Institutional Members: CEPR, NBER and Università Bocconi

WORKING PAPER SERIES

Extremal Incentive Compatible Transfers

Nenad Kos and Matthias Messner

Working Paper n. 359

First Version: December 2008, This Version: February 2011

IGIER – Università Bocconi, Via Guglielmo Röntgen 1, 20136 Milano –Italy
http://www.igier.unibocconi.it

Extremal Incentive Compatible Transfers*

Nenad Kos †
Dept. of Economics, IGIER
Bocconi University

Matthias Messner ‡
Dept. of Economics, IGIER
Bocconi University

First version: December 2008
This version: February 2011

Abstract

We characterize the boundaries of the set of transfers implementing a given allocation rule without imposing any assumptions on the agent's type space or utility function besides quasi-linearity. In particular, we characterize the pointwise largest and the pointwise smallest transfer that implement a given allocation rule and are equal to zero at some prespecified type (extremal transfers). Exploiting the concept of extremal transfers allows us to obtain an exact characterization of the set of all implementable allocation rules (the set of transfers is non-empty) and the set of allocation rules satisfying Revenue Equivalence (the extremal transfers coincide).

Furthermore, we show how the extremal transfers can be put to use in mechanism design problems where Revenue Equivalence does not hold. To this end we first explore the role of extremal transfers when the agents with type dependent outside options are free to participate in the mechanism. Finally, we consider the question of budget balanced implementation. We show that an allocation rule can be implemented in an incentive compatible, individually rational and ex post budget balanced mechanism if and only if there exists an individually rational extremal transfer scheme that delivers an ex ante budget surplus.

JEL Code: C72, D44, D82.
Keywords: Incentive Compatibility, Revenue Equivalence, Budget Balance, Mechanism Design

*For helpful discussions and comments we thank Rakesh Vohra and seminar participants at the European University Institute. We also benefited from comments of participants of the American Econometric Society Summer Meeting 2009, the European Econometric Society Meeting 2009, Capri 2009 and EARIE 2009. This paper subsumes some of the results previously circulated in the paper Incentive Compatible Transfers in Linear Environments.

†Department of Economics and IGIER, Bocconi University, Via Roentgen 1, I-20136 Milan, e-mail: nenad.kos@unibocconi.it

‡Department of Economics and IGIER, Bocconi University, Via Roentgen 1, I-20136 Milan, e-mail: matthias.messner@unibocconi.it

1 Introduction

In this paper we develop a new approach to mechanism design problems in quasi-linear environments with arbitrary type sets. Our analysis builds on the concept of *extremal transfers*. The extremal transfers for a given allocation rule and a prespecified type of the agent (the so called anchor type), describe (in a pointwise sense) the upper ($\bar{t}$) and the lower boundaries ($\underline{t}$) of the set of (interim expected) transfer schemes which i) are incentive compatible and ii) require a zero payment for the anchor type. We show that both of these boundaries constitute incentive compatible transfers, provided that the allocation rule to which they are associated is implementable. Consequently, they can be seen as the transfer payments that, respectively, maximize and minimize the payments that can be extracted from the agent subject to the constraint that the anchor type's transfer is zero.

With the use of extremal transfers we characterize the set of implementable allocation rules and the set of allocation rules satisfying Revenue Equivalence. In particular, we show that an allocation rule is implementable if and only if $\underline{t} \leq \bar{t}$ for every type of the agent. Moreover, an allocation rule satisfies Revenue Equivalence if and only if the associated extremal transfers coincide, i.e. if and only if $\underline{t} \equiv \bar{t}$. The characterization of implementability was first obtained in Rochet (1987), the characterization of revenue equivalence in Heydenreich, Müller, Uetz, and Vohra (2009). What is more, a careful reading of these papers reveals that in their proofs they make implicit use of what we characterize as extremal transfers. However, it is precisely the characterization of extremal transfers that enables us to obtain new results. While, for example, Heydenreich, Müller, Uetz, and Vohra (2009) focus on characterizing the conditions under which Revenue Equivalence obtains, we show how to deal with mechanism design problems regardless of whether the Revenue Equivalence is satisfied or not. As a technical tool in mechanism design revenue equivalence can be used to reduce the problem of choosing the optimal mechanism to the choice of the optimal nonmonetary allocation only. In Proposition 4 we show how this can be done using extremal transfers even when Revenue Equivalence does not necessarily hold.

To formalize the latter idea we first explore implementation under type dependent individual rationality constraints. We identify conditions under which the upper boundary of the set of transfer schemes implementing a given allocation rule in an incentive compatible and individually rational mechanism is given by some translation of an extremal transfer.

Next we show how extremal transfers can be used to characterize budget balanced implementation with individual rationality. In particular we show that an allocation rule can be implemented in an incentive compatible, individually rational and budget balanced mechanism if and only if it can be implemented through a translation of an extremal transfer that is individually rational and that generates an ex ante budget surplus. We apply our results to provide a simple characterization of budget balanced implementation in a linear environment with arbitrary compact type space. This characterization can, among other things, be applied to environments as in Myerson and Satterthwaite (1983) and Cramton, Gibbons, and Klemperer (1987) with more general type sets (i.e. type sets that are not necessarily intervals, but arbitrary compact sets).

In the final part of the paper we explore to which extent our results carry over to non quasi-linear environments. We show that a version of our characterization result regarding the set of implementable allocation rules goes through even when no restrictions on the payoff functions are imposed. In particular, we show that if a mechanism is incentive compatible then the indirect utility to which it gives rise lies between certain bounds $\underline{U}$ and $\bar{U}$. Conversely, if for some mechanism the condition $\underline{U} \leq \bar{U}$ holds and the indirect utility that it generates coincides with one of these two bounds then the mechanism is incentive compatible.

Related literature

In this paper we pick up several themes of the literature on mechanism design in quasi-linear environments. Since this literature is rather vast, a detailed discussion of all related papers is prohibitive. In what follows we therefore only mention the papers most closely related to ours. The organization of the discussion of the literature follows the order in which we present our results.

Characterization of the extremal transfers: Chung and Olszewski (2007) show for a very general setup, that the incentive compatible transfer that extracts the most money from the agent exists. We go beyond this result by i) showing that also a (pointwise) minimal transfer exists and by ii) providing a characterization of both of these extremal transfers.

Implementability: Among the papers that provide a characterization of the set of implementable allocation rules, the one that is most closely related to ours is Rochet (1987). He provides a 'no negative cycle'-condition which he shows to be necessary and sufficient for implementation in dominant strategies. In comparison to Rochet's approach our characterization offers the advantage that it not only allows us to make a statement about whether or not an allocation rule is implementable, but also delivers a description of the bounds of the set of transfers that implement the allocation rule.

Revenue Equivalence: The literature on Revenue Equivalence can be divided into two main strands. In a first and older strand Revenue Equivalence results are obtained in combination with explicit formulas for the transfers. That is, formulas that express the transfers in terms of the allocation rule which they are supposed to implement. The most notable early proponents of such results were Vickrey (1961) and Myerson (1981). For a more recent paper in this line of the literature see Milgrom and Segal (2002).

In more recent years there have been a number of papers (see for instance Chung and Olszewski (2007) and Heydenreich, Müller, Uetz, and Vohra (2009)) which focus their attention exclusively on the characterization of the conditions under which Revenue Equivalence obtains, without providing an explicit description of transfers. While our characterization of allocation rules for which Revenue Equivalence is satisfied is equally general as the ones obtained in the before mentioned papers, it has the advantages that it delivers a description of the corresponding transfers.

