIGALSKI, P., P. DWORCZAK, J. KRYSTA, AND F. TOKARSKI (2024): "Incentive Separability," *Journal of Political Economy Microeconomics (forthcoming)*.

DOVAL, L. AND V. SKRETA (2024): "Constrained Information Design," *Mathematics of Operations Research*, 49, 78–106.

DWORCZAK, P. AND E. MUIR (2024): "A mechanism-design approach to property rights," *Available at SSRN 4637366*.

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

FEREY, A., B. LOCKWOOD, AND D. TAUBINSKY (2021): "Sufficient statistics for nonlinear tax systems with general across-income heterogeneity," Tech. rep., National Bureau of Economic Research.

FIAT, A., K. GOLDNER, A. R. KARLIN, AND E. KOUTSOUPIAS (2016): "The FedEx Problem," in *Proceedings of the 2016 ACM Conference on Economics and Computation*, New York, NY, USA: Association for Computing Machinery, EC '16, 21–22.

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

GERRITSEN, A., B. JACOBS, A. V. RUSU, AND K. SPIRITUS (2020): "Optimal taxation of capital income with heterogeneous rates of return," Working Paper.

GOLOSOV, M. AND I. KRASIKOV (2023): "The optimal taxation of couples," Tech. rep., National Bureau of Economic Research.

GOLOSOV, M., M. TROSHKIN, A. TSYVINSKI, AND M. WEINZIERL (2013): "Preference heterogeneity and optimal capital income taxation," *Journal of Public Economics*, 97, 160–175.

HELLWIG, C. AND N. WERQUIN (2024): "Using Consumption Data to Derive Optimal Income and Capital Tax Rates," Working Paper.

JULLIEN, B. (2000): "Participation constraints in adverse selection models," *Journal of Economic Theory*, 93, 1–47.

KANG, M. AND C. Z. ZHENG (2022): "Optimal Design for Redistributions among Endogenous Buyers and Sellers," *Economic Theory*, 1141–1180.

KANG, Z. Y. (2023): "The Public Option and Optimal Redistribution," Working Paper.

——— (2024): "Optimal Indirect Regulation of Externalities," Working Paper.

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

LI, Y. (2021): "Mechanism design with financially constrained agents and costly verification," *Theoretical Economics*, 16, 1139–1194.

MIRRLEES, J. A. (1976): "Optimal tax theory: A synthesis," *Journal of public Economics*, 6, 327–358.

MOSER, C. AND P. OLEA DE SOUZA E SILVA (2019): "Optimal paternalistic savings policies," *Columbia Business School Research Paper*.

MYERSON, R. B. (1981): "Optimal auction design," *Mathematics of Operations Research*, 6, 58–73.

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

PAI, M. AND P. STRACK (2024): "Taxing Externalities Without Hurting the Poor," *SSRN Working Paper: https://ssrn.com/abstract=4180522*.

ROCHET, J.-C. AND P. CHONÉ (1998): "Ironing, sweeping, and multidimensional screening," *Econometrica*, 783–826.

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

SAXENA, R. R., A. SCHVARTZMAN, AND S. M. WEINBERG (2018): "Maximizing the spread of influence through a social network," in *In Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '18*.

SCHEUER, F. AND J. SLEMROD (2020): "Taxing our wealth," *National Bureau of Economic Research Working Paper Series*.

SCHULZ, K. (2021): "Redistribution of return inequality," Working Paper.

SPENCE, M. (1977): "Nonlinear prices and welfare," *Journal of Public Economics*, 8, 1–18.

SPIRITUS, K., E. LEHMANN, S. RENES, AND F. ZOUTMAN (2022): "Optimal taxation with multiple incomes and types," Working Paper.

WEITZMAN, M. L. (1977): "Is the price system or rationing more effective in getting a commodity to those who need it most?" *Bell Journal of Economics*, 8, 517–524.

A Proofs

A.1 Proof of Theorem 1

Consider an equivalent formulation of the problem in which the planner chooses consumption expenditure $e(\theta) := c(\theta) + k \cdot x(\theta)$ rather than numeraire $c(\theta)$. Let $\alpha \geq 0$ denote the

Lagrange multiplier on the resource constraint. The planner's problem, expressed as a Lagrangian, reads

$$\max_{(e,x,z):\Theta\to\mathbb{R}\times\mathbb{R}_+^{L+1}} \int \lambda(\theta)\, U((e(\theta)-k\cdot x(\theta),x(\theta),z(\theta)),\theta)\, dF(\theta)+\alpha\int [z(\theta)-e(\theta)]\, dF(\theta) \quad (16)$$

subject to incentive constraints

$$U((e(\theta)-k\cdot x(\theta),x(\theta),z(\theta)),\theta)\geq U((e(\theta')-k\cdot x(\theta'),x(\theta'),z(\theta')),\theta), \quad \forall\theta,\theta'\in\Theta. \quad (17)$$

Note that that $\alpha>0$. Otherwise, the objective could be improved while respecting the incentive constraints by increasing $e(\theta)$ such that $u(c(\theta))$ changes by a constant for all θ.

Consider a relaxed subproblem where we focus on the allocation rule for taste type $t_0\in\Theta_t$ in isolation and drop the incentive constraints requiring the agents to be truthful about their taste. We retain the incentive constraints ensuring that agents truthfully reveal their ability. Denote the cdf of ability conditional on taste t_0 as $F_{a|t}(a\mid t_0)$. The planner chooses an allocation rule $(e,x,z):\Theta_a\to\mathbb{R}\times\mathbb{R}_+^{L+1}$ to maximize the Lagrangian of the subproblem:

$$\mathcal{L}(t_0):=\int \lambda(t_0,a)\,[u(e(a)-k\cdot x(a))+v(x(a),t_0)-w(z(a),a)]+\alpha[z(a)-e(a)]\, dF_{a|t}(a\mid t_0) \quad (18)$$

subject to the incentive constraints in ability that must hold for all $a,a'\in\Theta_a$:

$$u(e(a)-k\cdot x(a))+v(x(a),t_0)-w(z(a),a)\geq u(e(a')-k\cdot x(a'))+v(x(a'),t_0)-w(z(a'),a). \quad (19)$$

In this subproblem agents are heterogeneous only in ability, and utility is weakly-separable in commodities (including numeraire c and goods x) and earnings. Thus, following the logic of the Atkinson-Stiglitz theorem, consumption choices should not be distorted. More formally, consumption choices (c,x) are incentive-separable as defined by Doligalski et al. (2024); then, by the application of their Lemma 1, for any allocation that distorts the choice of (c,x), there exists an alternative allocation which does not distort it, provide all agents with the same utility as before, and generates more revenue to the planner.[22] Since α is strictly positive, the value of the Lagrangian increases. Thus, the solution to the subproblem involves an efficient allocation of x.

By Assumption A1, the set of efficient choices of x is independent of expenditure and, thus,

[22]Note that Doligalski et al. (2024) define an undistorted choice with an expenditure-minimisation problem, while we do it with a utility-maximization problem. A strictly increasing $u(c)$ over $c\in\mathbb{R}$ guarantees the duality of the two approaches.

ability; we consider a solution in which all ability types receive the same vector of goods: $x(a) = x^*(t_0) \in X^*(t_0), \forall a$. Plug in the allocation of goods $x^*(t_0)$ into the Lagrangian and simplify by applying Assumptions A1–A3; the Lagrangian becomes

$$\mathcal{L}(t_0) = \int \left(\bar{\lambda}(a)[e(a) - w(z(a), a)] + \alpha[z(a) - e(a)]\right) dF_a(a) + K(t_0) \tag{20}$$

where $K(t_0) := v(x^*(t_0), t_0) - k \cdot x^*(t_0)$ is an additive term that depends only on taste.[23] The incentive constraints simplify to

$$e(a) - w(z(a), a) \geq e(a') - w(z(a'), a), \quad \forall a, a' \in \Theta_a. \tag{21}$$

The reformulated subproblem depends on the taste type t_0 only via the additive constant in the Lagrangian. Thus, if $(\tilde{e}, \tilde{z})$ denotes the solution to the relaxed subproblem for taste t_0, then $(\tilde{e}, \tilde{z})$ also solves the relaxed subproblem for any other taste type t.

We claim that $(\tilde{e}, \tilde{z})$ combined with the efficient choice of goods $x^*(t) \in X^*(t), \forall t$, solves the original (unrelaxed) problem. To prove it, we need to show that all incentive constraints from the original problem are satisfied. Indeed, for all $(t, a), (t', a') \in \Theta$,

$$\begin{aligned} \tilde{e}(a) - k \cdot x^*(t) + v(x^*(t), t) - w(\tilde{z}(a), a) &= [\tilde{e}(a) - w(\tilde{z}(a), a)] + [v(x^*(t), t) - k \cdot x^*(t)] \\ &\geq [\tilde{e}(a') - w(\tilde{z}(a'), a)] + [v(x^*(t'), t) - k \cdot x^*(t')] \\ &= \tilde{e}(a') - k \cdot x^*(t') + v(x^*(t'), t) - w(\tilde{z}(a'), a), \end{aligned}$$

where the inequality follows from the incentive constraint of the relaxed subproblem and the fact that the allocation of x is efficient. Since the solution to the relaxed problem satisfies all constraints of the original (non-relaxed) problem, it also solves the original problem.

Now we will show that the optimal allocation rule can be decentralized with a competitive goods market (pricing at marginal costs) and an income tax. Let $\tilde{T} : \mathbb{R}_+ \to \mathbb{R} \cup \{\infty\}$ be a tax on earnings satisfying $\tilde{T}(\tilde{z}(a)) = \tilde{z}(a) - \tilde{e}(a)$ for all $a \in \Theta_a$ and $\tilde{T}(z) = \infty$ for $z \notin \tilde{z}[\Theta_a]$.

