Optimal Design for Redistributions among Endogenous Buyers and Sellers*

Mingshi Kang† Charles Z. Zheng‡

June 15, 2022

Abstract

Given a large market of individuals entitled to equal shares of a limited resource, each allowed to buy or sell the shares, we characterize the interim incentive-constrained Pareto frontier subject to market clearance and budget balance. At most two prices—partitioning the type space into at most three tiers and using rations only on the middle tier—are needed to attain any interim Pareto optimum. When the virtual surplus function satisfies a single crossing condition without having to be monotone, the optimal mechanism reduces to a single, posted price and requires neither rationing nor lump sum transfers. We find which types gain, and which types lose, when the social planner chooses a rationing mechanism over the single-price solution, as well as the welfare weight of which type is crucial to the choice. The finding suggests a market-like mechanism to distribute Covid vaccines optimally within the same priority group.

JEL Classification: C61, D44, D82

Keywords: market design, redistribution, interim incentive efficiency, interim Pareto optimality, ironing, rationing, endogenous buyers and sellers, vaccine distribution

*Previously titled "Pareto Optimal Hierarchies." Latest version posted at this link. We thank Scott Kominers, two referees and the associate editor for comments and suggestions. Zheng acknowledges financial support from the Social Science and Humanities Research Council of Canada, Insight Grant R4809A04.

†School of Economics, Jinan University, Guangzhou, China, mingshikang92@gmail.com, https://mingshikang.weebly.com.

‡Department of Economics, University of Western Ontario, London, ON, Canada, charles.zheng@uwo.ca, https://economics.uwo.ca/faculty/zheng/.

The haves and have-nots
live on each other.
—Laotzu

1 Introduction

Ever since Myerson and Satterthwaite (1983) discovered the impossibility of fully efficient bilateral trades given asymmetric information, much has been done theoretically to characterize the mechanisms that achieve the incentive-constrained Pareto frontier. New characterization has been obtained by an emerging literature of redistribution-driven market design such as Akbarpour, Dworczak and Kominers (2020, henceforth ADK), Dworczak, Kominers and Akbarpour (2021, henceforth DKA), and Kang (2020).[1] Despite their large-market assumption (continuum of atomless individuals) that eliminates the market power of individuals, the general message from this literature is that a single market price is insufficient to implement allocations on the incentive-constrained Pareto frontier. An optimal mechanism needs to stratify the space of types (private valuations) into more than two tiers through tier-specific prices, augmented with *rationing* that restricts the quantity of demand or supply for individuals, as well as *lump sum transfers* among individuals to achieve redistribution objectives. Meanwhile, they observe upper bounds for the number of such instruments. In DKA (2021), there exists an optimal allocation that stratifies the type space into at most five tiers, implemented through rationing on at most two of the tiers together with a lump sum transfer and tier-specific prices. In Kang (2020), there exists an optimal allocation that stratifies the type space into at most four tiers, and rationing is necessary to attain optimality.[2]

However, all the studies cited above impose on the market some kind of sectorial restrictions, predetermined exogenously before the realization of types. An individual is ex-

[1] Kang and Zheng (2020), and Reuter and Groh (2020), consider redistribution-driven mechanism design without the large-market assumption.

[2] DKA (2021) observe that there exists an optimal mechanism whose allocation for the buyers, and allocation for the sellers, are each a monotone step function such that their total number of jumps or drops is at most four. In other words, if the lowest tier among the buyers and that among the sellers are combined into one tier (both excluded from trading), the total number of tiers, from the highest tier among the buyers to the highest tier among the sellers, is at most five.

Kang (2020) observes that there exists an optimal mechanism that partitions the public-sector buyers into at most three tiers. This combined with the private-market buyers means four tiers.

ogenously assigned the role of a buyer or that of a seller in Myerson and Satterthwaite (1983) and DKA (2021), or exogenously assigned to a group in ADK (2020). In Kang (2020), where individuals choose between the public and the private sectors to trade, the private sector is restricted to be operated under a competitive market price. If such exogenous restrictions are removed, would the policy instruments necessary to attain the Pareto frontier be simplified so that merely a single market price could suffice, or would they get complicated because the endogenous grouping of individuals now becomes an additional dimension in the policymaker's choice variable?

We therefore consider a large market with endogenous buyers and sellers. It is a continuum of individuals each allowed to buy or sell a good at a marginal utility or marginal cost equal to one's type. The planner has the same set of instruments for the buyers and for the sellers. As in the literature cited above, we consider the entire incentive-constrained Pareto frontier through examining a social planner's optimization problem where the given welfare weight can vary with individuals' types arbitrarily. Such type-dependent welfare weights capture a social planner's redistributive preferences for one type over another. Free of restrictions on such redistributive preferences, the observations would then be applicable to the various social welfare criteria according to which the planner designs the mechanism.

Across such arbitrary welfare weights, our characterization of optimal mechanisms turns out to be simpler than those in the above cited. We find a tighter upper bound, three, of the number of tiers that optimal mechanisms have to stratify the type space into (Theorem 2.i). Furthermore, when the virtual surplus function is constant only on a measure-zero set of types, we obtain the exact number of tiers, rather than only an upper bound thereof, in any optimal mechanism: The optimal allocation is unique, and the associated optimal number of tiers is equal to either two or three (Theorem 3). If the virtual surplus satisfies a single crossing condition, the optimal mechanism reduces to a competitive price—offering a single price to all types, be they sellers or buyers, without any other instrument such as rationing, lump sum transfers or tier-specific prices (Theorem 2.ii). In particular, this is true even when the virtual surplus function is non-monotone. In the related literature, by contrast, the only case where a single price is known to implement interim Pareto optimality is the exogenous buyer-seller bilateral trade model of DKA (2021), where the conditions they require together imply that the endogenous virtual surplus functions in their model be monotone.[3] Our optimality observation of a single market price without rationing is also opposite to the finding in Kang's (2020) model that rationing is in general necessary.

The simplicity of our characterization is due to a new observation, mainly driven by the endogeneity of one's buyer- or seller-role. The observation is that any incentive compatible and market clearing allocation can always be implemented in a budget-balanced (BB) and individually rational (IR) manner (Theorem 1). Thus, in contrast to the models of Myerson and Satterthwaite (1983), DKA (2021), and Kang (2020), the BB and IR constraints can be removed without loss in our model. That eliminates the two-sidedness (buyer- and seller-sides) of the information asymmetry, a main driving force of the Myerson-Satterthwaite impossibility theorem. Then, as in a monopolist's capacity-constrained second-degree price discrimination problem (Bulow and Roberts 1989; or the one-group special case in ADK 2020), or in the basic Bayesian persuasion problem (Kamenica and Gentzkow 2011), rationing is needed to attain optimality only when the market clearing cutoff type is interior to an interval on which the virtual surplus function has to be ironed.[4] If that happens, rationing is needed only on that interval, and the type space is stratified to at most three tiers. If that does not happen, a single market price suffices optimality, and the type space is stratified into only two tiers, one being all the buyers, the other all the sellers.

An implication of the characterization is that the welfare weight of the types near the buyer-seller cutoff in the posted-price system is crucial to the social planner's choice between rationing versus posting a single price. We find that such types are better-off in a mechanism with rationing than given a posted price, and the planner prefers the former to the latter if the welfare weight on such types is sufficiently heavy. We can also tell which among the other types are definitely worse-off under rationing than under the posted price based on the curvature of the type-distribution on the types near the said buyer-seller cutoff (Theorems 4 and 5). Our model can also be modified to capture a kind of externalities without altering any of the results (Corollary 4).

With endogenous buyers and sellers, our model is applicable to large-market exchange

[3] DKA's (2021) Theorem 2 requires quasi-convexity of virtual surplus functions and positive derivative of the virtual surplus function at the minimum type ("low same side inequality" in their language). The two together imply that the virtual surplus function is monotone for both buyers and sellers.

[4] The Bayesian consistency condition in Bayesian persuasion models corresponds to the market clearance in Bulow and Roberts's (1989) second-degree price discrimination problem, with the prior probability that a sender is supposed to split in the former corresponding to the market equilibrium quantity in the latter.

economies where individuals have equal entitlements to a limited resource and are heterogeneous in their willingness to pay for the access to the resource. Applied to such situations, our finding implies that, given any redistributive preferences, the social planner can attain optimality through a market-like mechanism for individuals to trade their shares. It would issue to each eligible individual a coupon that represents the person's initial equal access to the limited resource, and the coupon trading would eventually stratify the individuals into at most three tiers in terms of their final shares of the limited resource, those who give up their shares completely (the "have-nots"), those who max out their acquisition of shares (the "haves"), and those in between. When the virtual surplus satisfies the aforementioned conditions, even when the social planner has redistributive preferences across types, the coupon-trading mechanism reduces to a single competitive price for the coupon. We illustrate this in the context of Covid vaccine allocation within the same priority group (Section 6).

Our model shares a similarity with partnership dissolution models in treating the roles of buyers and sellers as endogenous (Cramton et al. 1987; Lu and Roberts 2001; Kittsteiner 2003; Chien 2007; Mylovanov and Tröger 2014; Segal and Whinston 2016; Loertscher and Cédric 2019). Theorem 1 can be extended to those models provided that the values are private and partners are not overly asymmetric ex ante (though we find no precedent thereof in that literature).[5] That is consistent with the observation of Cramton et al. (1987) that full efficiency can be attained in some partnership dissolution cases where the initial ownership is nearly equal across partners. However, the full efficiency result in partnership dissolution is based on a particular welfare weight that is neutral across types, while the counterpart in our model is valid for a nondegenerate set of welfare weights that may favor one type or another in various manners. In our model, it is trivial that a single market price implements optimality if our design objective is restricted to the neutral welfare weight. Recently, full efficiency is shown to be implementable by Yang et al. (2017) in their endogenous buyer-seller queuing model, where customers can trade their queuing positions.[6] Our model differs from their work in a similar way that ours differs from the partnership dissolution models.

[5]The extension requires that the sets of no-trade types according to an allocation should have nonempty intersection across all ex ante asymmetric partners.

[6]A model of exogenous buyers and sellers of queuing positions has been considered by Yang et al. (2021).

Our design objective, maximizing the integral of agents' interim expected payoffs across all types measured by any welfare weight distribution, is in the spirit of Holmström and Myerson's (1983) notion of interim incentive efficiency. This notion has been considered by a long strand of literature including Wilson (1985), Gresik (1996), Pérez-Nievas (2000), Laussel and Palfrey (2003), Ledyard and Palfrey (1999, 2007) and, recently, ADK (2020), Kang and Zheng (2020), Kang (2020), Reuter and Groh (2020), and DKA (2021). Our focus is the endogeneity of an agent's buyer- or seller-role. This has not been the focus of the literature except Kang and Zheng (2020), where we consider a design problem with finitely many players and without the market clearing condition.[7]

A main perspective of the above literature is that the welfare weight according to which the social planner maximizes the social welfare should be allowed to vary with individuals' types. The importance of this perspective is renewed by recent works on redistributive mechanisms such as ADK (2020), Kang (2020) and DKA (2021), where the social planner's redistributive preferences need not be aligned with the distribution of types across individuals. Moreover, as DKA (2021) have shown recently, even if the social planner is neutral across the fundamental characteristics of individuals, the planner would still be biased for some types against others when the type is not a sufficient statistic of the fundamental characteristics.

Allowing for all continuous welfare weight distributions, our characterization of the optimal mechanisms has the merit of being relatively value-free. Without making the absolute continuity assumption of the welfare weight distribution in the literature (ADK 2020; DKA 2021; Ledyard and Palfrey 1999 and 2007), our model allows for a larger variety of welfare weight distributions.

The next section defines the model. Section 3 observes that the budget balance constraint is never binding in our model. Section 4 characterizes the optimal mechanisms. Section 5 shows which types gain and which types lose when the planner chooses rationing over the posted-price solution, and whose welfare weight is crucial to the planner's choice between the two. Section 6 presents an application to a Covid vaccine distribution problem. Section 7 concludes.

[7]The model of Ledyard and Palfrey (2007) allows for such endogeneity but focuses on other topics.

2 The Model

There is a continuum of individuals, each characterized by a *type*. The type is distributed among the population according to a cdf F with support $[0,1]$ and density f positive and continuous on the support. An individual of type t can produce up to one unit of a good at a marginal cost equal to t, and can acquire up to B units of the good at a marginal utility equal to t, with parameter $B \in \mathbb{R}_{++}$. (The case $B = \infty$ is considered in Appendix J.)

By the revelation principle, a mechanism is modeled as a measurable function $(Q,p) : [0,1] \to [-1,B] \times \mathbb{R}$ such that $Q(t)$ is the *net* position of the good for any individual of type t, and $p(t)$ the expected value of the net money transfer from the individual to others, so that the expected payoff for anyone of type t who acts as type t' is equal to $tQ(t') - p(t')$.[8]

Of particular interest is a kind of payment rules such that $p(t)$ is a piecewise affine function of $Q(t)$ with the same constant term. That is, there exist $c \in \mathbb{R}$, integer n and mutually distinct $k_1, \ldots, k_n \in \mathbb{R}$ such that for any $t \in [0,1]$, $p(t) = c + k_i Q(t)$ for some $i \in \{1, \ldots, n\}$. Given such a payment rule, c is the *lump sum transfer* to all types, and n is the number of *prices*. The payment rule is called *posted price* iff $n = 1$, namely, it offers a constant per-unit price to all types, be they buyers or sellers. If Q is equal to a constant on some nondegenerate interval S of $[0,1]$ and the constant is neither -1 nor B, the mechanism is said to entail *rationing* on S.[9]

By *welfare weight distribution* we mean a cdf W with support $[0,1]$ that is continuous on $\mathbb{R}$. Given any welfare weight distribution W, the design problem is to maximize

$$\int_0^1 (tQ(t) - p(t))\, dW(t) \tag{1}$$

among all mechanisms (Q,p) subject to incentive compatibility (IC) that $tQ(t) - p(t) \geq tQ(t') - p(t')$ for any $t, t' \in [0,1]$, individual rationality (IR) that $tQ(t) - p(t) \geq 0$ for all t, budget balance (BB) that

$$\int_0^1 p(t) dF(t) \geq 0,$$

[8] More explicitly, any individual acting as type t' acquires a quantity $q_1(t')$ of the good, supplies a quantity $q_2(t')$ thereof, and delivers a payment $p(t')$. Consequently, given quasilinear preferences, the payoff to the individual of type t is equal to $(q_1(t') - q_2(t'))\,t - p(t')$, or $tQ(t') - p(t')$ with the notation $Q := q_1 - q_2$.

