Necessity of Auctions for Optimal Redistribution*

Mingshi Kang† Charles Z. Zheng‡

October 17, 2019

Abstract

Two items, one good, the other bad, may be assigned to n players, whose types determine their marginal rates of substitution of money. This paper characterizes the set of all interim Pareto optimal mechanisms. They are each in the form of auctions that may allocate the bad through rationing even when type-distributions are regular. When the Gini coefficient across types is above 1/2, Pareto optimality requires that the bad be assigned to someone sometimes, even though not assigning it at all is an option. Such assignment of the bad reduces inequality among types through giving larger surpluses to the types near the high and low ends than to those around the middle. The characterization of optimal mechanisms is derived from a class of nonlinear, concave functionals that we abstract from a player's countervailing incentives as his role endogenously switches between a buyer of the good and a receiver of the bad.

JEL Classification: C61, D44, D82

Keywords: mechanism design, optimal auction, redistribution, interim Pareto optimal mechanisms, countervailing incentive, ironing, Gini coefficient

---

*We thank Victor Aguiar, Roy Allen, Yi Chen, Rongzhu Ke, Scott Kominers, Vijay Krishna, Alexey Kushnir, Rohit Lamba, Greg Pavlov, Edward Schlee, Ron Siegel and the seminar participants at Penn State U., Ryerson U., CUHK, HK Baptist U., Lingnan U. College, the 2019 N. American ES Summer Meetings, the 2019 CETC, and the 2019 Stony Brook Game Theory Conference, for their questions and comments. Zheng acknowledges financial support from the Social Science and Humanities Research Council of Canada.

†Department of Economics, University of Western Ontario, London, ON, Canada, mkang94@uwo.ca.

‡Department of Economics, University of Western Ontario, London, ON, Canada, charles.zheng@uwo.ca, https://sites.google.com/site/charleszhenggametheorist/.

1 Introduction

This paper is motivated by the question how to induce Pareto improving wealth transfers across individuals. In order for wealth transfer to be Pareto improving, let us consider an environment where individuals may have different marginal rates of substitution of money. To induce voluntary wealth transfers, suppose that the social planner has two items, one good, the other bad, to assign to n players. For example, she wants to locate among n cities a high-tech giant's headquarter and an oil pipeline terminal. If the social planner sells the good to a player who values money less and uses the revenue to compensate another who values money more for receiving the bad, a Pareto improving wealth transfer is induced. Such a transfer, however, is only one instance among a large variety of redistribution that a social planner may deem Pareto improving. Depending on her value judgement, the social planner may favor one player against another, or favor one type of a player against another type of the same player, whether or not the former values money intrinsically more than the latter does. Thus, we assume no stand on interpersonal comparison, as one dollar for one type of a player may be deemed more valuable than one dollar for another type of the same or a different player. Rather we consider the entire set of interim Pareto optimal mechanisms, without assuming the existence of any rule according to which a social planner assigns welfare weights across players and across types of a player. That is, we shall find out the common pattern of all the Pareto optima not only among all players but also among all types of each player. The latter aspect makes this study relevant not only to mechanism design but also to macro settings where types in a continuum are interpreted as atomless individuals and players interpreted as sectors, regions, etc.

The model has n players, whose types are independently drawn from possibly different distributions. The positive values of the good, and the negative values of the bad, are commonly known. A player's type determines his marginal rate of substitution of money. Any mechanism committed to by the social planner is subject to the standard constraints: incentive compatibility (IC), (interim) individual rationality (IR) and (ex post) budget balance (BB). Interim Pareto improvement means an IC, IR and BB mechanism that makes a positive measure of some player's types better-off, and zero measure of every player's types worse-off, than the status quo. Interim Pareto optimality means IC, IR, BB and immunity to interim Pareto improvements. The problem is to characterize the set of all interim Pareto optimal mechanisms and identify their common features.

Each player's type drawn from a continuum, interim Pareto optimality is a design objective with infinite dimensions. In the mechanism design literature the design objective is usually one-dimension, such as the expected revenue, or a social welfare function aggregating individual preferences through exogenous welfare weights. It is rare to consider a design objective with finitely many dimensions such as Holström and Myerson's [5] incentive efficiency, let alone an objective with infinite dimensions and characterizing the optima thereof.[1]

Another feature of our design problem is each player's endogenously countervailing incentive: Depending on what the mechanism entails according to his realized type, a player may act as a buyer of the good sometimes, and as a recipient of the bad some other times. He would underreport his willingness to pay in the former event, and exaggerate his cost in the latter. By contrast, in the literatures of optimal auctions (Myerson [10]), optimal taxation (Mirrlees [9]) and bilateral trade (Myerson and Satterthwaite [12]), the role of a player is exogenous.

Our solution to this problem says that any interim Pareto optimal mechanism is necessarily in the form of auctions, with the winner-selection rule adjusted to the particularity of the optimum. First, for any interim Pareto optimum there is an associated *welfare weighting*, a profile of type-distributions across players, that aggregates individual preferences into a unidimensional social welfare function which the Pareto optimum maximizes subject to IC, IR and BB (Theorem 1). Second, the associated welfare weighting determines a rule to select the winner of the good, and another rule to select the "winner" of the bad, which the given Pareto optimum entails. These winner-selection rules together determine each player's expected value of money transfers in the Pareto optimal mechanism up to a constant, and the constant is determined by the expectation of the player's marginal rate of substitution of money weighed by the associated welfare weighting (Theorem 2).

This general characterization has implications on the optimal redistribution across players and that across types, suboptimality of the efficient allocation, and prevalence of rationing (Remarks 2–6). The most unexpected one is a relationship between the Gini

[1] Assuming infinite rather than finite type spaces is not just for the sake of technicalities. The finite-type assumption would undermine the relevance of the model to macro considerations of a continuum of agents and make it hard to relate to much of the mechanism design literature, where results are usually based on continuum types.

coefficient and the probability with which the bad is assigned to someone despite that the social planner could choose to not assign it at all. Theorem 3 says that, in any symmetric environment where the only heterogeneity among players is their types, any ex ante Pareto optimal mechanism assigns the bad with a strictly positive probability when the relative mean absolute difference (or twice the Gini index) among players is larger than one, despite the assumption that not assigning the bad at all is an option to the social planner.

We also find an inequality-reducing effect of assigning the bad. In the symmetric case, any Pareto optimal mechanism that assigns the bad with a positive probability gives larger surplus to the types near the high and low ends than to those around the middle (Figure 5). Thus, no matter how the types are permuted, being born with a high type does not imply ending with a high surplus (Remark 6).

One implication says that any interim Pareto optimum in our model is also an interim Pareto optimum in the matching environment that disallows a player to have both items (Remark 3). That complements the new literature of matching with transfers (Chiappori [1]) by showing that a particular auction mechanism achieves optimal matching.[2]

Our method to obtain these results has two novel aspects. The first is in the proof of Theorem 1, which reduces the infinite-dimensional objective to a unidimensional one that facilitates calculations of optimal mechanisms. Applying the Hahn-Banach theorem, this step resolves a dilemma, which often troubles infinite-horizon macro models (cf. Stokey, Lucas and Prescott [15, §15.4, §16.6]), between ensuring existence of a separating hyperplane and guaranteeing that the hyperplane can be properly represented. We resolve this dilemma because the separating hyperplane here needs only to be represented as a distribution rather than as an inner product operator, and because the hyperplane can be represented as a distribution due to a continuity observation in mechanism design.

The second novelty is a new method to integrate a player's information rent across his types that may switch between the countervailing incentives due to his allocation (Section 5.1.1). This method is encapsulated by a kind of nonlinear, *two-part operators* on allocations. Both the objective function and the joint constraint of IC, IR and BB for the optimal mechanism problem are reduced to such two-part operations. Being concave functionals of allocations, such operators guarantee that any optimal mechanism satisfies the saddle point condition (Lemma 5). Consequently, the associated Lagrangian is also reduced to an action

[2] According to Herodotus [4], auctions were used in ancient Babylon marriage matching markets.

on the allocations by a two-part operator. This two-part operator implies the formula for any maximizer of this Lagrangian and hence the formula for the optimal mechanism.

Deriving the formula for optimal mechanisms from the aforementioned two-part operator requires modifications on Myerson's [10] ironing technique because two-part operators are in general nonlinear in allocations. We bisect the Lagrangian maximization problem into two linear programmings, solve each while seting aside the other, and then show that the concatenation of the two solutions solves the original problem (Section 5.2.2 and Appendix B.6). We add a one-sided leveling operation to the ironing procedure in order to handle an additional constraint required by those linear problems (Appendix B.6.2).

Based on finitely many types but otherwise a highly general model, Myerson [11, §10.5] has characterized the set of incentive efficient mechanisms, or interim Pareto optima subject to IC constraints. The characterization is that any incentive efficient mechanism is a pointwise maximizer of the aggregate of virtual utility functions, each a function of the social choice outcome, the realized type profile, the welfare weights and the Lagrange multipliers for all the IC constraints with respect to the particular mechanism being characterized. To this profound, abstract perspective, our paper adds concrete, specific contents. Based on a continuum of types and a specific, allocation problem, our characterization is that an optimal mechanism allocates each item to a player whose realized type scores highest among all players whose realized types score above a threshold and that each player's score is a function of only his type. While the formula for the scoring functions in an optimal mechanism may depend on that mechanism, the auction-like pattern (allocating an item to a highest bidder), as well as some other implications of our characterization, is independent of which optimal mechanism is being characterized. Such disentanglement between the property of optima and the reference to a specific optimum is our main difference from Myerson's characterization.

Dworczak, Kominers and Akbarpour [3] have considered a model that captures wealth inequality by heterogeneous marginal rates of substitution (MRS) of money. They suggest that quasilinearity at the presence of wealth inequality is an appropriate local approximation when one's valuation of money is a smooth function of his wealth. Considering a bilateral trade environment, Dworczak et al. characterize the set of mechanisms that maximize the sum of the integrals across agents' utilities given exogenous welfare density functions of the agents (same as types in their model) such that the welfare density functions can be arbitrary. They use a novel technique and observe that the optimal design uses tax-like pricing mechanisms, with a wedge between the price for the buyers and that for the sellers. This paper builds upon their idea of capturing wealth inequality by heterogeneous MRS in a quasilinear setting. Our model differs from theirs in four aspects. First, we do not assume a unidimensional, utilitarian design objective such as a sum or an integral of utilities across types or agents; rather, the associated welfare weighting that aggregates preferences, across types and across players, is a consequence of the Pareto optimum under consideration (and the welfare weighting need not be representable as an inner product operator with its densities). Second, we have n players whose types are drawn from possibly different distributions; in their model, there is a continuum of i.i.d. buyers, and a continuum of i.i.d. sellers. Third, in our model a player's role—whether to be a seller or to be a buyer—is endogenous and hence has countervailing incentives, whereas in their model an agent's role is exogenously assumed. Fourth, the items in our model need not be assigned and hence the probability of assigning the good need not be equal to the probability of assigning the bad; in their model, market clearance requires that the aggregate probability of sales be equal to that of purchases.

Because of the first difference, this paper complements Dworczak et al. with our Theorem 1, which suggests that with a similar separating hyperplane argument their assumption of exogenous welfare densities might be relaxable. Because of the other differences, Pareto optima in our model are all auction-like mechanisms rather than the tax-like ones in their model. A player's bid in our model affects the type-cutoffs for other players to receive an item, whereas in their model an agent, atomless, has no influence on others. Because of the third difference, a player's surplus in our model is a non-monotone function of his type, while in theirs it is monotone. Applied to symmetric cases, this non-monotonicity implication says that our Pareto optimal mechanisms breaks the type-generated hierarchy through giving higher surpluses to types near the high and low ends than to those in the middle.

Countervailing incentives have been considered in the partnership dissolution literature, initiated by Cramton, Gibbons and Klemperer [2]. The focus of that literature is implementability of one particular winner-selection rule, the efficient allocation, which would be optimal if implementable and if the objective is the simple sum of surpluses across players. Loertscher and Wasser [6], differently, consider a design objective that is a convex combination between the auctioneer's expected revenue and the expected utility of the good for its final owner. Since the total money transfer from the players to the auctioneer is a plus rather than a negative in that objective,[3] their optimal mechanism squeezes the lowest surplus for each player down to the player's exogenous outside option. This outside option equation is crucial in Loertscher and Wasser's solution for the countervailing incentive problem. Our paper differs from the partnership dissolution literature by charactering the entire interim Pareto frontier. With players heterogenous in MRS of money, our counterpart of the efficient allocation is suboptimal even if it is implementable (Remark 5). The paper differs from Loertscher and Wasser also in the first and fourth aspects in which we differ from Dworczak et al. Consequently, our optimal mechanisms rebate surplus back to players and do not squeeze every player's lowest surplus down to his exogenous outside option. Hence Loertscher and Wasser's outside option equation is unavailable to us.

The following Section 2 illustrates why the bad is needed with a binary example. Then Section 3 defines the model and the design problem. Section 4 then presents the main results and implications. Section 5 sketches the proofs, with details relegated to the Appendix. The concept of two-part operators is introduced in Section 5.1.1. Our extension of the ironing technique is in Section 5.2.2 and Appendix B.6. Section 6 concludes.

2 Why the Bad is Needed: An Example

Consider within this section a binary example: There are only two players and the type for each is equal to either 1 or 6, each with probability 1/2; the value of the good is equal to 1, and that of the bad is equal to -1, to each player; given any type $t \in \{1, 6\}$, a player's expected payoff is equal to $x_A - x_B - y/t$ if he gets the good with probability x_A, the bad with probability x_B, and delivers monetary payment y (or receives payment $-y$), for any $(x_A, x_B, y) \in [0, 1]^2 \times \mathbb{R}$. Suppose within this section that the design objective is the *social surplus*, the sum of ex ante expected payoffs for both players.

If types were common knowledge then the bad is not needed at all to maximize the social surplus. To achieve the maximum, we simply assign the good always and, in the event where one player's type is high while the other's is low, transfer the maximum amount of money from the high-type player to the low-type player subject to the former's participation

[3] The expected utility of the good for its final owner (called social surplus by Loertscher and Wasser) is not equal to the total surplus among all players. That is because the total money transfer from the players to the auctioneer is not subtracted from the expected utility of the good for its final owner.

constraint, which means giving the former the good and transferring from him an amount equal to 6. That maximizes the social surplus to $1 + (1/2)(-6/6 + 6/1) = 7/2$. Assigning the bad with any positive probability can only lower this amount.

Suppose that types are private information. For simplicity of illustration within this example, a symmetric environment, let us consider only symmetric mechanisms. To demonstrate the necessity of the bad to maximize the social surplus, let us find out the constrained optimum subject to the restriction that the bad be never assigned. Given any incentive feasible symmetric revelation mechanism, let Q_t denote a player's expected probability of getting the good given his type being t, and let P_t denote the expected value of the money transfer from him to others. Incentive compatibility (IC) for both $t \in \{1, 6\}$ means

$$\begin{aligned} Q_6 - P_6/6 &\geq Q_1 - P_1/6, \\ Q_1 - P_1 &\geq Q_6 - P_6. \end{aligned}$$

Ex post budget balancing, combined with symmetry of the mechanism and equal probabilities of the two types, implies that $P_1 + P_6 = 0$. This, coupled with the IC conditions displayed above, implies

$$(Q_6 - Q_1)/2 \leq P_6 \leq 3(Q_6 - Q_1).$$

This implies $Q_6 - Q_1 \geq 0$ and hence $P_6 \geq 0$ and $P_1 = -P_6 \leq 0$. That means the individual rationality (IR) constraint for type 1, $Q_1 - P_1 \geq 0$, is non-binding. The IR for type 6, $Q_6 - P_6/6 \geq 0$, is also non-binding due to the above-displayed inequality. Thus, it is necessary for any constrained optimum that $P_6 = 3(Q_6 - Q_1)$. Hence the social surplus is equal to

$$2\left(\frac{1}{2}\left(Q_6 - \frac{P_6}{6}\right) + \frac{1}{2}(Q_1 - P_1)\right) = \frac{7}{2}Q_6 - \frac{3}{2}Q_1,$$

which is maximized when Q_6 is maximized and Q_1 minimized. Thus following is an optimum among symmetric mechanisms that do not assign the bad at all:

a. When both players report the high type, allocate the good randomly to one of them with equal probability and make no money transfer.

b. If both players report the low type, make no allocation and no money transfer.

c. If one player reports the high type and the other reports the low type, assign the good to the high-type player and have him transfer to the other player an amount equal to

$$2 \cdot 3(Q_6 - Q_1) = 6\left(\underbrace{\frac{1}{2}\cdot\frac{1}{2}+\frac{1}{2}}_{Q_6} - 0\right) = \frac{9}{2}.$$

Thus, given that the bad is never assigned, the optimal social surplus is equal to

$$\left(\frac{1}{2}\right)^2 + \underbrace{\frac{1}{2}\left(1 - \frac{9}{2}\cdot\frac{1}{6} + \frac{9}{2}\right)}_{\text{different types}} = \frac{21}{8}.$$

By contrast, consider another symmetric mechanism that assigns the bad sometimes. It stipulates the same rule as the previous mechanism except—

c*. If one player reports the high type and the other reports the low type, allocate the good to the high-type player and the bad to the low-type player, and have the high-type player transfer an amount of money equal to 6 to the low-type player.

The social surplus generated by this mechanism is equal to

$$\left(\frac{1}{2}\right)^2 + \underbrace{\frac{1}{2}\left(1 - 1 - \frac{6}{6} + 6\right)}_{\text{different types}} = \frac{11}{4},$$

which is larger than the social surplus from the previous mechanism. Thus, to enlarge the social surplus, it is necessary to assign the bad at least sometimes.[4]

3 The Model

3.1 The Good, the Bad, and n Players

Two items, named A and B, each indivisible, are to be allocated among n players ($n \geq 2$), each of whom can get one or both or none of the items. A social planner commits

[4] The bad cannot be replaced by "not winning the good." That is because the latter is already available in the constrained optimum obtained previously subject to the restriction of never assigning the bad, and we have seen that the constrained optimum can be improved upon by assigning the bad. Neither can the bad be replaced by splitting the good into two items, v_1 and v_2, such that $v_1 + v_2 = 1$ and $v_1, v_2 \in [0, 1]$. To see that, let q_t^1 denote the probability with which a type-t player gets the v_1 part of the good, and q_t^2 the probability with which he gets the v_2 part of the good. Then his interim expected payoff is equal to $v_1 q_s^1 + v_2 q_s^2 - P_s/t$ if he acts as type s given true type t. But then we can label $v_1 q_s^1 + v_2 q_s^2$ as Q_s to see that this setup is identical to the previous, one-good case with the restriction of never assigning the bad.

to a mechanism that may allocate the items to some players and may mamarginal rate of substitution of moneyndate money transfers among the players. The outcome to any player i takes the form (x_{iA}, x_{iB}, y_i), where x_{ij} $(\forall j \in \{A, B\})$ denotes the probability with which player i gets item j, and y_i the net money transfer from player i to others (with negative y_i signifying the transfer from others to i). After the mechanism is announced and before it is operated, each player, given his own private information, can opt out of the mechanism thereby getting the outcome $(0, 0, 0)$ for himself. Each player i's preference relation is represented by a vNM utility function

$$(x_{iA}, x_{iB}, y_i) \mapsto x_{iA} - cx_{iB} - \frac{y_i}{t_i}, \tag{1}$$

with $c \geq 0$ a constant across all players, and t_i player i's realized *type*. Thus, item A is interpreted as a good, and item B a bad, to all players; $1/t_i$ is interpreted as player i's marginal rate of substitution of money.