An individual facing the income tax T and goods prices k obtains the following utility

$$U^{dec}(T, (t, a)) := \max_{(x,z) \in \mathbb{R}_+^{L+1}} z - T(z) - k \cdot x + v(x, t) - w(z, a). \tag{22}$$

[23] Note that this step of the proof relies on additive separability of utility in consumption choices and earnings.

Then, $U^{dec}(\tilde{T},(t,a))$ corresponds to the utility implied by the optimal direct mechanism:

$$U^{dec}(\tilde{T},(t,a)) = \max_{z\in\tilde{z}[\Theta_a]}\{z - \tilde{T}(z) - w(z,a)\} + \max_{x\in\mathbb{R}^L_+}\{v(x,t) - k\cdot x\} \tag{23}$$

$$= \max_{a'\in\Theta_a}\{\tilde{e}(a') - w(\tilde{z}(a'),a)\} + v(x^*(t),t) - k\cdot x^*(t) \tag{24}$$

$$= \tilde{e}(a) - w(\tilde{z}(a),a) + v(x^*(t),t) - k\cdot x^*(t). \tag{25}$$

Here, we first separated the maximization over x and z, and dropped the earnings levels that are taxed prohibitively. Second, we changed the control variable from $z \in \tilde{z}[\Theta_a]$ to $a' \in \Theta_a$, applied the definition of $\tilde{T}$, and noted that the allocation of goods is efficient. The final equality follows from the incentive compatibility of the optimal allocation.

A.2 Proof of Lemma 1

It is immediate that labor supply must satisfy $z_h(t) = \bar{z}$ and $z_l(t) = 0$ in any efficient allocation. Since we restricted attention to allocations in which agents' utilities are finite, we can assume that $c_a(t) \geq \underline{c}$ for all a and t. It is also clear that the resource constraint (9) must be binding in any efficient allocation.

We will first prove that condition (10) is necessary. Suppose condition (10) fails for a positive mass of agents such that $c_a(t) > \underline{c}$ and $t \geq k$. Find $\epsilon > 0$ such that a strictly positive mass of agents have $c_a(t) \geq \underline{c} + \epsilon k$ and $x_a(t) \leq 1 - \epsilon$; then, decrease their $c_a(t)$ by ϵk, and increase their $x_a(t)$ by ϵ. This leaves the resource constraint unaffected and raises the utility of these agents (almost all of them strictly), which contradicts Pareto efficiency. Condition (10) does not restrict $x_a(t)$ for agents with $c_a(t) = \underline{c}$. Finally, suppose that condition (10) fails for a positive mass of agents such that $c_a(t) = \underline{c}$ and $t < k$. Then, an analogous argument as for the first case shows that the utility of a positive mass of such agents can be improved.

We will now prove that condition (10) is sufficient. Fix an allocation satisfying condition (10) and suppose that there is a Pareto improvement. Notice that there cannot be a Pareto improvement for agents for whom $c_a(t) > \underline{c}$ unless these agents consume more resources in total (understood as a decrease in the left-hand side of constraint (9)). Hence, if there is a Pareto improvement, there is also a Pareto improvement in which only agents with $c_a(t) = \underline{c}$ are affected. Similarly, the utility of agents with $c_a(t) = \underline{c}$ and $t < k$ can only be increased by giving them more resources, so we can find a Pareto improvement among agents with $c_a(t) = \underline{c}$ and $t \geq k$. Fix such a Pareto improvement, and denote the set of affected agent types by A. Let $\Delta c_a(t) \geq 0$, $\Delta x_a(t)$ denote the change in their allocation of c and x in the Pareto improvement. It must be that $\Delta c_a(t) + t\Delta x_a(t) \geq 0$ for all $(t,a) \in A$, with a strict inequality for a positive mass of agents within A. To preserve the resource constraint, it must be that, $\mathbb{E}[\Delta c_a(t) + k\Delta x_a(t)|A] \leq 0$, where the expectation is taken over (a, t) conditional on A. We have

$$0 \geq \mathbb{E}[\Delta c_a(t) + k\Delta x_a(t)|A] \geq \mathbb{E}[\underbrace{\Delta c_a(t) + t\min\{\Delta x_a(t), 0\}}_{\geq 0}|A] > 0,$$

where the last inequality is strict because $\Delta c_a(t) + t\Delta x_a(t) > 0$ for a positive mass of agents in the set A. Contradiction.

A.3 Proof of Lemma 2

First, suppose that $c_a(\bar{t}) > \underline{c}$ (which implies that, in an incentive-compatible mechanism, $c_a(t) > \underline{c}$ for all t). Then, Pareto efficiency requires that $x_a(t) = \mathbf{1}_{\{t \geq k\}}$. Fixing a, incentive compatibility implies that $c_a(t)$ jumps downward at $t = k$ by k, and is constant otherwise. In particular, only quality $q = 1$ is offered. Plugging this into the definition of the per-unit price, we obtain that $p_a(1) = k$, as required.

Next, let us assume that $c_a(\bar{t}) = \underline{c}$. By incentive-compatibility, there must exist a type $t^\star$ such that, for $t \in [t^\star, \bar{t}]$, $c_a(t) = \underline{c}$, while for types $t < t^\star$, $c_a(t) > \underline{c}$. For types $t < t^\star$, Pareto efficiency requires that $x_a(t) = \mathbf{1}_{\{t \geq k\}}$. For types $t \geq t^\star$, incentive compatibility requires that $x_a(t) = x_a(\bar{t})$, while Pareto efficiency requires that $t^\star \geq k$. However, if $t^\star > k$, then the resulting $x_a(t)$ would not be monotone on $[0, \bar{t}]$, which contradicts incentive-compatibility. We conclude that $x_a(t) = x_a(\bar{t})\mathbf{1}_{\{t \geq k\}}$. The rest of the proof is analogous to the previous case.

A.4 Proof of Lemma 3

It is without loss of generality to assume that $z_l(t) = 0$ (given the welfare objective function and the fact that $l = 0$). Thus, we only have to solve for the earnings choice of high-ability agents. We will prove that it is optimal to choose $z_h(t) = \bar{z}$ for all t. Suppose that it is not the case. Then, we can adjust all high types' allocations so that their $z_h(t)$ increases to $\bar{z}$ and their $c_h(t)$ increases just enough to make their overall utility unchanged. This adjustment does not affect the objective function and relaxes the resource constraint (as well as one of the IC constraints). Since the relaxation of the resource constraint is strict, and increasing the lump-sum payment increases social welfare, it is always strictly preferred to set $z_h(t) \equiv \bar{z}$.

A.5 Proof of Proposition 2

Consider the incentive constraints (8). Since all high-ability agents work (by Lemma 3), low-ability agents cannot mimic the high-ability agents. By standard arguments, their incentive constraint can be represented as a monotonicity constraint on the allocation $x_l(t)$ and an integral condition pinning down consumption $c_l(t)$ (up to a lump-sum payment). Letting $T_l := c_l(0)$ denote the lump-sum payment to low-ability agents, we have:

$$x_l(t) \text{ is non-decreasing, } c_l(t) + t x_l(t) = T_l + \int_0^t x_l(\tau) d\tau.$$

The constraint $c_l(t) \geq \underline{c}$ can only bind for the highest taste type $t = \bar{t}$, since the above representation implies that $c_l(t)$ is non-increasing in t. Therefore, the constraint $c_l(t) \geq \underline{c}$ for all $t \in [0, \bar{t}]$ is equivalent to requiring

$$T_l + \int_0^{\bar{t}} x_l(t) dt - \bar{t} x_l(\bar{t}) \geq \underline{c}.$$

The incentive constraint for high-ability agents preventing them from misreporting their taste type alone leads to a similar representation (let $T_h := c_h(0)$):

$$x_h(t) \text{ is non-decreasing, } c_h(t) + t x_h(t) = T_h + \int_0^t x_h(\tau) d\tau,\ T_h + \int_0^{\bar{t}} x_h(\tau) d\tau - \bar{t} x_h(\bar{t}) \geq \underline{c}.$$

However, we must additionally satisfy the incentive constraint that high-ability agents do not want to mimic one of the low-ability types, which can now be represented as

$$T_h + \int_0^t x_h(\tau) d\tau - \frac{\bar{\bar{z}}}{h} \geq T_l + \int_0^{t'} x_l(\tau) d\tau + (t - t') x_l(t'), \forall t, t' \in [0, \bar{t}].$$

Note that

$$\int_0^{t'} x_l(\tau) d\tau + (t - t') x_l(t') \leq \int_0^t x_l(\tau) d\tau$$

by the monotonicity of $x_l(t)$ in t, and hence—conditional on misreporting the ability type—it is optimal to report the taste type truthfully. Thus, the constraint simplifies to

$$T_h + \int_0^t x_h(\tau) d\tau - \frac{\bar{\bar{z}}}{h} \geq T_l + \int_0^t x_l(\tau) d\tau, \forall t. \tag{26}$$

Next, using the above formulas for $c_a(t)$, and after a few standard transformations (integration by part), we can rewrite the objective function (up to a term that is constant in the remaining choice variables) as

$$\sum_{a \in \{l, h\}} \mu_a \left(\bar{\lambda}_a T_a + \mu_a \int \Lambda_a(t) \gamma_a(t) x_a(t) dF_a(t) \right),$$

where $\gamma_a(t) = (1 - F_a(t))/f_a(t)$ is the inverse hazard rate and $\Lambda_a(t) = \mathbb{E}\left[\lambda_a(\tilde{t})|\tilde{t} \geq t, a\right]$ is the average welfare weight on types above t, conditional on ability a. Similarly, the resource constraint (9) can be rewritten as

$$\mu_h \bar{z} \geq G + \sum_{a \in \{l,h\}} \mu_a \left(T_a + \int_0^{\bar{t}} (k - J_a(t))\, x_a(t) dF_a(t) \right),$$

where $J_a(t) = t - \gamma_a(t)$ is the virtual surplus function, conditional on ability a.