[9] The term rationing makes sense because, by the envelope formula, the constancy of Q on S implies that $p(t)$ is an affine function of t on S. Thus, the per-unit price is constant while the type (marginal utility) ranges in S. Consequently, almost all types in S would like to either buy or sell up to full capacities, while they are allocated only a constant fraction of the full capacity.

and market clearance that

$$\int_0^1 Q(t)dF(t) = 0.$$

Any solution (Q^*, p^*) to this design problem is called optimal mechanism, and Q^* *optimal allocation*.

Comments First, the welfare weight distribution W reflects the social planner's redistributive preferences across types. It corresponds to the supporting hyperplane at a point on the interim incentive-constrained Pareto frontier, as in the interim incentive efficiency literature initiated by Holmström and Myerson (1983). DKA (2021) interpret the Radon-Nikodym derivative of W (with respect to the type distribution F) at any type t as the expected value of an individual's marginal utility of money (MU_m) conditional on that the marginal rate of substitution (MRS) of the good relative to money is equal to the type t. They show that a social planner whose objective is (1) subject to such welfare weights is equivalent to the planner who is neutral across the underlying individual characteristics that determine the MU_m and MRS. Thus, the planner prefers redistributions from types with low MU_m in expectation ("the rich") to types with high MU_m in expectation (the "poor").

Second, the continuity assumption of the welfare weight distribution W is consistent with the continuum-type (or large-market) model, as each type is supposed to be atomless. The assumption is weaker than its counterpart in the literature. For example, Ledyard and Palfrey (1999, 2007) and DKA (2021) assume absolute continuity of W. Allowing for singular W, we can consider situations where the planner cares only about a measure-zero set of types. Such an example is provided in Section 4.

3 The Budget Balance Condition

As is well-known in the market or mechanism design literature of bilateral trades (Myerson and Satterthwaite 1983; Ledyard and Palfrey 2007; DKA 2021; etc.), the main source of complication in characterizing the optimal mechanisms is a constraint that captures the BB condition (as well as IR and part of IC). In the literature, the possibility that the constraint is binding cannot be ruled out a priori. When it is binding, characterization of the optimal mechanisms depends on endogenous variables and hence in general cannot be described purely in terms of the primitives. In our model, by contrast, the constraint is never binding, as long as the allocation is incentive compatible (namely, weakly increasing) and market clearing:

Theorem 1 *For any weakly increasing allocation $Q : [0,1] \to \mathbb{R}$ that satisfies the market clearing condition, there exists a payment rule $p : [0,1] \to \mathbb{R}$ with which (Q, p) satisfies IR, IC and BB.*

Proved in Appendix A, Theorem 1 is driven by our assumption that each individual is free to choose between buying and selling. To understand the theorem, let us start with a routine in mechanism design that a mechanism (Q, p) is IC if and only if $Q(t)$ is a weakly increasing function of the type t and

$$p(t) = tQ(t) - \int_c^t Q(t')dt' - U(c) \tag{2}$$

for any types $t, c \in [0,1]$ and any $U(c) \in \mathbb{R}$, standing for the truth-telling expected payoff to type c. Let c_1 be the highest seller-type, and c_2 the lowest buyer-type, according to Q. That is, $c_1 := \sup\{t \in [0,1] \mid Q(t) < 0\}$ and $c_2 := \inf\{t \in [0,1] \mid Q(t) > 0\}$. From (2), one readily sees that (Q, p) satisfies IR if and only if $U(c) \geq 0$ for all $c \in [c_1, c_2]$.

Now let us rewind the complication about the BB constraint in the literature, where an individual is exogenously assigned the role of a buyer or that of a seller. As noted previously, we need only to consider any weakly increasing allocation. Define the c_1 and c_2 in this allocation as above. If $c_1 \leq c_2$ then the social planner can easily implement this allocation in a BB and IR manner. For example, she can use the payment rule according to (2) such that $U(c) = 0$ for all $c \in [c_1, c_2]$, which as noted above suffices IR. With this payment rule, the planner squeezes the minimum surplus among all types down to zero, and hence she does not sell the good to any buyer at a price lower than their minimum marginal utility c_2, nor buy the good from any seller at a price higher than their maximum marginal cost c_1. Thus the planner's profit is no less than c_2 times the aggregate demand subtracted by c_1 times the aggregate supply. With market clearance and $c_1 \leq c_2$, this profit is nonnegative and hence the planner's budget is balanced.

The problem, however, is that $c_1 \leq c_2$ cannot be guaranteed when individuals are not free to choose between buying and selling: When the set of buyer-types and that of seller-types are exogenous, an allocation has two functions, Q_1 defined on the buyer-types, and Q_2 defined on the seller-types. The monotonicity condition becomes Q_1 and Q_2 be each weakly increasing, which can be consistent with $c_1 > c_2$. Now that $c_1 \leq c_2$ cannot be guaranteed, the planner in considering a mechanism with $c_1 > c_2$ may need to pay more per unit of procurement than she charges per unit of sales. But that would break her budget unless the planner compromises on some other aspects of the allocation. Thus the BB constraint may be binding at an optimal mechanism. In our model, by contrast, individuals are free to switch between buying and selling, and so the allocation is a single function Q defined on all types. Then $c_1 > c_2$ would violate the monotonicity condition of Q, and the mechanism would fail to be IC: If $c_1 > c_2$, every buyer-type in (c_2, c_1) values the good less than some seller-types in (c_2, c_1), and so such buyers and sellers would rather switch roles. Now that $c_1 \leq c_2$ is guaranteed, the BB constraint is automatically satisfied as in the previous paragraph.

In a nutshell, Theorem 1 comes from the simple fact that, in a market where everyone is free to switch between buying and selling, any buyer's marginal value of the good is higher than any seller's marginal cost of supplying it. The social planner can therefore profit from buying the good from the sellers and selling it to the buyers.

Due to Theorem 1, our design problem is reduced to an optimization among allocations without the IR and BB constraints:

Corollary 1 *A mechanism (Q^*, p^*) is an optimal mechanism if and only if Q^* solves*

$$\begin{array}{ll} \max_Q & \int_0^1 QV dF \\ s.t. & Q : [0,1] \to [-1, B] \text{ is weakly increasing} \\ & \int_0^1 Q dF = 0, \end{array} \tag{3}$$

where $V : [0,1] \to \mathbb{R}$ is the virtual surplus *function defined by, for any $t \in [0,1]$,*

$$V(t) := t - \frac{W(t) - F(t)}{f(t)}. \tag{4}$$

To prove the corollary, use the routine of envelope theorem and integration by parts to show (Lemma 1, Appendix B) that a mechanism (Q^*, p^*) is an optimal mechanism (a solution to the design problem defined in Section 2) if and only if Q^* maximizes $\int_0^1 Q(t)V(t)dF(t)$ among all weakly increasing allocations $Q : [0,1] \to [-1, B]$ subject to two conditions: (i) market clears ($\int_0^1 Q dF = 0$), and (ii) there exists a payment rule that implements Q with respect to the IR and BB constraints. Condition (ii), by Theorem 1, is guaranteed by Condition (i) and the monotonicity of the allocation. Thus, the maximization problem is equivalent to Problem (3), and hence the corollary follows.

4 Optimal Mechanisms

Problem (3), with a harmless change of variables, is essentially the same as the single-market monopoly problem considered by Bulow and Roberts (1989), our market clearing condition being the counterpart to their capacity constraint. As has been understood in the literature, the virtual surplus $V(t)$ corresponds to the monopolist's marginal revenue extracted from type-t individuals. The problem can be solved by the standard ironing method.

If a single price τ per unit is offered to all individuals without quantity restrictions, so that every type above τ would buy B units of the good, and every type below τ would sell one unit thereof, then the market clearing condition $\int_0^1 Q dF = 0$ is satisfied iff $\tau = F^{-1}\left(\frac{B}{B+1}\right)$. Thus we call $F^{-1}\left(\frac{B}{B+1}\right)$ *market clearing price*. With the marginal revenue interpretation of $V(t)$, it is clear that a posted price equal to the market clearing price attains the optimality of (3) if the marginal revenue of any type below the market clearing price is no higher than the marginal revenue of any type above the market clearing price. In other words, the posted price is optimal if $V(\cdot) - V\left(F^{-1}\left(\frac{B}{B+1}\right)\right)$ is single-crossing on $[0, 1]$.

Without the single-crossing condition, there may be a type below the market clearing price that contributes a larger marginal revenue than some type above the price does. To exploit the larger marginal revenues of such lower types without violating the monotonicity (IC) condition of the allocation, the planner needs to find an appropriate interval $[a, b]$ that contains the market clearing price and treat the types in $[a, b]$ equally. As long as the average marginal revenue in $[a, b]$ is not less than the average marginal revenue in $[0, a)$, and not greater than that in $(b, 1]$, the planner can attain optimality through stratifying the types into at most three tiers:[10] Types in $[0, a)$ sell and types in $(b, 1]$ buy, each in full capacity, while types in (a, b) are rationed a constant quantity that clears the market. This characterization is formalized by the next theorem, proved in Appendix D.

Theorem 2 *(i) There exists an optimal mechanism consisting of an allocation*

$$Q^*(t) := \begin{cases} -1 & \text{if } 0 \le t < a \\ \frac{F(a) - B(1-F(b))}{F(b) - F(a)} & \text{if } a < t < b \\ B & \text{if } b < t \le 1, \end{cases} \tag{5}$$

[10]By a *tier* in an incentive compatible (and hence monotone) allocation $Q : [0, 1] \to \mathbb{R}$, we mean the inverse image $Q^{-1}(s)$ of some s in the range of Q such that $Q^{-1}(s)$ is a nondegenerate interval.

where $0 \leq a \leq F^{-1}\left(\frac{B}{B+1}\right) \leq b \leq 1$, *and a payment rule that has at most two prices. (ii) If the function* $V(\cdot) - V\left(F^{-1}\left(\frac{B}{B+1}\right)\right)$ *is single-crossing on* $[0, 1]$, *then*

$$Q^*(t) = \begin{cases} -1 & \text{if } 0 \leq t < F^{-1}\left(\frac{B}{B+1}\right) \\ B & \text{if } F^{-1}\left(\frac{B}{B+1}\right) < t \leq 1, \end{cases} \tag{6}$$

and the payment rule becomes a posted price equal to $F^{-1}\left(\frac{B}{B+1}\right)$ *without rationing or lump sum rebate.*

Theorem 2 implies that an optimal allocation exists and it is a tiered allocation consisting of at most three tiers. Moreover, when the single crossing condition of $V - V\left(F^{-1}\left(\frac{B}{B+1}\right)\right)$ is satisfied, the optimal allocation has only two tiers and is implemented by offering the market clearing price $F^{-1}\left(\frac{B}{B+1}\right)$ to everyone. Allocating the full capacity (-1 or B) to each buyer- or seller-type, the optimal allocation does not entail rationing. It is easy to check that the posted price yields zero profit for the planner (since the allocation satisfies the market clearing condition), and hence the optimal mechanism has no lump sum rebate. Note that the single-crossing condition can be satisfied by even non-monotone virtual surplus functions.

For the three-tier allocation (5), the interval (a, b) can be constructed from the primitives with the definition of ironing. As shown in the proof (Appendix D), when (6) is not optimal, the interval (a, b) for (5) to be optimal contains the market clearing price $F^{-1}\left(\frac{B}{B+1}\right)$ as an interior point. From the envelope formula one can derive the optimal payment rule that implements (5), described by the next corollary (proved in Appendix E).

Corollary 2 *For any* $a, b \in [0, 1]$ *such that* $a < F^{-1}\left(\frac{B}{B+1}\right) < b$, *any optimal mechanism that implements allocation (5) is equivalent to a mechanism that transfers a positive lump sum to every type, entails rationing only on* (a, b) *with the rationed quantity*

$$x := \frac{F(a) - B\left(1 - F(b)\right)}{F(b) - F(a)}, \tag{7}$$

and uses the following payment schedule p *for each type* t *to deliver a payment* $p(t)$ *(in addition to receiving the lump sum transfer):*

i. if $x \geq 0$,

$$p(t) = \begin{cases} -a & \text{if } t \in [0, a) \\ ax & \text{if } t \in [a, b] \\ bB - (b-a)x & \text{if } t \in (b, 1]; \end{cases}$$

ii. if $x \leq 0$,

$$p(t) = \begin{cases} -a + (b-a)x & \text{if } t \in [0,a) \\ bx & \text{if } t \in [a,b] \\ bB & \text{if } t \in (b,1]. \end{cases}$$

As is implied by the corollary, an optimal mechanism implementing the three-tier allocation (5) makes a positive lump sum transfer to all individuals, entails rationing only on the middle tier, and uses only two distinct prices. When the rationed quantity x is positive, the price is equal to either a, for the low and middle tiers to sell or buy the good, or equal to $b - (b-a)x/B$, for the high tier to buy the good (up to the quantity B). When x is negative, the price is equal to either b, for the middle and high tiers to sell or buy the good, or equal to $a - (b-a)x$, for the low tier to sell the good (up to the entire one unit).

The next theorem (proved in Appendix F) says that the optimal allocation characterized above is unique given a nondegenerate set of parameter values. Thus, not only is there no need to stratify the types into more than three tiers, often it is also suboptimal to do so. Combined with the previous theorem, Theorem 3 implies that, even when the virtual surplus function is non-monotone, there still exists a unique optimal allocation and it is implemented by a single posted price alone, requiring neither rationing nor lump sum transfers.

Theorem 3 *If there exists no positive-measure subset S of $[0,1]$ such that V is constant on S, the optimal allocation is unique (modulo measure zero).*

To see the role played by the non-constancy assumption of V, consider a case where the rationed interval (a,b) in (5) is a proper subset of another interval (a',b') in $[0,1]$ such that V restricted on $(a',b') \setminus (a,b)$ happens to be constantly equal to the average marginal revenue on (a,b). Then there may be a continuum of optimal allocations: Pick any (a'',b'') for which $a' \leq a'' \leq a < b \leq b'' \leq b'$, extend the rationed interval from (a,b) to (a'',b''), and vary the allocation for the types in $(a',a'') \cup (b'',b')$ in whatever fashion that satisfies the monotonicity and market clearing conditions. The allocation thereby obtained is optimal because it is the average marginal revenue within a set of types that determines how much the planner should prioritize the set (cf. the proof of Lemma 2, Appendix C). Such multiplicity of optimal allocations is ruled out by the non-constancy assumption in the theorem.