Budget balance: The issue of budget balance constraints in implementation goes back at least to D'Aspremont and Gérard-Varet (1975), D'Aspremont and Gérard-Varet (1979) and Walker (1980). While Walker shows that implementing efficient allocation rules in dominant strategies is typically incompatible with a (ex post) balanced budget, D'Aspremont and Gérard-Varet (1979) prove that when only Bayesian incentive compatibility is required efficient allocations can be implemented in a budget balanced way. Since Myerson and Satterthwaite (1983) it has been recognized, however, that even under Bayesian incentive compatibility there might be an unresolvable tension between budget balance and individual rationality. A more in depth treatment of mechanism design when both budget balance and individual rationality are required is given in Makowski and Mezzetti (1994), Krishna and Perry (1998) and Williams (1999). Our contribution to this literature is in that our treatment is more general. In particular, in order to derive our results we do not impose the standard conditions required for Revenue Equivalence.

In a recent paper Segal and Whinston (2010) explore whether efficient bargaining can arise for some status quo allocations. Most of their paper focuses on showing that efficient trade is possible when the status quo is the randomized allocation that has the same distribution as the equilibrium allocation in the mechanism. However, their Lemma 1 provides a general result on efficient budget balanced implementation. Our result, while being neither more nor less general then theirs provides certain advantages. First, we allow for more general outside options. Second, if one wanted to use their result to identify whether a certain allocation rule is implementable with budget balance, for a fixed status quo, one has to check all the transfer rules that implement a given allocation rule. Our result, on the other hand, provides the salient transfers, i.e. the extremal transfers with an appropriately chosen anchor type, and a characterization thereof, which enables us to express the necessary and sufficient condition as a single inequality written solely in the terms of the allocation rule.

In independent work, Rahman (2010) and Carbajal and Ely (2010) provide alternative characterizations of implementability. Rahman (2010) obtains his results using linear programming and duality. Carbajal and Ely (2010), after imposing additional structure (for example Lipschitz continuity), also derive bounds on the set of incentive compatible transfers. While the bounds they derive allow for a more convenient integral representation, they are - unlike ours - in general not tight. For a further discussion see Example 2. Hellwig (2010) provides a characterization of incentive compatible and individually rational mechanisms under not necessarily quasilinear environments on which several differentiability assumptions are imposed; see his Lemma 2.7.

2 Framework

We start with the description of the quasi-linear, single agent framework that will be at the center of our attention throughout most of the following sections. In terms of notation the extension to a multiple agent set up is straightforward. We will shortly comment on how our results for the single agent case carry over to multi agent environments at the end of Section 3.

An allocation is a pair $(x, \tau) \in X \times \mathbb{R}$, where x is a (non-monetary) choice variable and τ represents a monetary transfer (to be paid by the agent). With a slight abuse of terminology we will often use the term 'allocation' to indicate the non-monetary component x of a pair (x, τ). The agent's preferences over $X \times \mathbb{R}$ depend on his (preference-) type $v \in V$. We assume that they are quasi-linear for each $v \in V$. That is, type v's payoff at the allocation (x, τ) can be written in the form

$$u(x, v) - \tau,$$

where $u : X \times V \to \mathbb{R}$. We will refer to u as valuation function. It is important to point out that apart from the assumption of quasi-linearity of preferences, our framework is very general. In particular, for the time being we do not impose any assumptions at all on the set of decisions X, the set of preference types V or the valuation function u.[1]

We refrain from giving a definition of general mechanisms for this environment since due to the *Revelation Principle* it is without loss of generality to consider only direct mechanisms. A *direct mechanism* is a pair of functions, (q, t), where

$$q : V \to X,$$ [2]

and

$$t : V \to \mathbb{R}.$$

The function q, to which we will refer as *allocation rule* or *choice function*, associates with each type of the agent, $v \in V$, an allocation, $q(v) \in X$. Similarly, the *transfer rule* t specifies for each type of the agent a transfer to be paid by this type.

3 Incentive Compatible Transfers

The following definition recalls some standard terminology and concepts that we frequently use throughout the paper.

Definition 1 (Incentive compatibility, Implementability, Revenue Equivalence)**.**

[1]The implementation problem is non-trivial only if both X and V contain at least two elements.

[2]The restriction to deterministic mechanisms is made for notational convenience only. All our results extend to the case of non-deterministic allocation rules.

i) A direct mechanism (q, t) is **incentive compatible** *if*

$$u(q(v), v) - t(v) \geq u(q(v'), v) - t(v')$$

for all type pairs $v, v' \in V$.

ii) A choice function q is **implementable** *if there exists a transfer scheme t such that (q, t) is incentive compatible. Moreover, if (q, t) is incentive compatible then we say that the transfer rule t* **implements** *the choice function q.*

iii) A social choice function q satisfies **Revenue Equivalence** *if it is implementable and for any two incentive compatible mechanisms (q, t) and (q, t') there exists a constant $c \in \mathbb{R}$ such that*

$$t(v) = t'(v) + c,$$

for every $v \in V$.

It is straightforward to verify that the above definition of Revenue Equivalence coincides with the following one. A social choice function q satisfies Revenue Equivalence if it is implementable and for every $\hat{v} \in V$ and any two transfer schemes t and t' that implement q we have

$$t(v) - t(\hat{v}) = t'(v) - t'(\hat{v})$$

for all $v \in V$. We refer to the function $t(\cdot) - t(\hat{v})$ as the *transfer differential* at the type $\hat{v}$.

We are now ready to introduce the central concepts for our analysis. For any pair of types $v, \hat{v} \in V$ let $S(v; \hat{v})$ denote the set of all finite sequences in the type space V that start at $\hat{v}$ and end at v. That is, a finite sequence $\{v^j\}_{j=0}^n$ belongs to $S(v; \hat{v})$ if $v^j \in V$ for all $j = 0, \ldots, n$, $v^0 = \hat{v}$ and $v^n = v$. For any choice rule q and any $\hat{v} \in V$ we define the two functions $\bar{t}(\cdot; \hat{v}, q), \underline{t}(\cdot; \hat{v}, q) : V \to \overline{\mathbb{R}}$ as follows

$$\bar{t}(v; \hat{v}, q) \equiv \inf_{\{v^j\} \in S(v; \hat{v})} \sum_{j=1}^{n} \left[u(q(v^j), v^j) - u(q(v^{j-1}), v^j) \right], \qquad \text{and} \tag{1}$$

$$\underline{t}(v; \hat{v}, q) \equiv \sup_{\{v^j\} \in S(v; \hat{v})} \sum_{j=1}^{n} \left[u(q(v^j), v^{j-1}) - u(q(v^{j-1}), v^{j-1}) \right]. \tag{2}$$

Finally, we write $I(v; \hat{v}, q)$ for the interval $[\underline{t}(v; \hat{v}, q), \bar{t}(v; \hat{v}, q)]$. If $\bar{t}(v; \hat{v}, q) < \underline{t}(v; \hat{v}, q)$ then by convention we set $I(v; \hat{v}, q) = \emptyset$.