Assume that each player i's type t_i is independently drawn from a commonly known cumulative distribution function (CDF) F_i such that its support is $T_i := [a_i, b_i]$, its density f_i is positive on the support, and $a_i > 0$.

Remark 1 Assumption (1) plays an important role in the symmetric case where the social planner ranks all players and all types equally, as in the example of Section 2. There, were (1) replaced by the quasilinear private value assumption that equates a player i's payoff with $t_i (x_{iA} - cx_{iB}) - y_i$, money transfer would make no difference to the social surplus. Given (1), by contrast, money transfer changes the social surplus and, to induce the optimal amount of money transfer, the social planner needs to assign the bad to someone sometimes. By the same token, (1) is important to Theorem 3, which relates the dispersion among types to the necessity of the bad. However, (1) is not important to Theorems 1 or 2, which characterize Pareto optimality when welfare weights are unrestricted and hence money transfer may be deemed welfare improving even with the quasilinear private value assumption.[5]

[5] Assumption (1) can be generalized to $(x_{iA}, x_{iB}, y_i) \mapsto v_i x_{iA} - c_i x_{iB} - y_i/t_i$. Since our design objective, Pareto optimality, takes no stand on interpersonal comparison of utilities, there is no loss of generality to normalize v_i to one for all i. Thus we need only to extend (1) slightly to allow for different c_i across i. Our results can be extended to such heterogeneous c_i case.

3.2 Allocations and Mechanisms

For each player i, denote $T_{-i} := \prod_{j \neq i} T_j$, and let F_{-i} be the product measure on T_{-i} generated by $(F_j)_{j \neq i}$. An *ex post allocation* means a list $(q_{iA}, q_{iB})_{i=1}^n$ of functions such that $q_{iA}, q_{iB} : \prod_{j=1}^n T_j \to [0,1]$ for each i and, for each $t \in \prod_{j=1}^n T_j$,

$$\sum_i q_{iA}(t) \leq 1 \quad \text{and} \quad \sum_i q_{iB}(t) \leq 1.$$

An *ex post payment rule* means a list $(p_i)_{i=1}^n$ of functions such that $p_i : \prod_j T_j \to \mathbb{R}$ for each i. By the revelation principle, any equilibrium-feasible mechanism corresponds to a pair of ex post allocation $(q_{iA}, q_{iB})_{i=1}^n$ and ex post payment rule $(p_i)_{i=1}^n$, with $q_{ij}(t)$ interpreted as the probability with which item j ($j \in \{A, B\}$) is assigned to player i, and $p_i(t)$ the net money transfer from player i to others, when t is the profile of alleged types across players.

A list $(Q_i)_{i=1}^n$ of functions $Q_i : T_i \to \mathbb{R}$ ($\forall i$) is said *generated* by an ex post allocation $(q_{iA}, q_{iB})_{i=1}^n$ iff, for each $i = 1, \ldots, n$ and each $t_i \in T_i$,

$$Q_i(t_i) = \int_{T_{-i}} q_{iA}(t_i, t_{-i}) dF_{-i}(t_{-i}) - c \int_{T_{-i}} q_{iB}(t_i, t_{-i}) dF_{-i}(t_{-i}). \tag{2}$$

Likewise, a list $(P_i)_{i=1}^n$ of functions $P_i : T_i \to \mathbb{R}$ ($\forall i$) is said *generated* by an ex post payment rule $(p_i)_{i=1}^n$ iff, for each $i = 1, \ldots, n$ and each $t_i \in T_i$, $P_i(t_i) = \int_{T_{-i}} p_i(t_i, t_{-i}) dF_{-i}(t_{-i})$. Any list $(Q_i, P_i)_{i=1}^n$ such that $(Q_i)_{i=1}^n$ is generated by some ex post allocation $(q_{iA}, q_{iB})_{i=1}^n$, and $(P_i)_{i=1}^n$ generated by some ex post payment rule $(p_i)_{i=1}^n$, is called *reduced-form mechanism*, or *mechanism* for short.

3.3 Constraints

Given any (reduced-form) mechanism $(Q_i, P_i)_{i=1}^n$, it follows from the vNM utility function (1), and the shorthand (2), that the interim expected utility for any type $t_i \in T_i$ of player i to act type $\hat{t}_i$, given truthtelling from others, is equal to $Q_i(\hat{t}_i) - P_i(\hat{t}_i)/t_i$. Denote

$$U_i(t_i \mid Q, P) := \max_{\hat{t}_i \in T_i} Q_i(\hat{t}_i) - P_i(\hat{t}_i)/t_i. \tag{3}$$

Since $\inf T_i > 0$ by assumption, the maximization problem in (3) is equivalent to

$$\tilde{U}_i(t_i \mid Q, P) := \max_{\hat{t}_i \in T_i} t_i Q_i(\hat{t}_i) - P_i(\hat{t}_i). \tag{4}$$

Thus, as is routine in auction theory, *incentive compatibility* (IC) of $(Q_i, P_i)_{i=1}^n$ is equivalent to simultaneous satisfaction of two conditions for each player i: (i) Q_i is weakly increasing on T_i; (ii) for any $t_i, t_i^0 \in T_i$,

$$P_i(t_i) - P_i(t_i^0) = \int_{t_i^0}^{t_i} s dQ_i(s). \tag{5}$$

Since each player can opt out of a mechanism before it operates, his outside payoff is zero, hence $(Q_i, P_i)_{i=1}^n$ is said *individually rational* (IR) iff $U_i(t_i|Q, P) \geq 0$ for all i and all $t_i \in T_i$. By (3) and (4), $U_i(t_i|Q, P) = \tilde{U}_i(t_i \mid Q, P)/t_i$ for any $t_i \in T_i$, and it is routine to show that $\tilde{U}_i(\cdot \mid Q, P)$ is convex, with derivative almost everywhere equal to Q_i, which is weakly increasing by IC. Thus, $\tilde{U}_i(\cdot \mid Q, P)$ attains its minimum at

$$\tau(Q_i) := \inf \{t_i \in T_i : Q_i(t_i) \geq 0 \text{ or } t_i = b_i\}. \tag{6}$$

Consequently, $\tilde{U}_i(\tau(Q_i) \mid Q, P) \geq 0$ iff "$\tilde{U}_i(t_i \mid Q, P) \geq 0$ for all $t_i \in T_i$" iff "$U_i(t_i \mid Q, P) \geq 0$ for all $t_i \in T_i$." Thus, IR is equivalent to $\tilde{U}_i(\tau(Q_i) \mid Q, P) \geq 0$ for all players i.

For the society consisting of the n players to transfer wealth among themselves without relying on outside subsides, we require that a mechanism be always budget-balanced: $(Q_i, P_i)_{i=1}^n$ satisfies *budget balance* (BB) iff $(P_i)_{i=1}^n$ is generated by some ex post payment rule $(p_i)_{i=1}^n$ such that $\sum_i p_i(t) \geq 0$ for all $t \in \prod_i T_i$.

3.4 The Problem

To characterize a large class of Pareto optimal mechanisms, we use a strong notion of Pareto dominance based on interim, rather than ex ante, expected payoffs. A mechanism (Q^*, P^*) is *interim Pareto optimal* iff (i) (Q^*, P^*) is IC, IR and BB, and (ii) there does not exist any IC, IR and BB mechanism (Q, P) such that $u_i(\cdot \mid Q, P) \geq u_i(\cdot \mid Q^*, P^*)$ a.e. on T_i for all $i \in \{1, \ldots, n\}$ and, for some i, $u_i(\cdot \mid Q, P) > u_i(\cdot \mid Q^*, P^*)$ on a positive-measure subset of T_i. The problem is to characterize the set of all interim Pareto optimal mechanisms.

With interim Pareto optimality the welfare criterion, not only do we take no stand a priori regarding interpersonal comparison, we also allow for any inter-type comparison for the same player. That is, regardless of the *cardinal* interpretation of (1), the social planner may want to subsidize one player against another, or rank one type of a player higher than another type of the same player. Without even assuming the existence of such ranking rules, we shall find out the common feature of all interim Pareto optimal mechanisms.

4 The Solution

The common feature of all interim Pareto optimum, no matter how they rank across players and across types, is that they are necessarily in the form of auctions. This observation is formalized by Theorems 1 and 2 in Section 4.1. Remarks 2–5 then point out some implications. Section 4.2 focuses on the implication in the symmetric case and uncovers a link between inequality reduction and the necessity for allocating the bad at the optimal mechanism.

4.1 The Result: Common Features of All Interim Pareto Optima

Theorem 1, proved in Appendix A, reduces interim Pareto optimality, an objective with infinite dimensions, to a unidimensional, utilitarian objective. The operator that aggregates individual-type preferences to the utilitarian objective is encapsulated by a profile $(\lambda_i)_{i=1}^n$ of distribution functions associated with the Pareto optimum under consideration. By *distribution* on an interval $[a, b]$, we mean a real function on $\mathbb{R}$ that is weakly increasing on $\mathbb{R}$, right-continuous on (a, b), constant on (b, ∞), and equal to zero on $(-\infty, a)$.

Theorem 1 *For any interim Pareto optimal mechanism $(Q_i, P_i)_{i=1}^n$, there exists a profile $(\lambda_i)_{i=1}^n$ such that λ_i is a distribution on T_i for each i, $\lambda_i > 0$ on some positive-measure subset of T_i for some i, and $(Q_i, P_i)_{i=1}^n$ maximizes*

$$\sum_i \int_{T_i} U_i(\cdot \mid \tilde{Q}, \tilde{P}) d\lambda_i \tag{7}$$

among all IC, IR and BB mechanisms $(\tilde{Q}, \tilde{P})$.

The distribution profile $(\lambda_i)_{i=1}^n$ in Theorem 1 can be interpreted as the *welfare weighting*, across players and across types, that supports the given Pareto optimum as a maximum of the unidimensional social welfare (7) among all IC, IR and BB mechanisms $(\tilde{Q}, \tilde{P})$. Note that λ_i need not be absolutely continuous with respect to the prior distribution F_i, hence it need not have a derivative λ_i' for which the integral in (7) equals an inner product $\int_{T_i} U_i(\cdot \mid \tilde{Q}, \tilde{P})\lambda_i' dF_i$. Our characterization of optimal mechanisms does not need such inner product representation of λ_i, hence we make no assumption to force its absolute continuity.

Theorem 2, proved in Sections 5.1–5.2, characterizes all constrained maximizers of the social welfare function (7). It uses the following notations:

Q_i^+ **and** Q_i^-**:** For any $\phi : \mathbb{R} \to \mathbb{R}$, denote ϕ^+ and ϕ^- by $\phi^+(x) := \max\{\phi(x), 0\}$ and $\phi^-(x) := \max\{-\phi(x), 0\}$ for all $x \in \mathbb{R}$. Thus $\phi = \phi^+ - \phi^-$. (This definition applies only when $+$ or $-$ is the superscript and does not apply when they appear elsewhere such as in $Z_{i,+}$.)

$\overline{Z}_{i,+}$ **and** $\overline{Z}_{i,-}$**:** For any integrable function $\psi_i : T_i \to \mathbb{R}$, denote $\overline{\psi}_i$ for the weakly increasing function on T_i that results from ironing ψ_i according to Myerson [10, §6].

$\mathscr{A}\left((\overline{Z}_{i,+})_{i=1}^n\right)$ **and** $\mathscr{A}\left((-\overline{Z}_{i,-})_{i=1}^n\right)$**:** For any profile $(\varphi_i)_{i=1}^n$ of functions $\varphi_i : T_i \to \mathbb{R}$ $(\forall i)$, $\mathscr{A}\left((\varphi_i)_{i=1}^n\right)$ denotes the set of all profiles $(q_i)_{i=1}^n$ of functions $q_i : \prod_k T_k \to [0, 1]$ $(\forall i)$ such that $\sum_i q_i \le 1$ on $\prod_k T_k$ and, for almost every $(t_k)_{k=1}^n \in \prod_k T_k$ and for any i,

$$\begin{aligned} \varphi_i(t_i) > \max_{j \neq i} \varphi_j^+(t_j) \quad &\Rightarrow \quad q_i\left((t_k)_{k=1}^n\right) = 1, \\ \varphi_i(t_i) < \max_{j \neq i} \varphi_j^+(t_j) \quad &\Rightarrow \quad q_i\left((t_k)_{k=1}^n\right) = 0. \end{aligned}$$

Theorem 2 *For any profile $(\lambda_i)_{i=1}^n$ of distributions λ_i on T_i, if $(Q_i, P_i)_{i=1}^n$ maximizes (7) among all IC, IR and BB mechanisms, then there exists a profile $(Z_{i,+}, Z_{i,-})_{i=1}^n$ of integrable functions $Z_{i,+}, Z_{i,-} : T_i \to \mathbb{R}$ such that $Z_{i,+} \le Z_{i,-}$ and:*

a. for each i, $Q_i^+ = \int_{T_i} q_{iA}(\cdot, t_{-i}) dF_{-i}$ and $Q_i^- = c \int_{T_i} q_{iB}(\cdot, t_{-i}) dF_{-i}$ on T_i for some $(q_{iA})_{i=1}^n \in \mathscr{A}\left((\overline{Z}_{i,+})_{i=1}^n\right)$ and $(q_{iB})_{i=1}^n \in \mathscr{A}\left((-\overline{Z}_{i,-})_{i=1}^n\right)$;

b. $(P_i)_{i=1}^n$ is determined by $(Q_i)_{i=1}^n$ according to:

i. Eq. (5) for any i and any $t_i, t_i^0 \in T_i$;

ii. if $\int_{T_i} (1/s) d\lambda_i(s) < \int_{T_k} (1/s) d\lambda_k(s)$ for some $k \neq i$, then $\min_{t_i \in T_i} U_i(t_i \mid Q, P) = 0$;

iii. $\sum_{i=1}^n \int_{T_i} P_i(t_i) dF_i(t_i) = 0$.

Theorems 1 and 2 combined, any interim Pareto optimal mechanism is necessarily in form of auctions, encapsulated in a profile $(Z_{i,+}, Z_{i,-})_{i=1}^n$ of functions. These functions are jointly determined by the welfare weighting $(\lambda_i)_{i=1}^n$ supporting the Pareto optimum and the Lagrange multiplier for the constraint combining IR, IC and BB. Here $Z_{i,+}$ corresponds to the *virtual surplus* to the society contributed by player i who acts as a buyer of the good, and $Z_{i,-}$ the virtual surplus contributed by i who acts as a seller of the service of receiving the bad (Figure 1). A crucial feature, asserted by Theorem 2, is that $Z_{i,-}$ is above $Z_{i,+}$.

From $(Z_{i,+}, Z_{i,-})_{i=1}^n$, one finds out how the Pareto optimal mechanism allocates the two items according to Claim (a) of Theorem 2. First, obtain the ironed copy $\overline{Z}_{i,+}$ of $Z_{i,+}$

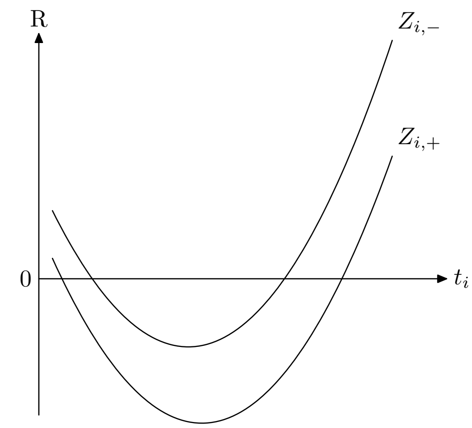

Figure 1: The bifurcated virtual surplus

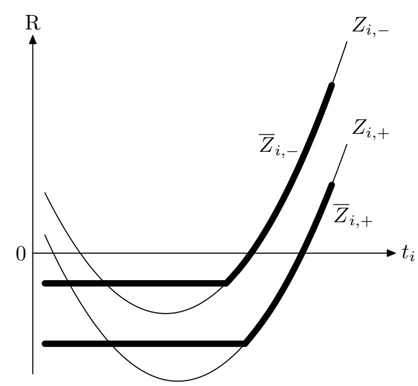

Figure 2: The thick curves: The ironed virtual surplus

and the ironed copy $\overline{Z}_{i,-}$ of $Z_{i,-}$, for each i (Figure 2). Second, assign the good (item A) by the rank of $\left(\overline{Z}_{i,+}\right)_{i=1}^{n}$ à la Myerson [10]: Score each player i's alleged type t_i according to the ironed function $\overline{Z}_{i,+}$, and assign the good to a player with the highest ironed Z score provided that it is positive. Likewise, assign the bad (item B) by the rank of $\left(-\overline{Z}_{i,-}\right)_{i=1}^{n}$: Score each player i's alleged type t_i according to the ironed function $\overline{Z}_{i,-}$, and assign the bad to a player with the lowest ironed Z score provided that it is negative.

These two assignments together generate the reduced form allocation $(Q_i)_{i=1}^{n}$ of the mechanism. From $(Q_i)_{i=1}^{n}$ the payment rule $(P_i)_{i=1}^{n}$ is derived via the envelope equation and minimum surplus condition in Claim (b) of Theorem 2. There, $\int_{T_i}(1/s)d\lambda_i(s)$ is player i's average weight in the social welfare that incorporates both the welfare weight λ_i on his various types and his marginal valuations of money given these types. Claim (b.ii) says that if this weight of i is less than that of someone else then player i has zero as his minimum surplus in the mechanism. This, coupled with Claim (b.iii), that the auctioneer retains no

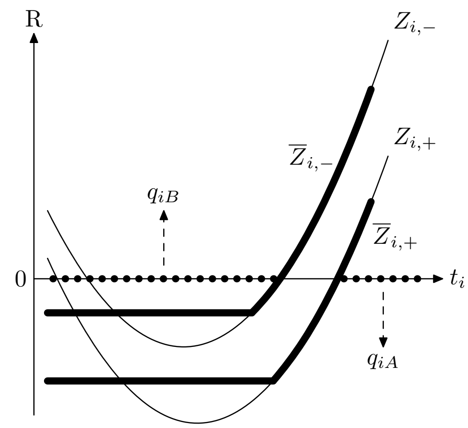

Figure 3: The optimal allocation

money surplus, specifies the payment rule.