Whether there can be multiple optimal allocations or not, any optimal allocation requires stratifying the type space into at least two tiers:

Corollary 3 *Egalitarian allocations ($Q = 0$ a.e. $[0, 1]$, or autarky) are never optimal.*

The proof of the corollary (Appendix E) uses an observation in the proof of Theorem 2. The intuition is simply that there is always a gain of trade between the sufficiently low types and the sufficiently high ones. When a type t near zero supplies a unit of the good, the cost to the society is $V(t) \approx 0$. When a type t' near the supremum type acquires a unit of the good, the social benefit is $V(t') \approx 1$.

Example: The Cantor Welfare Weight Distribution Suppose that the type distribution F is the uniform distribution $\mathrm{U}[0, 1]$ on $[0, 1]$, and that the welfare weight distribution W is the Cantor-Lebesgue function φ, so the support of the distribution is the (ternary) Cantor set.[11] By the well-known properties of the Cantor-Lebesgue function, φ is a continuous cdf that assigns positive welfare weights only to the (ternary) Cantor set, which is of zero (Lebesgue) measure, and φ increases at unbounded rates on the Cantor set. Thus the social planner cares only about a set of types of zero measure, and her redistributive preferences cannot be described by welfare *densities* (or the "Pareto weights" in DKA 2021). Plug $W = \varphi$ and $F = \mathrm{U}[0, 1]$ into (4) to obtain the virtual surplus function V:

$$V(t) = 2t - \varphi(t)$$

for all $t \in [0, 1]$. Obviously, V is not monotone, graphed in the left panel of Figure 1.

Nonetheless, our result applies. There are countably many intervals in $[0, 1]$ on which $V = \overline{V}$ because V is single-crossing at any point in those intervals. They are the intervals on which the graph in the right panel of Figure 1 has a positive slope. If the market clearing price $\frac{B}{B+1}$ belongs to any of such positively sloped intervals, the optimal allocation is uniquely the two-tier stratification with $\frac{B}{B+1}$ being the buyer-seller cutoff. Else, $\frac{B}{B+1}$ is interior to an interval where V needs to be ironed, and a three-tier allocation is optimal. Furthermore, this is the unique optimal allocation by Theorem 3, as the inverse image $V^{-1}(x)$ is of zero measure for any x in the range of V.

Although almost all types carry zero welfare weight according to the Cantor-Lebesgue function, the optimal mechanism gives positive surpluses to almost all types. That is because the IC condition requires that the surplus for any type be at least as large as the surplus for type $\frac{B}{B+1}$, the buyer-seller cutoff.

[11]See Royden and Fitzpatrick (2010, Section 2.7) for the definition and properties of the Cantor set and the Cantor-Lebesgue function.

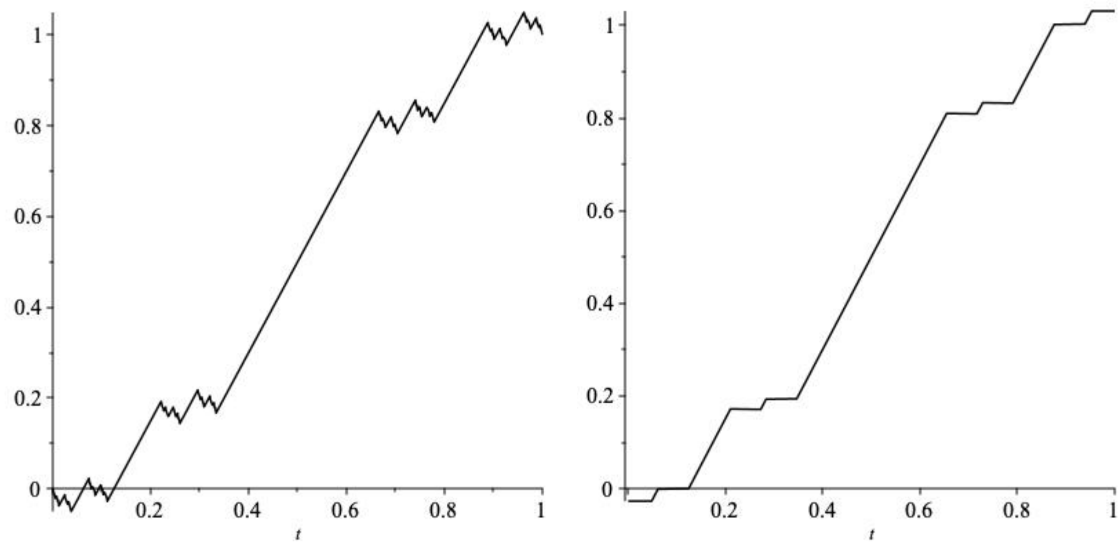

Figure 1: The virtual surplus (left panel) and its ironed copy (right panel) given the Cantor welfare weights and uniform distribution of types

5 Posted Price versus Rationing

The previous section shows that any optimal mechanisms—supported as a Pareto frontier point by some continuous welfare weight distribution—can be simplified to one of only two alternatives: It is either the posted-price system, implementing the two-tier allocation (6), or a rationing system that implements a three-tier allocation (5) and entails rationing on the middle tier. This section shows who gains, and who loses, when the mechanism switches from one kind to the other, each being Pareto optimal. We shall also see whose welfare weight plays a crucial role in the social planner's choice between the two alternatives.

5.1 Who Gains and Who Loses from Rationing

It is intuitive that the types near the market clearing price $F^{-1}\left(\frac{B}{B+1}\right)$ gain when the mechanism switches from the posted price to rationing. Given the posted price, which is equal to $F^{-1}\left(\frac{B}{B+1}\right)$, the type $F^{-1}\left(\frac{B}{B+1}\right)$ gets zero net payoff whether it buys or sells the good, as its valuation of the good is equal to the type. The type has no other source of surplus because the posted-price system, essentially a competitive equilibrium, yields no profit for the planner to rebate to the individuals. Given rationing, by contrast, the type has at least a positive lump sum rebate as part of its surplus. The lump sum is positive because the planner gets a positive profit from rationing on the middle tier (a,b) that contains the market clearing price $F^{-1}\left(\frac{B}{B+1}\right)$ as an interior point: The planner exploits her monopsony power in squeezing the range of her full-capacity procurement from $\left[0, F^{-1}\left(\frac{B}{B+1}\right)\right)$ to $[0,a)$, and monopoly power in squeezing the range of her full-capacity sales from $\left[F^{-1}\left(\frac{B}{B+1}\right), 1\right]$ to $(b,1]$. Thus, the type $F^{-1}\left(\frac{B}{B+1}\right)$ gains strictly when the mechanism changes from the posted price to rationing. By continuity, so do the nearby types.

The more complicated question is which types get hurt in order for such middle types to gain. While the general answer may depend on the parameter values, we can tell whether the high or the low types are definitely worse-off based on the curvature of the type distribution F around the market clearing price. According to the next theorem, if the distribution F of types is convex on the middle tier in a rationing mechanism, the low types—those who get to sell at full capacity in both mechanisms—are definitely worse-off when the posted price is replaced by the rationing mechanism: In Figure 2, on the set $[0,a)$ of low types, the red dotted line—the surplus given rationing—lies below the blue solid line—the surplus given the posted price. If F is concave on the middle tier, by contrast, the high types—those who get to buy at full capacity in both mechanisms—are definitely worse-off: In Figure 3, the red dotted line lies below the blue solid line on $(b,1]$, the set of high types.

Theorem 4 *If the allocation in an optimal mechanism switches from the two-tier (6) to a three-tier (5) that entails rationing on some (a,b) for which $0 < a < F^{-1}\left(\frac{B}{B+1}\right) < b < 1$, then:*

a. *all the types sufficiently near to the market clearing price $F^{-1}\left(\frac{B}{B+1}\right)$ are better-off;*

b. *if F is convex on (a,b), any type in $[0,a)$ is worse-off;*

c. *if F is concave on (a,b), any type in $(b,1]$ is worse-off.*

Theorem 4 is proved in Appendix G. To understand the less intuitive parts, Claims (b) and (c), let us consider a stochastic counterpart to the rationing allocation (5): For each individual, the allocation is independently and randomly selected so that it is

$$Q_a(t) := \begin{cases} -1 & \text{if } t \in [0,a] \\ B & \text{if } t \in (a,1] \end{cases}$$

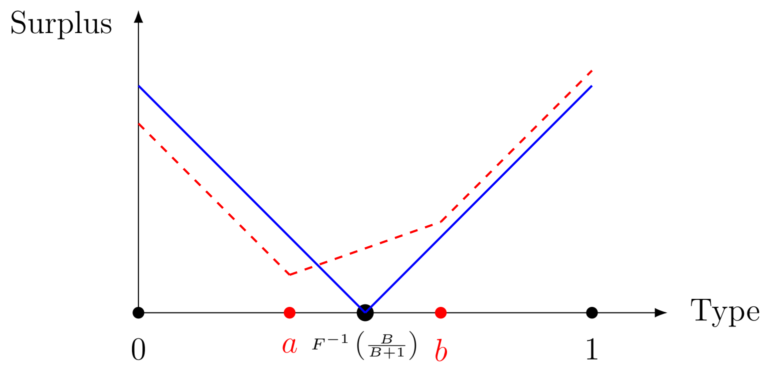

Figure 2: Posted-price (blue solid) vs. rationing (red dotted) given convex F

with probability $(1+x)/(B+1)$, and

$$Q_b(t) := \begin{cases} -1 & \text{if } t \in [0,b] \\ B & \text{if } t \in (b,1] \end{cases}$$

with probability $1-(1+x)/(B+1)$, where x is determined by (7). By the choice of x, $F(a)(1+x)/(B+1)+F(b)\left(1-(1+x)/(B+1)\right) = B/(B+1)$. That is, from the ex ante or the social planner's viewpoint, the expected quantity to procure from the individual is equal to $B/(B+1)$. By the same token, the expected quantity to sell to the individual is equal to $B/(B+1)$. Thus, when the same lottery is run independently for all individuals, supply is equal to demand at the aggregate level. Note that from each (privately informed) individual's viewpoint, the stochastic allocation is equivalent to the rationing allocation (5).

The stochastic allocation can be implemented by the corresponding stochastic payment rule: To each individual, if the lottery picks Q_a then offer him a per-unit price equal to a for the individual to buy or sell the good in full capacity; if the lottery picks Q_b then analogously offer him the price b per unit for buying and selling. This stochastic payment rule generates a negative expected profit for the social planner: When the lottery draws Q_a, the planner sells the good to balance the excess demand $B(1-F(a))-F(a)$ (which is positive because $B/(B+1) > F(a)$), and the average revenue she gets according to the payment rule is equal to a. When the lottery draws Q_b, the planner buys the good to balance the excess supply $F(b)-B(1-F(b))$ (>0 because $B/(B+1) < F(b)$), and the average price she pays is equal to b. Since $b > a$ and the market clears, the planner is losing profit.

Thus, under any optimal mechanism that implements the stochastic allocation, which is required to be budget balanced, the payment rule differs from the stochastic payment

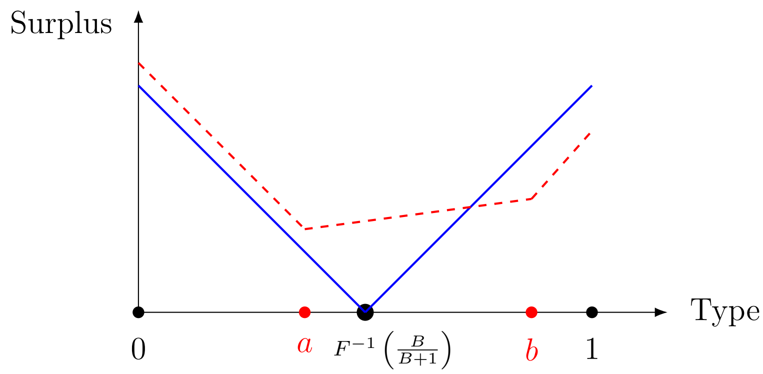

Figure 3: Posted-price (blue solid) vs. rationing (red dotted) given concave F

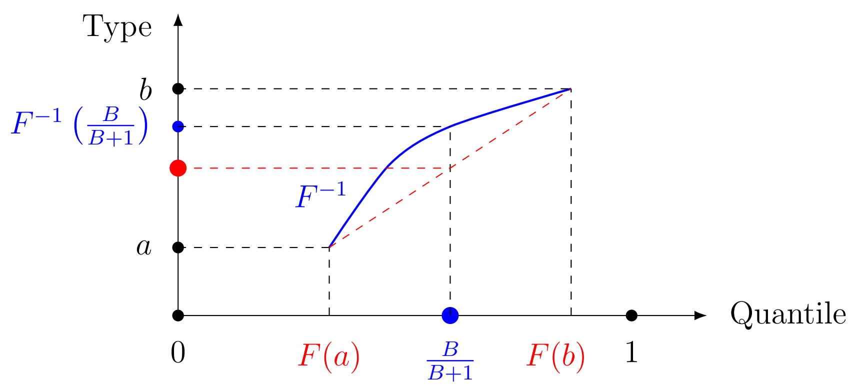

Figure 4: The red dot: Expected revenue upper bound for low types

rule, in expectation, only by a lump sum transfer to the individuals that is a nonpositive constant across types. Consequently, each individual's surplus under any optimal mechanism of the stochastic allocation, or equivalently the rationing allocation, is bounded from above by the surplus he receives from the stochastic payment rule, as the latter has yet to count the nonpositive lump sum transfer. That is, in any optimal mechanism that implements the rationing allocation Q^*, the surplus for any type $t \in [0, a)$ is bounded from above by the convex combination between a and b—the revenue received by type t in the stochastic payment rule—according to the probability mix in the lottery, plus $tQ^*(t)$. This convex combination is labeled by the red dot in Figure 4. As shown in that figure, when F is convex on (a, b) and hence F^{-1} concave on $(F(a), F(b))$, the convex combination is less than $F^{-1}\left(\frac{B}{B+1}\right)$. Since $F^{-1}\left(\frac{B}{B+1}\right)$ is the revenue received by any type in $[0, a)$ under the deterministic posted-price system without rationing, the type receives less expected revenue, and hence less surplus, in the rationing mechanism than in the posted-price system, as claimed in Part (b) of the theorem. Part (c) of the theorem is analogous from the perspective of the high types in $(b, 1]$.