We postpone a detailed comment on the above defined expressions since their meaning and interpretation will become clear from the following results. For now we limit ourself to making the following two observations. First, notice that the functions $\underline{t}$ and $\bar{t}$ are not necessarily real valued. We have specified the extended real numbers $\overline{\mathbb{R}}$ as their range since the expressions on the right hand sides of (1) and (2) may take on the values $-\infty$ and ∞, respectively. On the other hand, it is also easily seen that for each triple $(v, \hat{v}, q)$, $\bar{t}(v; \hat{v}, q) \leq u(q(v), v) - u(q(\hat{v}), v) < \infty$ and $\underline{t}(v; \hat{v}, q) \geq u(q(v), \hat{v}, q) - u(q(\hat{v}), \hat{v}) > -\infty$.

For later reference, it is convenient to point out that whenever $\bar{t}(\cdot; \hat{v}, q)$ and $\underline{t}(\cdot; \hat{v}, q)$ are real valued, they must be equal to zero at the type $\hat{v}$.

Lemma 1. *Let q be an allocation rule and $\hat{v} \in V$. Then*

$$\bar{t}(\hat{v}; \hat{v}, q) \in \{0, -\infty\} \quad \textit{and} \quad \underline{t}(\hat{v}; \hat{v}, q) \in \{0, +\infty\}.$$

Proof. First, observe that for the trivial sequence $\{\hat{v}, \hat{v}\}$ the sum on the rhs of (1) is equal to zero, which implies that $\bar{t}(\hat{v}; \hat{v}, q) \leq 0$. Next, suppose that $\bar{t}(\hat{v}; \hat{v}, q)$ is strictly negative. Then there has to be a finite sequence, $\{v^j\}_{j=1}^n$, that starts and ends at $\hat{v}$ (cycle) and satisfies the condition $\sum_{j=1}^n \left[u(q(v^j), v^j) - u(q(v^{j-1}), v^j)\right] < 0$. If we preattach this cycle or arbitrarily many repetitions of it to any finite sequence that starts at $\hat{v}$ and ends at v we again obtain a sequence which starts at $\hat{v}$ and ends at v. By adding more repetitions of the cycle to the sequence the corresponding sum, $\sum_{j=1}^n \left[u(q(v^j), v^j) - u(q(v^{j-1}), v^j)\right]$, can be made arbitrarily small. Hence, $\bar{t}(\hat{v}; \hat{v}, q) < 0$ implies $\bar{t}(v; \hat{v}, q) = -\infty$ for all v. By an analogous argument it follows that $\underline{t}(\hat{v}; \hat{v}, q) > 0$ implies $\underline{t}(v; \hat{v}, q) = \infty$ for all v. □

For the remainder of our analysis we refer to $\hat{v}$ as *anchor type*. The following lemma describes a first important result.

Lemma 2. *Let (q, t) be an incentive compatible mechanism. Then, for any given $\hat{v} \in V$*

$$\underline{t}(v; \hat{v}, q) \leq t(v) - t(\hat{v}) \leq \bar{t}(v; \hat{v}, q)$$

for every $v \in V$.

Proof. Fix $\hat{v} \in V$ and let $\{v^j\}_{j=0}^n$ be a sequence in $S(v; \hat{v})$. Incentive compatibility implies

$$u(q(v'), v'') - u(q(v''), v'') \leq t(v') - t(v'') \leq u(q(v'), v') - u(q(v''), v') \tag{3}$$

for any pair $v', v'' \in V$. Consider the first inequality in (3). Adding the inequalities for all pairs of consecutive elements of the sequence, (v^{j-1}, v^j), yields

$$\sum_{j=1}^n \left[u(q(v^{j-1}), v^j) - u(q(v^j), v^j)\right] \leq \sum_{j=1}^n \left[t(v^j) - t(v^{j-1})\right] = t(v) - t(\hat{v}).$$

Taking the supremum over all sequences in $S(v; \hat{v})$ we get $\underline{t}(v; \hat{v}, q) \leq t(v) - t(\hat{v})$. The second inequality is obtained in an analogous fashion. □

Lemma 2 tells us that $\underline{t}(\cdot;\hat{v},q)$ and $\bar{t}(\cdot;\hat{v},q)$ represent respectively, a lower and an upper bound of the transfer differentials at the anchor type $\hat{v}$ (that can be obtained from transfer functions implementing the given allocation rule q). Put differently, $\underline{t}(\cdot;\hat{v},q)$ and $\bar{t}(\cdot;\hat{v},q)$ bound the set of all transfer schemes which implement q and require the anchor type $\hat{v}$ to pay a zero transfer. Our next result shows that these bounds are tight. That is, $\underline{t}$ and $\bar{t}$ themselves constitute incentive compatible schemes, which - by construction - vanish at the anchor type $\hat{v}$.

Lemma 3. *Let q be such that for some $\hat{v} \in V$ we have $I(v;\hat{v},q) \neq \emptyset$ for all $v \in V$. Then both $(q, \underline{t}(\cdot;\hat{v},q))$ and $(q, \bar{t}(\cdot;\hat{v},q))$ are incentive compatible mechanisms.*

Proof. See the Appendix. □

The preceding two lemmata constitute the basis for the following theorem.

Theorem 1. *For any $\hat{v} \in V$ the following statements are true.*

i) An allocation rule q is implementable if and only if $I(v;\hat{v},q) \neq \emptyset$ for every $v \in V$.

ii) An allocation rule q satisfies Revenue Equivalence if and only if for all $v \in V$, $I(v;\hat{v},q)$ is a singleton.

iii) Let q be an implementable allocation rule and let $\hat{v} \in V$. If t implements q, and $t(\hat{v}) = 0$, then

$$\underline{t}(v;\hat{v},q) \leq t(v) \leq \bar{t}(v;\hat{v},q),$$

for every $v \in V$.

Proof. See the Appendix. □

The main novelty of Theorem 1 is that it provides a characterization of the boundaries of the set of incentive compatible transfers without imposing any topological structure or ordering assumptions on the environment. With expressions (1) and (2) two explicit formulas for the calculation of the extremal incentive compatible transfers are given. We will demonstrate how the extremal transfers are calculated in several examples below. A detailed treatment of the case of linear environments is given in Kos and Messner (2010a).

In addition to providing an explicit description of the boundaries of the set of incentive compatible transfers, the theorem unifies several existing results in the literature on mechanism design. The first statement of the theorem is formally equivalent to the characterization of implementability in Rochet (1987). According to Rochet's characterization an allocation rule q is implementable if and only if there does not exist a finite sequence of valuations $\{v^j\}_{j=0}^n \subset V$, which satisfies the conditions i) the sequence forms a cycle (i.e. $v^0 = v^n$, $v^j \neq v^k$ for all $j \neq k$) and ii) $\sum_{j=1}^{n} \left[u(q(v^j), v^j) - u(q(v^{j-1}), v^j)\right] < 0$. We will refer to the Rochet's conditions as cyclical monotonicity. Thus an allocation rule is implementable if and only if it is cyclically monotone.

The equivalence between cyclical monotonicity and our characterization is formalized in the next lemma.

Lemma 4. *Let q be an allocation rule and $\hat{v} \in v$. The following statements are equivalent. (a) q is cyclically monotone. (b) For every $v \in V$, $\underline{t}(v; \hat{v}, q) \leq \bar{t}(v; \hat{v}, q)$.*

Proof. If there exists a 'negative sum cycle' a la Rochet at type $\tilde{v}$, then there exists a sequence starting and ending at the anchor type $\hat{v}$ with a negative sum (just insert sufficiently many repetitions of this negative sum cycle in the middle of the cycle $\{\hat{v}, \tilde{v}, \hat{v}\}$). We have already observed earlier, that the existence of a negative sum sequence starting and ending at $\hat{v}$ implies that $\bar{t}(\cdot; \hat{v}, q)$ is equal to $-\infty$ everywhere.