Remark 2 (Redistribution across Players) Recall that $t_i U_i(t_i \mid Q, P) = \tilde{U}_i(t_i \mid Q, P)$ (by (3) and (4)), where $\tilde{U}_i$ denotes i's surplus in the units of money. If $\min_{t_i \in T_i} U_i(t_i \mid Q, P)$ is positive, even the type $\tau(Q_i)$ that receives the lowest surplus among i's types gets a money surplus in the amount of $\tau(Q_i) \min_{t_i \in T_i} U_i(t_i \mid Q, P)$. This amount corresponds to the lump sum transfer to player i in the optimal mechanism. By Claim (b.ii), the lump sum transfer goes only to those players i whose average welfare weights, $\int_{T_i}(1/s)d\lambda_i(s)$, are maximum among all players. Importantly, the direction of lump sum transfers in general is not tied to prior distributions $(F_i)_{i=1}^n$ of types, but rather by the welfare weighting $(\lambda_i)_{i=1}^n$ associated with the Pareto optimum. Thus, an interim Pareto optimal mechanism need not take money from the (ex ante) stochastically high-type players ("the rich") to subsidize the stochastically low-type ones ("the poor"). Only when the endogenous welfare weighting $(\lambda_i)_{i=1}^n$ is identical to the exogenous type-distribution profile $(F_i)_{i=1}^n$ would the optimal mechanism necessarily direct lump sum transfers from the stochastically rich to the stochastically poor.

Remark 3 (Exclusive Assignments) While we do not impose the condition that the good and the bad be assigned to different players, any interim Pareto optimal mechanism also satisfies that condition (Corollary 1, Appendix D.1). That is because, at any Pareto optimum, a player-type that has a positive probability to get the good has zero probability to get the bad, and vice versa: In Figure 3, since $Z_{i,-}$ is above $Z_{i,+}$, any type with a nonngegative ironed $\overline{Z}_{i,+}$ score is larger than any type with a negative $\overline{Z}_{i,-}$ score.

Remark 4 (Rationing) The virtual surplus functions $(Z_{i,+}, Z_{i,-})_{i=1}^n$ cannot be guaranteed monotone by restrictions on the exogenous, prior type-distributions (Corollary 2, Appendix D.2). Consequently, if players' types are i.i.d. and the welfare weighting treats the players equally, the optimal mechanism entails ironing for all realized types belonging to the lower sub-interval in the type support. That is, the bad is allocated through an egalitarian lottery among those players whose realized types belong to the lower sub-interval. We will provide a heuristic for such non-monotonicity in the next subsection, which identifies an inherently non-monotone component of the Z functions.

Remark 5 (Suboptimality of the Efficient Allocation) In our model, the efficient allocation means always allocating the good to a player with the highest realized type, and the bad to a player with the lowest realized type. With ironing prevalent (Remark 4), the efficient allocation in general does not belong to the Pareto frontier.

4.2 Gini Coefficient and the Social Value of the Bad

There is a link between the virtual surplus functions $(Z_i)_{i=1}^n$, the type-dispersion index, and the necessity for the optimal mechanism to sometimes allocate the bad to someone. To focus on the inequality among players in terms of their types, within this subsection we restrict attention to the *symmetric environment*, where every player i's type is independently drawn from the same distribution F, with support $[a, b]$ and density f. In this environment, players differ from one another only in their realized types t_i or, equivalently, in their realized marginal rates of substitution (MRS) of money, $1/t_i$. Given the CDF F of types, the corresponding CDF of MRS is

$$M_F(s) := 1 - F(1/s) \tag{8}$$

for all $s \in [1/b, 1/a]$. To capture the dispersion of MRS among the players, define

$$\Delta(F) := \frac{\int_{1/b}^{1/a} \int_{1/b}^{1/a} |x - y| dM_F(x) dM_F(y)}{\int_{1/b}^{1/a} s dM_F(s)}, \tag{9}$$

which is the relative mean absolute difference among all possible MRS. The significance of $\Delta(F)$ is that the bad should be assigned with a strictly positive probability if $\Delta(F) > 1$:

Theorem 3 *If $F_i = F$ for all players i and if $\Delta(F) > 1$, then for any $(\omega_i)_{i=1}^n \in \mathbb{R}_+^n \setminus \{\mathbf{0}\}$, the probability with which the bad is allocated to someone is strictly positive in any mechanism that maximizes $\sum_i \omega_i \int_a^b U_i(t_i \mid Q, P) dF(t_i)$ among all (Q, P) subject to IC, IR and BB.*

Theorem 3, proved in Appendix C, says that any ex ante Pareto optimal mechanism assigns the bad with a strictly positive probability when the type-distribution F is disperse enough for $\Delta(F) > 1$. We use *ex ante* Pareto optimality here—the welfare weight ω_i in the objective $\sum_i \omega_i \int_a^b U_i(t_i \mid Q, P) dF(t_i)$ being invariant to player i's types—in order to focus on the effect of the exogenous type-distribution F. Had the social planner ranked the types differently than F, whether the exogenous F is disperse or not would of course not necessarily matter. Allowing for different welfare weights across players, Theorem 3 is applicable to the case where the social planner regards player i's welfare more important than player j's and, equivalently, to the environment where players have heterogeneous, commonly known, valuations of their net consumptions of the good versus the bad (cf. Footnote 5).

Theorem 3 is due to an algebraic relationship between the index $\Delta(F)$ and the Gini coefficient on the distribution of MRS, and that between the Gini coefficient and the Z functions that determine the optimal mechanism according to Theorem 2. To uncover these relationships, define, for any $s \in [1/b, 1/a]$,

$$L_F(s) := \frac{\int_{1/b}^{s} r dM_F(r)}{\int_{1/b}^{1/a} r dM_F(r)}. \tag{10}$$

By the definition of L_F, the mapping $M_F(s) \mapsto L_F(s)$, from $[0, 1]$ to $[0, 1]$, corresponds to the Lorenz curve of the distribution of MRS, as in Figure 4.[6] Thus, $M_F(s) - L_F(s)$ is the

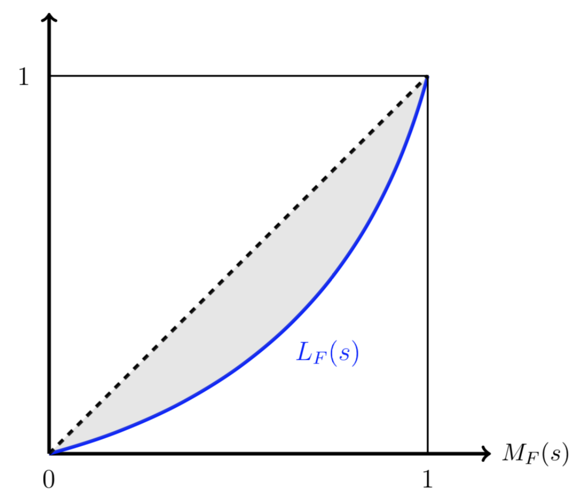

Figure 4: $M_F(s)$: The cumulative mass of individuals whose MRS $\leq s$

[6] One can verify that $L_F(1/b) = 0 = M_F(1/b)$, $L_F(1/a) = 1 = M_F(1/a)$ and $L_F(s) < M_F(s)$ for all $s \in (1/b, 1/a)$.

gap between the 45-degree line and the Lorenz curve in $[0,1]^2$, and the Gini coefficient is

$$G(F) := 2\int_{1/b}^{1/a} \left(M_F(s) - L_F(s)\right) dM_F(s). \tag{11}$$

The following equation is known to the inequality literature:[7]

$$\Delta(F) = 2G(F). \tag{12}$$

For convenience of illustration, let us assume equal welfare weights across all players for now, so that the design objective is $\sum_i \int_a^b U_i(t_i \mid Q, P) dF(t_i)$. Consider any optimal mechanism for this objective and let $(Z_{i,+}, Z_{i,-})_{i=1}^n$ be the associated function profile that determines the optimal mechanism according to Theorem 2. By symmetry across players, one can show (Appendix C) that there exists a $\beta > 0$ and a $Z_* : [a,b] \to \mathbb{R}$ such that, for all players i, $Z_{i,+} = Z_{i,-} = \beta Z_*$ on $[a,b]$ if the bad is not allocated at all; furthermore,

$$Z_*(t) = t - \frac{1}{f(t)}\left(M_F(1/t) - L_F(1/t)\right) \tag{13}$$

for all $t \in [a,b]$. By Theorem 2.a and the definition of ironing, one can prove (Appendix C) that the mechanism allocates the bad with a strictly positive probability if $\int_a^t Z_*(r)dF(r) < 0$ for some $t \in (a,b)$. This condition, with (10) and (13) plugged in, is true if

$$\int_{1/b}^{1/a} \left(M_F(s) - L_F(s)\right) dM_F(s) > \frac{\int_{1/b}^{1/a} sF(1/s) dM_F(s)}{\int_{1/b}^{1/a} r dM_F(r)}.$$

The left-hand side, by (11), is equal to $G(F)/2$; the right-hand side, one can show with (8), is equal to $1/2 - G(F)/2$. Thus the optimal mechanism assigns the bad with a strictly positive probability if $G(F) > 1/2$, which due to (12) is the same as $\Delta(F) > 1$.

Remark 6 (Inequality Reduction) To see how allocation of the bad reduces inequalities among individuals, recall the standard mechanism design models where a player's role, whether to be a buyer or to be a seller, is assumed a priori. Thus incentive compatibility there implies that his equilibrium surplus from any mechanism is necessarily monotone (weakly increasing or weakly decreasing). Hence the player's possible types can be permuted so that the mechanism gives higher types higher expected payoffs. That is, inequality in types begets inequality in payoffs. In our model, by contrast, a player's role is endogenous, and incentive

[7] Sen [14, p30].

compatibility implies that he acts as a seller—selling the service of receiving the bad—when his realized type is sufficiently low, and a buyer when his realized type is sufficiently high. Consequently, a player's expected surplus is decreasing when his type—functioning as the cost of receiving the bad—increases in a low interval, and is increasing when his type—functioning as the value of the good—increases in a high interval. In other words, in the symmetric environment, a player's interim expected payoff in the optimal mechanism is like the roughly U-shape curve in Figure 5 (formalized as Corollary 3, Appendix D.3). Thus the

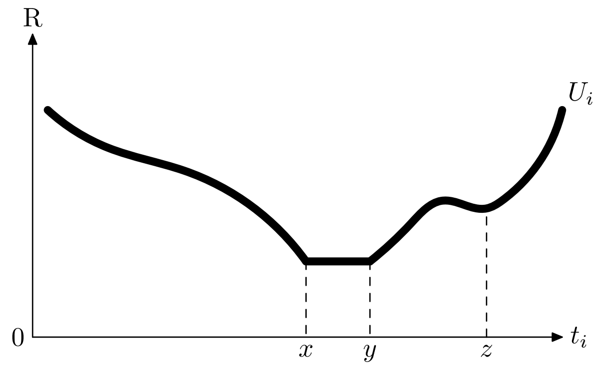

Figure 5: Non-monotone interim expected payoff functions

ranking of individuals in their types does not propagate to the ranking in their payoffs, no matter how types are permuted. What is crucial to such non-monotonicity is having two items of opposite values, one good, the other bad.

Remark 7 (Non-monotone Virtual Surplus and the Lorenz Gap) Eq. (13) suggests a heuristic to understand why the virtual surplus Z can be non-monotone even with regular distributions (e.g., uniform, exponential, square, Pareto, polynomial): On the right-hand side of (13), the term $M_F(1/t) - L_F(1/t)$, which is the gap between the 45-degree line and the Lorenz curve on $[0,1]^2$, is non-monotone, as in Figure 4. When the scalar $1/f(t)$ of this gap is large for all t in some interval of $[a,b]$, say around a, this non-monotone term may be scaled up to dominate the other components of Z at those t.

5 The Method

To characterize the Pareto frontier, our first step is to quantify the infinite-dimensional design objective, interim Pareto optimality, into a one-dimension, utilitarian, social welfare function (Theorem 1, proved in Appendix A). It uses the Hahn-Banach theorem to obtain a welfare weighting $(\lambda_i)_{i=1}^n$ that supports the Pareto optimal mechanism under consideration as a constrained maximum of the unidimensional objective (7). Since a mechanism corresponds to functions defined on a continuum, the choice of the function space requires care. On one hand, the space needs to satisfy the nonempty interior condition for existence of a linear functional on the space that supports the Pareto optimum. On the other hand, the space needs to guarantee that the linear functional be representable by a profile of distributions in the form $(\lambda_i)_{i=1}^n$. Our choice of the function space stems from an observation that the interim expected surplus generated by any IC mechanism is a continuous function of types. Hence our function space consists of continuous functions defined on compact intervals, conducive to the Hahn-Banach and representation theorems.

The second step (Theorem 2, proof summarized by Section 5.2.3) is to characterize any mechanism that is a constrained maximizer of some utilitarian social welfare function obtained in the first step. To that end, Section 5.1 incorporates part of the IC constraint and optimality condition into the social welfare function thereby obtaining a tractable optimization problem. Then Section 5.2 solves this optimization problem through bisecting the associated Lagrange problem into two linear programmings. Instrumental to the entire second step is a new kind of operators—two-part operators—that calculate a player's endogenously countervailing information rents (Section 5.1.1).

5.1 Calculating the Objective with Two-Part Operators

Theorem 1, coupled with (3), implies that any interim Pareto optimum is a maximum of

$$\sum_i \int_{T_i} Q_i(t_i) d\lambda_i(t_i) - \sum_i \int_{T_i} \frac{P_i(t_i)}{t_i} d\lambda_i(t_i) \tag{14}$$

among all mechanisms $(Q_i, P_i)_{i=1}^n$ subject to IC, IR and BB, given some profile $(\lambda_i)_{i=1}^n$ of distributions λ_i on T_i specified in Theorem 1. For any i and any t_i, define

$$\Lambda_i(t_i) := \int_{a_i}^{t_i} \frac{1}{s} d\lambda_i(s), \tag{15}$$

$$\beta_\lambda := \max_{i=1,\ldots,n} \Lambda_i(b_i). \tag{16}$$

By (15), $t_i \mapsto 1/t_i$ is the Radon-Nikodym derivative $\frac{d\Lambda_i}{d\lambda_i}$, hence (14) is equal to

$$\sum_i \int_{T_i} Q_i d\lambda_i - \sum_i \int_{T_i} P_i d\Lambda_i. \tag{17}$$

The integral $\int_{T_i} P_i d\Lambda_i$ in the second sum is the ex ante expected revenue—measured by the endogenous Λ_i rather than the exogenous F_i—that one can extract from player i after deducting the expected information rent necessary to incentivize i in the allocation Q_i. Because i's incentive is that of a buyer when $Q_i(t_i) > 0$, and that of a seller (recipient of the bad) when $Q_i(t_i) < 0$, the information rent calculation is an operation, introduced next, that bifurcates according to the sign of $Q_i(t_i)$ for each realized type t_i.

5.1.1 Two-Part Operators

For any player i and any three integrable functions $Q_i, \varphi_{i,+}, \varphi_{i,-} : T_i \to \mathbb{R}$, denote $\varphi_i := (\varphi_{i,+}, \varphi_{i,-})$, called *two-part function*, and define

$$\langle Q_i : \varphi_i| := \int_{T_i} Q_i^+(s)\varphi_{i,+}(s)ds - \int_{T_i} Q_i^-(s)\varphi_{i,-}(s)ds. \tag{18}$$

Thus the operation $Q_i \mapsto \langle Q_i : \varphi_i|$ acts on the function Q_i in two parts, one on the positive part Q_i^+ of Q_i, the other on the negative part, $-Q_i^-$. The asymmetric bracket of Q_i and φ_i is to highlight the asymmetry between the two arguments: Obviously, $\langle Q_i : \varphi_i|$ is not linear in Q_i unless $\varphi_{i,+} = \varphi_{i,-}$; by contrast, $\langle Q_i : \varphi_i|$ is a linear functional of φ_i. Any integrable function $g_i : T_i \to \mathbb{R}$ is the same as a two-part function $(g_{i,+}, g_{i,-})$ such that $g_{i,+} = g_{i,-} = g_i$.

A two-part function $\varphi_i := (\varphi_{i,+}, \varphi_{i,-})$ on T_i is said *well-ordered* iff $\varphi_{i,+} \leq \varphi_{i,-}$ a.e. on T_i. Next is an important property of well-ordered two-part functions (proved in Appendix B.1).

Lemma 1 *For any i and any two-part function φ_i that is well-ordered, $Q_i \mapsto \langle Q_i : \varphi_i|$ is a concave functional on the space of integrable functions defined on T_i.*

For any i and any distribution μ_i on T_i, define a two-part function $\rho(\mu_i) := (\rho_+(\mu_i), \rho_-(\mu_i))$ by letting, for any $t_i \in T_i$,

$$\rho_+(\mu_i)(t_i) := -\int_{T_i} d\mu_i + \int_{a_i}^{t_i} d\mu_i, \tag{19}$$

$$\rho_-(\mu_i)(t_i) := \int_{a_i}^{t_i} d\mu_i. \tag{20}$$

Obviously $\rho(\mu_i)$ is well-ordered. The next lemma says that $\int_{T_i} P_i d\Lambda_i$ in the design objective (17) amounts to an action on the allocation Q_i by the two-part function $\rho(\mu_i)$ when $\mu_i = \Lambda_i$. Thus $\rho_+(\mu_i)$ reflects i's information rent density when i acts as a buyer, and $\rho_-(\mu_i)$, i's information rent density when i acts as a seller, had i's type been measured by μ_i.

Lemma 2 *For any IC mechanism (Q,P), any player i and any distribution μ_i on T_i, with the notation in (4) and (6),*

$$\int_{T_i} P_i d\mu_i = \langle Q_i : \rho(\mu_i)| + \int_{T_i} t_i Q_i(t_i) d\mu_i(t_i) - \tilde{U}_i\left(\tau(Q_i)\right) \int_{T_i} d\mu_i. \tag{21}$$

5.1.2 The Budget Balance Condition Combined with IC and IR

Denote $\mathbb{I}$ for the identity map $s \mapsto s$ on $\mathbb{R}$. For any functions $g, h : \mathbb{R} \to \mathbb{R}$, denote gh for the pointwise product between g and h, so that $(gh)(s) = g(s)h(s)$ for all $s \in \mathbb{R}$.