Intuitively speaking, if F is the uniform distribution, the market clearing cutoff value $F^{-1}\left(\frac{B}{B+1}\right)$ would be equal to the quantile $B/(B+1)$. Now make F convex around the cutoff value. That means moving a mass of types from below the cutoff to above the cutoff. Consequently, the cutoff value $F^{-1}\left(\frac{B}{B+1}\right)$, whose quantile is supposed to be $B/(B+1)$, needs to be adjusted upward. Thus, the convexity of F enlarges the revenue received by the low types if there is no rationing. Rationing partially removes this advantage by averaging the revenue across types. That is the intuition of Part (b) of Theorem 4.

5.2 The Welfare Weight of the Market Clearing Cutoff Type

Since the middle types around the market clearing price gain from rationing (Part (a) of Theorem 4), it is natural that the heavier is the welfare weight on such middle types, the more is the social planner leaning towards a rationing mechanism. To formalize that, for any welfare weight distributions W and W_*, let us say—

- W_* is a *spread* of W away from $F^{-1}\left(\frac{B}{B+1}\right)$ iff $W_* \geq W$ on $\left[0, F^{-1}\left(\frac{B}{B+1}\right)\right)$ and $W_* \leq W$ on $\left(F^{-1}\left(\frac{B}{B+1}\right), 1\right]$;
- W_* is a *contraction* of W towards $F^{-1}\left(\frac{B}{B+1}\right)$ iff $W_* \leq W$ on $\left[0, F^{-1}\left(\frac{B}{B+1}\right)\right)$ and $W_* \geq W$ on $\left(F^{-1}\left(\frac{B}{B+1}\right), 1\right]$.

Intuitively speaking, a spread away from $F^{-1}\left(\frac{B}{B+1}\right)$ moves some welfare weights around the market clearing cutoff type to the higher and lower types, and a contraction towards $F^{-1}\left(\frac{B}{B+1}\right)$ does the opposite. The next theorem shows that the two operations have opposite effects on the optimality of the posted-price system.

Theorem 5 *Suppose that the posted-price system is optimal given a welfare weight distribution W. Then:*

a. if W_* is a spread of W away from $F^{-1}\left(\frac{B}{B+1}\right)$, then the posted-price system is optimal when the welfare weight distribution is W_* instead of W;

b. for any $\epsilon > 0$ there exists a contraction W_* of W towards $F^{-1}\left(\frac{B}{B+1}\right)$ such that $\|W_* - W\|_{\max} \leq \epsilon$ and the posted-price system is not optimal when the welfare weight distribution is W_* instead of W.

Theorem 5 is proved in Appendix H. Its intuition, as mentioned above, has been suggested by Part (a) of Theorem 4. We can get a more explicit intuition by adopting DKA's (2021) rich-vs-poor interpretation of the welfare weight distribution. According to them, the density of the welfare weight distribution at a type t corresponds to the average marginal utility of money among the individuals whose marginal rate of substitution of the good relative to money is equal to t (cf. Section 2). When W spreads the weight away from $F^{-1}\left(\frac{B}{B+1}\right)$, the types near $F^{-1}\left(\frac{B}{B+1}\right)$ are having lower marginal utilities of money in average and hence there is less a need for transferring money to such types through deviating from the zero-rebate posted-price system. When W contracts the weight towards $F^{-1}\left(\frac{B}{B+1}\right)$, by contrast, the nearby types value money more in average, which strengthens the need to transfer money to them through moving away from the posted-price system.

6 Application: Vaccine Allocation with Externalities

Our model is equivalent to the following exchange economy up to normalization. Every (atomless) individual is endowed with one unit of the good, individuals can sell any fraction of their endowments for money, and each can consume up to $B + 1$ units of the good. The good can be interpreted as the access to a limited public resource that everyone is equally entitled to. Individuals' types are their willingness to pay for the access to the resource. The social planner's welfare weight distribution W need not be aligned with the distribution F of the willingness to pay. Our model then applies and the planner's optimality can be achieved by a market-like mechanism where individuals trade their shares of the public resource given a menu containing at most three price-quantity contracts.

To be explicit, let us apply the idea to the allocation of Covid vaccines. While a social planner often has explicit preferences over who should receive the vaccines before others and hence might want to prioritize vaccine allocation across groups (Akbarpour et al. 2021; Sömet et al. 2021), it has often been reported that individuals of the same priority level (e.g., healthcare workers) are heterogeneous in their vaccine willingness or hesitancy. Thus, let us focus on the issue about the limited supply of Covid vaccines on one hand and the heterogeneous willingness to vaccination within the same priority group on the other.[12]

To capture this issue with a stylized model, normalize to one the measure of the population in a priority group, and suppose that the quantity of Covid vaccines available to this population is equal to $\alpha \in (0, 1)$. (In other words, only up to a fraction α of the population gets to be inoculated.) Suppose that an individual's willingness to get vaccinated is represented by a type $t \in [0, 1]$, drawn from a cdf F. Let $\psi : [0, 1] \to \mathbb{R}$ be a measurable function. Assume that, if q is the probability for an individual to get vaccinated, m his net monetary receipt, and each individual of type t' is allocated a probability $\tilde{q}(t')$ of getting vaccinated, then the type-t individual's gross payoff is equal to

$$qt + m + \psi(t) \int_0^1 \tilde{q}(t')dF(t'). \tag{8}$$

In (8), the term $\psi(t) \int_0^1 \tilde{q}(t')dF(t')$ captures a kind of externalities spilled over to the type-t individual: When an individual gets vaccinated with probability $\tilde{q}(t')$ (through acting as type t'), the externality spilled over to any other individual of type t, for any $t \in [0, 1]$, is equal to $\psi(t)\tilde{q}(t')$. Given truthtelling, the aggregate of the externalities spilled over to a type-t individual is therefore equal to $\psi(t) \int_0^1 \tilde{q}(t')dF(t')$. For instance, a strictly increasing ψ corresponds to situations where the more one is willing to get himself inoculated the more strongly he believes that his health is affected by the size of the vaccinated population. However, (8) does not apply to situations where an individual's evaluation of the externality spilled over to him depends on some other personal characteristics in addition to his willingness to pay.[13]

While externality is absent in our main model, changing the utility definition to (8) does not alter any of our results. That is because the aggregate externality spilled over to an individual of type t, given allocation $\tilde{q}$, is equal to $\psi(t) \int_0^1 \tilde{q}(t')dF(t')$, which is constant to the individual's own action. Thus, the design constraints remain the same as in the original model. Furthermore, mimicking the argument for Corollary 1, one can prove (Appendix I) that the planner's problem is exactly the same as (3):

[12] Akbarpour et al. (2021) consider both the issue of within-group heterogeneity and the planner's cross-group preferences in vaccine distributions. We consider only the within-group heterogeneity issue to focus on how it can be solved by simple market mechanisms once the quantity of vaccines available to a group has been determined.

[13] See Akbarpour et al. (2021) for a model that applies to such a situation.

Corollary 4 *In the modified model with externalities such that the utility for any individual of type t is defined by (8), an allocation is optimal if and only if it solves Problem (3).*

Since we consider the allocation within the same priority group, at the outset everyone is entitled to an equal access to vaccination. That is, any individual is entitled to a probability α of getting inoculated. Implicitly, each member of the population is initially issued a coupon, so that one coupon gives a person the probability α of getting vaccinated.

In the real world, Covid vaccines are often distributed to the individuals of the same priority group in an egalitarian manner without the possibility for individuals to trade their entitlements among themselves. That is never socially optimal, as we have observed in Corollary 3. Alternatively, a government should allow individuals to trade their vaccine entitlements, say, by issuing digital coupons to individuals, one unit to each, to represent their initial entitlements to inoculation. Then individuals can trade any fraction of their coupons so that anyone who holds a quantity q of coupons gets to be inoculated with a probability equal to $q\alpha$. To get vaccinated for sure, a person needs only to hold a quantity $1/\alpha$ of coupons. Thus, the quantity of coupons that a person needs to acquire, in addition to the one unit the person is endowed with, does not need to exceed $1/\alpha - 1$. That is,

$$B = \frac{1}{\alpha} - 1.$$

A mechanism can be represented by (Q, P) such that $Q(t)$ is the quantity of coupons that an individual acting as type t buys from others (which means a probability $(Q(t)+1)\alpha$ of getting vaccinated), and delivers a net payment $P(t)$.[14] By (8), a type-t individual's expected gross payoff from claiming to be type t' is equal to

$$\begin{aligned} & (Q(t')+1)\,\alpha t - P(t') + \psi(t)\int_0^1 \alpha\,(Q(s)+1)\,dF(s) \\ = \ & \alpha \underbrace{\left(Q(t')t - \frac{1}{\alpha}P(t')\right)}_{u(t',t|Q,P)} + \alpha t + \alpha\psi(t) + \alpha\int_0^1 \psi(t)Q(s)dF(s). \end{aligned}$$

Thus, any type t's decision is equivalent to maximizing the expected net payoff $u(t', t|Q, P)$ among all t'. This, coupled with Corollary 4, implies that the social planner's design objective is equivalent to $\int_0^1 u(t, t|Q, P)dW(t)$ for some welfare weight distribution W. Thus our result

[14]Negative $Q(t)$ means selling the corresponding quantity, and negative $P(t)$ means being paid.

applies and, if it entails (Q^*, p^*) as the mechanism, the planner would construct the payment rule P^* by $P^* := \alpha p^*$.

If $W = F$, namely, the social planner is neutral across types, then $V(t) = t$ for all t and the optimal mechanism is a single posted price equal to the market clearing cutoff $F^{-1}\left(\frac{B}{B+1}\right)$ $(= F^{-1}(1-\alpha))$. That is, if the planner has no redistribution bias across types, she would opt for the free-market solution to distribute vaccines within the same priority group.

In reality, however, the planner often puts heavier welfare weights on some types than on others. For instance, consider a situation where individuals' types represent their Covid comorbidities and that the type-distribution F is concave (say, due to the prevalence of the Omicron variant, more severe comorbidity occurs less frequently). Meanwhile, suppose that the social planner puts heavier weights on types of more severe comorbidities, because severe comorbidities entail heavy social costs for healthcare, so much so that the welfare weight distribution W is convex.[15] Thus, F is concave, and W convex. This, coupled with a technical condition that the Radon-Nikodym derivative w of W is bounded from above by 2, implies that the virtual surplus function is increasing and so the social planner would stay with the free-market solution (Theorem 2.ii). Even without the technical condition of w and the global concavity of F, as long as F is concave around the market clearing cutoff type, Part (c) of Theorem 4 would still imply that the planner is unlikely to forgo the free-market solution, because any other mechanism, entailing rationing, would hurt the types of severe comorbidities that she cares about.

Even when the free-market solution is not optimal, the planner can still achieve optimality through a market-like mechanism that uses at most two distinct prices and entails rationing on only one tier among the types (Theorem 2 and Corollary 2). For example, when the rationed quantity x in the optimal allocation (5) is positive, the planner can set the price per coupon to be equal to a for those who want to sell their coupons, and offer to those who want to acquire coupons a menu of two options, one to buy B coupons at the unit price equal to $b - x(b-a)/B$ (thereby getting vaccinated for sure), the other to buy x coupons at the unit price a (thereby getting vaccinated with probability $(x+1)\alpha$).

Contrary to the vaccine wastefulness problem of the mechanisms in current practice, none of the optimal allocations prescribed above leaves any vaccine unused. That is due to the market clearing condition satisfied by the optimal allocations. The bottom line is: A

[15] We thank the associate editor for suggesting comorbidity as a direction to interpret the welfare weights.

market-like Covid vaccine distribution mechanism, which sets at most two prices for vaccine entitlements to stratify the population of the same priority group into at most three tiers and uses rationing to at most one of the three, would outperform the current within-group egalitarian rationing mechanism.

7 Conclusion

It is common that individuals start on an equal footing and end with different outcomes, just because of the idiosyncrasy in one's ability, taste or pure luck. It is also common that such inequality in the outcomes, like it or not, is often class-oriented, grouping individuals into several tiers and treating the members of each roughly indiscriminately. This paper provides a mathematical fable for such stratification. It says that, even in the idealized situation where the society is framed in an interim incentive-constrained Pareto optimal manner, stratification is still unavoidable and, in fact, necessary for the social wellbeing. Meanwhile, our finding implies that stratification of more than three tiers is unnecessary, and often suboptimal. Consequently, while the people should be stratified into at least two tiers and, due to the market clearing condition, there should be at least one tier for the haves and another for the have-nots, oftentimes there should not be more than two subdivisions in either category. Thus, while the bisection of the rich into East Egg and West Egg, under the penetrating pen of F. Scott Fitzgerald, may be understood as part of a three-tier optimal allocation, any further subdivision of either Egg is likely suboptimal.

An insight from this study is that, in a large market where individuals are free to choose between buying and selling, a single competitive market price—without the help from any other instruments such as rationing, redistribution or tier-specific prices—is often capable of implementing interim Pareto optimality despite the presence of asymmetric information. It should be emphasized that such robustness of the competitive market is not an artifice of any specific social welfare criterion say a pro-market value system; but rather it holds true for a wide variety of welfare weight distributions that may favor one type or another, as long as the welfare weight is not overly contracted towards the market clearing cutoff type. From such a relatively value-free perspective, one could understand the institutional evolution in the United States regarding the allocation of the radio frequency spectrum, from hearings and lotteries to market-like auctions, as a movement towards the Pareto frontier that should still have happened even if the policymaker's objective is something other than to raise revenues or to develop the wireless industries. From the same perspective one could see a robust normative force towards market-oriented solutions to problems of prioritizing citizens for the access to limited resources, be they Covid vaccines or magnet schools.

A Proof of Theorem 1

Let $Q : [0,1] \to \mathbb{R}$ be weakly increasing and market clearing. Denote

$$\begin{aligned} c_1 &:= \sup\{t \in [0,1] \mid Q(t) < 0\}, &(9)\\ c_2 &:= \inf\{t \in [0,1] \mid Q(t) > 0\}. &(10) \end{aligned}$$

Since Q weakly increasing, we have $c_1 \leq c_2$. Furthermore, by the envelope theorem, one can construct a payment rule p that implements Q such that IC is satisfied and the surplus $cQ(c) - p(c)$ is equal to zero for some $c \in [c_1, c_2]$. With $c_1 \leq c_2$ and Q weakly increasing, $tQ(t) - p(t)$ is weakly decreasing on $[0, c]$ and weakly increasing on $[c, 1]$. Thus (Q, p) also satisfies IR. The rest of the proof shows that (Q, p) satisfies BB.