We will reparameterize the problem by denoting $T = T_l$ and $\Delta T = T_h - \frac{\bar{z}}{h} - T$. That is, T is the lump-sum payment to all agents, and ΔT is the additional monetary payment that high-ability agents receive on top of the lump-sum transfer enlarged by a compensation for disutility of labor. By the incentive constraint (26), $\Delta T \geq 0$. Intuitively, when $\Delta T = 0$, the post-tax wage received by high-ability agents who work is just enough to offset the disutility from labor that they incur (corresponding to case *(a)* in point 4 of Proposition 2); when $\Delta T > 0$, high-ability agents enjoy a strictly positive surplus from working (corresponding to case *(b)* in point 4 of Proposition 2).

We summarize the progress made so far by restating the full problem as

$$\max_{x_h(t), x_l(t), T, \Delta T \geq 0} T + \mu_h \bar{\lambda}_h \Delta T + \sum_{a \in \{l,h\}} \mu_a \int_0^{\bar{t}} \Lambda_a(t) \gamma_a(t) x_a(t) dF_a(t),$$

subject to

$$x_l(t) \text{ is non-decreasing, } T \geq \underline{c} + \bar{t} x_l(\bar{t}) - \int_0^{\bar{t}} x_l(\tau) d\tau,$$

$$x_h(t) \text{ is non-decreasing, } T + \frac{\bar{z}}{h} + \Delta T \geq \underline{c} + \bar{t} x_h(\bar{t}) - \int_0^{\bar{t}} x_h(\tau) d\tau,$$

$$\Delta T + \int_0^t x_h(\tau) d\tau \geq \int_0^t x_l(\tau) d\tau, \forall t.$$

$$\mu_h \bar{z}\left(1 - \frac{1}{h}\right) \geq G + T + \mu_h \Delta T + \sum_{a \in \{l,h\}} \mu_a \left(\int_0^{\bar{t}} (k - J_a(t))\, x_a(t) dF_a(t) \right).$$

Let us parameterize the problem by imposing an additional constraint $x_l(\bar{t}) \leq \bar{x}_l$ and $x_h(\bar{t}) \leq \bar{x}_h$, and optimizing separately over $\bar{x}_l$ and $\bar{x}_h$. Intuitively, the need to bound the allocation rule from above by a number less than 1 may come from the subsistence constraint. Note that, as long as constraint (26) and the subsistence constraint hold for the low-ability agents, we have

$$T + \frac{\bar{z}}{h} + \Delta T \geq \underline{c} + \bar{t} x_h(\bar{t}) - \int_0^{\bar{t}} x_h(\tau) d\tau + \frac{\bar{z}}{h} + \bar{t}(x_l(\bar{t}) - x_h(\bar{t})).$$

It follows that we can increase $x_h(\bar{t})$ to be at least at the level of $x_l(\bar{t})$ while preserving the subsistence constraint for the high-ability agents; thus, it is without loss of generality to assume that $\bar{x}_h \geq \bar{x}_l$. (This argument also establishes that the subsistence constraint is slack for high-ability agents if $\bar{z}/h \geq \bar{t}$.)

We solve the problem by introducing two Lagrange multipliers, $\eta_l \geq$ and $\eta_h \geq 0$, on the subsistence constraints for the low- and high-ability agents, respectively. The resource constraint must hold with equality at the optimal mechanism, which allows us to substitute T in the objective function. The Lagrangian—fixing $\bar{x}_l$ and $\bar{x}_h$—is then maximized over non-decreasing $x_l(t)$ and $x_h(t)$, as well as $\Delta T \geq 0$,

$$\max_{x_l(t)\leq\bar{x}_l,\, x_h(t)\leq\bar{x}_h,\, \Delta T\geq 0} \sum_{a\in\{l,h\}} \mu_a \int_0^{\bar{t}} \left[\Lambda_a(t)\gamma_a(t) + \frac{\eta_a}{\mu_a f_a(t)} + (1+\eta_l+\eta_h)(J_a(t)-k)\right] x_a(t)dF_a(t)$$

$$+ \left(\mu_h\bar{\lambda}_h + \eta_h - (1+\eta_l+\eta_h)\mu_h\right)\Delta T,$$

subject to a single constraint

$$\Delta T + \int_0^t x_h(\tau)d\tau \geq \int_0^t x_l(\tau)d\tau, \forall t.$$

First, we will derive the optimal $x_h(t)$ and ΔT holding fixed $x_l(t)$. Let

$$\phi_h(t) := \left(\Lambda_h(t)\gamma_h(t) + \frac{\eta_h}{\mu_h f_h(t)} + (1+\eta_l+\eta_h)(J_h(t)-k)\right)\mu_h f_h(t),$$

$$\psi := \mu_h\bar{\lambda}_h + \eta_h - (1+\eta_l+\eta_h)\mu_h,$$

so that this auxiliary problem can be written succinctly as

$$\max_{x_h(t)\leq\bar{x}_h,\, \Delta T\geq 0} \int_0^{\bar{t}} \phi_h(t)x_h(t)dt + \psi\Delta T$$

subject to

$$\Delta T + \int_0^t x_h(\tau)d\tau \geq \int_0^t x_l(\tau)d\tau, \forall t.$$

Note that we must have $\psi \leq 0$ as otherwise the problem would not have a solution.

The above problem is a linear mechanism design problem with a type-dependent outside option constraint pinned down by the allocation rule for the low-ability agents. Such a problem can be solved using existing techniques.

Lemma 4. *[Dworczak and Muir (2024)]. Define*

$$\Phi_h(t) = \int_t^{\bar{t}} \phi_h(\tau)d\tau \text{ and } \bar{\Phi}_h(t) = co(\Phi_h)(t),\ \bar{\phi}_h(t) = -\bar{\Phi}_h'(t),$$

where co stands for the concave closure of a function. Let t_0 be defined as the smallest solution to $\bar{\phi}_h(t_0) = \psi$ ($t_0 = 0$ if $\bar{\phi}_h(t) > \psi$ for all t), and let t_1 be defined as the largest solution to $\bar{\phi}_h(t_1) = 0$ ($t_1 = \bar{t}$ if $\bar{\phi}_h(t) < 0$ for all t). (Note that $t_0 \leq t_1$ because $\bar{\phi}_h(t)$ is non-decreasing.) Then,

$$\max_{x_h(t) \leq \bar{x}_h,\, \Delta T \geq 0} \left\{ \int_0^{\bar{t}} \phi_h(t)x_h(t)dt + \psi\Delta T \right\} = \int_{t_0}^{t_1} \bar{\phi}_h(t)x_l(t)dt + \bar{x}_h \int_{t_1}^{\bar{t}} \bar{\phi}_h(t)dt + \psi \int_0^{t_0} x_l(t)dt.$$

Moreover, the optimal solution is given by

$$\Delta T^\star = \int_0^{t_0} x_l(t)dt,$$

$$x_h^\star(t) = \begin{cases} 0 & t \leq t_0, \\ x_l(t) & t \in [a,b] \text{ for every maximal } [a,b] \text{ such that } \Phi_h \equiv \bar{\Phi}_h \text{ on } [a,b], \\ \frac{\int_a^b x_l(\tau)d\tau}{b-a} & t \in (a,b) \text{ for every maximal } (a,b) \text{ such that } \Phi_h < \bar{\Phi}_h \text{ on } (a,b), \\ \bar{x}_h & t \geq t_1. \end{cases}$$

Explaining Lemma 4 is beyond the scope of this paper.[24] The important take-aways for our purposes are that the problem of choosing the optimal $x_h(t)$ for a fixed $x_l(t)$ admits a closed-form solution characterized by two cutoffs, t_0 and t_1, and (possibly) a number of ironing intervals. Ignoring the possibility of ironing (formally, ironing is not needed if $\phi_h(t)$ is monotone), the intuition for the cutoffs t_0 and $t_1$is as follows. The designer chooses an allocation rule for high-ability agents to maximize welfare subject to delivering a certain minimal level of utility to high-ability agents, where the lower bound on utility comes from the possibility of mimicking a low-ability type. It is better to give a cash transfer $\Delta T^\star$ to types $t \leq t_0$ than to let these types consume the allocation for the low-ability agents. Note that $\Delta T^\star > 0$ only if low-ability agents of taste type below t_0 consume the good: $x_l(t) > 0$ for some $t < t_0$. These considerations (after endogenizing $x_l(t)$) ultimately determine whether or not high-ability agents receive strictly positive surplus from working. For types $t \in [t_0, t_1]$, it is optimal to satisfy the constraint by letting them consume what the low-ability agents with analogous taste types consume (again, this is further complicated if ironing is needed). Finally, types $t \geq t_1$ should consume the maximal amount $\bar{x}_h$ regardless of the outside option (here, we rely on the fact that $\bar{x}_h \geq \bar{x}_l \geq x_l(t)$ for all t).

[24]The reader is referred to Dworczak and Muir (2024) for a discussion.