Lemma 3 *For any allocation $(Q_i)_{i=1}^n$ such that Q_i is weakly increasing on T_i for any i, if $(Q_i)_{i=1}^n$ constitutes an IC, IR and BB mechanism then*

$$\sum_i \langle Q_i : \mathbb{I}f_i + \rho(F_i)| \geq 0; \tag{22}$$

conversely, if (22) holds then there exists an ex post payment rule $(p_i)_{i=1}^n$,

$$\sum_i p_i(t) = \sum_i \langle Q_i : \mathbb{I}f_i + \rho(F_i)| \tag{23}$$

for any $t \in \prod_j T_j$, which coupled with Q constitutes an IC, IR and BB mechanism such that IR is binding at some $t_i \in T_i$ for every i.

Proof By Lemma 2, IC of (Q,P) implies (21). Plug $\mu_i = F_i$ into (21) (and recall $f_i = F_i'$) to obtain

$$\begin{aligned} \int_{T_i} P_i dF_i &= \langle Q_i : \rho(F_i)| + \int_{T_i} t_i Q_i(t_i) f_i(t_i) dt_i - \tilde{U}_i\left(\tau(Q_i)\right) \\ &= \langle Q_i : \rho(F_i)| + \langle Q_i : \mathbb{I}f_i| - \tilde{U}_i\left(\tau(Q_i)\right), \end{aligned}$$

where the second line comes from the notations of $\mathbb{I}$ and pointwise product $\mathbb{I}f_i$, and the fact that $\int_{T_i} Q_i(s)\psi(s)ds = \langle Q_i : \psi|$ for any integrable function $\psi : T_i \to \mathbb{R}$, as ψ is a special two-part function such that $\psi_+ = \psi_- = \psi$. Then, since $\varphi \mapsto \langle Q_i : \varphi|$ is linear,

$$\int_{T_i} P_i dF_i = \langle Q_i : \mathbb{I}f_i + \rho(F_i)| - \tilde{U}_i\left(\tau(Q_i)\right).$$

Sum this equation across $i = 1, \ldots, n$ to get

$$\sum_i \int_{T_i} P_i dF_i = \sum_i \langle Q_i : \mathbb{I}f_i + \rho(F_i)| - \sum_i \tilde{U}_i\left(\tau(Q_i)\right).$$

BB implies that the left-hand side is nonnegative and hence

$$\sum_i \langle Q_i : \mathbb{I}f_i + \rho(F_i)| \geq \sum_i \tilde{U}_i\left(\tau(Q_i)\right), \tag{24}$$

which coupled with IR (Section 3.3) implies (22). Thus (22) is implied by IC, IR and BB. The proof of the converse is routine and hence relegated to Appendix B.3. ■

5.1.3 The Objective with Optimal Payment Rules

Now we calculate the objective (14) by incorporating (21) and optimality of the payment rule. Denote $\mathscr{Q}$ for the set of all (reduced-form) allocations $(Q_i)_{i=1}^n$, each generated by some ex post allocation according to (2). Let $\mathscr{Q}_{\text{mon}}$ be the set of all $(Q_i)_{i=1}^n \in \mathscr{Q}$ such that Q_i is weakly increasing for every i.

Lemma 4 *For any profile $\lambda := (\lambda_i)_{i=1}^n$ of distributions specified in Theorem 1, denote Λ and β_λ by (15)–(16); then maximization of (7) subject to IC, IR and BB is equivalent to*

$$\begin{array}{cc} \max\limits_{Q\in\mathscr{Q}_{\text{mon}}} & \sum_i \langle Q_i : \beta_\lambda\left(\mathbb{I}f_i + \rho(F_i)\right) - \rho(\Lambda_i)| \\ s.t. & \sum_i \langle Q_i : \mathbb{I}f_i + \rho(F_i)| \geq 0. \end{array} \tag{25}$$

Proof By Sections 3.3 and Lemma 3, the constraints $Q \in \mathscr{Q}_{\text{mon}}$ and $\sum_i \langle Q_i : \mathbb{I}f_i + \rho(F_i)| \geq 0$ together constitute the choice set for the problem. We still need to show that the objective in (25) is equal to (7), i.e., equal to (17). By Lemma 2, IC of (Q, P) implies (21). Plug $\mu_i = \Lambda_i$ into (21) and note $d\Lambda_i(s) = (1/s)d\lambda_i$ by (15) to obtain

$$\begin{aligned} \int_{T_i} P_i d\Lambda_i &= \langle Q_i : \rho(\Lambda_i)| + \int_{T_i} sQ_i(s)(1/s)d\lambda_i(s) - \tilde{U}_i\left(\tau(Q_i)\right)\int_{T_i} d\Lambda_i \\ &= \langle Q_i : \rho(\Lambda_i)| + \int_{T_i} Q_i d\lambda_i - \tilde{U}_i\left(\tau(Q_i)\right)\int_{T_i} d\Lambda_i. \end{aligned}$$

Sum this across i and plug the equation obtained thereof into (17) to see that the objective (14) is equal to

$$\sum_i \tilde{U}_i\left(\tau(Q_i)\right)\int_{T_i} d\Lambda_i - \sum_i \langle Q_i : \rho(\Lambda_i)|. \tag{26}$$

By (16) and (24),

$$\sum_i \tilde{U}_i\left(\tau(Q_i)\right)\int_{T_i} d\Lambda_i \leq \beta_\lambda \sum_i \tilde{U}_i\left(\tau(Q_i)\right) \leq \beta_\lambda \sum_i \langle Q_i : \mathbb{I}f_i + \rho(F_i)|.$$

Furthermore, the right end of this inequality can be attained: Pick a player i_* for whom $\int_{T_{i_*}} d\Lambda_{i_*} = \beta_\lambda$; for any realized type profile $t \in \prod_i T_i$ and any $i \neq i_*$, set the money transfer $p_i^*(t)$ from i to others to be $p_i(t)$, with p_i being the ex post payment rule in (23); set the money transfer $p_{i_*}^*(t)$ from i_* to others as $p_{i_*}(t) - \sum_i \langle Q_i : \mathbb{I} f_i + \rho(F_i)|$. Given $(p_i^*)_{i=1}^n$, BB follows from (23), and $\tilde{U}_i(\tau(Q_i)) = 0$ for all $i \neq i_*$, while $\tilde{U}_i(\tau(Q_{i_*})) = \sum_i \langle Q_i : \mathbb{I} f_i + \rho(F_i)|$. Thus, given the allocation Q, when the payment is optimized, we have

$$\sum_i \tilde{U}_i(\tau(Q_i)) \int_{T_i} d\Lambda_i = \beta_\lambda \sum_i \langle Q_i : \mathbb{I} f_i + \rho(F_i)| .$$

Hence there is no loss of generality to assume that (7), or (26), is equal to

$$\beta_\lambda \sum_i \langle Q_i : \mathbb{I} f_i + \rho(F_i)| - \sum_i \langle Q_i : \rho(\Lambda_i)| .$$

This, by linearity of $\varphi_i \mapsto \langle Q_i : \varphi_i|$, is equal to

$$\sum_i (\langle Q_i : \beta_\lambda (\mathbb{I} f_i + \rho(F_i))| - \langle Q_i : \rho(\Lambda_i)|) = \sum_i \langle Q_i : \beta_\lambda (\mathbb{I} f_i + \rho(F_i)) - \rho(\Lambda_i)| ,$$

the objective in (25). ■

Remark 8 (Optimal Choice of Transfer Rules) By (24), any payment rule $\hat{p}$ that renders $\tilde{U}_i(\tau(Q_i)) > 0$ while $\int_{T_i} d\Lambda_i < \beta_\lambda$ makes $\sum_i \tilde{U}_i(\tau(Q_i)) \int_{T_i} \Lambda_i < \beta_\lambda \sum_i \langle Q_i : \mathbb{I} f_i + \rho(F_i)|$. Since the proof of Lemma 4 has shown that the right-hand side of this inequality is attainable, the payment rule $\hat{p}$ is suboptimal. This, coupled with the definitions of $\tau(Q_i)$ and $\tilde{U}_i$, implies Claim (b.ii) of Theorem 2. Claim (b.iii) is obvious. If it does not hold, there is a positive expected money surplus, which can be equally distributed to the players independently of their types thereby achieving Pareto improvement, contradiction.

5.2 Solving the Constrained Optimization Problem

To solve the optimization problem (25), first we reformulate it through the saddle point condition, which delivers the profile $(Z_{i,+}, Z_{i,-})_{i=1}^n$ of functions stated in Theorem 2. Then we show that any maximum of the associated Lagrangian is the combination of two auctions according to $(Z_{i,+}, Z_{i,-})_{i=1}^n$. Maximization of this Lagrangian a nonlinear programming problem. We solve it through bisecting it into two linear programmings, solving each with the other set aside, and then showing that the two solutions are compatible with each other and together attain the maximum of the original problem.

5.2.1 Deriving the $(Z_i)_{i=1}^n$ Functions through the Saddle Point Condition

Recall that $\mathscr{Q}$ denotes the space of all allocations $(Q_i)_{i=1}^n$, each generated by some ex post allocation according to (2). It is easy to verify that $\mathscr{Q}$ belongs to a normed linear space. Endow $\mathscr{Q}$ with such a norm.[8] Also recall $\mathscr{Q}_{\text{mon}}$ as the set of $(Q_i)_{i=1}^n \in \mathscr{Q}$ such that Q_i is weakly increasing for any i.

Denote ν for the Lagrange multiplier of the constraint $\sum_i \langle Q_i : \mathbb{I}f_i + \rho(F_i)| \geq 0$ in (25). The Lagrangian associated with (25) is

$$\begin{aligned} \mathscr{L}(Q,\nu) &:= \sum_i \langle Q_i : \beta_\lambda \left(\mathbb{I}f_i + \rho(F_i)\right) - \rho(\Lambda_i)| + \nu \sum_i \langle Q_i : \mathbb{I}f_i + \rho(F_i)| \\ &= \sum_i \langle Q_i : \beta_\lambda \left(\mathbb{I}f_i + \rho(F_i)\right) - \rho(\Lambda_i) + \nu \left(\mathbb{I}f_i + \rho(F_i)\right)| \\ &= \sum_i \left\langle Q_i : \left((\beta_\lambda + \nu)\left(\mathbb{I} + \frac{\rho(F_i)}{f_i}\right) - \frac{\rho(\Lambda_i)}{f_i} \right) f_i \right|, \end{aligned} \tag{27}$$

with the second line due to linearity of $\varphi_i \mapsto \langle Q_i : \varphi_i|$, and the third line due to the fact that two-part functions constitute an algebra (allowing for multiplications and divisions).

Lemma 5 *Q^* is a solution for (25) if and only if there exists a $\nu^* \in \mathbb{R}_+$ such that (Q^*, ν^*) is a saddle point in the sense that, for all $Q \in \mathscr{Q}_{\text{mon}}$ and all $\nu \in \mathbb{R}_+$,*

$$\mathscr{L}(Q^*, \nu) \geq \mathscr{L}(Q^*, \nu^*) \geq \mathscr{L}(Q, \nu^*). \tag{28}$$

Proof The "if" part is trivial. To prove the "only if" part, it suffices to verify the conditions corresponding to those in Luenberger [7, Corollary 1, p219]. To that end, we start with two claims about each player i, which are proved in Appendix B.5: First, the two-part functions $\mathbb{I}f_i + \rho(F_i)$ and $\beta_\lambda \left(\mathbb{I}f_i + \rho(F_i)\right) - \rho(\Lambda_i)$ are each well-ordered. Second,

$$\langle Q_i : \mathbb{I}f_i + \rho(F_i)| \geq \tau(Q_i) \int_{T_i} Q_i dF_i \tag{29}$$

for any weakly increasing $Q_i : T_i \to \mathbb{R}$.

By the first claim and Lemma 1, both $\sum_i \langle Q_i : \mathbb{I}f_i + \rho(F_i)|$ and $\sum_i \langle Q_i : \beta_\lambda \left(\mathbb{I}f_i + \rho(F_i)\right) - \rho(\Lambda_i)|$ are concave functions of $(Q_i)_{i=1}^n$. Thus,

$$\left\{ (Q_i)_{i=1}^n \in \mathscr{Q} : \sum_i \langle Q_i : \mathbb{I}f_i + \rho(F_i)| \geq 0 \right\}$$

[8] For example, for each player i let $L^2(T_i)$ be the L^2-space of measurable real functions defined on T_i, endowed with the measure F_i. Clearly $\mathscr{Q} \in \prod_i L^2(T_i)$. Define the norm for $\prod_i L^2(T_i)$ by $\|Q\| := \sum_i \|Q_i\|_2$ for any $Q := (Q_i)_{i=1}^n \in \prod_i L^2(T_i)$.

is a convex set, and the objective in (25) is concave in the choice variable. This, coupled with convexity of $\mathscr{Q}_{\text{mon}}$ (Appendix B.4), means that the proof is complete if there exists a $(Q_i)_{i=1}^n \in \mathscr{Q}_{\text{mon}}$ such that $\sum_i \langle Q_i : \mathbb{I} f_i + \rho(F_i)| > 0$. Such $(Q_i)_{i=1}^n$ exists: always assign the good to player 1 and never assign the bad at all. That is, $Q_1 = 1$, hence $\tau(Q_1) = a_1$, and $Q_i = 0$ for all $i \neq 1$. Note $(Q_i)_{i=1}^n \in \mathscr{Q}_{\text{mon}}$. By (29),

$$\sum_i \langle Q_i : \mathbb{I} f_i + \rho(F_i)| = \langle Q_1 : \mathbb{I} f_1 + \rho(F_1)| \geq a_1 \int_{T_1} Q_1 dF_1 = a_1 > 0.$$

Now that all conditions are verified, the saddle point characterization follows. ■

Coupled with Theorem 1, Lemma 5 implies that any Pareto optimal mechanism is necessarily a solution of $\max_{Q \in \mathscr{Q}_{\text{mon}}} \mathscr{L}(Q, \nu)$ with $\mathscr{L}(Q, \nu)$ defined by (27) for some profile $(\lambda_i)_{i=1}^n$ of distributions specified in Theorem 1, and some $\nu \in \mathbb{R}_+$. For each i, denote

$$Z_i := (\beta_\lambda + \nu) \left(\mathbb{I} + \frac{\rho(F_i)}{f_i} \right) - \frac{\rho(\Lambda_i)}{f_i}. \tag{30}$$

Then (27) is the same as

$$\mathscr{L}(Q, \nu) = \sum_i \langle Q_i : Z_i f_i| \,. \tag{31}$$

The Z_i defined in (30) is exactly the two-part function that constitutes the profile $(Z_i)_{i=1}^n$ in Theorem 2. Plugging (19) and (20) into (30), and recalling $\mathbb{I}$ as the notation for the identity map, we obtain the explicit formula for the functions $Z_{i,+}$ and $Z_{i,-}$: for all i and all $t_i \in T_i$,

$$Z_{i,+}(t_i) = \beta_\lambda t_i + \frac{\beta_\lambda F_i(t_i) - \Lambda_i(t_i)}{f_i(t_i)} - \frac{\beta_\lambda - \Lambda_i(b_i)}{f_i(t_i)} + \nu \left(t_i - \frac{1 - F_i(t_i)}{f_i(t_i)} \right), \tag{32}$$

$$Z_{i,-}(t_i) = \beta_\lambda t_i + \frac{\beta_\lambda F_i(t_i) - \Lambda_i(t_i)}{f_i(t_i)} + \nu \left(t_i + \frac{F_i(t_i)}{f_i(t_i)} \right). \tag{33}$$

By (32) and (33), $Z_{i,+}(t_i) - Z_{i,-}(t_i) = -(\beta_\lambda - \Lambda_i(b_i) + \nu) / f_i(t_i) \leq 0$, with "$\leq$" due to $\beta_\lambda - \Lambda_i(b_i) \geq 0$ and $\nu \geq 0$. Thus, $Z_{i,+} \leq Z_{i,-}$, as asserted in Theorem 2.

5.2.2 Maximizing the Lagrangian through Bisection

It follows that any interim Pareto optimal mechanism maximizes $\sum_i \langle Q_i : Z_i f_i|$, defined in (31), among all $Q \in \mathscr{Q}_{\text{mon}}$. To solve this Lagrange problem, we use the above-proved fact that $Z_{i,+} \leq Z_{i,-}$ for all i to bisect the problem into two independent linear programmings.

Let $\mathscr{Q}_+$ be the set of all $(Q_i)_{i=1}^n \in \mathscr{Q}$ such that $Q_i \geq 0$ for all i, and $\mathscr{Q}_-$ the set of all $(Q_i)_{i=1}^n \in \mathscr{Q}$ such that $Q_i \leq 0$ for all i. By (18), $\max_{Q\in\mathscr{Q}_{\text{mon}}} \sum_i \langle Q_i : Z_i f_i|$ is equivalent to

$$\max_{(Q_i)_{i=1}^n \in \mathscr{Q}_{\text{mon}}} \left(\sum_i \int_{T_i} Q_i^+ Z_{i,+} dF_i + \sum_i \int_{T_i} \left(-Q_i^-\right) Z_{i,-} dF_i \right) \tag{34}$$

$$\leq \max_{(Q_i)_{i=1}^n \in \mathscr{Q}_{\text{mon}}} \sum_i \int_{T_i} Q_i^+ Z_{i,+} dF_i + \max_{(Q_i)_{i=1}^n \in \mathscr{Q}_{\text{mon}}} \sum_i \int_{T_i} \left(-Q_i^-\right) Z_{i,-} dF_i \tag{35}$$

$$= \max_{(Q_i)_{i=1}^n \in \mathscr{Q}_{\text{mon}} \cap \mathscr{Q}_+} \sum_i \int_{T_i} Q_i Z_{i,+} dF_i \tag{36}$$

$$+ \max_{(Q_i)_{i=1}^n \in \mathscr{Q}_{\text{mon}} \cap \mathscr{Q}_-} \sum_i \int_{T_i} Q_i Z_{i,-} dF_i. \tag{37}$$

Thus, to solve (34), it suffices to first solve (36) and (37) individually and then construct from the two solutions a $Q^* \in \mathscr{Q}_{\text{mon}}$ given which the objective in (34) attains the sum of the maximands in (36) and (37). The next two lemmas are proved in Appendix B.6.