First, we claim that (Q, p) satisfies BB if the following condition holds:

$$t < c_1 \leq c_2 < t' \Longrightarrow \frac{p(t)}{Q(t)} \leq c_1 \leq c_2 \leq \frac{p(t')}{Q(t')}. \tag{11}$$

By the definitions of c_1 and c_2, $Q = 0$ on (c_1, c_2). Thus, $tQ(t) - p(t) = 0$ and $p(t) = 0$ for all $t \in (c_1, c_2)$ by the envelope theorem and $cQ(c) - p(c) = 0$. It follows that

$$\begin{aligned} \int_0^1 p(t)dF(t) &= \int_0^{c_1} p(t)dF(t) + \int_{c_2}^1 p(t)dF(t) \\ &= \int_0^{c_1} \frac{p(t)}{Q(t)}Q(t)dF(t) + \int_{c_2}^1 \frac{p(t)}{Q(t)}Q(t)dF(t) \\ &\geq \int_0^{c_1} c_1Q(t)dF(t) + \int_{c_2}^1 c_2Q(t)dF(t) \\ &\geq c_1\left(\int_0^{c_1} Q(t)dF(t) + \int_{c_2}^1 Q(t)dF(t)\right) \\ &= 0, \end{aligned}$$

where the third line is due to (11) and the fact that $Q < 0$ on $[0, c_1)$ and $Q > 0$ on $(c_2, 1]$, the fourth line due to $c_2 \geq c_1$ and $Q > 0$ on $(c_2, 1]$, and the last line due to market clearing. Thus, (Q, p) satisfies BB if (11) holds.

To prove (11), Pick any $t \in [0, c_1)$, IC implies

$$0 = c_1Q(c_1) - p(c_1) \geq c_1Q(t) - p(t),$$

where the equality follows from extending the equation $tQ(t) - p(t) = 0$, which we have proved for all $t \in (c_1, c_2)$, to the boundary point c_1 by continuity of the surplus function.

The formula displayed above implies $p(t)/Q(t) \leq c_1$ because $Q(t) < 0$ for all $t \in [0, c_1)$. Analogously, for any $t \in (c_2, 1]$, IC implies $0 = c_2 Q(c_2) - p(c_2) \geq c_2 Q(t) - p(t)$, which implies $p(t)/Q(t) \geq c_2$ since $Q(t) > 0$ for all $t \in (c_2, 1]$. Thus (11) is true, as desired.

B Derivation of the Virtual Surplus Function

While various forms of the routine have appeared numerously in the literature, it is helpful to formalize it so as to clarify the role of the constraints.

Lemma 1 *For any nonempty subset S of $\mathbb{R}$, the problem*

$$\begin{array}{ll} \max_{(Q,p)} & \int_0^1 (tQ(t) - p(t))\, dW(t) \\ s.t. & Q : [0,1] \to S \text{ is weakly increasing} \\ & p(t') - p(t) = \int_t^{t'} s dQ(s) \quad (\forall t, t' \in [0,1]) \\ & \int_0^1 p dF \geq 0 \end{array} \tag{12}$$

is equivalent to

$$\begin{array}{ll} \max_{(Q,p)} & \int_0^1 Q(t) V(t) dF(t) \\ s.t. & Q : [0,1] \to S \text{ is weakly increasing} \\ & p(t') - p(t) = \int_t^{t'} s dQ(s) \quad (\forall t, t' \in [0,1]) \\ & \int_0^1 p dF = 0. \end{array} \tag{13}$$

Proof First, there is no loss of generality to replace the constraint $\int_0^1 p dF \geq 0$ in (12) by $\int_0^1 p dF = 0$: If $\int_0^1 p dF > 0$, we can modify the payment rule by rebating the positive money surplus $\int_0^1 p dF$ back to the types uniformly. That enlarges $tQ(t) - p(t)$ for all $t \in [0,1]$ and hence enlarges the objective $\int_0^1 (tQ(t) - p(t))\, dW(t)$ because the distribution W assigns a positive measure on $[0,1]$. Thus, in any optimum, $\int_0^1 p dF > 0$ does not hold.

Second, with $\int_0^1 p dF = 0$, we show that the objective in (12) is equal to that in (13).[16] Denote $U(t) := tQ(t) - p(t)$ for all t. By the envelope theorem, $dU(t) = Q(t) dt$. This coupled with integration-by-parts gives

$$\int_0^1 U dF = U(1) - \int_0^1 F dU = U(1) - \int_0^1 F(t) Q(t) dt.$$

[16] We thank the associate editor for suggesting the following short proof. Our previous proof is longer and suitable to asymmetric models where individuals' types are drawn from different distributions.

Likewise,

$$\int_0^1 U dW = U(1) - \int_0^1 W(t)Q(t)dt.$$

Plug the expression of $U(1)$ from the former equation into the latter equation to obtain

$$\begin{aligned}\int_0^1 U dW &= \int_0^1 U dF + \int_0^1 F(t)Q(t)dt - \int_0^1 W(t)Q(t)dt \\ &= \int_0^1 tQ(t)dF(t) - \int_0^1 p(t)dF(t) - \int_0^1 \frac{W(t)-F(t)}{f(t)}Q(t)dF(t) \\ &= \int_0^1 Q(t)V(t)dF(t),\end{aligned}$$

with the last equality due to $\int_0^1 pdF = 0$ and the definition of V. ■

C Lemmas of Ironing

Define for each $s \in [0,1]$

$$H(V)(s) := \int_0^s V\left(F^{-1}(r)\right) dr. \tag{14}$$

Denote $\widetilde{H}(V)$ for the convex hull of $H(V)$ on $[0,1]$ (cf. Myerson 1981). Then the *ironed virtual surplus* $\overline{V} : [0,1] \to \mathbb{R}$ is defined by

$$\overline{V}(t) = \frac{d}{ds}\left(\widetilde{H}(V)\right)(s)\Big|_{s=F(t)} \tag{15}$$

whenever $\widetilde{H}(V)$ is differentiable at $F(t)$, and extended to all of $[0,1]$ by one-sided continuity.

If $0 \le a < b \le 1$, (a,b) is called *ironed interval* iff $\widetilde{H}(V) < H(V)$ on $(F(a), F(b))$, $\widetilde{H}(V)(F(a)) = H(V)(F(a))$ and $\widetilde{H}(V)(F(b)) = H(V)(F(b))$. That is, an ironed interval is an inclusion-maximal open interval on which $H(V)(F(\cdot)) > \widetilde{H}(V)(F(\cdot))$. As is well-known, on an ironed interval the monotonicity condition of Q is binding, and $\overline{V}$ is constant.[17]

Lemma 2 *For any $a, b \in [0,1]$ such that $a \le F^{-1}\left(\frac{B}{B+1}\right) \le b$, if $\overline{V}$ is constant on (a,b) (unless $(a,b) = \varnothing$) and neither a nor b is an interior point of any ironed interval, then the Q^* defined in (5) is an optimal solution for (3) and, for any Q that is feasible to (3), $\int_0^1 QVdF < \int_0^1 Q^*VdF$ in any of the following three cases:*

[17] While $\overline{V}$ is constant on any ironed interval, an interval on which $\overline{V}$ is constant need not be an ironed interval, as it is possible that $\overline{V} = V$ on some interval where V happens to be constant.

i $a \leq a' < b' \leq b$, (a', b') *is an ironed interval, and* Q *is not constant on* (a', b')*;*

ii $Q \neq Q^*$ *on a positive-measure subset* S *of* $[0, a)$ *such that* $\overline{V} < \inf_{t>a} \overline{V}(t)$ *on* S*;*

iii $Q \neq Q^*$ *on a positive-measure subset* S *of* $(b, 1]$ *such that* $\overline{V} > \sup_{t<b} \overline{V}(t)$ *on* S*.*

Proof By (5), $\int_0^1 Q^* dF = 0$. Since $a \leq F^{-1}\left(\frac{B}{B+1}\right) \leq b$, it follows again from (5) that Q^* is weakly increasing. Thus Q^* is feasible to (3). To prove that it is optimal for (3), use Myerson's (1981) equation (from (14), (15), and integration by parts)

$$\int_0^1 Q(t)V(t)dF(t) = \int_0^1 Q(t)\overline{V}(t)dF(t) - \int_0^1 \Big(H(V)\left(F(t)\right) - \widetilde{H}(V)\left(F(t)\right)\Big)\, dQ(t) \quad (16)$$

for any weakly increasing function Q on $[0, 1]$. Observe that the second integral on the right-hand side of (16) is nonnegative as Q is weakly increasing, and that it is strictly positive if and only if Q is not constant on some ironed interval. It then follows from the definition of Q^* that the said integral is zero when $Q = Q^*$, because Q^* by construction has only a and b as jump points and, by the hypothesis of the lemma, neither a nor b is interior to any ironed interval. Thus, to prove optimality of Q^* it suffices to show $\int_0^1 Q^*\overline{V} dF \geq \int_0^1 Q\overline{V} dF$ for any Q feasible to (3). To show that, note

$$\int_0^1 Q^*\overline{V} dF - \int_0^1 Q\overline{V} dF = \underbrace{\int_0^a (Q^* - Q)\overline{V} dF}_{X} + \underbrace{\int_a^b (Q^* - Q)\overline{V} dF}_{Y} + \underbrace{\int_b^1 (Q^* - Q)\overline{V} dF}_{Z}.$$

Let $v := \inf\{\overline{V}(t) \mid t > a\}$. By the the hypothesis of the lemma, $\overline{V} = v$ on (a, b) if $(a, b) \neq \varnothing$. Note: $X \geq v \int_0^a (Q^* - Q)dF$ because $Q^* - Q = -1 - Q \leq 0$ and $\overline{V} \leq v$ on $[0, a)$; $Y = v \int_a^b (Q^* - Q)dF$ because either $a = b$, or $\overline{V} = v$ on (a, b); and $Z \geq v \int_b^1 (Q^* - Q)dF$ because $Q^* - Q = B - Q \geq 0$ and $\overline{V} \geq v$ on $(b, 1]$. Thus,

$$\int_0^1 Q^*\overline{V} dF - \int_0^1 Q\overline{V} dF \geq v \int_0^a (Q^* - Q)dF + v \int_a^b (Q^* - Q)dF + v \int_b^1 (Q^* - Q)dF = 0, \quad (17)$$

with the equality due to $\int Q^* dF = 0 = \int Q dF$. Thus, Q^* is optimal for (3).

To prove the rest of the lemma, pick any Q feasible to (3). Then $Q : [0, 1] \to [-1, B]$ is weakly increasing and $\int Q dF = 0$. In Case (i), $a' < b'$ and Q is not constant on (a', b'). Then Q, weakly increasing, is strictly increasing on a positive-measure subset of (a', b'). Thus, the distribution induced by Q assigns a positive measure on (a', b'). This, coupled with the hypothesis that (a', b') is an ironed interval, implies that the second integral on the right-hand side of (16) is strictly positive given Q. By contrast, the integral given Q^* is zero. This, coupled with $\int_0^1 Q^*\overline{V}dF \geq \int_0^1 Q\overline{V}dF$ proved above, implies $\int_0^1 VQdF < \int_0^1 VQ^*dF$.

In Case (ii), since $Q^* = -1$ on $[0, a)$, the hypothesis $Q \neq Q^*$ on $S \subseteq [0, a)$ implies that $Q^* - Q < 0$ on the positive-measure subset S of $[0, a)$. This, combined with $Q^* - Q \leq 0$ on $[0, a)$ and $\overline{V} < V(a) \leq v$ on S, implies $\int_0^a (Q^* - Q)\,\overline{V}dF > v\int_0^a (Q^* - Q)\,dF$. Thus the inequality in (17) is strict. Case (iii) is analogous to Case (ii). ■

Since (15) defines $\overline{V}$ only at t for which $\widetilde{H}(V)$ is differentiable at $F(t)$, let us specify the extension of $\overline{V}$ to the two endpoints:

$$\overline{V}(1) := \sup_{t'\uparrow 1} \overline{V}(t') \quad \text{and} \quad \overline{V}(0) := \inf_{t'\downarrow 0} \overline{V}(t'). \tag{18}$$

Lemma 3 $\overline{V}(0) < \overline{V}(1)$.

Proof First, we observe that $\overline{V}(0) \leq 0$. To see that, note from the definition of ironing that $\overline{V}(0)$ is the slope of the supporting line at the point 0 of the epigraph of $H(V)$. Since $V(0) = 0$ by the definition of V and V is continuous by the assumption that both f and W are continuous, the right-derivative of $H(V)$ at point 0 is well defined and is equal to 0. Thus, the slope $\overline{V}(0)$ of the supporting line of $H(V)$ at point 0 is less than or equal to 0.

Now that $\overline{V}(0) \leq 0$, we need only to show $\overline{V}(1) > 0$. Suppose not, then $\overline{V} \leq 0$ on $[0, 1]$ by its monotonicity. Then

$$0 \geq \int_0^1 \overline{V}(t)dF(t) = \widetilde{H}(V)(1) = H(V)(1) = \int_0^1 V(t)dF(t),$$

where the first equality is due to (15), the absolute continuity of $\widetilde{H}(V)$ and $\widetilde{H}(V)(0) = 0$, the second equality due to $\widetilde{H}(V)$ being the convex hull of $H(V)$ on $[0, 1]$ and $H(V)(0) = 0$, and the last equality due to (14). Thus, $0 \leq \int_0^1 V(t)dF(t)$ and hence, by the definition of V,

$$\int_0^1 W(t)dt \geq \int_0^1 tdF(t) + \int_0^1 F(t)dt = 1,$$

with the equality due to integration by parts. Since W is a cdf that is supported by $[0, 1]$ and continuous on $\mathbb{R}$, $W \leq 1$ on $[0, 1]$ and strictly so on a positive-measure subset thereof. Thus $\int_0^1 W(t)dt < 1$ and the above-displayed inequality is impossible, which leads to the desired contradiction. ■

Lemma 4 $\overline{V}$ *is continuous at the points 0 and 1.*

Proof We shall prove that $\overline{V}$ is continuous at point 1. The case of point 0 is symmetric.

If $\widetilde{H}(V) = H(V)$ on $(F(1-\delta), 1]$ for some $\delta > 0$, then by (15) $\overline{V} = V$ on $(1-\delta, 1]$, and hence the continuity $\overline{V}$ at point 1 follows from the continuity of V. If $\widetilde{H}(V) < H(V)$ on $(F(1-\delta), 1)$ for some $\delta > 0$, then $(1-\delta, 1)$ is contained in an ironed interval, so $\overline{V}$ is constant on $(1-\delta, 1)$ and hence $\sup_{t'\uparrow 1} \overline{V}(t')$ is equal to this constant. Then (18) implies that $\overline{V}(1)$ is equal to the constant; thus again $\overline{V}$ is continuous at 1.