Note that the definition of t_0 and t_1 does not depend on $x_l(t)$. The closed-form expression for the maximized objective function thus allows us to maximize over $x_l(t)$ in the next step. Let

$$\phi_l(t) := \left[\Lambda_l(t)\gamma_l(t) + \frac{\eta_l}{\mu_l f_l(t)} + (1+\eta_l+\eta_h)(J_l(t)-k)\right]\mu_l f_l(t).$$

Then, the problem of maximizing over $x_l(t)$ (assuming that $x_h(t)$ and ΔT are chosen optimally for any $x_l(t)$, as described by Lemma 4), becomes (fixing $\bar{x}_l$)

$$\max_{x_l(r)\leq \bar{x}_l} \int_0^{\bar{t}} \phi_l(t)x_l(t)dt + \int_{t_0}^{t_1} \bar{\phi}_h(t)x_l(t)dt + \bar{x}_h \int_{t_1}^{\bar{t}} \bar{\phi}_h(t)dt + \psi \int_0^{t_0} x_l(t)dt.$$

This problem is linear in $x_l(t)$ with no additional constraints (other than monotonicity of $x_l(t)$), so there exists an optimal solution that takes the form $x_l(t) = \bar{x}_l \mathbf{1}_{\{t\geq t_l\}}$ for some t_l (formally, these are the extreme points of the set of non-decreasing functions on $[0, \bar{t}]$ bounded below by 0 and above by $\bar{x}_l$.)

Finally, consider maximizing the Lagrangian over $\bar{x}_h$ and $\bar{x}_l$ (at the optimal solution, without loss of generality, $x_l(\bar{t}) = \bar{x}_l$ and $x_h(\bar{t}) = \bar{x}_h$):

$$\max_{\bar{x}_l, \bar{x}_h} \left\{ \bar{x}_l \int_{t_l}^{\bar{t}} \phi_l(t)dt + \bar{x}_l \int_{t_0}^{t_1} \bar{\phi}_h(t)dt + \bar{x}_h \int_{t_1}^{\bar{t}} \bar{\phi}_h(t)dt + \psi \bar{x}_l (t_0 - t_l)_+ - \eta_l \bar{x}_l - \eta_h \bar{x}_h \right\}. \tag{27}$$

The problem is linear. There are two possibilities. First, the subsistence constraint could be slack for types a, in which case $\eta_a = 0$ and $\bar{x}_a = 1$. Moreover, if the subsistence constraint is slack for low-ability agents, then it is also slack for high-ability agents. Second, the subsistence constraint could bind (for low-ability types, or both types). In that case, η_a is set so that the coefficient on $\bar{x}_a$ in the Lagrangian (27) is zero; this allows us to choose $\bar{x}_a$ to satisfy the subsistence constraint with equality (by assumption, we restricted attention to cases in which we can satisfy the subsistence constraint when agents do not consume the good, so there is some intermediate level of consumption that satisfies the constraint with equality). In either case, we conclude that the solution described above is a solution to the original problem for *some* choice of η_l and η_h.

We are now ready to finish the proof of Proposition 2. Part 1 follows from the fact that the optimal $x_l(t)$ is a cutoff allocation rule (we set $q_l = \bar{x}_l$). Part 2 follows from the following observation: Since the optimal $x_l(t)$ is a cutoff allocation rule, the optimal allocation rule $x_h(t)$—as predicted by Lemma 4—can take on at most one value, which we call q_i, other than 0, $q_l = \bar{x}_l$, and $q_h := \bar{x}_h$. Specifically, $q_i \in (0, q_l)$ if and only if $t_l \in (a, b) \subseteq (t_0, t_1)$ for some maximal interval (a, b) such that $\Phi_h < \bar{\Phi}_h$ on (a, b): then, $q_i = q_l(b - t_l)/(b - a)$. In this sense, q_i is a result of ironing that is required when the objective function $\phi_h(t)$ is not monotone (so that Φ_h lies below its concave closure $\bar{\Phi}_h$ on some interval). Part 3 follows from the analysis of Lagrange multipliers η_l and η_h above. Finally, to prove part 4, let us separately analyze the form of the optimal mechanism when *(a)* $t_l \geq t_0$, *(b)* when $t_l < t_0$.

In case *(a)*, by Lemma 4, $\Delta T^\star = 0$. This means that $T_h = T_l + \bar{z}/h$ or $c_h(0) - \bar{z}/h = c_l(0)$. By incentive compatibility, $p_l(q_l) = t_l$ since type t_l is the cutoff type consuming quality q_l. To determine the average prices paid by high-ability type, we consider three cases:

(i) If $t_1 \leq t_l$, then $x_h(t) = \bar{x}_h(t)\mathbf{1}_{\{t \geq t_1\}}$, so $\mathrm{Im}(x_h) = \{0, \bar{x}_h\}$ and $p_h(q_h) = t_1 \leq t_l = p_l(q_l)$. If $q_l \in \mathrm{Im}(x_h)$, then it follows that $q_l = q_h$ and hence $p_h(q_l) \leq p_l(q_l)$.

(ii) If $t_1 > t_l$ and no ironing is required (q_i is not offered), then it follows that $x_h(t) = x_l(t) = \mathbf{1}_{\{t \geq t_l\}}$ for all $t \in (t_0, t_1)$, and hence $p_l(q_l) = p_h(q_l) = t_l$.

(iii) If $t_1 > t_l$ but ironing is required, then we have $x_h(t) = q_i$ for $t \in [a, b)$, $x_h(t) = q_l$ for $t \in [b, t_1)$, and $x_h(t) = \bar{x}_h$ for $t \geq t_1$, for some $t_0 \leq a \leq t_l \leq b \leq t_1$. In this case, quality q_i must be offered at the average price a, and type b must be indifferent between buying quality q_i at a per-unit price of a, or buying q_l at a per-unit price of $p_h(q_l)$:

$$(b - a)q_i = (b - p_h(q_l))q_l \iff t_l = p_h(q_l).$$

We conclude that in all cases when q_l is offered to high-ability agents ($q_l \in \mathrm{Im}(x_h)$), we have $p_h(q_l) \leq p_l(q_l)$.

In case *(b)*, $\Delta T^\star = \bar{x}_l(t_0 - t_l) > 0$, by Lemma 4. A further consequence of the lemma is that—since x_l is constant in $[t_0, t_1]$—x_h must be equal to $\bar{x}_l$ on $[t_0, t_1]$, and hence $\mathrm{Im}(x_h) \subseteq \{0, q_l, q_h\}$. Since $t_l < t_0$, high-ability agents must face a higher per-unit price for consuming q_l.

Cases *(a)* and *(b)* above thus correspond to the analogous cases in Proposition 2, which finishes its proof.

A.6 Proof of Theorem 2

We will construct a solution (based on Proposition 2) in which the subsistence constraint is slack for high-ability agents and binds for low-ability agents (the condition on the aggregate resources in the statement of the theorem ensures that we will be able to verify that property). Using the notation from the proof of Proposition 2, we set $\eta = \eta_l$ and $\eta_h = 0$.

First, we prove a technical lemma showing that under our regularity conditions, ironing is not required in the optimal allocation rule for high-ability agents.

Lemma 5. *Under the assumptions of Theorem 2, the solution described by Lemma 4 does not involve ironing ($\phi_h = \bar{\phi}_h$ in the relevant range).*

Proof. Under the current assumptions, we have

$$\phi_h(t) \equiv \left((t-k)(1+\eta)\mu_h + \psi\frac{1-F(t)}{f(t)}\right)f(t),$$

$$\psi = -\mu_h(1+\eta-\lambda_h),$$

where λ_h is the (constant) welfare weight on high-ability agents. We will first show that $\phi_h(t)$ is non-decreasing over $[t_0, t_1]$, so that ironing is not required in the optimal mechanism, where t_0 and t_1 are defined by

$$(t_1-k)(1+\eta)\mu_h + \psi\frac{1-F(t_1)}{f(t_1)} = 0,$$

$$(t_0-k)(1+\eta)\mu_h - \psi\frac{F(t_0)}{f(t_0)} = 0.$$

(We will later verify that these definitions coincide with the definition in Lemma 4.) Note that our regularity assumptions imply that t_0 and t_1 are uniquely defined (recall that $\psi \leq 0$).

We need to show that, for $t \in (t_0, t_1)$,

$$((1+\eta)\mu_h + \psi\gamma'(t))\,f(t) + ((t-k)(1+\eta)\mu_h + \psi\gamma(t))\,f'(t) > 0.$$

We know that, in the relevant range,

$$\psi\frac{F(t)}{f(t)} < (t-k)(1+\eta)\mu_h < -\psi\frac{1-F(t)}{f(t)}.$$

When $f'(t) > 0$, we have

$$((1+\eta)\mu_h + \psi\gamma'(t))\,f(t) + ((t-k)(1+\eta)\mu_h + \psi\gamma(t))\,f'(t) > \left(\psi\frac{F(t)}{f(t)} + \psi\frac{1-F(t)}{f(t)}\right)f'(t) > 0.$$

When $f'(t) < 0$, we have

$$((1+\eta)\mu_h + \psi\gamma'(t))\,f(t) + ((t-k)(1+\eta)\mu_h + \psi\gamma(t))\,f'(t) > \left(-\psi\frac{1-F(t)}{f(t)} + \psi\frac{1-F(t)}{f(t)}\right)f'(t) = 0.$$

This shows that $\phi_h(t)$ is non-decreasing over $[t_0, t_1]$.

Next, notice that $\phi_h(t)$ crosses zero once from below, and hence $\phi_h(t) \geq 0$ for all $t \geq t_1$.