Similarly, by inverting the order of the negative sum cycle at $\tilde{v}$ we obtain a positive sum cycle at $\tilde{v}$ which in turn means that a positive sum cycle exists everywhere, in particular at $\hat{v}$. By our earlier observations we know that this means that $\underline{t}(v; \hat{v}, q) = \infty$ for all v. But if $\underline{t}(\cdot; \hat{v}, q) \equiv \infty$ and $\bar{t}(\cdot; \hat{v}, q) \equiv -\infty$, then $I(v, \hat{v}, q) = \emptyset$ for all v.

Conversely, suppose that $\underline{t}(v; \hat{v}, q) \leq \bar{t}(v; \hat{v}, q)$ for all $v \in V$. Then, in particular, $\underline{t}(\hat{v}; \hat{v}, q) \leq \bar{t}(\hat{v}, \hat{v}, q)$. By construction we have that $\underline{t}(\hat{v}; \hat{v}, q) \geq 0$ and $\bar{t}(\hat{v}; \hat{v}, q) \leq 0$. Combining these observations implies that whenever q is implementable then $\underline{t}(\hat{v}; \hat{v}, q) = 0 = \bar{t}(\hat{v}; \hat{v}, q)$. Since $\bar{t}(\hat{v}; \hat{v}, q)$ is the lower bound of all values that can be written as a sum of the form $\sum_{j=1}^{n} \left[u(q(v^j), v^j) - u(q(v^{j-1}), v^j)\right]$ for some finite sequence that starts and ends at $\hat{v}$, the sum for no such sequence can be negative. But then there can be no negative-value cycle for any other type either since we could always include (sufficiently many replica of) it into a sequence which starts and ends at $\hat{v}$ in order to get a negative sum sequence at $\hat{v}$.

□

The characterization of Revenue Equivalence here is formally equivalent to the one given in Heydenreich et al. (2009). While Heydenreich et al. (2009) provide a graph theoretic characterization, we formulate our result in direct economic terms (the extremal transfers).

The most novel result of Theorem 1, is the one contained in statement iii) in which we describe the boundaries of the set of transfer differentials for implementable allocation rules. By the preceding lemma we also know that these boundaries themselves constitute incentive compatible transfer rules. The two boundaries correspond to the transfers that extract the most money ($\bar{t}$) and the least money ($\underline{t}$), respectively, from the agent, given that the transfer of the anchor type $\hat{v}$ is zero. That is, any transfer rule which i) implements the given allocation rule and ii) requires a zero payment from the anchor type $\hat{v}$ is pointwise dominated by (dominates) the transfer scheme $\bar{t}$ ($\underline{t}$).

Finally, we should point out that Theorem 1 does not say that condition $\underline{t}(\cdot;\hat{v},q) \leq t(\cdot) - t(\hat{v}) \leq \bar{t}(\cdot;\hat{v},q)$ is a sufficient condition for incentive compatibility of t. In fact, it is rather easy to construct examples of transfer rules which violate incentive compatibility but still do lie within the bounds defined by the extremal transfers.

Multiple agents

Extending the preceding analysis to a framework with multiple agents whose valuations are independently distributed does not generate any major conceptual difficulties. The only difference between the single agent case that we have considered so far and a multi agent environment with independently distributed types lies in the fact that in the latter the agents' incentive constraints are defined in terms of agents' expected utilities (where the expectation is taken with respect to the other agents' types). Consequently, the bounds on transfer differentials in the multi agent case are to be interpreted as bounds on (interim) expected transfers.

In order to see this, suppose there are I agents, indexed by $i = 1, \ldots, I$. Let V_i be the set of types of player i. We denote type profiles by v and write V for the set of all such profiles, i.e. $V = \times_i V_i$. Each agent's type is private information. The designer and the agents share the common prior belief that types are distributed according to the probability measure F. Types are independent in the sense that i) F is the product of its marginal probability measures on V_i (i.e. if F_i is the marginal of F on V_i then $F = \times F_i$) and ii) for each i the marginal probability measure on V_{-i}, F_{-i} is given by the product $\times_{j \neq i} F_j$.

Player i's valuation function is allowed to depend not only on his own type, v_i, but also on the type profile of all other players, v_{-i}. Thus we write $u_i(x, v) - \tau_i$ for the payoff of agent i when the type profile is v, the choice is x and agent i's transfer is τ_i.

In this multi agent environment a direct mechanism is a pair of functions (Q, T) that specify for each profile of valuations, v, an allocation, $Q(v) \in X$, and a profile of transfers, $T(v) \in \mathbb{R}^I$, respectively. A mechanism (Q, T) is said to be (Bayesian) incentive compatible if for each player $i = 1, \ldots, I$ we have

$$E_{v_{-i}}[u_i(Q(v_i, v_{-i}), v_i, v_{-i}) - T_i(v_i, v_{-i})] \geq E_{v_{-i}}[u(Q(v_i', v_{-i}), v_i, v_{-i}) - T_i(v_i', v_{-i})], \qquad \forall\, v_i, v_i' \in V_i,$$

where $E_{v_{-i}}$ is the expectation operator with respect to the probability measure F_{-i} on V_{-i}.

Denoting

$$t_i(v_i) = E_{v_{-i}}[T_i(v_i, v_{-i}]$$

and

$$\tilde{u}_i(v_i', v_i; Q) = E_{v_{-i}}[u_i(Q(v_i', v_{-i}), v_i, v_{-i})]$$

this condition can be rewritten as

$$\tilde{u}_i(v_i, v_i; Q) - t_i(v_i) \geq \tilde{u}_i(v_i', v_i; Q) - t_i(v_i'), \qquad \forall\, v_i, v_i' \in V_i. \tag{4}$$

Notice that (4) has the same form as the incentive constraint in the single agent case. Thus, using the functions $\tilde{u}_i$ we can define for each allocation rule Q and any $\hat{v}_i \in V_i$ upper and lower bounds on the (interim) *expected* transfer differential $t_i(\cdot) - t_i(\hat{v}_i)$. Extending the characterization result from Theorem 1 to the profile of these (interim) expected transfer differentials is simply a matter of adding the appropriate quantifiers for i.

3.1 Examples

In what follows we illustrate the use of our tools by analyzing two simple environments where revenue equivalence does not necessarily hold. The two examples will also serve as an introduction to the next section, for they bring to the forefront issues that emerge in optimal mechanism design problems that include individual rationality constraints.

Example 1. Let the agent's valuation function be (bi-)linear, i.e. $u(x, v) = vx$, $X = [0, 1]$, and the set of types a compact subset of $\mathbb{R}$. In his seminal paper Myerson (1981) has shown that if the support of the type distribution is an interval and q is an implementable allocation rule, then any transfer scheme t that implements q can be written as

$$t(v) = t(0) + vq(v) - \int_0^v q(x)dx.$$

The objective of this example is to show that our approach leads to a simple integral representation of transfers even when the support of the type distribution is not necessarily an interval but an arbitrary compact subset of $\mathbb{R}$. Here we show how extremal transfers can be represented in an integral form. For the characterization of all transfers the reader is referred to Kos and Messner (2010b).