Thus, suppose that neither of the previous cases hold. That is, there exists an ironed interval (a_1, b_1) such that $0 \leq a_1 < b_1 < 1$, there exists another ironed interval (a_2, b_2) for which $b_1 \leq a_2 < b_2 < 1$, and furthermore for any ironed interval (a_k, b_k) for which $b_k < 1$, there exists another ironed interval (a_{k+1}, b_{k+1}) for which $b_k \leq a_{k+1} < b_{k+1} < 1$. Thus, by recursion, $[0, 1]$ is partitioned by

$$0 \leq a_1 < b_1 \leq a_2 < b_2 \leq a_3 < \cdots \leq a_k < b_k \leq a_{k+1} < \cdots < 1$$

such that

$$\lim_{k\to\infty} a_k = \lim_{k\to\infty} b_k = 1,$$
$$\forall k\, \exists v_k \in \mathbb{R} : \left[\overline{V} = v_k \text{ on } (a_k, b_k)\right],$$

with the last line due to (a_k, b_k) being an ironed interval. Since $\overline{V}$ is weakly increasing,

$$v_1 \leq v_2 \leq v_3 \leq \cdots v_k \leq v_{k+1} \leq \cdots.$$

Within this case, to prove the continuity of $\overline{V}$ at point 1, we start by observing that

$$\lim_{k\to\infty} v_k = V(1). \tag{19}$$

To show that, for each k pick any $t_k \in [b_k, a_{k+1}]$. Then $\overline{V}(t_k) = V(t_k)$. With $\overline{V}$ weakly increasing,

$$v_1 \leq V(t_1) \leq v_2 \leq V(t_2) \leq v_3 \leq V(t_3) \leq \cdots.$$

Thus $\lim_{k\to\infty} v_k = \lim_{k\to\infty} V(t_k) = V(1)$, with the second equality due to $t_k \to 1$ and V being continuous at 1.

Next, pick any sequence $(t'_j)_{j=1}^{\infty}$ converging to 1 such that $t'_1 \leq t'_2 \leq t'_3 \leq \cdots$. For each j, either $t'_j \in (a_{k_j}, b_{k_j})$ for some k_j, or $t'_j \in [b_{k_j}, a_{k_j+1}]$ for some k_j. In the former case, $\overline{V}(t'_j) = v_{k_j}$; in the latter, $v_{k_j} \leq \overline{V}(t'_j) \leq v_{k_j+1}$. Both cases considered,

$$\lim_{j\to\infty} \overline{V}(t'_j) = \lim_{j\to\infty} v_{k_j} = V(1),$$

with the second equality due to (19). Since $(t'_j)_{j=1}^{\infty}$ can be any sequence converging to 1 from below, the above equation also implies $\lim_{t'\uparrow 1}\overline{V}(t') = V(1)$. Consequently, $\sup_{t'\uparrow 1}\overline{V}(t') = \lim_{t'\uparrow 1}\overline{V}(t') = V(1)$. This coupled with (18) implies $\overline{V}(1) = \lim_{t'\uparrow 1}\overline{V}(t')$, namely, $\overline{V}$ is continuous at 1. ■

Lemma 5 *If $V(1) > V(t)$ for all $t \in [0,1)$, the followings are true for $\overline{V}$:*

a. there is no $x \in [0,1)$ for which $\overline{V}(t) \neq V(t)$ for all $t \in (x,1)$;

b. $\overline{V}(1) = V(1)$;

c. $\overline{V}(1) > \overline{V}(t)$ for all $t \in [0,1)$.

Proof Proof of (a): Suppose, to the contrary, that there exists an $x \in [0,1)$ for which $\overline{V}(t) \neq V(t)$ for all $t \in (x,1)$. Then $(x,1)$ is contained in some ironed interval say $(x_*,1)$ such that $\widetilde{H}(V)(F(t)) < H(V)(F(t))$ for all $t \in (x_*,1)$, $\widetilde{H}(V)(F(x_*)) = H(V)(F(x_*))$, $\widetilde{H}(V)(F(1)) = H(V)(F(1))$, and $\widetilde{H}(V)$ has a constant slope β on $[F(x_*),1]$. Since $\widetilde{H}(V)(F(t)) < H(V)(F(t))$ for all $t \in (x_*,1)$, for any $t < 1$ sufficiently close to 1,

$$\frac{1}{1-F(t)}\int_t^1 V(s)dF(s) = \frac{1}{1-F(t)}\left(H(V)(F(1)) - H(V)(F(t))\right) \leq \beta.$$

Taking the limit of the inequality as $t \to 1$ and noting continuity of V at 1, we have $V(1) \leq \beta$. Meanwhile, since $\widetilde{H}(V)(F(t)) < H(V)(F(t))$ for all $t \in (x_*,1)$, there exists $t' \in (x_*,1)$ for which the slope of $H(V)$ at $F(t')$ is greater than β. That is, $V(t') > \beta$, which coupled with $V(1) \leq \beta$ implies $V(t') > V(1)$, contradicting the hypothesis that V is maximized at 1.

Proof of (b): Note, from the proof of Lemma 4, that $\overline{V}(1) = V(1)$ unless $\widetilde{H}(V) < H(V)$ on $(F(1-\delta),1)$ for some $\delta > 0$, namely, $(1-\delta,1)$ is contained in an ironed interval. Thus $\overline{V} \neq V$ on $(x,1)$ for some $x \in (1-\delta,1)$, contradicting (a). Thus (b) holds.

Proof of (c): Suppose, to the contrary, that $\overline{V}(1) \leq \overline{V}(t_0)$ for some $t_0 \in [0,1)$. Then, with $\overline{V}$ weakly increasing, $\overline{V}(t) = \overline{V}(1)$ for all $t \in [t_0,1]$. By (a), there exists $t_1 \in (t_0,1)$ for which $\overline{V}(t_1) = V(t_1)$. Since $t_1 \in (t_0,1)$, $\overline{V}(t_1) = \overline{V}(1)$. Then (b) implies $V(1) = \overline{V}(1) = \overline{V}(t_1) = V(t_1)$, contradicting the hypothesis that $V(1) > V(t)$ for all $t \in [0,1)$. ■

D Proof of Theorem 2

Recall the definition of ironed interval from Appendix C.

Lemma 6 *For any welfare weight distribution W, the two-tier allocation (6) is optimal if and only if $F^{-1}\left(\frac{B}{B+1}\right)$ is not interior to any ironed interval.*

Proof If $F^{-1}\left(\frac{B}{B+1}\right)$ is not interior to any ironed interval, then Lemma 2 applies to the case where $a = b = F^{-1}\left(\frac{B}{B+1}\right)$ so that the Q^* defined in (5) specializes to (6), the two-tier allocation implemented by the posted-price system. Thus Lemma 2 implies that (6) is optimal, and the "if" part of the claim is true. To prove the "only if" part, suppose that $F^{-1}\left(\frac{B}{B+1}\right)$ is interior to some ironed interval. Then Part (i) in Lemma 2 implies that no optimal allocation has a jump point in the ironed interval and hence the allocation (6), whose jump point is $F^{-1}\left(\frac{B}{B+1}\right)$, is not optimal. ■

Lemma 7 *If the function $V(\cdot) - V\left(F^{-1}\left(\frac{B}{B+1}\right)\right)$ is single-crossing on $[0,1]$, then $F^{-1}\left(\frac{B}{B+1}\right)$ is not interior to any ironed interval.*

Proof Denote $m := F^{-1}\left(\frac{B}{B+1}\right)$ and suppose that $V(\cdot) - V(m)$ is a single-crossing function on $[0,1]$. By the definition of ironing, it suffices to prove that $F(m)$ is a *convex point* of $H(V)$ in the sense that at no point below $F(m)$ is $H(V)$ steeper than it is at $F(m)$, and at no point above $F(m)$ is $H(V)$ less steep than it is at $F(m)$. Since $V(\cdot) - V(m)$ single-crossing on $[0,1]$,

$$s < F(m) < s' \Longrightarrow V\left(F^{-1}(s)\right) \leq V\left(F^{-1}(F(m))\right) = V(m) \leq V\left(F^{-1}(s')\right).$$

By (14) the definition of $H(V)$, the derivative of $H(V)$ at any $s \in (0,1)$ is $D\left(H(V)\right)(s) = V\left(F^{-1}(s)\right)$. Plug this into the above-displayed formula to obtain

$$s < F(m) < s' \Longrightarrow D\left(H(V)\right)(s) \leq D\left(H(V)\right)(F(m)) \leq D\left(H(V)\right)(s').$$

Thus, $F(m)$ is a convex point of $H(V)$, as desired. ■

Proof of Theorem 2 Let a be the infimum, and b the supremum, of

$$\left\{t \in [0,1] \mid \overline{V}(t) = \overline{V}\left(F^{-1}(B/(B+1))\right)\right\}. \tag{20}$$

Then neither a nor b is interior to an ironed interval, and hence Lemma 2 implies that the allocation (5) is an optimal allocation. In the case where $V(\cdot) - V\left(F^{-1}\left(\frac{B}{B+1}\right)\right)$ is single-crossing on $[0,1]$, Lemmas 6 and 7 together imply that the two-tier allocation (6) is optimal. The payment rules that the theorem asserts implement (5) and (6) respectively can be derived from the allocations according to the envelope formula, as explained in the comments around the theorem and the proof of Corollary 2. ■

E Proofs of Corollaries 2 and 3

Corollary 2 First, consider the case $x \geq 0$. By $a < F^{-1}\left(\frac{B}{B+1}\right)$ and (7), $x < B$ and hence the allocation (5) is weakly increasing. Thus, the allocation can be implemented by a payment rule. Consider the one that maximizes the planner's profit among all that implement the allocation. By the envelope formula, one can show that this payment rule is the same as the one described in Part (i) of the corollary. For instance, an individual of type $t \geq b$ gets the surplus $(b-a)x + (t-b)B$ by the envelope theorm and hence needs to deliver a total payment equal to $bB - (b-a)x$ for the quantity B of the good. That implies the per-unit price $b - (b-a)x/B$. Note $b - (b-a)x/B > a$ (because $a < F^{-1}\left(\frac{B}{B+1}\right)$). Thus, the profit generated by this payment rule is greater than

$$-aF(a) + a\left(F(b) - F(a)\right)x + a\left(1 - F(b)\right)B = 0,$$

with the equality due to the market clearing condition. By the envelope formula, the payment rules that implement the allocation differ from one another only by a constant. Thus, since the planer would rebate all her profit to the individuals to achieve the optimality of (1), the optimal payment rule that implements (5) is the profit-maximizing one among those that implement (5), augmented with a lump sum transfer to the individuals. Since the allocation restricted to (a, b) is equal to the constant x and $0 \leq x < B$, the mechanism entails rationing on (a, b). Thus the corollary is true in the case $x \geq 0$.

The case $x \leq 0$ is symmetric. By $F^{-1}\left(\frac{B}{B+1}\right) < b$ and (7), we have $-1 < x$ and hence (5) is weakly increasing and entails rationing on (a, b). The profit-maximizing payment rule follows similarly from the envelope formula. Since $-1 < x$, the per-unit price $a - (b-a)x$ for the seller-types in $[0, a)$ is strictly less than b. This coupled with the market clearing condition implies that the profit generated by the payment rule is positive and hence the optimal payment rule makes a positive lump sum transfer to the individuals. ■

Corollary 3 By Lemmas 3 and 4 (Appendix C), the ironed virtual surplus function $\overline{V}$ is continuous at both points 0 and 1, and $\overline{V}(0) < \overline{V}(1)$. This fact implies that the conditions (ii) and (iii) in Lemma 2 are true when the egalitarian allocation Q is compared to the optimal allocation Q^*, and hence Q is strictly outperformed by Q^*. ■

F Proof of Theorem 3

Theorem 3 follows directly from the next lemma, as the non-constancy assumption of V in the theorem implies the condition (21) in the lemma. Recall the definition of ironed interval in Appendix C.

Lemma 8 *There exists at most one (modulo measure zero) optimal allocation if*

$$\overline{V} = \overline{V}\left(F^{-1}\left(\tfrac{B}{B+1}\right)\right) \text{ on } (a,b) \neq \varnothing \Longrightarrow (a,b) \text{ is a subset of an ironed interval.} \tag{21}$$

Proof Since $\overline{V}$ is weakly increasing, the set (20) defined in Apendix D is an interval. By (21), the interior of (20) is either empty or an ironed interval. The $\overline{V}$-value on the set (20) is higher than those of the types below the infimum of the set, and lower than those of the types above the supremum thereof. This, coupled with the fact that the interior of the set is either empty or an ironed interval, implies that the conditions (i), (ii) and (iii) in Lemma 2 hold for any weakly increasing and market clearing allocation Q that differs from the Q^* defined in (5) by a positive measure. Any such Q is therefore strictly outperformed by Q^*. Thus the optimal allocation is unique. ■

G Proof of Theorem 4

Lemma 9 *In any optimal mechanism where the allocation is the three-tier (5) that rations a quantity x—defined by (7)—on some (a,b) for which $0 < a < F^{-1}\left(\frac{B}{B+1}\right) < b < 1$:*

a. the surplus for type zero is equal to $a + (b-a)(B-x)(1-F(b))$;

b. the surplus for type one is equal to $(1-b)B + (b-a)(1+x)F(a)$.

Proof To prove Claim (a), consider first the case where $x \geq 0$. Recall from Corollary 2 for the payment rule in this case. The surplus for type zero is equal to a (the revenue the type receives from selling his one unit endowment) plus the lump sum rebate from the planner. The lump sum rebate is equal to the planner's profit from implementing (5) through the profit-maximizing payment rule. Note that the planner cannot profit from selling the good to the types in (a,b), as the per-unit revenue extracted from them is equal to a, which is equal to the per-unit cost from procuring the good (from the seller-types in $[0,a)$). Thus, the planner can profit only from the sales to the buyer-types in $(b, 1]$. The per-unit profit is the price difference $b-(b-a)x/B-a$ between the price $b-(b-a)x/B$ offered to $(b, 1]$ and the price a to the seller-types. Since the amount of sales to $(b, 1]$ is $(1-F(b))B$, the profit is equal to

$$(b-(b-a)x/B-a)\,(1-F(b))\,B=(b-a)(B-x)\,(1-F(b))\,. \tag{22}$$

Thus the surplus for type zero is equal to a plus the above expression, as in Claim (a).