Similarly, we want to show that $\phi_h(t) \le \psi$ for all $t \le t_0$. For $t \le t_0$,we have

$$\left((t-k)(1+\eta)\mu_h + \psi\frac{1-F(t)}{f(t)}\right)f(t) = \left(\underbrace{(t-k)(1+\eta)\mu_h - \psi\frac{F(t)}{f(t)}}_{\le 0}\right)f(t) + \psi \le \psi.$$

We have thus shown that $\phi_h(t) = \bar{\phi}_h(t)$ over $[t_0, t_1]$, and moreover that t_0 and t_1 defined above coincide with those defined in Lemma 4. It follows that no ironing is needed: For a fixed x_l, the optimal allocation rule for high-ability agents is given by

$$x_h^\star(t) = \begin{cases} 0 & t < t_0, \\ x_l(t) & t \in [t_0, t_1), \\ \bar{x}_h & t \ge t_1. \end{cases}$$

□

Combining Lemma 5 with Proposition 2, we conclude that the optimal solution is parameterized by: t_l, t_0, t_1, $\bar{x}_l$ (which we keep fixed for now), and $\bar{x}_h$ (which we conjecture will be equal to 1). Note that as long as $t_0 < t_l$, the value of t_0 does not affect the mechanism (since $x_l(t) = 0$ for $t \le t_l$). Thus, it is without loss of generality to assume that $t_0 \ge t_l$. Under that assumption, we have $\Delta T^\star = \bar{x}_l(t_0 - t_l)$. The resulting Lagrangian—which is maximized over t_l, t_0, and t_1—takes the form:

$$\begin{aligned} &-(1+\eta)\left[\mu_h\bar{x}_l\int_{t_0}^{t_1}(k-J(t))\,dF(t) + \mu_h\int_{t_1}^{\bar{t}}(k-J(t))\,dF(t) + \mu_l\bar{x}_l\int_{t_l}^{\bar{t}}(k-J(t))\,dF(t)\right] - \eta\bar{x}_l t_l \\ &-(1-\lambda_h+\eta)\mu_h\bar{x}_l(t_0-t_l) + \mu_l\lambda_l\bar{x}_l\int_{t_l}^{\bar{t}}\gamma(t)dF(t) + \mu_h\lambda_h\bar{x}_l\int_{t_0}^{t_1}\gamma(t)dF(t) + \mu_h\lambda_h\int_{t_1}^{\bar{t}}\gamma(t)dF(t). \end{aligned} \tag{28}$$

We will argue that $t_0 = t_l$ in the optimal mechanism. Since we know that $t_0 \ge t_l$, towards a contradiction, suppose that $t_0 > t_l$; then, the first-order conditions for optimal t_0 and t_l must hold, which would require (after some transformations, and in particular substituting $\lambda_l = (1-\mu_h\lambda_h)/\mu_l$):

$$(1+\eta)(t_0-k)f(t_0) + (1+\eta-\lambda_h)F(t_0) = 0,$$

$$(1+\eta)(t_l-k)f(t_l) + (1+\eta-\lambda_l)F(t_l) = 0.$$

The first condition states that $\phi_h(t_0) = \psi$, and since t_0 is the smallest solution to this equation, we know that $\phi_h(t) < \psi$ for all $t < t_0$, and thus in particular,

$$(1+\eta)(t_l - k)f(t_l) + (1+\eta-\lambda_h)F(t_l) < 0.$$

But this clearly contradicts the second condition (since $\lambda_h \le \lambda_l$).

Thus, we have proven that $t_0 = t_l$. In particular, $\Delta T^\star = 0$, so we are in case *(a)* in part 4 of Proposition 2. The solution is characterized (up to pinning down $\bar{x}_l$ and confirming that $\bar{x}_h = 1$) by the first-order conditions for optimal t_0 and t_1:

$$[(1+\eta)(k - J(t_0)) - h(t_0)]\, f(t_0) = \eta,$$

$$(1+\eta)(k - J(t_1)) = \lambda_h h(t_1).$$

We can rewrite the FOCs as

$$t_0 = k - \frac{\eta}{1+\eta}\frac{F(t_0)}{f(t_0)}.$$

$$t_1 = k + \left(1 - \frac{\lambda_h}{1+\eta}\right)\frac{1-F(t_1)}{f(t_1)}.$$

Note that, as long as $\eta > 0$, we have $t_0 < k < t_1$. In the indirect implementation, the per-unit price $p_l = t_l = t_0$ for quality $\bar{x}_l = q_l$ is thus below marginal cost. The total price $p_h(1)$ for the good with quality 1 must make type t_1 indifferent:

$$t_1 - p_h(1) = q_l(t_1 - p_l) \implies p_h(1) = p_l q_l + t_1(1 - q_l),$$

which verifies point *1* of Theorem 2 if we define $p_h = t_1$ (note that p_h is the revenue-maximizing price if $\lambda_h = 0$). Point *2* of Theorem 2 follows from the fact that $\Delta T^\star = 0$ (the income tax makes high-ability agents indifferent between working or not). Finally, the binding subsistence constraint for low-ability agents implies that $c_l(0) = \underline{c} + p_l q_l$, verifying point *3*.

It remains to verify that *(i)* the subsistence constraint binds for low-ability agents (so that $\eta > 0$) and *(ii)* the subsistence constraint is slack for high-ability agents (which will verify our conjecture that $\eta_h = 0$ and $\bar{x}_h = 1$).

The resource constraint states that

$$\mu_h \bar{z}\left(1 - \frac{1}{h}\right) = G + T - \mu_h(1-\bar{x}_l)(t_1 - k)(1 - F(t_1)) - \bar{x}_l(t_0 - k)(1 - F(t_0)).$$

Towards a contradiction, suppose that $\eta = 0$. Then, it is optimal to set $\bar{x}_l = 1$, $t_0 = t_1 = k$, and the resource constraint becomes

$$\mu_h \bar{z}\left(1-\frac{1}{h}\right) \geq G+T.$$

Since low-ability agents do not work but can afford to buy one unit of the good at price k, it must be that $T \geq \underline{c}+k$. Thus, we must have

$$\mu_h \bar{z}\left(1-\frac{1}{h}\right) \geq G+\underline{c}+k,$$

which is ruled out by the condition assumed in Theorem 2.

Finally, we make sure that in the solution we have constructed consumption of the high ability agents exceeds the subsistence level. We know that $T=\underline{c}+t_0\bar{x}_l$. Thus, it suffices to show that

$$\underline{c}+t_0\bar{x}_l+\frac{\bar{z}}{h} \geq \underline{c}+t_0\bar{x}_l+t_1(1-\bar{x}_l) \iff \frac{\bar{z}}{h} \geq t_1(1-\bar{x}_l).$$

A sufficient condition is that $\bar{z}/h \geq \bar{t}$, which is what we assumed.

A.7 Proof of Theorem 3 and Proposition 1

Since we assumed that

$$\phi_h(t) \equiv \left(t-k-(1-\Lambda_h(t))\gamma_h(t)\right) f_h(t)$$

is non-decreasing whenever it is negative, it follows from Lemma 4 that no ironing is required to describe the optimal $x_h(t)$. Moreover, combining this observation with Proposition 2, we conclude that the allocation rule $x_l(t)$ takes the form $\mathbf{1}_{\{t\geq t_l\}}$, from which it follows that $x_h(t)=\mathbf{1}_{\{t\geq t_h\}}$, for some t_h. It remains to characterize t_h and t_l.

The optimization problem—based on the derivation in the proof of Proposition 2—becomes

$$\max_{t_h,t_l} \sum_{a\in\{l,h\}} \mu_a \int_{t_a}^{\bar{t}} \left(J_a(t)-k+\Lambda_a(t)\gamma_a(t)\right) dF_a(t)-(1-\bar{\lambda}_h)\mu_h(t_h-t_l)_+.$$

The FOCs for an interior solution (in particular, when $t_h \neq t_l$) are

$$\text{FOC } t_h: \; -(1-\bar{\lambda}_h)\mathbf{1}_{\{t_h\geq t_l\}}-\left(t_h-k-(1-\Lambda_h(t_h))\gamma_h(t_h)\right) f_h(t_h)=0,$$

$$\text{FOC } t_l: \; \frac{\mu_h}{\mu_l}(1-\bar{\lambda}_h)\mathbf{1}_{\{t_h\geq t_l\}}-\left(t_l-k-(1-\Lambda_l(t_l))\gamma_l(t_l)\right) f_l(t_l)=0.$$

We argue that these conditions can never hold with $t_h < t_l$. Indeed, assumption (14) guarantees that

$$t - k - (1 - \Lambda_l(t))\gamma_l(t) \geq t - k - (1 - \Lambda_h(t))\gamma_h(t).$$

Thus, we must have $t_h \geq t_l$.

We will consider the two cases, *(i)* $t_h = t_l$ and *(ii)* $t_h > t_l$, separately.

In case *(i)*, we immediately obtain that $\Delta T^\star = 0$ (which means that high-ability agents get no utility surplus from working or, equivalently, that income is taxed at the rate $1 - 1/h$ per unit of earnings) and that all agents face the same price p in the market. This price p must be equal to $t_h = t_l$. Since the same mechanism is offered to low- and high-ability agents, we can use the unconditional distribution F of taste types. Let us also denote by $\Lambda(p)$ the unconditional (over ability types) expectation of the welfare weight on agents with taste type above p. The FOC for that price p is

$$p - k - (1 - \Lambda(p))\gamma(p) = 0.$$

which gives us the formula from point 1 in Theorem 3.

In case *(ii)*, we conclude that $\Delta T^\star > 0$, so that high-ability agents receive a strictly positive surplus from working. In this case, the two FOCs must hold, and thus (using the fact that $\bar{\lambda}_h \leq 1$)

$$t_h - k - (1 - \Lambda_h(t_h))\gamma_h(t_h) \leq 0,$$

$$t_l - k - (1 - \Lambda_l(t_l))\gamma_l(t_l) \geq 0.$$

This gives us the string of inequalities on the prices from point 2 in Theorem 3.