Proposition 1. *Let q be an implementable allocation rule. Then for any $\hat{v} \in V$,*

$$\bar{t}(v; \hat{v}, q) = \int_{\hat{v}}^{v} s d\bar{q}(s; \hat{v}) \qquad \text{and} \qquad \underline{t}(v; \hat{v}, q) = \int_{\hat{v}}^{v} s d\underline{q}(s; \hat{v}),$$

for all $v \in V$, where

$$\underline{q}(v; \hat{v}) = \begin{cases} \inf\{q(\tilde{v}) : \tilde{v} \in V, \tilde{v} \geq v\} & \text{if } v \geq \hat{v} \\ \sup\{q(\tilde{v}) : \tilde{v} \in V, \tilde{v} \leq v\} & \text{if } v < \hat{v} \end{cases} \quad \text{and} \quad \bar{q}(v; \hat{v}) = \begin{cases} \sup\{q(\tilde{v}) : \tilde{v} \in V, \tilde{v} \leq v\} & \text{if } v \geq \hat{v} \\ \inf\{q(\tilde{v}) : \tilde{v} \in v, \tilde{v} \geq v\} & \text{if } v < \hat{v} \end{cases}.$$

Proof. See the Appendix. □

The functions $\bar{q}$ and $\underline{q}$ extend the allocation rule q from V to the smallest interval that contains V. Since q is implementable it is non-decreasing and this property is preserved by $\bar{q}$ and $\underline{q}$. When V is convex q, $\bar{q}$, and $\underline{q}$ coincide yielding the Revenue Equivalence result of Myerson (1981). The other well studied case is the one with a finite number of types. There it is easy to see that the extremal transfers extracting the most money are obtained by making all the downward adjacent IC constraints for the types above the anchor type, and all the upward adjacent IC constraints for the types below the anchor type binding. For further details see Kos and Messner (2010b).

We would like to point out that the characterization of $\bar{t}(v; \hat{v}, q)$ for $\hat{v} = \min V$ was previously provided by Skreta (2006) who was studying optimal auctions under arbitrary (not necessarily connected) valuation spaces.

Example 2. Let $X = V = [0, 1]$ and

$$u(x, v) = -k(x) \left| x - v \right|,$$

where $k(x) = 1$ for $x \in X \cap \mathbb{Q}$ and $k(x) = k$, for $k > 1$ otherwise. We focus on the efficient allocation rule $q(v) = v$. Define $t(v) = 0$ for all v and notice that (q, t) is incentive compatible since the agent gets utility zero from reporting truthfully and a negative utility if he deviates. Hence, q is implementable.

In what follows we characterize the largest transfer that implements the allocation rule q with the anchor type $\hat{v} = 0$. Let $v \in V$ and $\epsilon > 0$. Pick an increasing sequence $\{v^j\}_{j=0}^n$ with $v^0 = 0$ and $v^n = v$, such that $v^j \in \mathbb{Q}$ for $j = 0$ to $n - 1$ and $v^n - v^{n-1} < \epsilon$. Then $\sum u(q(v^j), v^j) - u(q(v^{j-1}), v^j) < v + k\epsilon$. Using this inequality, the definition of $\bar{t}(v; 0, q)$ and the fact that $\bar{t}(v; 0, q) = 0$ one obtains

$$\bar{t}(v; 0, q) \leq v + k\epsilon.$$

Since epsilon can be chosen arbitrarily small $\bar{t}(v; 0, q) \leq v$. Finally, it is easy to verify that the transfer scheme defined by $t(v) = v$ implements q. Therefore

$$\bar{t}(v; 0, q) = v.$$

In the same manner one shows $\underline{t}(v; 0, q) = -v$.

Carbajal and Ely (2010) also provide a bound on incentive compatible transfers; see Example 1 and Theorem 2 in their paper. Their bound has an advantage in that it has a convenient integral representation. However, the present example shows that their is not necessarily tight.

Applying their analysis, see Theorem 2 in Carbajal and Ely (2010), one obtains that for any two transfers t and t' implementing the given allocation rule and vanishing at $v = 0$

$$\left|t(v) - t'(v)\right| \leq 2kv. \tag{5}$$

Our bound, on the other hand, yields

$$\left|t(v) - t'(v)\right| \leq 2v, \tag{6}$$

and is tight, i.e. there exist such t and t' implementing q that achieve it. For k large their bound is rather imprecise.

A fruitful procedure might be to combine the techniques of Carbajal and Ely (2010), when the assumptions of their paper are satisfied, with the ones developed here. One could compute the bounds on transfers as proposed in Carbajal and Ely (2010). If those bounds represent incentive compatible transfers they constitute extremal transfers. If they do not, the extremal transfers can be computed from the characterization of Theorem 1 in the present paper.

4 Extremal transfers and individual rationality

In many economic applications the designer does not have the power to force the agent to participate in the mechanism but needs to make sure that participation is voluntary or - in more standard terminology - *individually rational.* In the previous section we have seen how to characterize the boundaries of the set of transfers which implement a given allocation rule. In this section we address the question under what conditions the concept of extremal transfers can also be used to describe the boundaries of the set of admissible transfers when on top of incentive compatibility also individual rationality is required.

The answer to this question has important implications for all mechanism design problems where the designer needs to determine the largest transfer scheme by which a certain allocation rule can be implemented under voluntary participation. Notice that this is the case whenever the designer maximizes his expected revenue (auctions). In the next section we will argue that the same is true also in other applications, like optimal mechanism design problems where the designer maximizes some weighted sum of the participating players' payoffs, or implementation problems where the designer faces an ex post budget constraint.

Most papers in the existing literature which deal with problems of optimal mechanism design impose conditions which imply that Revenue Equivalence holds for all allocation rules. The reason for doing so is that under Revenue Equivalence the choice of the allocation rule also determines (up to a constant) the transfer scheme. Hence, under Revenue Equivalence the designer's problem effectively reduces to choosing an allocation rule only.

An important insight that can be derived from our results is that often optimal mechanism design problems can be ‘reduced’ by the transfer dimension even when Revenue Equivalence does not hold. If for every allocation rule the largest feasible transfer that implements it is a (translation of an) extremal transfer, then again the problem of finding the optimal transfer is rather trivial once we have determined the optimal allocation rule.

We start our analysis by formally defining the individual rationality constraint. We denote the payoff that an agent of type v realizes when he opts out of the mechanism by $\phi(v)$. Of course, an agent is willing to participate in a mechanism if and only if doing so allows him to earn a payoff that is no smaller than the value of his outside option.

Definition 2 (Individual Rationality)**.** *The mechanism* (q, t) *is* **individually rational** *(henceforth, IR) if*

$$u(q(v), v) - t(v) \geq \phi(v) \qquad \forall\, v \in V. \tag{7}$$

Later on we will refer again to multiple agent environments. In those settings we require individual rationality to hold at the *interim* stage. That is, a mechanism (Q, T) is individually rational if for all i and all $v_i \in V_i$ we have

$$E_{v_{-i}}[u(Q(v_i, v_{-i}), v_i, v_{-i}) - T_i(v_i, v_{-i})] \geq E_{v_{-i}}[\phi_i(v)].$$

Suppose that for a mechanism (q, t) there is a type $\tilde{v}$ who realizes the lowest net benefit from participation in the mechanism, i.e.

$$u(q(\tilde{v}), \tilde{v}) - t(\tilde{v}) - \phi(\tilde{v}) \leq u(q(v), v) - t(v) - \phi(v).$$

Then (q, t) is individually rational if and only if condition (7) holds for type $\tilde{v}$. We will therefore refer to $\tilde{v}$ as *IR-critical* type for the mechanism (q, t).