Next consider the other case, $x<0$. Again recall from Corollary 2 for the payment rule in this case. The planner can profit only from the quantity she procures from the seller-types in $[0, a)$. The per-unit profit from this quantity is the price difference $b-(a-(b-a)x)$ between the sales price b and the procurement price $a-(b-a)x$. The quantity is equal to the mass $F(a)$ of $[0, a)$. Thus the profit is equal to

$$(b-(a-(b-a)x))\,F(a). \tag{23}$$

Note that the revenue a type zero receives from selling his one unit of the good is equal to $a-(b-a)x$. Therefore, the surplus for type zero is equal to

$$\begin{aligned} a-(b-a)x+(b-(a-(b-a)x))\,F(a) &= a-(b-a)x+(b-a)(1+x)F(a) \\ &= a+(b-a)(B-x)\,(1-F(b))\,, \end{aligned}$$

where the second line is equivalent to (7), the definition of x. Thus the surplus for type zero in the case $x<0$ is also equal to the expression in Claim (a). Hence Claim (a) is true.

To prove Claim (b), consider first the case $x<0$. Given the payment rule characterized in Corollary 2 for the mechanism of (5), type one buys the quantity B of the good at the price b per unit and receives a lump sum rebate, which has been shown to be equal to (23) in the proof of Claim (a). Thus, the surplus for type one given (5) is equal to

$$(1-b)B+(b-(a-(b-a)x))\,F(a)$$

when $x<0$, as asserted by Claim (b).

Next consider the other case, $x \geq 0$. Given the payment rule characterized in Corollary 2 for the mechanism of (5), type one buys the quantity B of the good at the price $b-(b-a)x/B$ per unit and receives a lump sum rebate, which has been shown to be equal to (22) in the proof of Claim (a). Thus, the surplus for type one given (5) is equal to

$$\begin{aligned} & (1-(b-(b-a)x/B))\,B+(b-a)(B-x)\,(1-F(b)) \\ = \; & (1-b)B+(b-a)x+(b-a)(B-x)\,(1-F(b)) \\ = \; & (1-b)B+(b-a)(1+x)F(a), \end{aligned}$$

with the last line equivalent to (7), the definition of x. Thus, the surplus for type one in the case $x \geq 0$ is also equal to the expression asserted by Claim (b). Hence Claim (b) is true. ■

Proof of Theorem 4 Claim (a) is intuitive. By Corollary 2, the mechanism of allocation (5) transfers a (strictly) positive lump sum rebate to all types, and hence the surplus for the type $F^{-1}\left(\frac{B}{B+1}\right)$ given allocation (5) is positive. By contrast, the surplus for the type $F^{-1}\left(\frac{B}{B+1}\right)$ is equal to zero in the mechanism of the allocation (6): By the envelope theorem, one readily sees that the surplus function given allocation (6) attains its minimum at the type equal to $F^{-1}\left(\frac{B}{B+1}\right)$. Meanwhile, it is easy to show that any payment rule that implements a market clearing two-tier allocation such as (6) generates zero profit for the planner and hence zero lump sum rebate to the individuals. Thus type $F^{-1}\left(\frac{B}{B+1}\right)$ gets zero surplus under the allocation (6). It follows that the surplus for type $F^{-1}\left(\frac{B}{B+1}\right)$ given allocation (5) is greater than that given (6). By continuity of surplus as a function of types, this strict inequality extends to types sufficiently near to $F^{-1}\left(\frac{B}{B+1}\right)$, and hence Claim (a) is true.

To prove Claims (b) and (c), note from (7), the definition of x, that

$$-F(a)+B(1-F(b))+x(F(b)-F(a))=0,$$

or equivalently,

$$\frac{B}{B+1}=\frac{1+x}{B+1}F(a)+\frac{B-x}{B+1}F(b). \tag{24}$$

Since $(1+x)/(B+1)$ and $(B-x)/(B+1)$ are between zero and one and sum up to one, $B/(B+1)$ is a convex combination between $F(a)$ and $F(b)$. When F is convex on (a,b), F^{-1} is concave on $(F(a),F(b))$ because F is strictly increasing by assumption. Thus, Jensen's inequality implies

$$F^{-1}\left(\frac{B}{B+1}\right)\geq\frac{1+x}{B+1}a+\frac{B-x}{B+1}b=a+\frac{B-x}{B+1}\left(b-a\right). \tag{25}$$

By Lemma 9.a, the surplus for type zero under the optimal mechanism of the rationing allocation (5) is equal to

$$\begin{aligned} a + (b-a)(B-x)(1-F(b)) &< a + (b-a)\frac{B-x}{B+1} \\ &\leq F^{-1}\left(\frac{B}{B+1}\right), \end{aligned}$$

where the first inequality follows from $F(b) > B/(B+1)$, and the second inequality, from (25). Thus, since the surplus for type zero given the optimal mechanism of allocation (6) is equal to the market clearing price $F^{-1}\left(\frac{B}{B+1}\right)$, type zero is worse-off in the rationing mechanism of allocation (5) than in the mechanism of the allocation (6). Since the allocations (5) and (6) are identically equal to -1 for all types in $[0,a)$, the envelope theorem implies that in both allocations, the surplus decreases at the same rate -1 when the type increases from zero to a. Consequently, Claim (b) of the theorem follows.

Similarly, when F is concave on (a,b), F^{-1} is convex on $(F(a),F(b))$. Thus

$$F^{-1}\left(\frac{B}{B+1}\right) \leq \frac{1+x}{B+1}a + \frac{B-x}{B+1}b = b - \frac{1+x}{B+1}(b-a). \tag{26}$$

By Lemma 9.b, the surplus for type one under the optimal mechanism of the rationing allocation (5) is equal to

$$\begin{aligned} (1-b)B + (b-a)(1+x)F(a) &= B - B\left(b - (1+x)(b-a)\frac{F(a)}{B}\right) \\ &< B - B\left(b - \frac{1+x}{B+1}(b-a)\right) \\ &\leq B\left(1 - F^{-1}\left(\frac{B}{B+1}\right)\right), \end{aligned}$$

where the first inequality follows from $F(a) < B/(B+1)$, and the second inequality from (26). Thus, since the surplus for type one given allocation (6) is equal to $B\left(1 - F^{-1}\left(\frac{B}{B+1}\right)\right)$, type one is worse-off in the allocation (5) than in the allocation (6). Since the allocations (5) and (6) are identically equal to B for all types in $(b,1]$, the envelope theorem implies that in both allocations, the surplus increases at the same rate B when the type increases from b to 1. Consequently, Claim (c) of the theorem follows. ■

H Proof of Theorem 5

Denote V for the virtual surplus function when the welfare weight distribution is W, and V_* the virtual surplus function when the said distribution is W_*. Let $m := F^{-1}\left(\frac{B}{B+1}\right)$.

Part (a) By the definition of ironing (Appendix C), m is interior to an ironed interval given W if and only if $\widetilde{H}(V)(F(m)) < H(V)(F(m))$, which in turn holds if and only if there exist $a, b \in [0,1]$ for which $a < m < b$ and

$$\begin{aligned} H(V)(F(m)) &> H(V)(F(a)) + \frac{F(m)-F(a)}{F(b)-F(a)}\left(H(V)(F(b)) - H(V)(F(a))\right), \\ &= \frac{F(b)-F(m)}{F(b)-F(a)}H(V)(F(a)) + \frac{F(m)-F(a)}{F(b)-F(a)}H(V)(F(b)). \end{aligned}$$

The above condition is equivalent to that, for some $a < m < b$ ($m = F^{-1}\left(\frac{B}{B+1}\right)$),

$$\frac{F(b)-F(m)}{F(b)-F(a)}\int_a^m V(s)dF(s) - \frac{F(m)-F(a)}{F(b)-F(a)}\int_m^b V(s)dF(s) > 0,$$

which one can simplify, by dividing $(F(b)-F(m))(F(m)-F(a))/(F(b)-F(a))$, to

$$\frac{1}{F(m)-F(a)}\int_a^m V(s)dF(s) - \frac{1}{F(b)-F(m)}\int_m^b V(s)dF(s) > 0.$$

It follows that m is not interior to any ironed interval given V if and only if, for any $a \in [0,m)$ and any $b \in (m,1]$,

$$\frac{1}{F(m)-F(a)}\int_a^m V(s)dF(s) - \frac{1}{F(b)-F(m)}\int_m^b V(s)dF(s) \leq 0. \tag{27}$$

By the hypothesis in the theorem that the posted-price system is optimal given W, Lemma 6 implies that m is not interior to any ironed interval given V, and hence (27) holds for any $a \in [0,m)$ and any $b \in (m,1]$. Now let W_* be any spread of W away from m, namely, $W_* \geq W$ on $[0,m)$ and $W_* \leq W$ on $(m,1]$. Then, by (4) the definition of virtual surplus,

$$V_*(t) - V(t) = \frac{W(t)-W_*(t)}{f(t)} \begin{cases} \leq 0 & \text{if } t \in [0,m) \\ \geq 0 & \text{if } t \in (m,1]. \end{cases}$$

Thus, when the V in (27) is replaced by V_*, the inequality (27) remains to be true for any $a \in [0,m)$ and any $b \in (m,1]$. In other words, m is not interior to any ironed interval given W_*. Then Lemma 6, applied to the case of W_*, implies that the posted-price system is optimal given W_*.

Part (b) Let $\epsilon > 0$. Let

$$\begin{aligned} a &:= \inf\{t \in [0,m) \mid W(t) > W(m) - \epsilon/2\}, \\ b &:= \sup\{t \in (m,1] \mid W(t) < W(m) + \epsilon/2\}. \end{aligned}$$

Since welfare weight distributions are assumed continuous on $\mathbb{R}$, $W(a) < W(m)$ and there is a positive-measure subset of (a, m) on which $W > W(a)$. For any $t \in \mathbb{R}$ define

$$W_*(t) := \begin{cases} W(t) & \text{if } t < a \\ W(a) & \text{if } a \le t < m \\ W(b) & \text{if } m \le t < b \\ W(t) & \text{if } t \ge b. \end{cases}$$

By the construction of W_* and the monotonicity of W, W_* is a cdf, $W_* \le W$ on $[0, m)$, $W_* \ge W$ on $(m, 1]$, and $\|W_* - W\|_{\max} \le W(b) - W(a) \le \epsilon$. Since W_* has a jump at $t = m$, the virtual surplus function V_* given W_*, by (4), has a drop at $t = m$. Thus one can modify W_* into a continuous function (just to satisfy our assumption that welfare weight distributions are continuous) by replacing the jump at $t = m$ with a sufficiently steep affine segment, so that V_* after modification remains to be decreasing strictly at $t = m$. It follows from the definition of ironing that m is interior to an ironed interval given W_* (with or without the continuity modification of the jump). Then Lemma 6 implies that no optimal allocation given W_* can be implemented by the posted-price system.

I Proof of Corollary 4

In the modified model, denote e for each person's endowed share of the public resource (e.g., the probability of vaccination that each individual is entitled to). Given any mechanism $(\tilde{q}, p) : [0, 1] \to [e, 1] \times \mathbb{R}$ in the modified model, denote $Q(t) := \tilde{q}(t) - e$ for all types t, so that if an individual of type t acts as type t', his expected payoff is equal to

$$tQ(t') - p(t') + t + \psi(t) \int_0^1 (Q(s) + e) dF(s).$$

Thus, a mechanism $(\tilde{q}, p)$ in the modified model corresponds to a mechanism (Q, p) in the main model, and the objective for any type-t individual is equal to a constant plus $tQ(t') - p(t')$, which is the same as that in our main model. It follows that the IC, IR and BB constraints are the same in the modified model as in the main model. Thus we need only to show that the objectives to the social planner in the two models are equivalent. In the modified model, the social planner's objective is equal to

$$\int_0^1 \left(tQ(t) - p(t) + \int_0^1 \psi(t) Q(s) dF(s) \right) dW(t) \tag{28}$$

plus a constant that remains invariant when the mechanism varies. Note that (28) is equal to (1) plus

$$\int_0^1 \int_0^1 \psi(t)Q(s)dF(s)dW(t).$$

Switch the order of integration to rewrite this double integral as

$$\int_0^1 Q(s) \int_0^1 \psi(t)dW(t)dF(s).$$

Thus, following the same routine of envelope theorem and integration by parts that results in Corollary 1, one can prove that an allocation is optimal if and only if it solves Problem (3) such that the objective $\int_0^1 QVdF$ therein is replaced by $\int_0^1 Q\tilde{V}dF$ where

$$\tilde{V}(t) := V(t) + \int_0^1 \psi(s)dW(s)$$

for all $t \in [0, 1]$. Since the additional integral is a constant, the two optimization problems are equivalent.

J Unbounded Acquisition

The main model assumes that the upper bound B for acquisition quantity per type is finite. Here we consider an extension where $B = \infty$. This case reflects a world with severe inequalities and insatiable demands for the good. For example, it could be an exchange economy where the endowment is an individual's initially acquired tract of land when a group of colonists arrive at a new, unoccupied place, or one's own private information in digital format that can be traded off for convenience, or a citizen's initial voting power in a fledging republic say the early Roman Republic. The following extension sheds light on the tendency that such resources are concentrated to a tiny few of the society.

Now that there is no upper bound on the quantity that a type is allowed to acquire, the buyer-types in this case should only be those types that maximize the ironed virtual surplus $\overline{V}$—selling the good to any type with lower $\overline{V}$-value would be a waste—and all other types should be sellers. The outcome in this case is therefore intuitive. Either the ironed virtual surplus $\overline{V}$ attains its maximum at a unique point (the highest type, as $\overline{V}$ is weakly increasing by construction), or $\overline{V}$ is maximized by multiple points, which constitute an upper interval in the type space. In the former case, all members of the society supply the good to the single, highest type. In the latter case, the optimal allocation entails two tiers, the "haves" consisting of the $\overline{V}$-maximizers, and the "have-nots" consisting of all the other types.