Lemma 6 *$(\hat{Q}_i)_{i=1}^n \in \mathscr{Q}_{\text{mon}} \cap \mathscr{Q}_+$ solves (36) if and only if, for some $(\hat{q}_i)_{i=1} \in \mathscr{A}\left((\overline{Z}_{i,+})_{i=1}^n\right)$ and for each i, $\hat{Q}_i = \int_{T_{-i}} \hat{q}_i(\cdot, t_{-i}) dF_{-i}(t_{-i})$ on T_i.*

Lemma 7 *$(\check{Q}_i)_{i=1}^n \in \mathscr{Q}_{\text{mon}} \cap \mathscr{Q}_-$ solves (37) if and only if, for some $(\check{q}_{iB})_{i=1} \in \mathscr{A}\left((-\overline{Z}_{i,-})_{i=1}^n\right)$ and for each i, $\check{Q}_i = -c \int_{T_{-i}} \check{q}_i(\cdot, t_{-i}) dF_{-i}(t_{-i})$ on T_i.*[9]

Pick any element $(\hat{q}_{iA})_{i=1}^n$ of $\mathscr{A}\left((\overline{Z}_{i,+})_{i=1}^n\right)$ such that $\hat{q}_{iA}(t_i, \cdot) = 0$ on T_{-i} whenever $\overline{Z}_{i,+}(t_i) \leq 0$, and any $(\check{q}_{iB})_{i=1}^n$ of $\mathscr{A}\left((-\overline{Z}_{i,-})_{i=1}^n\right)$ such that $\check{q}_{iB}(t_i, \cdot) = 0$ on T_{-i} whenever $\overline{Z}_{i,-}(t_i) \geq 0$. For each i, let $\hat{Q}_i$ be the marginal of $\hat{q}_{iA}$, and $\check{Q}_i$ the marginal of $-c\check{q}_{iB}$. By Lemmas 6 and 7, $(\hat{Q}_i)_{i=1}^n$ solves (36), and $(\check{Q}_i)_{i=1}^n$ solves (37). We observe that the support of $(\hat{Q}_i)_{i=1}^n$ and that of $(\check{Q}_i)_{i=1}^n$ have no overlapped interior: For each i,

$$\hat{Q}_i(t_i) \neq 0 \iff \check{Q}_i(t_i) = 0. \tag{38}$$

That is because, by the choice of $(\hat{q}_{iA})_{i=1}^n$ and $(\check{q}_{iB})_{i=1}^n$,

$$\check{Q}_i(t_i) \neq 0 \;\Rightarrow\; \check{Q}_i(t_i) < 0 \;\Rightarrow\; \check{q}_i(t_i, \cdot) \not\equiv 0 \text{ on } T_{-i} \;\Rightarrow\; t_i \leq \sup\left\{\tau_i \in T_i : \overline{Z}_{i,-}(\tau_i) < 0\right\},$$
$$\hat{Q}_i(t_i) \neq 0 \;\Rightarrow\; \hat{Q}_i(t_i) > 0 \;\Rightarrow\; \hat{q}_i(t_i, \cdot) \not\equiv 0 \text{ on } T_{-i} \;\Rightarrow\; t_i \geq \inf\left\{\tau_i \in T_i : \overline{Z}_{i,+}(\tau_i) > 0\right\},$$

and, because Z_i is well-ordered ($Z_{i,+} \leq Z_{i,-}$ for all i), one can prove (Appendix B.7) that

$$\sup\left\{\tau_i \in T_i : \overline{Z}_{i,-}(\tau_i) < 0\right\} \leq \inf\left\{\tau_i \in T_i : \overline{Z}_{i,+}(\tau_i) \geq 0\right\}. \tag{39}$$

[9] Recall (2) for the role of the coefficient c.

Thus, the following function Q_i^* is well-defined and weakly increasing on T_i:

$$Q_i^*(t_i) := \begin{cases} \check{Q}_i(t_i) & \text{if } \overline{Z}_{i,-}(t_i) < 0 \\ \hat{Q}_i(t_i) & \text{if } \overline{Z}_{i,+}(t_i) > 0 \\ 0 & \text{else.} \end{cases} \tag{40}$$

Because of (38), $(Q_i^*)^+ = \hat{Q}_i$ and $(Q_i^*)^- = -\check{Q}_i$ for any i. It follows that $(Q_i^*)_{i=1}^n$ is a solution for both problems in (35) simultaneously. By (39), (40) and monotonicity of $\hat{Q}_i$ and $\check{Q}_i$, each Q_i^* is weakly increasing; thus $(Q_i^*)_{i=1}^n$ is a feasible choice for (34), an upper bound of which is the maximand of (35), attained by $(Q_i^*)_{i=1}^n$. Thus, $(Q_i^*)_{i=1}^n$ is a solution of (34).

5.2.3 Proof of Theorem 2

Given any profile $(\lambda_i)_{i=1}^n$ of distributions, λ_i on T_i for each i, the objective (7) is defined. Let (Q, P) be a mechanism that maximizes (7) subject to IC, IR and BB. Then Q solves (25) and P obeys Claim (b) of the theorem with respect to Q (Lemma 4 and Remark 8). We still need to prove Claim (a) of the theorem. To do that, note from Q being a solution of (25) that (Q, ν) is a saddle point for some $\nu \geq 0$ with respect to the Lagrangian $\mathscr{L}$ defined by $(\lambda_i)_{i=1}^n$ via (30) and (31) (Lemma 5). Thus, Q solves (34). Section 5.2.2 has shown that the maximand of (34) is equal to the sum of the maximands in (35). Consequently, for Q to solve (34) it must solve the two problems in (35) simultaneously. That is, $(Q_i^+)_{i=1}^n$ solves (36) and $(-Q_i^-)_{i=1}^n$ solves (37). For $(Q_i^+)_{i=1}^n$ to solve (36), Lemma 6 requires that Q_i^+ be the marginal of some $\hat{q}_{iA}$, for each i, such that $(\hat{q}_{iA})_{i=1}^n \in \mathscr{A}\left((\overline{Z}_{i,+})_{i=1}^n\right)$; for $(-Q_i^-)_{i=1}^n$ to solve (37), Lemma 7 requires that $-Q_i^-$ be the marginal of some $-c\check{q}_{iB}$, for each i, such that $(\check{q}_{iB})_{i=1}^n \in \mathscr{A}\left((-\overline{Z}_{i,-})_{i=1}^n\right)$. That proves Claim (a) of the theorem. ■

6 Conclusion

Although the literature has long recognized auctions as the optimal means to allocate scarce resources to multiple individuals, the role of auctions has yet to be recognized, and sometimes deemed morally repugnant, to issues about redistribution among individuals.[10] Given wealth

[10] Recall the indignant outcry expressed in the mass media when news broke that the locations of some international games were chosen through bidding, or the negative media coverage on less developed countries being paid to receive toxic, recycled materials.

inequality, auctions are feared to benefit the rich and impoverish the poor. This paper, by contrast, argues that the role of auctions is essential to achieve redistributive optimality. Any interim Pareto optimum, no matter where it is located on the Pareto frontier, whether it weighs the poor more than it does the rich, or the rich more than the poor, is necessarily in the form of auctions, with the winner-selection rule adjusted to reflect the particular welfare weighting associated with the particular Pareto optimum. Instead of mandating wealth transfers from one individual to another, whose idiosyncrasies are uncertain to regulators, a social planner could have used auctions to induce the right amount of wealth transfers among the right types of individuals.

Our finding goes beyond just the characterization that all Pareto optimal mechanisms are auctions. We also obtain specific properties of such auctions. Among them is the inequality-reducing effect of allocating an item that is commonly disliked by all individuals. We have uncovered an explicit link between an inequality index among individual types and the necessity for a socially optimal mechanism to assign the bad item to someone sometimes.

This paper makes a methodology contribution to the mechanism design literature. Rather than assuming a utilitarian, one-dimension, design objective, we start with interim Pareto optimality, an objective with infinite dimensions, and show that any optimum in this infinite dimension space corresponds to a constrained optimization of a utilitarian objective obtained through the endogenous welfare weighting associated with the optimum. We introduce a new class of nonlinear operators to systematically keep track of each player's countervailing incentive of playing the role of a buyer sometimes and the role of a seller some other times. We devise a bisection technique to solve an optimal mechanism problem that has a nonlinear objective and multiple binding constraints.

Our model is relevant to matching theory in the case where one side of the matching market has both desirable and undesirable items (e.g., toxic assets that need to be absorbed by other financial institutions; enrollment of schools in undesirable neighborhoods; thankless tasks to be carried out by some team members; donation of one's own kidney). While much of the matching theory literature assumes that money transfers are banned, our result suggests that it is suboptimal to ban money transfers from matching markets.

A Proof of Theorem 1

For each $i \in \{1, \ldots, n\}$ denote $C(T_i)$ for the space of continuous real functions defined on the closed, bounded interval T_i, with the maximum norm $\|\cdot\|_{\max}$. Let

$$\mathscr{C} := \prod_{i=1}^{n} C(T_i)$$

and endow $\mathscr{C}$ with the maximum norm such that $\|(\varphi_i)_{i=1}^n\|_{\max} := \max_i \|\varphi_i\|_{\max}$ for all $(\varphi_i)_{i=1}^n \in \mathscr{C}$. Thus, $\mathscr{C}$ is a normed linear space. Define the utility possibility set

$$\mathbb{U} := \{(W_i)_{i=1}^n \in \mathscr{C} : \exists \text{ IC, IR \& BB } (Q,P)\, [\forall i\, \forall t_i \in T_i\, [W_i(t_i) \le U_i(t_i \mid Q,P))]]\}. \tag{41}$$

Note that $(U_i(\cdot \mid Q,P))_{i=1}^n \in \mathscr{C}$ for any IC mechanism (Q,P).

Lemma 8 $\mathbb{U}$ *is convex.*

Proof Pick any $(W_i^1)_{i=1}^n, (W_i^2)_{i=1}^n \in \mathbb{U}$. Thus, for some IC, IR and BB mechanisms $(Q_i^1, P_i^1)_{i=1}^n$ and $(Q_i^2, P_i^2)_{i=1}^n$ we have, for each $i = 1, \ldots, n$, each $k = 1, 2$, and any $t_i \in T_i$,

$$W_i^k(t_i) \le Q_i^k(t_i) - \frac{P_i^k(t_i)}{t_i}, \tag{42}$$

$$P_i^k(t_i') = P_i^k(t_i) + \int_{t_i}^{t_i'} s dQ_i^k(s) \quad (\forall t_i' \in T_i), \tag{43}$$

$$0 \le t_i Q_i^k(t_i) - P_i^k(t_i), \tag{44}$$

$$P_i^k(t_i) = \int_{T_{-i}} p_i^k(t_i, t_{-i}) dF_{-i}(t_{-i}), \tag{45}$$

$$0 \le \sum_i p_i^k(t) \quad (\forall t \in \prod_i T_i), \tag{46}$$

and $Q^k \in \mathscr{Q}_{\text{mon}}$ and W_i^k a continuous function on T_i for each k. Here (43) coupled with $Q^k \in \mathscr{Q}_{\text{mon}}$ is equivalent to IC, (44) is equivalent to IR, (45) is the definition of expected payment, and (46) means BB. For any $\gamma \in [0,1]$, define for each i

$$Q_i := \gamma Q_i^1 + (1-\gamma) Q_i^2,$$

$$p_i := \gamma p_i^1 + (1-\gamma) p_i^2.$$

Then it follows from (45) that, for any i and any $t_i \in T_i$,

$$P_i = \gamma P_i^1 + (1-\gamma) P_i^2.$$

We shall show that $(Q_i, P_i)_{i=1}^n$ satisfies IC, IR and BB. Immediately from the definition of p_i and (46), BB follows. IR is proved by combining together the definition of Q_i, the fact $P_i = \gamma P_i^1 + (1-\gamma)P_i^2$, and (44) for both $k = 1, 2$. To verify IC, first note that $\gamma Q^1 + (1-\gamma)Q^2 \in \mathscr{Q}_{\text{mon}}$ by convexity of $\mathscr{Q}_{\text{mon}}$ (Appendix B.4). Second, by (43),

$$\gamma P_i^1(t_i') + (1-\gamma)P_i^2(t_i') = \gamma P_i^1(t_i) + (1-\gamma)P_i^2(t_i) + \int_{t_i}^{t_i'} s d\left(\gamma Q_i^1(s) + (1-\gamma)Q_i^2(s)\right)$$

for any $t_i, t_i' \in T_i$ and any i. Hence $(Q_i, P_i)_{i=1}^n$ is IC. Thus, $(Q_i, P_i)_{i=1}^n$ satisfies IC, IR and BB. Finally, plug $Q_i = \gamma Q_i^1 + (1-\gamma)Q_i^2$ and $P_i = \gamma P_i^1 + (1-\gamma)P_i^2$ into (42) to obtain

$$\gamma W_i^1(t_i) + (1-\gamma)W_i^2(t_i) \le Q_i(t_i) - \frac{P_i(t_i)}{t_i}$$

for each i and any $t_i \in T_i$. This coupled with continuity of $\gamma W_i^1 + (1-\gamma)W_i^2$ implies $(\gamma W_i^1 + (1-\gamma)W_i^2)_{i=1}^n \in \mathbb{U}$, as desired. ∎

To prove Theorem 1, pick any interim Pareto optimal mechanism (Q^*, P^*). Denote $u_i^* := U_i(\cdot \mid Q^*, P^*)$ for each i. Then $(u_i^*)_{i=1}^n \in \mathbb{U}$. Denote

$$\mathbb{V}((u_i^*)_{i=1}^n) := \{(u_i)_{i=1}^n \in \mathscr{C} \ : \ \forall i\, [u_i \ge u_i^* \text{ a.e. } T_i]\, ; \exists i\, [u_i > u_i^* \text{ on a positive-measure } S_i \subseteq T_i]\}\,.$$

Obviously, $\mathbb{V}((u_i^*)_{i=1}^n)$ is convex.

Lemma 9 *There exists a continuous linear functional ϕ on $\mathscr{C}$, not identically zero, such that for all $(u_i)_{i=1}^n \in \mathbb{U}$,*

$$\phi\left((u_i)_{i=1}^n\right) \le \phi\left((u_i^*)_{i=1}^n\right). \tag{47}$$

Proof First, $\mathbb{U}$ is convex by Lemma 8, and $\mathbb{V}((u_i^*)_{i=1}^n)$ convex by its definition. Second, $\mathbb{U}$ contains an interior point: Consider the mechanism that gives away the good A for free with probability $1/2$, else assigns neither item to anyone, and, in the former event, randomly assigns the good A (for free) to one of the n players with equal probability. This mechanism is IC, IR and BB, and it generates for everyone an interim expected payoff constantly equal to $1/(2n)$. Thus, this payoff profile belongs to $\mathbb{U}$. Now consider another mechanism that differs from the former only by that it assigns the good with probability $1/2 + \epsilon$. The mechanism is also IC, IR and BB, and generates an expected payoff profile larger than the former in every dimension by ϵ/n. Since this is true for all $\epsilon \in (0, 1/2]$, the payoff profile generated by the former mechanism is an interior point of $\mathbb{U}$ with respect to the max norm. Third, $\mathbb{V}\left((u_i^*)_{i=1}^n\right)$ contain no interior point of $\mathbb{U}$; otherwise, such an interior point, by definition of $\mathbb{V}\left((u_i^*)_{i=1}^n\right)$, interim Pareto dominates $(u_i^*)_{i=1}^n$, contradiction. Thus, by the Hahn-Banach theorem, there exists a continuous linear functional ϕ on $\mathscr{C}$, not identically zero, such that, for some constant w, for any $(u_i)_{i=1}^n \in \mathbb{U}$ and any $(\hat{u}_i)_{i=1}^n \in \mathbb{V}((u_i^*)_{i=1}^n)$,

$$\phi\left((u_i)_{i=1}^n\right) \leq w \leq \phi\left((\hat{u}_i)_{i=1}^n\right). \tag{48}$$

For any $\epsilon > 0$, the profile $(u_i^* + \epsilon)_{i=1}^n \in \mathbb{V}((u_i^*)_{i=1}^n)$. Thus

$$w \leq \phi\left((u_i^* + \epsilon)_{i=1}^n\right) = \phi\left((u_i^*)_{i=1}^n\right) + \epsilon\phi(\mathbf{1}),$$

with the equality due to linearity of ϕ, and $\mathbf{1}$ denoting the unit vector of $\mathscr{C}$. Since continuous linear functionals are bounded, $\epsilon\phi(\mathbf{1}) \to 0$ as $\epsilon \to 0$. Hence $w \leq \phi\left((u_i^*)_{i=1}^n\right)$. This coupled with the fact $(u_i^*)_{i=1}^n \in \mathbb{U}$ implies $\phi\left((u_i^*)_{i=1}^n\right) \leq w \leq \phi\left((u_i^*)_{i=1}^n\right)$, hence $\phi\left((u_i^*)_{i=1}^n\right) = w$. Plug this into (48) to obtain (47) and hence the claim. ■

For each $i \in \{1, \ldots, n\}$ and any $u_i \in C(T_i)$ let

$$\phi_i(u_i) := \phi\left(0, \ldots, 0, u_i, 0, \ldots, 0\right),$$

that is, the action of ϕ on the profile of payoff functions whose components are constantly zero except the one corresponding to player i's payoff function. By linearity of ϕ,

$$\phi\left((u_i)_{i=1}^n\right) = \sum_{i=1}^n \phi_i(u_i) \tag{49}$$

for all $(u_i)_{i=1}^n \in \mathscr{C}$. Obviously, for each i, ϕ_i is a continuous linear functional on $C(T_i)$. Thus ϕ_i is also a bounded functional on $C(T_i)$.

Lemma 10 *For each $i \in \{1, \ldots, n\}$, ϕ_i is positive.*[11]

Proof Suppose, to the contrary, that $\phi_i(u_i) < 0$ for some $u_i \in C(T_i)$ such that $u_i \geq 0$ on T_i. Then $\left(u_i^* - u_i, (u_j^*)_{j \neq i}\right) \in \mathbb{U}$ by definition of $\mathbb{U}$, hence Lemma 9 implies

$$\phi\left((u_j^*)_{j=1}^n\right) \geq \phi\left(\left(u_i^* - u_i, (u_j^*)_{j \neq i}\right)\right) = \sum_{j=1}^n \phi_j(u_j^*) - \phi_i(u_i) > \sum_{j=1}^n \phi_j(u_j^*) = \phi\left((u_j^*)_{j=1}^n\right),$$

contradiction. ■

[11] A functional ϕ_i on $C(T_i)$ is *positive* iff $\phi_i(u_i) \geq 0$ for any $u_i \in C(T_i)$ such that $u_i \geq 0$ on T_i.