Finally, we prove Proposition 1. Under the additional assumptions we made, the first-order conditions described above are necessary and sufficient (under the convention that the equality becomes an inequality at the boundaries 0 or $\bar{t}$). Therefore, mechanism 2 is optimal if there exists a solution to the system of equations

$$-(1 - \bar{\lambda}_h) - \left(t_h - k - (1 - \bar{\lambda}_h\bar{\Lambda}(t_h))\gamma_h(t_h)\right) f_h(t_h) = 0,$$

$$\frac{\mu_h}{\mu_l}(1 - \bar{\lambda}_h) - (t_l - k - (1 - \Lambda_l(t_l))\gamma_l(t_l))\, f_l(t_l) = 0,$$

that satisfies $t_h > t_l$. Our goal is to show that if a solution exists for some $\bar{\lambda}_h$, then it must also exist for all higher $\bar{\lambda}_h$. In the second condition, if $\bar{\lambda}_h$ becomes larger, then $\frac{\mu_h}{\mu_l}(1 - \bar{\lambda}_h)$ gets smaller, so the term $(t_l - k - (1 - \Lambda_l(t_l))\gamma_l(t_l))\, f_l(t_l)$ must also get smaller to satisfy the condition (since we assumed that this term is non-decreasing whenever it is positive, we can lower t_l until the condition holds or t_l hits 0). Because t_l gets lower, we preserve the constraint that $t_l < t_h$ (as long as we can show that t_h weakly increases). In the first condition, if $\bar{\lambda}_h$ becomes larger, then $-(1-\bar{\lambda}_h)$ gets larger, so the term $\left(t_h - k - (1-\bar{\lambda}_h\bar{\Lambda}(t_h))\gamma_h(t_h)\right) f_h(t_h)$ must get larger (less negative) to satisfy the condition. By the monotonicity assumptions, we can always increase t_h until the first-order condition holds (or t_h hits $\bar{t}$). However, since the whole term $\left(t_h - k - (1-\bar{\lambda}_h\bar{\Lambda}(t_h))\gamma_h(t_h)\right) f_h(t_h)$ increases when $\bar{\lambda}_h$ increases, we have to make sure that t_h overall increases (to make sure that we preserve the constraint $t_h > t_l$). By the implicit function theorem

$$\frac{\partial t_h}{\partial \bar{\lambda}_h} = \frac{1 - \bar{\Lambda}(t_h)(1-F_h(t_h))}{\frac{\partial}{\partial t_h}\left(\left(t_h - k - (1-\bar{\lambda}_h\bar{\Lambda}(t_h))\gamma_h(t_h)\right) f_h(t_h)\right)} \geq 0$$

as long as

$$1 \geq \bar{\Lambda}(t_h)(1-F_h(t_h)) = \int_{t_h}^{\bar{t}} \bar{\lambda}(t)dF_h(t).$$

But

$$\int_{t_h}^{\bar{t}} \bar{\lambda}(t)dF_h(t) \leq \int_0^{\bar{t}} \bar{\lambda}(t)dF_h(t) = 1,$$

so this always holds.

We conclude that there exists a cutoff λ_h^0 such that mechanism 2 is optimal if $\bar{\lambda}_h > \lambda_h^0$ and mechanism 1 is optimal if $\bar{\lambda}_h < \lambda_h^0$. (Note that it is possible that $\lambda_h^0 \in \{0, 1\}$ in which case one of the mechanisms might never be optimal.)

Finally, when $\bar{\lambda}_h < 1$, when μ_l is low enough, the term $\frac{\mu_h}{\mu_l}(1-\bar{\lambda}_h)$ becomes arbitrarily large, so the second condition cannot hold and thus mechanism 1 must be optimal. On the other hand, when $\bar{\lambda}_h = 1$, mechanism 2 is optimal as long as inequality (14) is strict for all interior t.

B Results under curvature in the utility function

In this section, we examine the robustness of findings from the model with subsistence constraints (Theorem 2) to a utility function that is smooth and strictly concave in numeraire. We also allow for a strictly concave utility from the good x and a strictly convex disutility from working. Specifically, assume the utility of type (t, a) is given by

$$u(c) + v(x,t) - (\mathbb{1}_{a=h}w(z) + \mathbb{1}_{a=l}\bar{w}z) \tag{29}$$

that is twice continuously differentiable in all arguments and where: $u(c)$ is strictly increasing, strictly concave and either $c \in \mathbb{R}$ or $c \geq 0$ and $\lim_{c\to 0} u'(c) = \infty$; $v(x,t)$ is concave in

$x \in \mathbb{R}_+$ and satisfies the single-crossing property: $v_{xt}(x,t) > 0$ for all $t \in [0,\bar{t}]$ and $x \geq 0$; $w(z)$ is strictly increasing and strictly convex in $z \in \mathbb{R}_+$ and $\bar{w}$ is high enough that low-ability types neither work nor prefer to mimic high-ability types in the optimum. The rest of the model is the same as in Section 4.3.

B.1 Preliminary results

No earnings distortion. Define the efficient choice of earnings of high-ability agents given numeraire c as $z^*(c) := w'^{-1}(u'(c))$. Suppose there exists type (t,h) with distorted earnings: $z_h(t) \neq z^*(c_h(t))$. Perturb $z_h(t)$ towards $z^*(c_h(t))$ and adjust $c_h(t)$ to keep the utility of this type constant. The perturbation improves the planner's objective: It relaxes the resource constraint and preserves all incentive constraints, since high-ability types can be mimicked only by other high-ability agents, who are indifferent to this alteration. Thus, earnings of high-ability types are undistorted at the optimum.

Given this result, it will be convenient to define the utility from numeraire net of disutility from working as

$$\tilde{u}(c) := u(c) - w(z^*(c)).$$

Summarizing incentive constraints. For brevity, we will refer to an incentive constraint as IC. Take some $t, t' \in \Theta_t$ and assume that the IC of type (t,h) mimicking (t,l) and of (t,l) mimicking (t',l) are satisfied. Then

$$u(c_h(t)) + v(x_h(t),t) - w(z_h(t)) \geq u(c_l(t)) + v(x_l(t),t) \geq u(c_l(t')) + v(x_l(t'),t). \tag{30}$$

Comparing the left-hand and the right-hand sides, we see that type (t,h) has no incentives to mimic (t',l). Thus, provided that other ICs are satisfied, the ICs corresponding to joint deviations in ability and taste are redundant.

Denote the utility level of type (t,a) by $U_a(t) = u(c_a(t)) + v(x_a(t),t) - \mathbb{1}_{a=h} w(z_h(t))$. The downward ICs in ability can be written as:

$$U_h(t) \geq U_l(t), \forall t \in \Theta_t. \tag{31}$$

Regarding the ICs in taste dimension, given the single-crossing assumption, it is standard to summarize them as

$$U_a(t) = U_a(0) + \int_0^t v_t(x_a(t),t)\,dt, \quad \forall t \in \Theta_t, a \in \{h,l\}, \tag{32}$$

combined with a requirement that $x_l(\cdot)$ and $x_h(\cdot)$ are non-decreasing. Note that $U_a'(t) = v_t(x_a(t), t)$, whenever it exists.

Note that ICs (in taste) imply that $c_l(t)$ must be non-increasing, and strictly decreasing whenever $x_l(t)$ is strictly increasing. The same is true for high-ability agents. To see that, suppose that $x_h(t') \geq x_h(t)$ and $c_h(t') > c_h(t)$ for some $t' > t$. Since earnings are undistorted, $z_h(t') \leq z_h(t)$. Thus, type (t, h) strictly gains from mimicking (t', h)—a contradiction.

Reformulating the resource constraint. Let $u_a(t)$ represent the utility from numeraire net of the cost of working of type (t, a). The resource constraint can be written as a function of $\{u_a(\cdot), x_a(\cdot)\}_{a\in\{h,l\}}$:

$$\int \left(\mu_h(z^*(\tilde{u}^{-1}(u_h(t))) - \tilde{u}^{-1}(u_h(t)) - kx_h(t)) - \mu_l(u^{-1}(u_l(t)) + kx_l(t))\right) dF(t) \geq G. \quad (33)$$

Furthermore, $u_a(t)$ is pinned down by $U_a(0)$ and $x_a(\cdot)$:

$$u_a(t) := U_a(t) - v_t(x_a(t), t) = U_a(0) + \int_0^t v_t(x_a(t'), t')\, dt' - v_t(x_a(t), t). \quad (34)$$

Thus, we effectively expressed the resource constraint as a function of $\{U_a(0), x_a(\cdot)\}_{a\in\{h,l\}}$.

Reformulating the objective. Incorporate the downward incentive constraints in ability (31) into the objective function by forming a Lagrangian:

$$\mathcal{L} = \lambda_h\mu_h \int U_h(t)\, dF(t) + \lambda_l\mu_l \int U_l(t)\, dF(t) + \int (U_h(t) - U_l(t))\, d\Gamma(t) \quad (35)$$

where $\Gamma(t)$ stands for the value of marginally relaxing the downward incentive constraints (in ability) for all types in the interval $[0, t]$—see Jullien (2000) for an analogous formulation in the model with type-dependent outside options. We assume that $\Gamma(t)$ corresponding to the optimal mechanism exists. Note that $\Gamma(t)$ is non-negative and non-decreasing, equal to zero for $t < 0$ and constant for $t \geq \bar{t}$. The multiplier $\Gamma(t)$ can be discontinuous. For instance, $\Gamma(\bar{t}) = \Gamma(0) > 0$ means that the IC in ability binds only for the lowest taste type $t = 0$, while $\Gamma(\bar{t}) > 0$ and $\Gamma(t) = 0, \forall t < \bar{t}$, means that this constraint binds only for the highest taste type $t = \bar{t}$. An intermediate case, with $\Gamma(t)$ increasing over the interval of types, is also possible.