The former case, the utmost form of inequalities, needs to be formalized because the corresponding optimal allocation is not a real function. We say that the optimal allocation is *singular* iff there exists a sequence $(Q^n)_{n=1}^{\infty}$ of functions $Q^n : [0,1] \to [-1, \infty)$, each weakly increasing and market clearing (and hence budget balancing by Theorem 1, which remains intact when $B = \infty$), such that Q^n converges pointwise to the extended-real function Q^{∞} defined by

$$Q^{\infty}(t) := \begin{cases} -1 & \text{if } t \in [0,1) \\ \infty & \text{if } t = 1 \end{cases} \tag{29}$$

and, for any function $Q : [0,1] \to [-1, \infty)$ that is weakly increasing and market clearing (and hence budget-balancing), there exists N for which Q^n outperforms Q in terms of the design objective in Section 2 for all $n \geq N$. Then one can prove (Appendix K) the following characterization of the optimal allocation.[18]

Theorem 6 *When $B = \infty$:*

a. *if $V(1) > V(t)$ for all $t \in [0,1)$, then the optimal allocation is singular;*

b. *else then there exists an optimal allocation and it is a two-tier allocation defined by*

$$Q^*(t) := \begin{cases} -1 & \text{if } 0 \leq t < c^* \\ \frac{F(c^*)}{1-F(c^*)} & \text{if } c^* < t \leq 1, \end{cases} \tag{30}$$

where

$$c^* := \inf \left(\arg\max_{[0,1]} \overline{V} \right).$$

Clearly, rationing is needed to implement the optimal allocation (30) in case (b). Such necessity of rationing among the "haves" makes sense realistically: unchecked concentration begets social upheavals. Nonetheless, case (a) in a sense corresponds to the optimality of the posted-price system: Since $V(1) > V(t)$ for all $t \in [0,1)$, one can construct a sequence $(B_n, Q^n)_{n=1}^{\infty}$ such that $Q^n \to Q^{\infty}$ pointwise, $B_n \to_n \infty$, and for each n, Q^n is the optimal allocation in the basic model given upper bound B_n, implemented by posting the market clearing price $F^{-1}\left(\frac{B_n}{B_n+1}\right)$. Since Q^n in the sequence attains the optimality given B_n, Q^{∞}

[18] With a condition similar to (21), one can also establish a uniqueness claim of the optimal allocation.

can be viewed as the limit of the optimum-implementing posted-price system when the acquisition cap rises without bound.

A sufficient condition to rule out the singularity case in Theorem 6 is that V be strictly decreasing at 1, which means $2 < w(1)$ if w is the Radon-Nikodym derivative of the welfare weight distribution W with respect to F. Intuitively speaking, had the optimal allocation been singular, the surplus for a type-1 player, whose type is the highest, would be infinitesimal, since the price for the good converges to one. But if the designer rations the quantity to an interval $(c, 1]$, the trading price is $c < 1$, and so the type-1 player gets a strictly positive surplus. Thus if the welfare weight density on type one is sufficiently large, the optimal allocation is to ration the good to some interval $(c^*, 1]$.

K Proof of Theorem 6

First, note that Theorem 1 applies to the $B = \infty$ case. Second, adopt the proof of Lemma 2 to obtain the following fact analogous to the lemma: For any $a, b \in [0, 1]$ such that $\overline{V}$ is constant on (a, b) (unless $(a, b) = \varnothing$) and neither a nor b is an interior point of any ironed interval, for any $-1 \leq x \leq y$ such that the allocation

$$Q^*(t) := \begin{cases} -1 & \text{if } 0 \leq t < a \\ x & \text{if } a < t < b \\ y & \text{if } b < t \leq 1 \end{cases} \tag{31}$$

satisfies market clearing, and for any $Q : [0, 1] \to [-1, y]$ that is weakly increasing and satisfies market clearing, we have $\int_0^1 Q^* V dF \geq \int_0^1 Q V dF$, and the inequality is strict if at least one of the conditions (i), (ii) and (iii) listed in Lemma 2 holds.

Claim (a) Assume the premise of this claim, that $V(1) > V(a)$ for all $a \in [0, 1)$. To satisfy the condition for singularity, we start by constructing a sequence of allocations that converges to Q^∞. Since $\overline{V}$ is monotone, for any $x \in \mathbb{R}$ the inverse image $\overline{V}^{-1}(x)$ is a nondegenerate interval if it contains more than one point. There are at most countably many such nondegenerate intervals. Thus, either $\overline{V} = V$ and is strictly increasing on $[0, 1]$, or $[0, 1]$ is partitioned by a sequence $(\tau_k, \theta_k)_{k=1}^K$, for some $K \in \{1, 2, 3, \ldots\} \cup \{\infty\}$, such that

$$0 \leq \tau_1 < \theta_1 \leq \tau_2 < \theta_2 \leq \tau_3 < \theta_3 \leq \cdots \leq 1,$$

for each k there is $v_k \in \mathbb{R}$ for which $\overline{V} = v_k$ on (τ_k, θ_k), and $k < j \Rightarrow v_k < v_j$. Note that any ironed interval is contained in $[\tau_k, \theta_k]$ for some k. For any $n = 2, 3, 4, \ldots$, define

$$Q^n(t) := \begin{cases} -1 & \text{if } 0 \le F(t) < 1 - 1/n \\ n-1 & \text{if } 1 - 1/n < F(t) \le 1. \end{cases}$$

For each n, Q^n is weakly increasing and market clearing by construction and hence is also budget balancing by Theorem 1. Clearly $(Q^n)_{n=2}^{\infty}$ converges to Q^{∞} pointwise. We shall extract an infinite subsequence of $(Q^n)_{n=2}^{\infty}$ whose jump points do not belong to the interior of any ironed interval, which is contained by $[\tau_k, \theta_k]$ for some k. Start with the smallest n for which $F^{-1}(1 - 1/n) \in (\tau_k, \theta_k)$ for some k. Replace the jump point $F^{-1}(1 - 1/n)$ for Q^n by θ_k, and raise the level of Q^n on $(\theta_k, 1]$ to $F(\theta_k)/(1 - F(\theta_k))$ to preserve market clearing. Remove all the Q^m in the original sequence such that

$$F^{-1}(1 - 1/(n-1)) < F^{-1}(1 - 1/m) < \theta_k.$$

Since $V(1) > V(a)$ for all $a \in [0, 1)$ by hypothesis, $\theta_k < 1$ (Lemma 5.a), thus there exists an integer M that is the largest among such m. Then, starting from Q^{M+1}, modify the sequence $(Q^n)_{n=M+1}^{\infty}$ as we do $(Q^n)_{n=2}^{\infty}$. By recursion, we obtain an infinite subsequence $(Q^{n_j})_{j=1}^{\infty}$ of $(Q^n)_{n=2}^{\infty}$ such that for any j and any k the jump point of Q^{n_j} does not belong to (τ_k, θ_k).

To show that the optimal allocation is singular, pick any weakly increasing and market clearing allocation Q, and we shall prove that Q is outperformed by the Q^{n_j} in $(Q^{n_j})_{j=1}^{\infty}$ for all sufficiently large j. Since the elements in the sequence are both market clearing and budget balancing, and their jump points are not interior to any ironed interval, the observation at the start of this section applies and we need only to prove that condition (iii) listed in Lemma 2 holds, where the role of Q^* is played by Q^{n_j} for some sufficiently large j. To that end, recall from Lemmas 4 and 5 that $\overline{V}$ is continuous at $t = 1$ and

$$\forall t < 1 : \overline{V}(t) < \overline{V}(1) = V(1). \tag{32}$$

This coupled with $V(1) = 1 > 0$, implies

$$\exists \delta > 0 : \forall t \in (1 - \delta, 1] : \overline{V}(t) > 0. \tag{33}$$

Since the range of Q is contained in $[-1, \infty)$, the market clearing condition implies that $Q > -1$ on a positive-measure subset of $[0, 1]$. Consequently, with Q weakly increasing,

$$\theta := \inf \{t \in [0, 1] \mid Q(t) > -1\} < 1.$$

Since the range of Q is $[-1, \infty)$ and Q is weakly increasing, $\max_{[0,1]} Q = Q(1) < \infty$. Thus there exists J such that for any $j \geq J$ we have

$$\max\{\theta, 1 - \delta\} < F^{-1}\left(1 - \frac{1}{n_j}\right) \quad \text{and} \quad n_j - 1 > Q(1).$$

For any $j \geq J$, denote the jump point of Q^{n_j} by x_j. Then either $x_j = F^{-1}(1 - 1/n_j)$, or $x_j = \theta_k$ such that θ_k is the right endpoint of the interval (τ_k, θ_k) to which $F^{-1}(1 - 1/n_j)$ belongs. Let $v := \overline{V}(x_j)$. Thus, $1 > x_k \geq F^{-1}(1 - 1/n_j)$ and $v \geq \overline{V}\left(F^{-1}(1 - 1/n_j)\right)$. Since $1 - \delta < F^{-1}\left(1 - 1/n_j\right)$, (33) implies $v > 0$. With $\overline{V}$ weakly increasing, $\overline{V}(t) \geq v$ for all $t \in [x_j, 1]$. Furthermore, (32) implies $\overline{V}(1) > v$; since $\overline{V}$ is continuous at $t = 1$, there exists a positive-measure subset E of $[x_j, 1]$ such that $\overline{V}(t) > v$ for all $t \in E$. This, coupled with the fact $Q^{n_j} \geq n_j - 1 > Q$ on $(x_j, 1]$ (by the construction of Q^{n_j} and the choice of J), implies that condition (iii) listed in Lemma 2 holds when $Q^* = Q^{n_j}$, with x_j here playing the role of a and b there and, by construction of Q^{n_j}, not interior to any ironed interval. Thus, by the observation at the start of this section, $\int_0^1 Q^{n_j} V dF > \int_0^1 QV dF$, as desired.

Claim (b) Since $\overline{V}$ is weakly increasing, $\arg\max_{[0,1]} \overline{V}$ is equal to either $[c^*, 1]$ or $(c^*, 1]$ for some $c^* \leq 1$. Since $V(1) \leq V(a)$ for some $a < 1$, $c^* < 1$. Thus, the allocation Q^* is well-defined by (30). It is a a two-tier allocation because $c_* > 0$ due to the fact $\overline{V}(0) < \overline{V}(1)$ (Lemma 3). By (30), Q^* is market clearing. It is also budget balancing by Theorem 1. Thus, it suffices to show that Q^* maximizes $\int_0^1 QV dF$ among all weakly increasing $Q : [0, 1] \to [-1, \infty)$ subject to the market clearing condition. Pick any such Q. Note that Q^* corresponds to the special case of the Q^* defined in (31) where $a = c^*$ and $b = 1$. By the definition of c^*, c^* is not interior to any ironed interval. Thus the observation at the start of this section applies, and hence $\int_0^1 Q^* V dF \geq \int_0^1 QV dF$, as desired.

References

[1] Akbarpour, M., Budish,E., Dworczak, P., Kominers, S.D.: An economic framework for vaccine prioritization. Working Paper, December 18, 2021

[2] Akbarpour, M., Dworczak, P., Kominers, S.D.: Redistributive allocation mechanisms. Working Paper, December 16, 2020

[3] Bulow, J., Roberts, J.: The simple economics of optimal auctions. Journal of Political Economy 97(5), 1060–1090 (1989)

[4] Chien, H.-C.: Incentive efficient mechanism for partnership. Mimeo, June 1, 2007

[5] Cramton, P., Gibbons, R., Klemperer, P.: Dissolving a partnership efficiently. Econometrica 55(3), 615–632, (1987).

[6] Dworczak, P., Kominers, S.D., Akbarpour, M.: Redistribution through markets. Econometrica 89(4), 1665–1698 (2021)

[7] Gresik, T.: Incentive-efficient equilibria of two-party sealed-bid bargaining games. Journal of Economic Theory 68, 26–48 (1996)

[8] Holmström, B., Myerson, R.: Efficient and durable decision rules with incomplete information. Econometrica 51(6), 1799–1820 (1983)

[9] Kamenica, E., Gentzkow, M.: Bayesian persuasion. American Economic Review 101, 2590–2615 (2011)

[10] Kang, M., Zheng, C.Z.: Pareto optimality of allocating the bad. Working Paper. https://economics.uwo.ca/faculty/zheng/research/AllocatingTheBad.pdf (2020)

[11] Kang, Z.Y.: Optimal public provision of private goods. Mimeo, Stanford USB, December 2020.

[12] Kittsteiner, T.: Partnerships and double auctions with interdependent valuations. Games and Economic Behavior 44, 54–76 (2003)

[13] Laussel, D., Palfrey, T.: Efficient equilibria in the voluntary contributions mechanism with private information. Journal of Public Economic Theory 5, 449–478 (2003)

[14] Ledyard, J., Palfrey, T.: A characterization of interim efficiency with public goods. Econometrica 67, 435–448 (1999)

[15] Ledyard, J., Palfrey, T.: A general characterization of interim efficient mechanisms for independent linear environments. Journal of Economic Theory 133, 441–466 (2007)

[16] Loertscher, S., Wasser, C.: Optimal structure and dissolution of partnerships. Theoretical Economics 14, 1063–1114 (2019)

[17] Lu, H., Robert, J.: Optimal trading mechanisms with ex ante unidentifiable traders. Journal of Economic Theory 97, 50–80 (2001)

[18] Myerson, R.: Optimal auction design. Mathematics of Operations Research 6(1), 58–73 (1981)

[19] Myerson, R., Satterthwaite, M.A.: Efficient mechanisms for bilateral trading. Journal of Economic Theory 29, 265–281 (1983)

[20] Mylovanov, T., Tröger, T.: Mechanism design by an informed principal: Private values with transferable utility. Review of Economic Studies 81, 1668–1707 (2014)

[21] Pérez-Nievas, M.: Interim efficient allocation mechanisms. Working Paper 00-20, Departmento de Economia, Universidad Carlos III de Madrid, February 2000

[22] Reuter, M., Groh, C.-C.: Mechanism design for unequal societies. Discussion Paper No. 228, November 2020

[23] Royden, H.L., Fitzpatrick, P.M.: Real Analysis. Pearson Education, Boston, 4th edition (2010)

[24] Segal, I., Whinston, M.: Property rights and the efficiency of bargaining. Journal of the European Economic Association 14, 1287–1328 (2016)

[25] Sömet, T., Pathak, P.A., Ünever, U., Persad, G., Truog, R.D., White, D.B.: Categorized priority systems. Humanities: Vantage 159(3), 1294–1299 (2021)

[26] Robert Wilson. Incentive efficiency of double auctions. *Econometrica*, 53:1101–1105, 1985.

[27] Yang, L., Debo, L., Gupta, V.: Trading time in a congested environment. Management Science 63, 2377–2395 (2017)

[28] Yang, L., Wang, Z., Cui, S.: A model of queue scalping. Management Science 67 (2021). https://doi.org/10.1287/mnsc.2020.3865