For any i, since ϕ_i is a bounded linear functional on $C(T_i)$, with $T_i = [a_i, b_i]$ a closed, bounded interval, the Riesz representation theorem in its original version (Royden and Fitzpatrick [13, p468]) implies that there exists a unique function $\phi_i : T_i \to \mathbb{R}$, of bounded variation on $[a_i, b_i]$, continuous on the right on (a_i, b_i), and vanishing at a_i, such that

$$\phi_i(u_i) = \int_{T_i} u_i d\lambda_i$$

for all $u_i \in C(T_i)$. This, combined with (47) and (49), delivers Theorem 1 if (i) λ_i is also weakly increasing, (ii) its range belongs to $\mathbb{R}_+$, and (iii) $\lambda_i > 0$ on some positive-measure subset of T_i for some i.

Property (ii) follows from property (i) because $\lambda_i(a_i) = 0$. Then (iii) follows from (ii): Otherwise (ii) implies $\lambda_i = 0$ for all i, hence ϕ is identically zero on $\mathscr{C}$, contradiction to Lemma 9. Thus it suffices to prove (i).

To that end, suppose, to the contrary, that $t_i < t'_i$ and $\lambda_i(t_i) > \lambda_i(t'_i)$ for some $t_i, t'_i \in (a_i, b_i)$.[12] Then, since λ_i is right-continuous on (a_i, b_i), there exists a sufficiently small $\epsilon > 0$ such that for any $\delta \in (0, \epsilon)$, $\lambda_i(t_i + \delta) > \lambda_i(t'_i + \delta)$. It is easy to construct a continuous function $u_i : T_i \to [0, 1]$ whose support is contained by $[t_i, t'_i + \epsilon]$ such that $u_i = 1$ on $[t_i + \epsilon/2, t'_i + \epsilon/2]$. Then $u_i \geq 0$ on T_i, $u_i \in C(T_i)$, and yet

$$\phi_i(u_i) = \int_{T_i} u_i d\lambda_i = \int_{t_i}^{t'_i+\epsilon} d\lambda_i \leq \int_{t_i+\epsilon/2}^{t'_i+\epsilon/2} d\lambda_i = \lambda_i(t'_i + \epsilon/2) - \lambda_i(t_i + \epsilon) < 0,$$

contradicting Lemma 10. That proves property (i) of λ_i. Thus Theorem 1 follows.

B Details in Theorem 2

B.1 Proof of Lemma 1

For any integrable function $Q_i : T_i \to \mathbb{R}$ and any well-ordered two-part function $\varphi_i := (\varphi_{i,+}, \varphi_{i,-})$, use the definition of two-part operators and the fact $Q_i = Q_i^+ - Q_i^-$ to obtain

$$\begin{aligned}
\langle Q_i : \varphi_i| &= \int_{T_i} Q_i^+(t_i)\varphi_{i,+}(t_i)dF_i(t_i) - \int_{T_i} Q_i^-(t_i)\varphi_{i,-}(t_i)dF_i(t_i) \\
&= \int_{T_i} Q_i(t_i)\varphi_{i,-}(t_i)dF_i(t_i) + \int_{T_i} Q_i^+(t_i)\left(\varphi_{i,+}(t_i) - \varphi_{i,-}(t_i)\right)dF_i(t_i).
\end{aligned}$$

[12] There is no need to consider $t_i = a_i$ and $t'_i = b_i$ because we already have $\lambda_i(a_i) = 0$, and it is immaterial to change the value of λ_i at the singleton b_i.

On the second line, the first integral is linear in Q_i; and the second integral concave in Q_i because $Q_i(t_i) \mapsto Q_i^+(t_i)$ is a convex mapping and, because $\varphi_{i,+} - \varphi_{i,-} \leq 0$ a.e. on T_i (φ being well-ordered) and hence $Q_i(t_i) \mapsto Q_i^+(t_i)\left(\varphi_{i,+}(t_i) - \varphi_{i,-}(t_i)\right)$ is a concave mapping for almost every t_i in T_i. Thus $\langle Q_i : \varphi_i|$ is concave in Q_i.

B.2 Proof of Lemma 2

Denote $t_i^0 := \tau(Q_i)$. Since (Q_i, P_i) is IC, (5) implies

$$\begin{aligned}\int_{T_i} P_i d\mu_i &= \int_{T_i}\left(t_i Q_i(t_i) - \int_{t_i^0}^{t_i} Q_i(s)ds - \tilde{U}_i(t_i^0)\right) d\mu_i(t_i)\\ &= \int_{T_i} t_i Q_i(t_i) d\mu_i(t_i) - \tilde{U}_i(t_i^0)\int_{T_i} d\mu_i - \int_{T_i}\int_{t_i^0}^{t_i} Q_i(s) ds d\mu_i(t_i).\end{aligned}$$

Decompose the last double integral to obtain

$$\begin{aligned}\int_{T_i}\int_{t_i^0}^{t_i} Q_i(s)dsd\mu_i(t_i) &= \int_{a_i}^{t_i^0}\int_{t_i^0}^{t_i} Q_i(s)dsd\mu_i(t_i) + \int_{t_i^0}^{b_i}\int_{t_i^0}^{t_i} Q_i(s)dsd\mu_i(t_i)\\ &= -\int_{a_i}^{t_i^0}\int_{t_i}^{t_i^0} Q_i(s)dsd\mu_i(t_i) + \int_{t_i^0}^{b_i}\int_{t_i^0}^{t_i} Q_i(s)dsd\mu_i(t_i)\\ &= -\int_{a_i}^{t_i^0}\int_{a_i}^{s} Q_i(s)d\mu_i(t_i)ds + \int_{t_i^0}^{b_i}\int_{s}^{b_i} Q_i(s)d\mu_i(t_i)ds\\ &= -\int_{a_i}^{t_i^0} Q_i(s)\int_{a_i}^{s} d\mu_i(t_i)ds + \int_{t_i^0}^{b_i} Q_i(s)\int_{s}^{b_i} d\mu_i(t_i)ds\\ &= \int_{a_i}^{t_i^0} Q_i^-(s)\int_{a_i}^{s} d\mu_i(t_i)ds + \int_{t_i^0}^{b_i} Q_i^+(s)\left(\int_{a_i}^{b_i} d\mu_i(t_i) - \int_{a_i}^{s} d\mu_i(t_i)\right)ds\\ &= \int_{a_i}^{t_i^0} Q_i^-(s)\rho_-(\mu_i)(s)ds - \int_{t_i^0}^{b_i} Q_i^+(s)\rho_+(\mu_i)(s)ds\\ &= -\langle Q_i : \rho(\mu_i)|,\end{aligned}$$

with the third equality due to Fubini's theorem, the second last equality due to (19) and (20), and the last equality due to (18). Plugging $\int_{T_i}\int_{t_i^0}^{t_i} Q_i(s)dsd\mu_i(t_i) = -\langle Q_i : \rho(\mu_i)|$ into the equation of $\int_{T_i} P_i d\mu_i$ displayed above, we get (21).

B.3 Proof of the Sufficiency of (22)

For each player i, denote $t_i^0 := \tau(Q_i)$ (τ defined in (6)). For each player i, define

$$c_i := t_i^0 Q_i(t_i^0) - \int_{a_i}^{t_i^0} s dQ_i(s) + \frac{1}{n-1} \sum_{j \neq i} \int_{a_j}^{b_j} s \left(1 - F_j(s)\right) dQ_j(s) \tag{50}$$

and, for any $(t_i, t_{-i}) \in T_i \times T_{-i}$, let the money transfer from i to others be equal to

$$p_i(t_i, t_{-i}) := c_i + \int_{a_i}^{t_i} s dQ_i(s) - \frac{1}{n-1} \sum_{j \neq i} \int_{a_j}^{t_j} s dQ_j(s). \tag{51}$$

Integrating $p_i(t_i, t_{-i})$ across t_{-i} gives the envelope equation (5), which coupled with the monotonicity hypothesis of Q_i implies IC. The integration also implies $\tilde{U}_i(t_i^0) = 0$, hence IR follows. To complete the proof, we prove BB: It suffices to prove (23), $\sum_i p_i(t) = \sum_i \langle Q_i : \mathbb{I} f_i + \rho(F_i)|$ for all $t \in \prod_i T_i$, for then BB follows from (22). Hence pick any $t := (t_i)_{i=1}^n \in \prod_i T_i$. By (51),

$$\sum_i p_i(t) = \sum_i c_i + \sum_i \int_{a_i}^{t_i} s dQ_i(s) - \frac{1}{n-1} \sum_i \sum_{j \neq i} \int_{a_j}^{t_j} s dQ_j(s) = \sum_i c_i.$$

Thus, by (50),

$$\begin{aligned}
\sum_i p_i(t) &= \sum_i t_i^0 Q_i(t_i^0) - \sum_i \int_{a_i}^{t_i^0} s dQ_i(s) + \frac{1}{n-1} \sum_i \sum_{j \neq i} \int_{a_j}^{b_j} s \left(1 - F_j(s)\right) dQ_j(s) \\
&= \sum_i t_i^0 Q_i(t_i^0) - \sum_i \int_{a_i}^{t_i^0} s dQ_i(s) + \sum_i \int_{a_i}^{b_i} s \left(1 - F_i(s)\right) dQ_i(s) \\
&= \sum_i \left(t_i^0 Q_i(t_i^0) - \int_{a_i}^{t_i^0} s dQ_i(s) + \int_{a_i}^{b_i} s \left(1 - F_i(s)\right) dQ_i(s) \right).
\end{aligned}$$

Calculate the two integrals in the last line through integration by parts and then combine terms to obtain

$$\begin{aligned}
\sum_i p_i(t) &= \sum_i \left(\int_{a_i}^{t_i^0} Q_i(s) ds - \int_{a_i}^{b_i} Q_i(s) \left(1 - F_i(s) - s f_i(s)\right) ds \right) \\
&= \sum_i \left(\int_{a_i}^{t_i^0} Q_i(s) \left(1 - \left(1 - F_i(s) - s f_i(s)\right)\right) ds - \int_{t_i^0}^{b_i} Q_i(s) \left(1 - F_i(s) - s f_i(s)\right) ds \right) \\
&= \sum_i \langle Q_i : \mathbb{I} f_i + \rho(F_i)| ,
\end{aligned}$$

with the last line due to $t_i^0 = \tau(Q_i)$, (6), (18) and the definition of $\rho(F)$ (Eqs. (19) and (20)). That proves (23) and hence BB.

B.4 Convexity of $\mathscr{Q}_{\text{mon}}$

Let $\gamma \in [0,1]$ and $Q, \hat{Q} \in \mathscr{Q}_{\text{mon}}$. Since $Q \in \mathscr{Q}_{\text{mon}}$, it is generated by a $(q_{iA}, q_{iB})_{i=1}^n$ with $\sum_i q_{iA}(\cdot) \leq 1$ and $\sum_i q_{iB}(\cdot) \leq 1$ via (2), and Q_i is weakly increasing for all i. Likewise, $\hat{Q} = (\hat{Q}_i)_{i=1}^n$ is generated by a $(\hat{q}_{iA}, \hat{q}_{iB})_{i=1}^n$ with each $\hat{Q}_i$ weakly increasing. Then $\sum_i (\gamma q_{iA} + (1-\gamma)\hat{q}_{iA}) \leq 1$ and $\sum_i (\gamma q_{iB} + (1-\gamma)\hat{q}_{iB}) \leq 1$; furthermore, for each i, $\gamma Q_i + (1-\gamma)\hat{Q}_i$ satisfies (2) with respect to $(\gamma q_{iA} + (1-\gamma)\hat{q}_{iA}, \gamma q_{iB} + (1-\gamma)\hat{q}_{iB})$, and is weakly increasing because both Q_i and $\hat{Q}_i$ are so. Thus $\left(\gamma Q_i + (1-\gamma)\hat{Q}_i\right)_{i=1}^n \in \mathscr{Q}_{\text{mon}}$.

B.5 Details in the Proof of Lemma 5

Claim 1 *$\mathbb{I}f_i + \rho(F_i)$ is well-ordered.*

Proof By (19) and (20)), $\rho_+(F_i) \leq \rho_-(F_i)$. This, coupled with the fact $(\mathbb{I}f_i)_+ = (\mathbb{I}f_i)_- = \mathbb{I}f_i$, implies $\mathbb{I}f_i + \rho_+(F_i) \leq \mathbb{I}f_i + \rho_-(F_i)$. ■

Claim 2 *$\beta_\lambda (\mathbb{I}f_i + \rho(F_i)) - \rho(\Lambda_i)$ is well-ordered.*

Proof Since $(\mathbb{I}f_i)_+ = (\mathbb{I}f_i)_- = \mathbb{I}f_i$, and β_λ is a coefficient, it suffices to show that $\beta_\lambda \rho(F_i) - \rho(\Lambda_i)$ is well-ordered. To that end, let $t_i \in T_i$. By (15), (16), (19) and (20),

$$\begin{aligned} \beta_\lambda (\rho_+(F_i))(t_i) - (\rho_+(\Lambda_i))(t_i) &= \beta_\lambda (-1 + F_i(t_i)) - (-\Lambda_i(b_i) + \Lambda_i(t_i)) \\ &= \beta_\lambda F_i(t_i) - \Lambda_i(t_i) - (\beta_\lambda - \Lambda_i(b_i)) \\ &\leq \beta_\lambda F_i(t_i) - \Lambda_i(t_i) \\ &= \beta_\lambda (\rho_-(F_i))(t_i) - (\rho_-(\Lambda_i))(t_i), \end{aligned}$$

with the inequality due to (16). Thus, $\beta_\lambda \rho(F_i) - \rho(\Lambda_i)$ is well-ordered, as desired. ■

Claim 3 *Eq. (29) is true.*

Proof Let $Q : T_i \to \mathbb{R}$ be weakly increasing. Denote $t_i^0 := \tau(Q_i)$ for any i. Recall that

$(\mathbb{I}f_i)(t_i) = t_i f_i(t_i)$ for all t_i; use (19) and (20) to calculate $\rho(F)$; then use (18) to obtain

$$\begin{aligned}
\langle Q_i : \mathbb{I}f_i + \rho(F_i)| &= \int_{a_i}^{b_i} Q_i^+(t_i) t_i dF_i(t_i) - \int_{a_i}^{b_i} Q_i^+(t_i)\left(1 - F_i(t_i)\right) dt_i \\
&\quad + \int_{a_i}^{b_i} Q_i^-(t_i) t_i dF_i(t_i) - \int_{a_i}^{b_i} Q_i^-(t_i) F_i(t_i) dt_i \\
&\overset{(6)}{=} -\int_{t_i^0}^{b_i} Q_i(t_i) d\left(t_i(1 - F_i(t_i))\right) + \int_{a_i}^{t_i^0} Q_i(t_i) d\left(t_i F_i(t_i)\right) \\
&= Q(t_i^0) t_i^0 + \int_{t_i^0}^{b_i} t_i (1 - F_i(t_i)) dQ_i(t_i) - \int_{a_i}^{t_i^0} t_i F_i(t_i) dQ_i(t_i) \\
&\geq Q(t_i^0) t_i^0 + t_i^0 \left[\int_{t_i^0}^{b_i} (1 - F_i(t_i)) dQ_i(t_i) - \int_{a_i}^{t_i^0} F_i(t_i) dQ_i(t_i)\right] \\
&= Q(t_i^0) t_i^0 + t_i^0 \left[\int_{a_i}^{b_i} Q_i(t_i) dF_i(t_i) - Q(t_i^0)\right] \\
&= t_i^0 \int_{T_i} Q_i(t_i) dF_i(t_i),
\end{aligned}$$

with the third and fourth equalities due to integration by parts, and the inequality due to Q_i being weakly increasing. ■

B.6 Proofs of Lemmas 6 and 7

The proofs of these lemmas are similar to Myerson's [10, pp. 68–70] proof except for a modification. The modification is necessary because Problem (36) has a nonnegativity constraint $Q \in \mathscr{Q}_+$, and Problem (37) a nonpositivity constraint $Q \in \mathscr{Q}_-$.

B.6.1 Myerson's Definition of Ironing

For any continuous function $\varphi : [0,1] \to \mathbb{R}$, denote the convex hull of φ by $\operatorname{conv}\varphi$. For any integrable function $\psi_i : T_i \to \mathbb{R}$, define $H_i(\psi_i) : [0,1] \to \mathbb{R}$ by

$$(H_i(\psi_i))(s) := \int_0^s \psi_i\left(F_i^{-1}(r)\right) dr \tag{52}$$

for any $s \in [0,1]$. Thus $H_i(\psi_i)$ is continuous, and the convex hull $\operatorname{conv} H_i(\psi_i)$ of $H_i(\psi_i)$ is well-defined. By the definition of ironing, the ironed copy $\overline{\psi}_i$ of ψ_i satisfies

$$\overline{\psi}_i(t_i) = \frac{d}{ds}\left((\operatorname{conv} H_i(\psi_i))(s)\right)\Big|_{s=F_i(t_i)} \tag{53}$$

for any $t_i \in T_i$ such that $\operatorname{conv} H_i(\psi_i)$ is differentiable at $F_i(t_i)$.