Integrate the objective by parts, starting with high-ability agents:

$$\begin{aligned}\int \lambda_h\mu_h U_h(t)\, dF(t) + \int U_h(t)\, d\Gamma(t) &= (\lambda_h\mu_h + \Gamma(\bar{t}))U_h(\bar{t}) - \int (\lambda_h\mu_h F(t) + \Gamma(t))U_h'(t)\, dt \\ &= (\lambda_h\mu_h + \Gamma(\bar{t}))U_h(0) + \int (\lambda_h\mu_h(1 - F(t)) + \Gamma(\bar{t}) - \Gamma(t))v_t(x_h(t), t)\, dt,\end{aligned}$$

and similarly for the low-ability agents:

$$\int \lambda_l \mu_l U_l(t)\, dF(t) - \int U_l(t)\, d\Gamma(t) = (\lambda_l \mu_l - \Gamma(\bar{t}))U_l(\bar{t}) - \int (\lambda_l \mu_l F(t) - \Gamma(t))U_l'(t)\, dt$$
$$= (\lambda_l \mu_l - \Gamma(\bar{t}))U_l(0) + \int (\lambda_l \mu_l (1 - F(t)) + \Gamma(t) - \Gamma(\bar{t}))v_t(x_l(t), t)\, dt,$$

where we used $U_a'(t) = v_t(x_a(t), t)$, implied by the local ICs in taste.

Planner's problem. We can write the planner's problem as

$$\max_{\{U_a(0), x_a(\cdot)\}_{a\in\{h,l\}}} (\lambda_h \mu_h + \Gamma(\bar{t}))U_h(0) + \int (\lambda_h \mu_h (1 - F(t)) + \Gamma(\bar{t}) - \Gamma(t))v_t(x_h(t), t)\, dt$$
$$+ (\lambda_l \mu_l - \Gamma(\bar{t}))U_l(0) + \int (\lambda_l \mu_l (1 - F(t)) + \Gamma(t) - \Gamma(\bar{t}))v_t(x_l(t), t)\, dt \qquad (36)$$

subject to the resource constraint (33) and the monotonicity constraints that require $x_h(\cdot)$ and $x_l(\cdot)$ to be non-decreasing. We define a relaxed problem as the planner's problem with the monotonicity constraints dropped.

FOCs of the relaxed problem. It will be convenient to define $g_a(t) := 1/u'(c_a(t))$. Note that $g_a(t)$ is a strictly increasing transformation of $c_a(t)$. Thus, $g_a(t)$ is non-increasing in t, and strictly decreasing when $x_a(t)$ is strictly increasing.

$g_a(t)$ represents a resource benefit of marginally lowering the utility from numeraire (net of labor cost) of an agent with type (t, a). For the low-ability agents, this can be verified by differentiating the resource constraint (33) with respect to $u_l(t)$. For the high-ability types, the resource impact of perturbing $u_h(t)$ is given by:

$$\frac{d[z^*(\tilde{u}^{-1}(u_h(t))) - \tilde{u}^{-1}(u_h(t))]}{du_h(t)} = \frac{dz^*(c(t))}{dc(t)} \frac{1}{\tilde{u}'(c(t))} - \frac{1}{\tilde{u}'(c(t))}. \qquad (37)$$

Furthermore, $\tilde{u}'(c(t)) = u'(c(t)) - w'(z^*(c(t))\frac{dz^*(c)}{dc}$. If earnings are on the boundary and $\frac{dz^*(c)}{dc} = 0$, then it follows that the resource impact is $g_h(t)$. Otherwise, given that earnings are undistorted, we have $u'(c) = w'(z^*(c))$, which implies $\frac{dz^*(c)}{dc} = \frac{u''(c)}{w''(z^*(c))}$. Plugging these in, we obtain

$$\frac{d[z^*(\tilde{u}^{-1}(u_h(t))) - \tilde{u}^{-1}(u_h(t))]}{du_h(t)} = -\frac{1}{u'(c_h(t))} = -g_h(t). \qquad (38)$$

Intuitively, since earnings are undistorted, the planner is indifferent between adjusting $c_h(t)$ or $z_h(t)$ to achieve a given change of $u_h(t)$.

The first-order conditions of the relaxed problem with respect to $U_h(0)$ and $U_l(0)$ are:

$$\lambda_h \mu_h + \Gamma(\bar{t}) - \alpha \mu_h \int g_h(t) dF(t) = 0, \tag{39}$$

$$\lambda_l \mu_l - \Gamma(\bar{t}) - \alpha \mu_l \int g_l(t) dF(t) = 0. \tag{40}$$

Summed, they pin down the multiplier on the resource constraint α:

$$\frac{1}{\alpha} = \mathbb{E}_{a,t}\left[g_a(t)\right] =: \bar{g}. \tag{41}$$

We can also rewrite them as

$$\Gamma(\bar{t}) = \mu_h \int \left(\frac{g_h(t)}{\bar{g}} - \lambda_h\right) dF(t) = -\mu_l \int \left(\frac{g_l(t)}{\bar{g}} - \lambda_l\right) dF(t). \tag{42}$$

The first-order conditions with respect to $x_a(t), a \in \{h, l\}$, accounting for the potential corner solution at $x_a(t) = 0$, require

$$\begin{aligned}(\lambda_h \mu_h (1 - F(t)) + \Gamma(\bar{t}) - \Gamma(t)) v_{tx}(x_h(t), t) \\ + \alpha \left[\left(\frac{v_x(x_h(t), t)}{u'(c_h(t))} - k\right) \mu_h f(t) - v_{tx}(x_h(t), t) \mu_h \int_t^{\bar{t}} g_h(t') dF(t')\right] \leq 0\end{aligned} \tag{43}$$

and

$$\begin{aligned}(\lambda_l \mu_l (1 - F(t)) + \Gamma(t) - \Gamma(\bar{t})) v_{tx}(x_l(t), t) \\ + \alpha \left[\left(\frac{v_x(x_l(t), t)}{u'(c_l(t))} - k\right) \mu_l f(t) - v_{tx}(x_l(t), t) \mu_l \int_t^{\bar{t}} g_l(t') dF(t')\right] \leq 0.\end{aligned} \tag{44}$$

Define the good x wedge as $\tau_a(t) := \frac{v_x(x_a(t), t)}{u'(c_a(t))} - k$. A positive (respectively, negative) value of the wedge implies that allocation x is distorted downwards (reps., distorted upwards, provided that $x_a(t) > 0$). We can express the FOCs as

$$\tau_h(t) \frac{\mu_h f(t)}{v_{tx}(x_h(t), t)} \frac{1}{\bar{g}} \leq \mu_h \int_t^{\bar{t}} \left(\frac{g_h(t')}{\bar{g}} - \lambda_h\right) dF(t') - \Gamma(\bar{t}) + \Gamma(t) \tag{45}$$

$$\tau_l(t) \frac{\mu_l f(t)}{v_{tx}(x_l(t), t)} \frac{1}{\bar{g}} \leq \mu_l \int_t^{\bar{t}} \left(\frac{g_l(t')}{\bar{g}} - \lambda_l\right) dF(t') + \Gamma(\bar{t}) - \Gamma(t). \tag{46}$$

Sum them up and multiply by $\bar{g}$ to get

$$\tau_h(t)\frac{\mu_h f(t)}{v_{tx}(x_h(t),t)} + \tau_l(t)\frac{\mu_l f(t)}{v_{tx}(x_l(t),t)} \leq (1-F(t))\mathbb{E}_{a,t'}\left[g_a(t') - \bar{g} \mid t' \geq t\right]. \tag{47}$$

Since $g_a(t)$ is non-increasing with taste, the right-hand side is (weakly) negative. Thus, either $\tau_h(t)$ or $\tau_l(t)$ must be (weakly) negative for any $t \in \Theta_t$.

B.2 Optimal goods distortions

The following proposition characterizes the optimal goods market distortions with curvature in the utility function. To rule out an uninteresting case, we assume that in the optimum a positive measure of agents receive $x > 0$. We discuss this proposition and provide intution in the main body of the paper (Section 4.3.1).

Proposition 3. *Suppose that Assumptions A2 and A3 hold, and that agents' preferences are given by formula (29). The optimal mechanism has the following properties:*

1. *Distortions to good x are optimal: There can be no interval of taste types $[i_1, i_2] \subseteq [0, \bar{t}]$ where, for all $t \in [i_1, i_2]$, $\max\{x_h(t), x_l(t)\} > 0$ and both $x_h(t)$ and $x_l(t)$ are undistorted.*

2. *Assume that the optimum does not require ironing.*[25] *The optimal allocation of good x of the low-ability types is either distorted upwards or undistorted.*

3. *Assume $\lambda_h = 0$ and that at the optimum $x_h(t) > x_l(t)$ for all $t \geq t_0$. The optimal allocation of good x of the high-ability types with taste $t \in (t_0, \bar{t}]$ coincides with the solution to the one-dimensional monopolistic screening problem (with the reservation value given by the utility of type (t_0, h) and the lower bound on feasible allocations of x given by $x_h(t_0)$).*

Proof. **Part 1.** Consider an optimal allocation rule. Suppose there exist $i_1, i_2 \in \Theta_t$, $i_2 > i_1$, such that for all $t \in [i_1, i_2]$ both $x_h(t)$ and $x_l(t)$ are undistorted and at least one of them is strictly positive.

We will start by showing that when $x_a(t)$ is undistorted and strictly positive over the taste interval $[i_1, i_2]$ then $x_a(t)$ is strictly increasing and $c_a(t)$ strictly decreasing in t over this interval, for all $a \in \{h, l\}$. This is useful since it means that the monotonicity constraints are slack for any $t \in (i_1, i_2)$ and the FOC from the relaxed problem must hold at the optimum. Suppose that $x_a(t_1) = x_a(t_2) = \bar{x} > 0$ for some $i_1 \leq t_1 < t_2 \leq i_2$. Since markets are not distorted

$$ku'(c_a(t_2)) - ku'(c_a(t_1)) = v_x(\bar{x}, t_2) - v_x(\bar{x}, t_1) = \int_{t_1}^{t_2} v_{xt}(\bar{x}, t)dt > 0, \tag{48}$$

[25] That is, we can drop the monotonicity constraints on $x_l(t)$ and $x_h(t)$ without affecting the solution.

which means that $c_a(t_2) < c_a(t_1)$. however, then type (t_2, a) would mimic (t_1, a), which is a contradiction. Furthermore, if $x_a(t_2) > x_a(t_1)$ then $c_a(t_2) < c_a(t_1)$, since otherwise type (t_1, a) would mimic (t_2, a).