Definition 3 (IR-critical types)**.** *Type* $\tilde{v}$ *is an* **IR-critical** *type for the mechanism* (q, t) *if*

$$u(q(\tilde{v}), \tilde{v}) - t(\tilde{v}) - \phi(\tilde{v}) \leq u(q(v), v) - t(v) - \phi(v) \qquad \forall v \in V.$$

The following assumption is of central importance for the remainder of this section. Its role will become clear from the next proposition. It is therefore convenient to postpone a more detailed discussion of the assumption after the presentation of the proposition.

Assumption 1. q is implementable and there exists a type v_q, such that v_q is IR-critical for the mechanism $(q, \bar{t}(\cdot; v_q, q))$.

We are now ready to state our first important result regarding implementation problems with individual rationality constraints.

Proposition 2. *Suppose q satisfies Assumption 1. Let v_q be IR-critical for $(q, \bar{t}(\cdot; v_q, q))$ and let $\tilde{t} : V \to \mathbb{R}$ be defined by $\tilde{t}(v) = \bar{t}(v; v_q, q) + u(q(v_q), v_q) - \phi(v_q)$. The following is true:*

i) $(q, \tilde{t})$ is incentive compatible and individually rational.

ii) For all transfer schemes t such that (q, t) is incentive compatible and individually rational,

$$\tilde{t}(v) \geq t(v) \quad \text{for all } v \in V.$$

Proof. Notice that the payoff of type v_q from the mechanism $(q, \bar{t}(\cdot; v_q, q) + u(q(v_q), v_q) - \phi(v_q))$ is exactly $\phi(v_q)$. This means that the IR constraint for this type is satisfied with equality. Since type v_q is an IR-critical type for $(q, \bar{t}(\cdot; v_q, q))$ it must be so also for all translations of $\bar{t}(\cdot; v_q, q)$ and thus the mechanism $(q, \tilde{t})$ is individually rational.

The second statement of the proposition follows readily from Theorem 1: If (q, t) is incentive compatible and individually rational, then there must be a $\delta \geq 0$ such that $u(q(v_q), v_q) - t(v_q) - \delta = \phi(v_q)$. $\tilde{t}$ is the pointwise largest payment scheme in the set of all transfers that are incentive compatible and that guarantee type v_q a payoff that is just equal to the value of his outside option, $\phi(v_q)$. Since this set includes the transfer $t(\cdot) + \delta$ it follows that

$$\bar{t}(v; v_q, q) + u(q(v_q), v_q) - \phi(v_q) \geq t(v) + \delta \geq t(v) \qquad \text{for all } v \in V.$$

□

While the requirement that there exists a type v_q which is IR-critical when q is implemented through the extremal transfer defined with respect to v_q, is a rather weak condition, it is not an assumption on the fundamentals of the environment but involves endogenous/derived objects like extremal transfers. Moreover, it is an assumption that can be difficult to verify as it requires to find a fixed point of the correspondence $f_q : V \Rightarrow V$, defined by

$$f_q(v) = \arg\min_{\tilde{v} \in V} u(q(\tilde{v}), \tilde{v}) - \bar{t}(\tilde{v}; v, q) - \phi(\tilde{v}).$$

Identifying such a fixed point might well be a non-trivial problem. Of course, if q is an allocation rule which satisfies Revenue Equivalence then only the level but not the shape of the extremal transfers can depend on the anchor type. But then f_q is constant (provided that it is well defined). Thus, the fixed point problem reduces to a minimization problem.[3]

In general the complexity of the task of finding a fixed point of f_q depends on the exact properties of the functions u and ϕ. In the next assumption we formulate a condition on the

[3]Cramton, Gibbons, and Klemperer (1987) calculate the IR-critical types for an environment where both the valuation functions and the outside options of the agents are linear in their types.

agent's 'net valuation' function $u(x, v) - \phi(v)$ which not only guarantees that for any allocation rule q the correspondence f_q does admit a fixed point but also allows to directly identify this fixed point.

Assumption 2. There exists a type $\hat{v} \in V$ such that

$$u(x, \hat{v}) - \phi(\hat{v}) \leq u(x, v) - \phi(v) \tag{8}$$

for all $(x, v) \in X \times V$.

Notice that unlike Assumption 1, Assumption 2 neither refers to a specific allocation rule nor does it involve any endogenous objects, but is defined only in terms of the fundamentals of the environment. It is not difficult to see that for any implementable allocation rule q the set of fixed points of f_q contains all types that satisfy condition (8).

Lemma 5. *Suppose that Assumption 2 holds. If q is implementable and $\hat{v} \in V$ satisfies (8), then $\hat{v}$ is IR-critical for $(q, \bar{t}(\cdot; \hat{v}, q))$.*

Proof. The result is obtained by combining incentive compatibility (which implies the first inequality of the following expression) and Assumption 2 (second inequality):

$$\begin{aligned} u(q(v), v) - \bar{t}(v; \hat{v}, q) - \phi(v) &\geq u(q(\hat{v}), v) - \bar{t}(\hat{v}; \hat{v}, q) - \phi(v) \\ &\geq u(q(\hat{v}), \hat{v}) - \bar{t}(\hat{v}; \hat{v}, q) - \phi(\hat{v}). \end{aligned}$$

□

Notice that Assumption 2 is satisfied by most standard models with one-dimensional type sets like Myerson (1981) and Myerson and Satterthwaite (1983).

A framework that violates Assumption 2 is considered in Cramton et al. (1987). There the authors consider the problem of how to allocate a good between two (or more) agents who are privately informed about their valuations and who already own a share $\theta_i > 0$ of the good. The valuation function and the outside option are assumed to be linear. In particular, the net valuation is $u_i(x, v_i) - \phi_i(v_i) = v_i(x_i - \theta_i)$ (where x_i is the share of the object assigned to agent i). This function is minimized by the lowest type, $v_i = \underline{v}_i$ if $\theta_i \leq x_i$; whenever instead $\theta_i > x_i$ then the minimizing type is the highest type of agent i, $v_i = \bar{v}_i$.

Despite the fact that Assumption 2 does not hold in this set up, for every implementable allocation rule Q the associated fixed point problems, $v_i \in f_{q_i}(v_i)$, do have a solution. Cramton et al. (1987) show this in a linear framework where agents' types are drawn from a distribution that admits a strictly positive density on the compact support $[\underline{v}, \bar{v}]$, which is the typical assumption under which Revenue Equivalence holds. In Kos and Messner (2010a) we show how this result is extended to situations where type sets are disconnected (but compact).

5 Budget balance

In many applications of mechanism design the designer not only has to satisfy the constraints of incentive compatibility and individual rationality, but also faces resource constraints. For instance, in a bilateral trade setting it is most natural to require that every trade has to be financed exclusively by transfers among the involved individuals. Similarly, in a public good environment it is reasonable to impose that the public good has to be financed exclusively through the transfers collected from the agents.

In the mechanism design literature resource constraints of the just described type run under the label of *ex post budget balance constraints*. Several papers address the question whether certain 'desirable' allocation rules (typically, efficient ones) can be implemented in a way that respects individual rationality and ex post budget balance.[4] Most of the papers take as their starting point a set of assumptions that are strong enough to imply Revenue Equivalence for all allocation rules. In what follows we use the concept of extremal transfers in order to generalize these results along two dimensions. First, we do not restrict our attention to environments where Revenue Equivalence holds for all allocation rules. Second, our findings are not limited to efficient allocation rules only.

We start our analysis with a more formal definition of the concept of budget balance.