B.6.2 Left- and Right-Hand Leveling of an Ironed Function

This new construct is to handle the aforementioned nonnegativity and nonpositivity constraints. For any continuous function $\varphi : [0, 1] \to \mathbb{R}$, define

$$(\mathrm{conv_L}\,\varphi)(s) \quad := \quad \begin{cases} \min_{[0,1]} \varphi & \text{if } s \le \inf\left(\arg\min_{r\in[0,1]} \varphi(r)\right) \\ (\mathrm{conv}\,\varphi)(s) & \text{else,} \end{cases} \tag{54}$$

$$(\mathrm{conv_R}\,\varphi)(s) \quad := \quad \begin{cases} \min_{[0,1]} \varphi & \text{if } s \ge \sup\left(\arg\min_{r\in[0,1]} \varphi(r)\right) \\ (\mathrm{conv}\,\varphi)(s) & \text{else.} \end{cases} \tag{55}$$

For any integrable function $\psi_i : T_i \to \mathbb{R}$, $H_i(\psi_i) : [0, 1] \to \mathbb{R}$, defined by (52), continuous. Hence $\mathrm{conv_L}\, H_i(Z_{i,+})$ and $\mathrm{conv_R}\, H_i(Z_{i,-})$ are defined by (54) and (55). Each a convex function, their derivatives are defined for almost every $t_i \in T_i$, and weakly increasing on the set of these points:

$$\overline{\overline{Z}}_{i,+}(t_i) \quad := \quad \frac{d}{ds}\left(\left(\mathrm{conv_L}\, H_i(Z_{i,+})\right)(s)\right)\bigg|_{s=F_i(t_i)}, \tag{56}$$

$$\overline{\overline{Z}}_{i,-}(t_i) \quad := \quad \frac{d}{ds}\left(\left(\mathrm{conv_R}\, H_i(Z_{i,-})\right)(s)\right)\bigg|_{s=F_i(t_i)}. \tag{57}$$

Extend the definitions to all $t_i \in T_i$ to keep $\overline{\overline{Z}}_{i,+}$ and $\overline{\overline{Z}}_{i,-}$ monotone. By (53), (56) and (57), for any i and almost every $t_i \in T_i$,

$$\begin{aligned} \overline{\overline{Z}}_{i,+}(t_i) \quad &= \quad \max\left\{0, \overline{Z}_{i,+}(t_i)\right\}, \\ \overline{\overline{Z}}_{i,-}(t_i) \quad &= \quad \min\left\{0, \overline{Z}_{i,-}(t_i)\right\}. \end{aligned} \tag{58}$$

B.6.3 Allocation by Ranks

Recall the notation $\mathscr{A}\left((\varphi_i)_{i=1}^n\right)$ defined prior to Theorem 2. Denote $\mathscr{S}$ for the set of all profiles $(\pi_i)_{i=1}^n$ of functions $\pi_i : \prod_k T_k \to [0, 1]$ $(\forall i)$ such that $\sum_i \pi_i \le 1$ on $\prod_k T_k$. By the definition of $\mathscr{A}\left((\varphi_i)_{i=1}^n\right)$,

$$\mathscr{A}\left((\varphi_i)_{i=1}^n\right) = \arg\max_{(\pi_i)_{i=1}^n \in \mathscr{S}} \int_{\prod_i T_i} \sum_{i=1}^n \varphi_i(t_i)\pi_i\left((t_k)_{k=1}^n\right) dF_1(t_1)\cdots dF_n(t_n) \tag{59}$$

for any profile $(\varphi_i)_{i=1}^n$ of integrable functions $\varphi_i : T_i \to \mathbb{R}$ $(\forall i)$. Also note that $\mathscr{A}\left((\varphi_i)_{i=1}^n\right)$ contains an element $(\pi_i)_{i=1}^n$ such that $\pi_i(t_i, \cdot) = 0$ on T_{-i} whenever $\varphi_i(t_i) \le 0$.

B.6.4 Proof of Lemma 6

Recall that any $(Q_i)_{i=1}^n \in \mathscr{Q}$ is generated by some ex post allocation $(q_{iA}, q_{iB})_{i=1}^n$ via (2). For any $(q_{iA}, q_{iB})_{i=1}^n$ and any i, denote

$$q_i := q_{iA} - cq_{iB}. \tag{60}$$

Then (36) is equivalent to

$$\begin{aligned}
\max_{(q_i)_{i=1}^n \in \prod_i [-c,1]^{T_i}} \quad & \textstyle\sum_i \int_{T_i} \int_{T_{-i}} q_i(t_i, t_{-i}) Z_{i,+}(t_i) dF_{-i}(t_{-i}) dF_i(t_i) && (61)\\
\text{s.t.} \quad & \left(\textstyle\int_{T_{-i}} q_i(\cdot, t_{-i}) dF_{-i}(t_{-i})\right)_{i=1}^n \in \mathscr{Q}_{\text{mon}}, \\
& -c \le \textstyle\sum_i q_i \le 1 \quad \text{on} \quad \textstyle\prod_i T_i, \\
& \textstyle\int_{T_{-i}} q_i(t_i, \cdot) dF_{-i} \ge 0 \quad \forall t_i \in T_i \; \forall i \in \{1, \ldots, n\}. && (62)
\end{aligned}$$

For any $(Q_i)_{i=1}^n \in \mathscr{Q}_{\text{mon}} \cap \mathscr{Q}_+$, by the definitions of the two sets there exists $(q_i)_{i=1}^n \in \prod_i [-c, 1]^{T_i}$ such that, for each i, Q_i is the marginal of q_i, Q_i is weakly increasing, and $Q_i \ge 0$ on T_i. Thus, $(q_i)_{i=1}^n \in \prod_i [-c, 1]^{T_i}$ satisfies all the constraints in Problem (61). Denote $G_i := H_i(Z_{i,+})$, $G_i^L := \text{conv}_\text{L}\, H_i(Z_{i,+})$, $T := \prod_i T_i$, $t := (t_i)_{i=1}^n$ and $F := \prod_i F_i$. The objective in (61) given the feasible choice $(q_i)_{i=1}^n$ is equal to

$$\begin{aligned}
&\int_T \sum_i q_i(t) \overline{\overline{Z}}_{i,+}(t_i) dF(t) - \sum_i \int_{T_i} \left(G_i\left(F_i(t_i)\right) - G_i^L\left(F_i(t_i)\right)\right) dQ_i(t_i) \\
+ \; &\sum_i Q_i(b_i)\left(G_i(1) - G_i^L(1)\right) - \sum_i Q_i(a_i)\left(G_i(0) - G_i^L(0)\right)
\end{aligned}$$

by (56) and integration by parts. Here the third sum is zero because $G_i^L(1) = G_i(1)$ by definition of G_i^L, (54), for all i. Thus the objective in (61) is equal to

$$\underbrace{\int_T \sum_i q_i(t) \overline{\overline{Z}}_{i,+}(t_i) dF(t)}_{=:I((q_i)_{i=1}^n)} - \underbrace{\sum_i \int_{T_i} \left(G_i\left(F_i(t_i)\right) - G_i^L\left(F_i(t_i)\right)\right) dQ_i(t_i)}_{=:J((q_i)_{i=1}^n)} - \underbrace{\sum_i Q_i(a_i)\left(G_i(0) - G_i^L(0)\right)}_{=:K((q_i)_{i=1}^n)}.$$

Suppose $(\hat{Q}_i)_{i=1}^n \in \mathscr{Q}_{\text{mon}} \cap \mathscr{Q}_+$ and, for some $(\hat{q}_i)_{i=1}^n \in \mathscr{A}\left((\overline{Z}_{i,+})_{i=1}^n\right)$ and for each i, $\hat{Q}_i$ is the marginal of $\hat{q}_i$. Then the objective in (61) given $(\hat{q}_i)_{i=1}^n$ is equal to the above-displayed expression, where the roles of q_i and Q_i are played by $\hat{q}_i$ and $\hat{Q}_i$. Combine (58) with (59) to observe that, for any $(Q_i)_{i=1}^n \in \mathscr{Q}_{\text{mon}} \cap \mathscr{Q}_+$ and its associated $(q_i)_{i=1}^n$,

$$I\left((\hat{q}_i)_{i=1}^n\right) \ge I\left((q_i)_{i=1}^n\right)$$

and that the inequality is strict if $(q_i)_{i=1}^n \notin \mathscr{A}\left((\overline{Z}_{i,+})_{i=1}^n\right)$. By definition of G_i^L, $G_i \geq G_i^L$ on $[0,1]$; by the monotonicity constraint in (61), Q_i and $\hat{Q}_i$ are weakly increasing, and $\hat{Q}_i$ by definition of $\mathscr{A}\left((\overline{Z}_{i,+})_{i=1}^n\right)$ is constant on any interval where $G_i > G_i^L$. Thus,

$$J\left((\hat{q}_i)_{i=1}^n\right) = 0 \leq J\left((q_i)_{i=1}^n\right).$$

By (62), $Q_i(a_i) \geq 0$. If $\min \arg\min_{r\in[0,1]} \left(H_i(Z_{i,+})\right)(r) > 0$, then $\overline{Z}_{i,+} < 0$ on a neighborhood of a_i; since $(\hat{q}_i)_{i=1}^n \in \mathscr{A}\left((\overline{Z}_{i,+})_{i=1}^n\right)$, that means $\hat{q}_i(t_i,\cdot) = 0$ on T_{-i} for all t_i sufficiently near a_i. Thus, with $(\hat{q}_i)_{i=1}^n$ generating $\hat{Q}$, we have $\hat{Q}_i(a_i) = 0$. If, on the other hand, $\min \arg\min_{r\in[0,1]} \left(H_i(Z_{i,+})\right)(r) = 0$, then by the definition of G_i^L and (54), we have $G_i(0) = G_i^L(0)$. Thus, in either case,

$$K\left((\hat{q}_i)_{i=1}^n\right) \leq K\left((q_i)_{i=1}^n\right).$$

It follows that $I\left((\hat{q}_i)_{i=1}^n\right) - J\left((\hat{q}_i)_{i=1}^n\right) - K\left((\hat{q}_i)_{i=1}^n\right) \geq I\left((q_i)_{i=1}^n\right) - J\left((q_i)_{i=1}^n\right) - K\left((q_i)_{i=1}^n\right)$, with the inequality strict if $(q_i)_{i=1}^n \notin \mathscr{A}\left((\overline{Z}_{i,+})_{i=1}^n\right)$. Since the objective in (61) is equal to the objective in (36), we have proved that $(\hat{Q}_i)_{i=1}^n$ solves (36).

Conversely, if there is no element of $\mathscr{A}\left((\overline{Z}_{i,+})_{i=1}^n\right)$ whose marginals are $(Q_i)_{i=1}^n$, then the $(q_i)_{i=1}^n$ for which $(Q_i)_{i=1}^n \in \mathscr{Q}_{\text{mon}} \cap \mathscr{Q}_+$ are marginals is not an element of $\mathscr{A}\left((\overline{Z}_{i,+})_{i=1}^n\right)$. Then by the previous paragraph the objective given $(q_i)_{i=1}^n$ is strictly less than the objective given an element of $\mathscr{A}\left((\overline{Z}_{i,+})_{i=1}^n\right)$, which obviously exists. Thus $(Q_i)_{i=1}^n$ does not solve (36). That completes the proof. ■

B.6.5 Proof of Lemma 7

This is analogous to the proof of Lemma 6, where constraint (62) is replaced here by

$$\int_{T_{-i}} q_i(t_i,\cdot)dF_{-i} \leq 0 \quad \forall t_i \in T_i \; \forall i \in \{1,\ldots,n\}.$$

For any $(\check{Q}_i)_{i=1}^n \in \mathscr{Q}_{\text{mon}} \cap \mathscr{Q}_-$ such that, for some $(\check{q}_i)_{i=1}^n \in \mathscr{A}\left((-\overline{Z}_{i,-})_{i=1}^n\right)$ and for all i, $\check{Q}_i$ is the marginal of $-c\check{q}_i$, $(c\check{q}_i)_{i=1}^n$ satisfies all the constraints in the counterpart of (61). By the same token as the previous proof, $(c\check{q}_i)_{i=1}^n$ solves the counterpart of (61): Denote $G_i := H_i(Z_{i,-})$ and $G_i^R := \text{conv}_{\text{R}}\, H_i(Z_{i,-})$. We need only to make the following changes in the proof of Lemma 6: (i) $\sum_i Q_i(a_i)\left(G_i(0) - G_i^R(0)\right) = 0$ because $G_i(0) = G_i^R(0)$; (ii) $\sum_i Q_i(b_i)\left(G_i(1) - G_i^R(1)\right)$ is maximized by $\check{Q}$ because $\check{Q}_i(b_i) = 0$ when

$$\max\left(\arg\min_{r\in[0,1]} H_i(Z_{i,-})(r)\right) < 1,$$

and $G_i(1) = G_i^R(1)$ when the inequality does not hold. ■

B.7 Proof of (39)

The following general observation on ironing (defined in (52) and (53)) implies (39).

Lemma 11 *For any two integrable functions φ and ϕ defined on T_i, if $\varphi \geq \phi$ on T_i then*

$$\sup \{t \in T_i : \overline{\varphi}(t) < 0\} \leq \inf \left\{t \in T_i : \overline{\phi}(t) \geq 0\right\}. \tag{63}$$

Proof Note from (52) and (53) that the left-hand side of (63) is equal to

$$\inf \left(\arg\min_{t \in T_i} \left(H_i(\varphi)\right)\left(F_i(t)\right) \right),$$

and the right-hand side of (63) equal to

$$\inf \left(\arg\min_{t \in T_i} \left(H_i(\phi)\right)\left(F_i(t)\right) \right).$$

By (52), for any $t' > t$ the difference $\left(H(\varphi)\right)\left(F_i(t')\right) - \left(H(\varphi)\right)\left(F_i(t)\right) = \int_t^{t'} \varphi(s) dF_i(s)$ increases when φ increases pointwise. Thus, with $\varphi \geq \phi$, $\arg\min_{t \in T_i} \left(H(\varphi)\right)\left(F_i(t)\right)$ is less than $\arg\min_{t \in T_i} \left(H(\phi)\right)\left(F_i(t)\right)$ in strong-set order (Milgrom and Shannon [8]), implying (63). ■

C Proof of Theorem 3

Since the environment is assumed symmetric in this theorem, we suppress the subscripts in F_i, f_i, t_i, a_i, b_i and the like. For convenience we shall first assume that the welfare weights $(\omega_i)_{i=1}^n$ in the hypothesis of the theorem are identical across all i, and we shall indicate at the end the modifications (which are obvious) to remove the equal-weight assumption.

Thus let (Q, P) be any optimal mechanism that maximizes $\sum_i \int_a^b U_i(t_i \mid \tilde{Q}, \tilde{P}) dF(t_i)$ among all $(\tilde{Q}, \tilde{P})$ subject to IC, IR and BB in the symmetric environment F. By Theorem 2.a, Q allocates the bad according to $(Z_{i,-})_{i=1}^n$, which is given by (33). That equation, with the environment assumed symmetric, becomes

$$Z_{i,-}(t) = \beta Z_*(t) + \nu \left(t + \frac{F(t)}{f(t)} \right) \tag{64}$$

for all players i and all $t \in [a, b]$, where ν is the Lagrange multiplier for the joint constraint of IC, IR and BB, and

$$\begin{aligned}\beta &= \int_a^b \frac{1}{s} dF(s), && (65)\\ Z_*(t) &= t + \frac{F(t) - \Lambda(t)/\beta}{f(t)}, && (66)\\ \Lambda(t) &= \int_a^t \frac{1}{s} dF(s).\end{aligned}$$

Claim: Q allocates the bad with a strictly positive probability if

$$\exists t \in (a,b) : \int_a^t Z_*(s) dF(s) < 0. \tag{67}$$

Suppose, to the contrary, that (67) holds while $Q_i \geq 0$ on $[a,b]$ for all players i. Then the joint constraint of IC, IR and BB is non-binding (Lemma 12, Appendix D). That means, by the saddle point condition (28), $\nu = 0$. Thus, for all i, $Z_{i,-} = \beta Z_*$ and, by (52) and (67),

$$\left(H_i(Z_{i,-})\right)(F(t)) = \beta \int_0^{F(t)} Z_*(s) dF(s) < 0.$$

This, coupled with the fact $\left(H_i(Z_{i,-})\right)(0) = 0$ (due to (52)) implies that the convex hull of $H_i(Z_{i,-})$ is negatively sloped on $[0, F(t)]$. Then (53) implies $\overline{Z}_{i,-} < 0$, and hence $Q_i < 0$, on the nondegnerate interval $[a,t]$, contradicting the supposition that $Q_i \geq 0$ for all i.

By the claim proved above, the rest of the proof is to show that (67) is implied by the hypothesis $\Delta(F) > 1$. Use the definition of Z_* and integration by parts to obtain

$$\int_a^t Z_*(s) dF(s) = \int_a^t s dF(s) - \int_a^t \left(\frac{\Lambda(s)}{\beta} - F(s)\right) ds = tF(t) - \int_a^t \frac{\Lambda(s)}{\beta} ds.$$

By the definition of Λ and Fubini's theorem,

$$\begin{aligned}\int_a^t \frac{\Lambda(s)}{\beta} ds &= \frac{1}{\beta}\int_a^t \int_a^s \frac{1}{r} dF(r) ds = \frac{1}{\beta}\int_a^t \int_r^t \frac{1}{r} ds dF(r)\\ &= \frac{1}{\beta}\int_a^t \left(\frac{t}{r} - 1\right) dF(r) = \frac{t}{\beta}\Lambda(t) - \frac{F(t)}{\beta}.\end{aligned}$$

Thus, the inequality in (67) is equivalent to $tF(t) - \frac{t}{\beta}\Lambda(t) + \frac{F(t)}{\beta} < 0$, i.e.,

$$\frac{\Lambda(t)}{\beta} - F(t) > \frac{F(t)}{\beta t}.$$

By the definitions of Λ and β,

$$\begin{aligned}\frac{\Lambda(t)}{\beta} - F(t) &= 1 - F(t) - \frac{1}{\beta}\left(\int_a^b \frac{1}{s} dF(s) - \int_a^t \frac{1}{s} dF(s)\right)\\ &= M_F(1/t) + \frac{1}{\beta}\int_b^t \frac{1}{s} dF(s) = M_F(1/t) - \frac{1}{\beta}\int_{1/b}^{1/t} r d\left(1 - F(1/r)\right)\\ &= M_F(1/t) - L_F(1/t), && (68)\end{aligned}$$

with the second and third lines using (8) and (10). Thus, (67) is equivalent to

$$\exists s \in (1/b, 1/a) : M_F(s) - L_F(s) > \frac{F(1/s)s}{\beta}.$$

By continuity, the above condition is true if

$$\int_{1/b}^{1/a} (M_F(s) - L_F(s))\, dM_F(s) > \frac{1}{\beta}\int_{1/b}^{1/a} sF(1/s)dM_F(s). \tag{69}$$

The right-hand side, by (8), (65) and hence $\beta = \int_{1/b}^{1/a} sdM_F(s)$, is

$$\frac{1}{\beta}\int_{1/b}^{1/a} sF(1/s)dM_F(s) = 1 - \frac{1}{\beta}\int_{1/b}^{1/a} sM_F(s)dM_F(s). \tag{70}$$

Furthermore, by (68) and the definition of M_F,

$$\begin{aligned}
\int_{1/b}^{1/a} (M_F(s) - L_F(s))\, dM_F(s) &= \int_{1/b}^{1/a} \left(\frac{\Lambda(1/s)}{\beta} - F(1/s)\right) dM_F(s) \\
&= \int_{1/b}^{1/a} \frac{\Lambda(1/s)}{\beta} dM_F(s) - \int_{1/b}^{1/a} F(1/s)dM_F(s) \\
&= \frac{1}{\beta}\int_{1/b}^{1/a}\int_{s}^{1/a} rdM_F(r)dM_F(s) + \int_{1/b}^{1/a} M_F(s)dM_F(s) - 1 \\
&= \frac{1}{\beta}\int_{1/b}^{1/a}\int_{1/b}^{r} rdM_F(s)dM_F(r) - \frac{1}{2} \\
&= \frac{1}{\beta}\int_{1/b}^{1/a} rM_F(r)dM_F(r) - \frac{1}{2},
\end{aligned}$$

where the third line also uses the definition of Λ, and the fourth line is integration by parts. Plug the above equation into the right-hand side of (70) to obtain

$$\frac{1}{\beta}\int_{1/b}^{1/a} sF(1/s)dM_F(s) = \frac{1}{2} - \int_{1/b}^{1/a} (M_F(s) - L_F(s))\, d(M_F(s)).$$

Plugging this into the right hand side of (69), we see that (69) is equivalent to

$$\int_{1/b}^{1/a} (M_F(s) - L_F(s))\, dM_F(s) > \frac{1}{4},$$

which, by (11), is equivalent to $G(F) > 1/2$. This, by (12), is the same as $\Delta(F) > 1$.