Suppose that low-ability types with taste above threshold t_l, where $t_l < i_2$, consume a positive amount of good x. Define $\tilde{t} := \max\{i_1, t_l\}$. Then the high-ability types with taste from $[\tilde{t}, i_2]$ must also consume x in a positive quantity, which follows from their allocation of x being undistorted. Since both $x_h(t)$ and $x_l(t)$ are strictly increasing over $[\tilde{t}, i_2]$, the monotonicity constraints are slack for all $t \in (\tilde{t}, i_2)$ and the FOCs from the relaxed problem inform us of the welfare impact of a small perturbation within this open interval. Consider (47). Since $g_a(t)$ is a monotone transformation of $c_a(t)$, it is strictly increasing over $[\tilde{t}, i_2]$, and the right-hand side of (47) is strictly negative for any $t \in (\tilde{t}, i_2)$. On the other hand, the left-hand side is equal to zero, since good x is undistorted. The FOC is violated and the planner can improve the allocation by perturbing both $x_l(t)$ and $x_h(t)$ upward for all $t \in (\tilde{t}, i_2)$.

Next suppose that within the taste interval $[i_1, i_2]$ none of the low-ability types and all of the high-ability types consume a positive amount of x. Recall that that $U_a(t) = v_t(x_a(t), t)$, for all $a \in \{h, l\}$, where the right-hand side is increasing in $x_a(t)$. Since $U_h(i_1) \geq U_l(i_1)$ and $x_h(t) > 0 = x_l(t)$ for all $t \geq i_1$, it follows that $U_h'(t) > U_l'(t)$ and $U_h(t) > U_l(t)$ for all $t > i_1$. Thus, $\Gamma(t) = \Gamma(i_1)$ for all $t \in (i_1, i_2)$. The FOC for $x_h(t)$ becomes

$$\tau_h(t) \frac{\mu_h f(t)}{v_{tx}(x_h(t), t)} \frac{1}{\bar{g}} = \mu_h \int_t^1 \left(\frac{g_h(t')}{\bar{g}} - \lambda_h \right) dF(t') - \Gamma(\bar{t}) + \Gamma(i_1). \tag{49}$$

The derivative of the right-hand side with respect to t is proportional to $\lambda_h - \frac{g_h(t)}{\bar{g}}$. Given that $g_h(t)$ is strictly decreasing over the interval (i_1, i_2), the right-hand side can be either strictly decreasing or strictly increasing or first strictly decreasing and then strictly increasing. Either way, there are at most two values of t for which the right-hand side is zero. Thus, for almost all $t \in (i_1, i_2)$ the first-order condition is violated and the planner can improve the allocation by distorting $x_h(t)$.

Part 2. Consider the relaxed problem. Define

$$\phi(t) := \mu_l \int_0^t \left(\lambda_l - \frac{g_l(t)}{\bar{g}} \right) dF(t) - \Gamma(t). \tag{50}$$

Combining FOCs with respect to $x_l(t)$ and $U_l(0)$ yields

$$\phi(t) \geq \tau_l(t) \frac{\mu_l f(t)}{v_{tx}(x_l(t), t)} \frac{1}{\bar{g}}. \tag{51}$$

Let's characterize the behavior of $\phi(t)$. From the FOC with respect to $U_l(0)$ we know that

$$\mu_l \int_0^{\bar{t}} \left(\lambda_l - \frac{g_l(t)}{\bar{g}}\right) dF(t) = \Gamma(\bar{t}) \geq 0. \tag{52}$$

Since $g_l(t)$ is non-increasing, there exists a threshold $\tilde{t} \geq 0$ such that $\lambda_l \geq \frac{g_l(t)}{\bar{g}}$ for all $t \geq \tilde{t}$. It follows that the first term of $\phi(t)$ is continuous in t, equal to 0 at $t = 0$, and (weakly) decreasing until $\tilde{t}$, at which point it becomes (weakly) increasing, eventually reaching $\Gamma(\bar{t})$. The second term, $-\Gamma(t)$, is right-continuous (which follows from the definition of $\Gamma(t)$) and (weakly) decreasing in t, eventually reaching $-\Gamma(\bar{t})$. Thus, $\phi(0) \leq 0$ and $\phi(\bar{t}) = 0$. In addition, $\phi(t)$ can be strictly positive only when $t > \tilde{t}$, i.e. in the region where the first term is increasing.

Now, suppose there exists $t_0 \in \Theta_t$ for which $\tau_l(t_0) > 0$, which would contradict the proposition. That requires $\phi(t_0) > 0$. Since $\phi(\cdot)$ is right-continuous and it eventually reaches the value $\phi(\bar{t}) = 0$, there must exist $t_1 > t_0$ such that $\phi(t) > 0$ for all $t \in (t_0, t_1)$ and $\phi(t_1^-) > \phi(t)$ for all $t > t_1$. Since $\phi(\cdot)$ is strictly decreasing at t_1, $\Gamma(\cdot)$ is strictly increasing at this point, implying $U_h(t_1) = U_l(t_1)$.

Given that $U_h(t_0) \geq U_l(t_0)$, it follows that

$$U_h(t_1) - U_h(t_0) \leq U_l(t_1) - U_l(t_0), \tag{53}$$

or

$$\int_{t_0}^{t_1} U_h'(s)ds \leq \int_{t_0}^{t_1} U_l'(s)ds. \tag{54}$$

Thus, there must exists $t' \in (t_0, t_1)$ such that $U_h'(t') < U_l'(t')$ or, equivalently, $x_h(t') \leq x_l(t')$.

If $x_l(t') > 0$, then the FOC with respect to $x_l(t')$ holds as an equality and $\tau_l(t') > 0$. Since $x_h(t') \leq x_l(t')$ and $c_h(t') > c_l(t')$—which must hold, as otherwise type (t, h) would mimic (t, l)—it follows that $\tau_h(t') > 0$. That contradicts (47), which requires that either $\tau_h(t')$ or $\tau_l(t')$ is non-positive.

If $x_l(t') = x_h(t') = 0$, then, by monotonicity, $x_l(t_0) = x_h(t_0) = 0$. Similarly as in the previous case, $\tau_l(t_0) > 0$ and $x_h(t_0) \leq x_l(t_0)$ implies that $\tau_h(t_0) > 0$, which contradicts (47).

Part 3. We will consider the planner's subproblem of choosing the allocation of high ability types with taste $t > t_0$ taking as given the rest of the allocation rule. We will show that it can be written as a one-dimensional monopolistic screening problem with non-quasi-linear utilities.

By assumption, the welfare weight is 0 for the high-ability types. Furthermore, the marginal value of public funds is always positive. Thus, the planner's objective with respect to the high-ability types is to maximize revenue.

We will show that the downward ICs in ability are slack for $t > t_0$. Note that the ICs in taste require that $U_a'(t) = v_t(x_a(t), t)$, with the right-hand side strictly increasing in $x_a(t)$. Thus, given that the optimal allocation involves $U_h(t_0) \geq U_l(t_0)$ and $x_h(t) > x_l(t)$ for all $t \geq t_0$, it follows that $U_h'(t) > U_l'(t)$ and $U_h(t) > U_l(t)$ for all $t > t_0$.

Define $p_h(t) := z^*(c_h(t)) - c_h(t)$ as a transfer from type (t, h) to the planner. Let $P(c) := z^*(c) - c$, which is strictly decreasing, and define the disutility from transfer as $d(p) := -\tilde{u}(P^{-1}(p))$. It follows that the utility of type (t, h) from allocation (c, x, z) where $z = z^*(c)$ can be described as $v(x, t) - d(p)$ where $p = z - c$.

Now, we can write the planner's subproblem over $\{x_h(t), p_h(t)\}_{t \in (t_0, \bar{t}]}$, taking the allocation of remaining high-ability types as given, as:

$$\max_{\{x_h(t), p_h(t)\}_{t \in (t_0, \bar{t}]}} \int_{t_0}^{\bar{t}} (p_h(t) - kx_h(t))\, dF(t) \tag{55}$$

subject to incentive-compatibility constraints

$$v(x_h(t), t) - d(p_h(t)) \geq v(x_h(t'), t) - d(p_h(t')), \quad \forall t, t' \in [0, \bar{t}]. \tag{56}$$

The single-crossing condition implies that the local incentive constraints are sufficient and we only need to keep track of incentives to deviate to within the set $[t_0, \bar{t}]$. We can summarize these incentive constraints as

$$U_h(t) = U_h(t_0) + \int_{t_0}^{t} v_t(x_h(t), t), \quad \forall t \in (t_0, \bar{t}], \tag{57}$$

together with the requirement that $x_h(t)$ is non-decreasing over $[t_0, \bar{t}]$. Note that $U_h(t_0)$ and $x_h(t_0)$, which are taken as given, play the roles of the reservation value and the lower bound on the feasible allocation of x, respectively. Thus, we can rewrite the problem as

$$\max_{\{x_h(t), p_h(t)\}_{t \in (t_0, \bar{t}]}} \int_{t_0}^{\bar{t}} (p_h(t) - kx_h(t))\, dF(t) \tag{58}$$

subject to (57), $x_h(\cdot)$ being non-decreasing and $x_h(t) \geq x_h(t_0), \forall t \in (t_0, \bar{t}]$, which concludes the proof. □