Definition 4. *A mechanism* (Q, T) *is (ex post)* **budget balanced** *if*

$$\sum_{i=1}^{I} T_i(v) = 0$$

for every $v \in V$. *It is* **ex ante budget balanced** *if*

$$E_v \left[\sum_{i=1}^{I} T_i(v) \right] = 0.$$

Finally, we say that it **runs an ex ante budget surplus** *if*

$$E_v \left[\sum_{i=1}^{I} T_i(v) \right] \geq 0.$$

The following proposition refers to the set of all allocation rules that satisfy Assumption 1. For this class of allocation rules an exact criterion for individually rational and ex post budget balanced implementability is given. In particular, it shows that an allocation rule Q in this set can be implemented with individual rationality and ex post budget balance if and only if there is an extremal transfer (or translation thereof) which is compatible with voluntary participation and generates an ex ante budget surplus.

[4]See among others Myerson (1981), Makowski and Mezzetti (1994) and Krishna and Perry (1998).

Proposition 3. *Suppose that for the allocation rule Q there is a type profile, $\hat{v}$, which is IR-critical for the mechanism $(Q, \bar{t}(\cdot; \hat{v}, Q))$. Then there exists a transfer rule T such that (Q, T) is ex post budget balanced, incentive compatible and individually rational if and only if the transfer scheme $\bar{T}$ defined by*

$$\bar{T}_i(v) = \tilde{u}_i(\hat{v}_i, \hat{v}_i; Q) - \phi_i(\hat{v}) + \bar{t}_i(v_i; \hat{v}_i, Q),$$

for all i, runs an ex ante budget surplus.

Proof. See the Appendix □

The result follows from the following two observations. First, if $\hat{v}$ is IR-critical for $(Q, \bar{t}(\cdot; \hat{v}, Q))$, then the transfers $\bar{t}_i(\cdot)$ are the (pointwise) largest (interim expected) transfers which implement Q in an individually rationally way. Consequently, there can be no other incentive compatible and individually rational transfer scheme that generates a larger ex ante budget surplus. Second, the set of all transfers which implement Q in an individually rational way, contains an ex post budget balanced transfer scheme if and only if it contains a transfer scheme that generates an ex ante budget surplus. In proving this second fact we build on arguments that are familiar in the existing literature; see Makowski and Mezzetti (1994).

Observe that the proposition applies to any allocation rule satisfying Assumption 1. In particular, the allocation rules are not required to be efficient as it is the case in most of the existing literature. Still though, efficient allocation rules are of particular interest. A characterization of the conditions under which efficient allocation rules are implementable in a way which respects ex post budget balance and individual rationality is obtained as a straightforward corollary of Proposition 3.

Corollary 1. *Let Q^* be an ex post efficient allocation rule and suppose that there exists a $\hat{v}$ which is IR-critical for $(Q^*, \bar{t}(\cdot; \hat{v}, Q^*))$. Moreover, let $\bar{T}$ be defined by $\bar{T}_i(v) = \tilde{u}_i(\hat{v}_i, \hat{v}_i; Q^*) - \phi_i(\hat{v}_i) + \bar{t}_i(v_i; \hat{v}_i, Q^*)$. There exists a transfer T such that (Q^*, T) is budget balanced, individually rational and incentive compatible if and only if $(Q^*, \bar{T})$ runs a budget surplus.*

Notice that for fixed type sets and net valuation functions, the condition formulated in the corollary can be interpreted as a condition on the distribution of the agents' types. This perspective is adopted in Kos and Manea (2009) who consider bilateral trade environments. They obtain an exact characterization of all type distributions with finite support such that the efficient allocation rule can be implemented with individual rationality and ex post budget balance.

To demonstrate the use of our tools we now apply them to a simple bilateral trade setting with binary type sets. In their seminal paper Myerson and Satterthwaite (1983) have shown that in environments where type sets are overlapping intervals the efficient allocation cannot be implemented by a mechanism that is individually rational and ex post budget balanced. They also provided an example that shows that their impossibility result does not hold in settings with finitely many types. For environments with binary type sets Matsuo (1989) characterizes the type distributions for which efficient voluntary and budget balanced trade is feasible. Kos and Manea (2009) generalize this characterization result to environments with arbitrary finite type spaces. In the following example we show how the result of Matsuo (1989) follows from Corollary 1.

Example 3: Linear environments

In this subsection we discuss how Proposition 3 can be put to work in linear environments; more detailed treatment of the linear case can be found in Kos and Messner (2010b). The payoff of each agent is given by

$$xv - t,$$

where x is the probability of getting the object and t the transfer to be paid. The assumption that for each agent a critical type exists then amounts to requiring that for each i there is a $\hat{v}_i \in V_i$ such that

$$\phi_i(v_i) - \phi_i(\hat{v}_i) \leq v_i q_i(v_i) - \hat{v}_i q_i(\hat{v}_i) - \bar{t}_i(v_i; \hat{v}_i, q_i).$$

Using $\bar{t}_i(v_i; \hat{v}_i, q_i) = \int_{\hat{v}_i}^{v_i} s d\bar{q}_i(s, \hat{v}_i)$ and integrating by parts this inequality can be further simplified to

$$\phi_i(v_i) - \phi_i(\hat{v}_i) \leq \int_{\hat{v}_i}^{v_i} \bar{q}_i(s; \hat{v}_i) ds, \qquad \forall v_i \in V_i. \tag{9}$$

A simple sufficient condition for existence of a critical type is linearity of ϕ_i; for details see Kos and Messner (2010b). Proposition 3 can now be expressed in the following form.

Corollary 2. *Let Q be an implementable allocation rule such that for all i there exists a $\hat{v}_i \in V_i$ for which condition (9) holds. Then Q can be implemented in an incentive compatible, individually rational and ex post budget balanced mechanism if and only if*

$$\sum_i \left[\hat{v}_i \bar{q}_i(\hat{v}_i; \hat{v}_i) - \phi_i(\hat{v}_i) - E_{v_i} \left[\int_{\hat{v}_i}^{v_i} s d\bar{q}_i(s, \hat{v}_i) \right] \right] \geq 0. \tag{10}$$

We now apply the above provided tools to a simple bilateral trade setting. The seller's valuation for the good is either S_L (with probability r) or S_H; the buyer instead values the good either B_L (with probability p) or B_H. The value of the buyer's outside option is equal to zero, while the seller's outside option is equal to his valuation of the good.

We assume $B_H > S_H > B_L > S_L$ so that the efficient allocation rule, Q^*, is to always trade the good, except for the case where a high valuation seller meets a low valuation buyer.

It is straightforward to verify that condition (9) is satisfied for the types B_L and S_H and the condition (10) can be computed to say that the efficient allocation rule is implementable if and only if

$$(1-r)(1-p)B_H + rB_L - S_H(1-p) - prS_L \geq 0.$$ [5]

Segal and Whinston (2011) also derive a condition under which an allocation rule q is implementable with individual rationality and ex post budget balance that is formulated in terms of the ex ante budget surplus of incentive compatible transfers (see their Lemma 1). In particular, they affirm that the existence of some transfer that implements q in an individually rational way and delivers an ex ante budget surplus is a sufficient condition for implementability with an individual rational and ex post budget balanced mechanism. We go beyond this result in that we also describe which transfers need to be checked and characterize this transfer, which enables us to express the necessary and sufficient condition solely in terms of the allocation rule.