Thus the theorem is proved in the case where the welfare weights $\omega_i = \omega_j > 0$ for all players i and j. To extend it to the general case of $(\omega_i)_{i=}^n \in \mathbb{R}_+^n \setminus \{\mathbf{0}\}$, observe that any welfare weights $(\omega_i)_{i=}^n \in \mathbb{R}_+^n$ amount to a profile $(\lambda_i)_{i=1}^n$ such that $\lambda_i = \omega_i F$ for each i.

Thus, by (33), Eq. (64) remains valid after the β and Z_* in (64) are replaced by β_* and $Z_{i,*}$ according to the following equations instead of (65) and (66):

$$\begin{aligned}\beta_* &= \omega_* \beta = \omega_* \int_a^b \frac{1}{s} dF(s), \\ Z_{i,*}(t) &= t + \frac{F(t) - \omega_i \Lambda(t) / (\omega_* \beta)}{f(t)},\end{aligned}$$

where $\omega_* := \max_{i=1,\ldots,n} \omega_i$. The rest of the proof is almost identical because (67), a sufficient condition for the conclusion, now becomes

$$\exists i : \exists t \in (a,b) : \int_a^t Z_{i,*}(s) dF(s) < 0.$$

It suffices to prove this inequality when $i = i_*$ for which $i_* \in \arg\max_{k=1,\ldots,n} \omega_k$. Since $\omega_{i_*} = \omega_*$ and hence $Z_{i_*,*} = Z_*$, the rest of the proof is the same as in the previous case.

D The Corollaries

Lemma 12 *For any solution Q of (25), if $Q_i \geq 0$ on $(a_i, b_i]$ for any i, then $\sum_i \langle Q_i : \mathbb{I} f_i + \rho(F_i) | > 0$.*

Proof For any i, $Q_i \geq 0$ on $(a_i, b_i]$ means that $\tau(Q_i) = a_i$ (with $\tau(Q_i)$ defined in (6)). Then (29) implies

$$\sum_i \langle Q_i : \mathbb{I} f_i + \rho(F_i) | \geq \left(\min_i a_i\right) \sum_i \int_{T_i} Q_i(t_i) dF_i(t_i).$$

We claim that $\sum_i \int_{T_i} Q_i(t_i) dF_i(t_i) > 0$. Otherwise, since $Q_i \geq 0$ for all i, $Q_i = 0$ for all i. Consequently, the objective in (25) is equal to

$$\sum_i \langle 0 : \beta_\lambda (\mathbb{I} f_i + \rho(F_i)) - \rho(\Lambda_i) | = 0.$$

Thus, by Lemma 4, the social welfare $\sum_{i=1}^n \int_{T_i} U_i(\cdot \mid Q, P) d\lambda_i = 0$. But then Q is suboptimal because assigning the good to any i for free, for whom $\lambda_i > 0$ on a positive-measure subset of T_i (such i exists because, by Theorem 1, λ_k's are not identically zero), generates a positive social welfare. Thus $\sum_i \int_{T_i} Q_i(t_i) dF_i(t_i) > 0$, hence $\sum_i \langle Q_i : \mathbb{I} f_i + \rho(F_i) | > 0$. ■

D.1 Corollary 1

A mechanism satisfies *assignment exclusivity* iff, for the underlying ex post allocation rule $(q_{iA}, q_{iB})_{i=1}^n$, $q_i^A(t)q_i^B(t) = 0$ for almost every $t \in \prod_k T_k$ and all i.

Corollary 1 *If $c > 0$ then any interim Pareto optimal mechanism is also an interim Pareto optimal mechanism subject to not only IC, IR and BB but also assignment exclusivity.*

Proof Theorems 1 and 2 combined, any interim Pareto optimal mechanism (Q, P) satisfies (a) in Theorem 2. That is, Q is generated by an ex post allocation $(q_{iA}, q_{iB})_{i=1}^n$ such that, for each i, the marginal of q_{iA} is equal to Q_i^+, and the marginal of cq_{iB}, Q_i^-. Thus, for any $t_i \in T_i$, if $q_{iA}(t_i, \cdot) > 0$ on a positive-measure subset of T_{-i} then

$$0 < \int_{T_{-i}} q_{iA}(t_i, \cdot)dF_{-i} = Q_i^+(t_i)$$

and then, by definition of Q_i^+ and Q_i^-,

$$0 = Q_i^-(t_i) = c\int_{T_{-i}} q_{iB}(t_i, \cdot)dF_{-i},$$

which, since $c > 0$ by hypothesis, implies $q_{iB}(t_i, \cdot) = 0$ a.e. on T_{-i}. Analogously, if $q_{iB}(t_i, \cdot) > 0$ on a positive-measure subset of T_{-i} then $q_{iA}(t_i, \cdot) = 0$ a.e. on T_{-i}. Thus, $q_{iA}q_{iB} = 0$ a.e. on $\prod_k T_k$ for all i, i.e., the mechanism (Q, P) satisfies assignment exclusivity, as desired. ■

D.2 Corollary 2

Corollary 2 *If f_i is continuously differentiable at a_i for each i, then in any interim Pareto optimum $(Q_i, P_i)_{i=1}^n$ for which the supporting welfare weighting $(\lambda_i)_{i=1}^n$ has the property that, for any i, λ_i has a Radon-Nikodym derivative $\frac{d\lambda_i}{dF_i}$ such that*

$$\forall i \in \{1, \dots, n\} : \liminf_{t_i \downarrow a_i} \lambda_i'(t_i) > 2\beta_\lambda a_i, \tag{71}$$

there exists a player i for whom Q_i is constant on a neighborhood of a_i.[13]

[13] The absolute continuity assumption in this corollary allows for the simple case, often assumed in mechanism design, where the Radon-Nikodym derivative of λ_i with respect to F_i is constantly equal to one so that the social welfare function is a simple sum of players' surpluses.

Dworczak, Kominers and Akbarpour [3] also obtain a conclusion similar to that of the corollary. Interestingly, our sufficient condition (71) looks similar to theirs, despite differences in our models.

Proof Let $(Q_i, P_i)_{i=1}^n$ be any Pareto optimum specified by the hypothesis. By Theorem 2, $(Q_i, P_i)_{i=1}^n$ is determined by some $(Z_{i,+}, Z_{i,-})_{i=1}^n$, with the latter given by (32) and (33), where ν denotes the Lagrange multiplier for the constraint $\sum_i \langle Q_i : \mathbb{I} f_i + \rho(F_i)| \geq 0$. By (32) and (33) and continuous differentiability of f_i at a_i, and $\lambda_i' = d\lambda_i/dF_i$, one can show, for each i, that $\frac{d}{dt_i} Z_{i,+}$ and $\frac{d}{dt_i} Z_{i,-}$ are continuous at a_i and

$$\begin{aligned} \frac{d}{dt_i} Z_{i,+}(a_i) &= 2(\beta_\lambda + \nu) - \frac{\lambda_i'(a_i)}{a_i} + \frac{f_i'(a_i)}{f_i^2(a_i)} (\beta_\lambda - \Lambda_i(b_i)) + \frac{\nu f_i'(a_i)}{f_i^2(a_i)}, \\ \frac{d}{dt_i} Z_{i,-}(a_i) &= 2(\beta_\lambda + \nu) - \frac{\lambda_i'(a_i)}{a_i}. \end{aligned}$$

Case (i): $\nu = 0$. Then, by (71), $\frac{d}{dt_i} Z_{i,-}(a_i) < 0$ for any i, and $\frac{d}{dt_{i_*}} Z_{i_*,+}(a_{i_*}) < 0$ for the i_* that maximizes $\Lambda_i(b_i)$ among all i (so $\beta_\lambda - \Lambda_{i_*}(b_{i_*}) = 0$). Thus, since $\frac{d}{dt_i} Z_{i,+}$ and $\frac{d}{dt_i} Z_{i,-}$ are continuous at a_i, both $Z_{i_*,+}$ and $Z_{i_*,-}$ are strictly decreasing on $[a_{i_*}, a_{i_*} + \delta)$ for some $\delta > 0$. Then $H_{i_*}(Z_{i_*,+})$ and $H_{i_*}(Z_{i_*,-})$ by (52) are strictly concave, and hence their convex hulls are affine, on $[F_{i_*}(a_{i_*}), F_{i_*}(a_{i_*} + \delta))$. Thus $\overline{Z}_{i_*,+}$ and $\overline{Z}_{i_*,-}$ are constant on $[a_{i_*}, a_{i_*} + \delta)$. Then Claims (a) of Theorem 2 implies that Q_{i_*} is constant on this neighborhood.

Case (ii): $\nu > 0$. We claim that there exists some i for whom $Q_i < 0$ on a neighborhood in T_i. Otherwise, the constraint $\sum_i \langle Q_i : \mathbb{I} f_i + \rho(F_i)| \geq 0$ is non-binding (Lemma 12), which coupled with the saddle point condition (28) implies that $\nu = 0$, contradiction. Now that $Q_i < 0$ on a neighborhood in T_i for some i, it follows from monotonicity (IC) of Q_i that $Q_i < 0$ on $[a_i, a_i + \eta)$ for some $\eta > 0$. Then Theorem 2.a implies $\overline{Z}_{i,-} < 0$ on $[a_i, a_i + \eta)$. By definition of ironing, if $\overline{Z}_{i,-}$ is not constant on a neighborhood of a_i, then $\overline{Z}_{i,-} = Z_{i,-}$ on that neighborhood and then $Z_{i,-} < 0$ on that neighborhood, contradicting the fact that $Z_{i,-}(a_i) = (\beta_\lambda + \nu) a_i > 0$ and that $Z_{i,-}$ is continuous. Thus, on a neighborhood of a_i, $\overline{Z}_{i,-}$ is constant, and hence so is Q_i. ■

D.3 Corollary 3

Corollary 3 *If $F_i = F$ for all i, with $[a, b]$ the support of F, if $\Delta(F) > 1$ and if (Q, P) is a symmetric mechanism that maximizes $\sum_i \int_a^b U_i(t_i|\tilde{Q}, \tilde{P}) dF(t_i)$ subject to IC, IR and BB, then there exists (x, y, z) for which $a < x < y < z < b_i$ and, for any player i, $U_i(\cdot \mid Q, P)$ is strictly decreasing on $[a, x)$, constant on (x, y), and strictly increasing on $(z, b]$.*

Proof By symmetry, (32) and (33) imply that $(Z_{i,+}, Z_{i,-}, Q_i)$ is identical across i. Thus we shall suppress the subscript i from these functions.

First, we claim that $Q > 0$ on $(b-\delta, b]$: By (32), $Z_+(b) = (\beta_\lambda + \nu)\, b > 0$. Thus, by continuity, $Z_+ > 0$ on $(b-\delta, b]$ for some $\delta > 0$. Then $H(Z_+)$ is strictly increasing on $(F(b-\delta), 1]$; thus, since 1 is the maximum of the domain for $H(Z_+)$, its convex hull $\operatorname{conv} H(Z_+)$ is also strictly increasing on $(F(b-\delta), 1]$. It follows that $\overline{Z}_+ > 0$ on $(b-\delta, b]$, hence Theorem 2.a implies that $Q > 0$ on $(b-\delta, b]$.

Second, by the hypothesis $\Delta(F) > 1$, Theorem 3 implies that the bad is assigned with strictly positive probability. This, coupled with symmetry of the mechanism, means that $Q < 0$ on some nondegenerate interval for all players. By IC, Q is weakly increasing, hence $Q < 0$ on $[a, a+\epsilon]$ for some $\epsilon > 0$.

The first and second observations combined, there exists $x, y \in (a, b)$ for which

$$Q(t_i) \begin{cases} < 0 & \text{if } t_i \in (a, x) \\ = 0 & \text{if } t_i \in (x, y) \\ > 0 & \text{if } t_i \in (y, b). \end{cases}$$

Recall the notation $\tilde{U}_i(\cdot \mid Q, P)$ from (4). By the envelope theorem, $\frac{d}{dt_i}\tilde{U}_i(t_i \mid Q, P) = Q(t_i)$ for almost every t_i, and $\tilde{U}_i(\cdot \mid Q, P)$ is absolutely continuous. Thus, $\tilde{U}_i(\cdot \mid Q, P)$ is strictly decreasing on $[a, x)$, constant on (x, y), and strictly increasing on $(y, b]$. Recall from (3) and (4) that

$$U_i(t_i \mid Q, P) = \frac{1}{t_i}\tilde{U}_i(t_i \mid Q, P)$$

for all $t_i \in [a, b]$. Thus, $U_i(\cdot \mid Q, P)$ is absolutely continuous on $[a, b]$[14] and, since $t_i \geq a > 0$, $U_i(\cdot \mid Q, P)$ is strictly decreasing on $[a, y)$. To complete the proof, we need only to show that $U_i(\cdot \mid Q, P)$ is strictly increasing on $(b-\delta, b]$ for some $\delta > 0$. To that end, pick any $t_i \in (a, b)$ at which $\tilde{U}_i(\cdot \mid Q, P)$ is differentiable and note

$$\frac{d}{dt_i}U_i(t_i \mid Q, P) = \frac{d}{dt_i}\left(\frac{\tilde{U}_i(t_i \mid Q, P)}{t_i}\right) = \frac{1}{(t_i)^2}\left(t_i Q(t_i) - \tilde{U}_i(t_i \mid Q, P)\right) = \frac{1}{(t_i)^2}P(t_i),$$

[14] It suffices to prove that $U_i(\cdot \mid Q, P)$ is Lipschitz on $[a, b]$: For any $t_i, t_i' \in [a, b]$, with $U_i := U_i(\cdot \mid Q, P)$,

$$\begin{aligned} |U_i(t_i') - U_i(t_i)| &= \left|\frac{1}{t_i'}\tilde{U}_i(t_i') - \frac{1}{t_i}\tilde{U}_i(t_i)\right| = \left|\frac{1}{t_i' t_i}\left(t_i\left(\tilde{U}_i(t_i') - \tilde{U}_i(t_i)\right) + \tilde{U}(t_i)\,(t_i - t_i')\right)\right| \\ &= \left|\frac{1}{t_i' t_i}\left(t_i \int_{t_i}^{t_i'} Q(s)ds + \tilde{U}(t_i)\,(t_i - t_i')\right)\right| \leq \left|\frac{1}{t_i' t_i}\left(t_i\,|t_i' - t_i| \max_{T_i}|Q| + \left|\tilde{U}(t_i)\right| |t_i - t_i'|\right)\right| \\ &\leq \frac{b_i}{a_i^2}\left(\max\{1, c\} + 1\right)|t_i' - t_i|, \end{aligned}$$

with the last inequality due to $-c \leq Q \leq 1$ and $0 \leq \tilde{U}_i \leq b_i$. Hence $U_i(\cdot \mid Q, P)$ is Lipschitz.

with the last equality due to (4). By the envelope equation (5), P is continuous and weakly increasing on $[a, b]$, hence $\lim_{t_i \uparrow b_i} P(t_i) = P(b_i) = \max_{[a,b]} P$. We claim that $P(b) > 0$, otherwise by Theorem 2.b.iii we have $P = 0$ on $[a, b]$, which contradicts (5), as Q has been proved to be nonzero on positive-measure subsets of $[a, b]$. Now that $P(b) > 0$, $\lim_{t_i \uparrow b} P(t_i) = P(b)$ means that $P > 0$ on $(b - \delta, b]$ for some $\delta > 0$. Thus $\frac{d}{dt_i} U_i(t_i \mid Q, P) > 0$ at any differentiable point t_i in this interval. This, coupled with absolute continuity of $U_i(\cdot \mid Q, P)$, implies that $U_i(\cdot \mid Q, P)$ is strictly increasing on this interval, as desired. ■

References

[1] Pierre-André Chiappori. *Matching with Transfers: The Economics of Love and Marriage*. Princeton University Press, 2017. 1

[2] Peter Cramton, Robert Gibbons, and Paul Klemperer. Dissolving a partnership efficiently. *Econometrica*, 55(3):615–632, May 1987. 1

[3] Piotr Dworczak, Scott Duke Kominers, and Mohammad Akbarpour. Redistribution through markets. Mimeo, March 28, 2019. 1, 13

[4] Herodotus. *The Histories*. Penguin Books: Baltimore, 1954. Translated by Aubrey de Sélincourt; with an introduction and notes by A. R. Burn. 2

[5] Bengt Holmström and Roger B. Myerson. Efficient and durable decision rules with incomplete information. *Econometrica*, 51(6):1799–1820, 1983. 1

[6] Simon Loertscher and Cédric Wasser. Optimal structure and dissolution of parternships. *Theoretical Economics*. Forthcoming. 1

[7] David G. Luenberger. *Optimization by Vector Space Methods*. John Wiley & Sons, 1969. 5.2.1

[8] Paul Milgrom and Chris Shannon. Monotone comparative statics. *Econometrica*, 62(1):157–180, 1994. B.7

[9] J. A. Mirrlees. An exploration in the theory of optimum income taxation. *Review of Economic Studies*, 38(2):175–208, 1971. 1

[10] Roger B. Myerson. Optimal auction design. *Mathematics of Operations Research*, 6(1):58–73, February 1981. 1, 4.1, 4.1, B.6

[11] Roger B. Myerson. *Game Theory: Analysis of Conflict*. Harvard University Press, 1991. 1

[12] Roger B. Myerson and Mark A. Satterthwaite. Efficient mechanisms for bilateral trading. *Journal of Economic Theory*, 29:265–281, 1983. 1

[13] H.L. Royden and P.M. Fitzpatrick. *Real Analysis*. Pearson Education, Boston, 4th edition, 2010. A

[14] Amartya Sen. *On Economic Inequality*. Oxford University Press, expanded 1997 edition, 1974. 7

[15] Nancy L. Stokey, Robert E. Lucas, Jr., and Edward C. Prescott. *Recursive Methods in Economic Dynamics*. Harvard Univerity Press, 1989. 1