Welfare maximization

Our findings on implementation under individual rationality have important implications for optimal mechanism design problems where the planner maximizes a weighted sum of the participating players' payoffs subject to the constraints of incentive compatibility, individual rationality and budget balance. In the preceding section we have argued that ex post budget balance imposes the same restrictions on implementability as ex ante budget constraints, provided that Assumption 1 holds. In what follows we therefore only consider the case of an ex ante budget constraint.

Assume that the mechanism designer's objective function is given by

$$U(Q,T) = E_v \left\{ \sum_{i=1}^{I} \lambda_i [u_i(Q(v), v) - T_i(v)] \right\}, \tag{11}$$

where $\lambda = (\lambda_1, \ldots, \lambda_I) \geq 0$. For convenience also assume that agents are indexed according to their welfare weights, i.e. $\lambda_i \geq \lambda_j$ if and only if $i \leq j$. Suppose that the planner maximizes (11) subject to incentive compatibility, individual rationality and the ex ante budget constraint $\sum_i E_{v_i}[t_i(v_i)] \geq 0$. The following proposition shows that whenever there is an optimal allocation rule that satisfies Assumption 1, then one of the optimal transfer schemes can be described in terms of extremal transfers, meaning that in such a case it is without loss of generality to consider only extremal transfers to start with.

Proposition 4. *Suppose that the planner maximizes (11) subject to IC, IR and the ex ante budget constraint* $\sum_i E_{v_i}[t_i(v_i)] \geq 0$. *Moreover, assume that* (Q^*, T^*) *is a solution of the designer's*

[5] For further intuition see Kos and Manea (2009).

problem. If there exists a profile of types, $(v_1^*, \dots, v_I^*)$, such that v_i^* is IR-critical for agent i when Q^* is implemented through $\bar{t}(\cdot; v^*, Q^*)$, then

$$U(Q^*, T^*) = U(Q^*, \bar{t}(\cdot; v^*, Q^*) + c^*),$$

where $c_i^* = \tilde{u}_i(v_i^*, v_i^*) - \phi_i(v_i^*)$ for all $i > 1$ and $c_1^* = -\sum_{i=1}^{I} E_{v_i} \bar{t}_i(v_i; v_i^*, Q^*) - \sum_{i=2}^{I} c_i^*$.

Proof. See the Appendix. □

The intuition behind this result is rather straightforward. Given an optimal allocation rule Q^*, the planner's priority in designing the transfers is to keep the payments of the agent with the largest welfare weight as low as possible. Since he has to respect the resource constraint, doing so requires him to extract as much rent from the other agents as possible and to redistribute it to agent 1.

It is important to point out that the class of welfare maximization problems considered in the preceding paragraphs includes as special case the standard revenue maximization problem, where the designer maximizes the ex ante expected value of the sum of the transfers extracted from the agents (his revenue) subject to incentive compatibility and individual rationality. In order to see this consider the following specification of the welfare maximization problem: The first agent's type set, V_1, is a singleton; his valuation function is constant and equal to zero, i.e. $u_1(x, v) = 0$ for all $(x, v) \in X \times V$; his outside option, ϕ_1 is low enough so that any mechanism (Q, T) which satisfies IC, budget balance and the individual rationality constraints of all agents $i > 1$ with equality, also satisfies the individual rationality constraint of the first agent; the welfare weights are given by $\lambda_1 = 1$ and $\lambda_i = 0$ for all $i > 1$.

Under this specification agent one can be interpreted as the designer who has no relevant private information. The fact that the welfare weights of the other agents are zero guarantees that only the first agent's payoff is maximized. Agent 1's payoff is given by his transfer only since his valuation is zero. We have already observed that at the optimum the budget balance constraint must be satisfied with equality, which means that the first agent's transfer must be equal to (minus) the sum of the other agents' transfers. Given that the first agent's valuation is constant implies that his incentive compatibility constraint is always trivially satisfied. Thus it can be eliminated from the problem. The same is true for the first agent's individual rationality constraint since his low outside option implies that it can never be binding at the optimum.

6 Beyond quasi-linearity

In Section 3 we used the notion of extremal transfers in order to characterize the set of implementable allocation rules in quasi-linear environments. In this section we argue that the fundamental ideas behind our characterization result carry over to settings with more general payoff functions.

In what follows we will no longer distinguish between monetary and non-monetary components of allocations. In order to avoid any confusion it is convenient to adapt the notation that we have introduced in Section 2. In particular, we now denote the set of all social decisions by Y. We write $u(y, v)$ for the payoff of type $v \in V$ when the social decision $y \in Y$ is implemented. A direct mechanism, μ, is a function mapping V into Y. $\mu : V \to Y$ is (Bayesian) incentive compatible if

$$u(\mu(v), v) \geq u(\mu(v'), v) \qquad \text{for all } v, v' \in V.$$

In deriving the extremal transfers we started out from the observation that incentive compatibility imposes bounds on the variation of the transfers once the non-monetary component is fixed. While in the more general environment that we are considering in this section it is no longer meaningful to distinguish between monetary and non-monetary components of the allocation, it is still true that incentive compatibility imposes constraints on the variation of the agent's indirect utility. We will show now how this insight can be developed in order to obtain a characterization result for the set of implementable allocation rules which parallels our results from Section 3.

Let μ be some incentive compatible mechanism and consider the two types v and $\hat{v}$ in V. By incentive compatibility

$$u(\mu(v), v) \geq u(\mu(\hat{v}), v).$$

Subtracting from both sides of this inequality the indirect utility of type $\hat{v}$ delivers

$$u(\mu(v), v) - u(\mu(\hat{v}), \hat{v}) \geq u(\mu(\hat{v}), v) - u(\mu(\hat{v}), \hat{v}).$$

Thus, the expression $u(\mu(\hat{v}), v) - u(\mu(\hat{v}), \hat{v})$ constitutes a lower bound on the difference between the two types' indirect utility. If v' is a third type, then by the above observation it follows that

$$\begin{aligned} u(\mu(v), v) - u(\mu(\hat{v}), \hat{v}) &= [u(\mu(v), v) - u(\mu(v'), v')] + [u(\mu(v'), v') - u(\mu(\hat{v}), \hat{v})] \\ &\geq [u(\mu(v'), v) - u(\mu(v'), v')] + [u(\mu(\hat{v}), v')] - u(\mu(\hat{v}), \hat{v})]. \end{aligned}$$

By iterating on this argument we obtain

$$u(\mu(v), v) - u(\mu(\hat{v}), \hat{v}) \geq \underline{U}(v; \hat{v}, \mu),$$

where

$$\underline{U}(v; \hat{v}, \mu) = \sup_{\{v^j\} \in S(v;\hat{v})} \sum_{j=1}^{n} [u(\mu(v^{j-1}), v^j) - u(\mu(v^{j-1}), v^{j-1})].$$

By an analogous argument it can be shown that incentive compatibility of μ implies that $u(\mu(v), v) - u(\mu(\hat{v}), \hat{v})$ is bounded above by

$$\bar{U}(v; \hat{v}, \mu) = \inf_{\{v^j\} \in S(v;\hat{v})} \sum_{j=1}^{n} [u(\mu(v^j), v^j) - u(\mu(v^j), v^{j-1})].$$

It follows that a necessary condition for the incentive compatibility of a mechanism μ is that for some type $\hat{v}$ the function $\underline{U}(\cdot;\hat{v},\mu)$ lies below the function $\bar{U}(\cdot;\hat{v},\mu)$, for otherwise the condition

$$\underline{U}(v;\hat{v},\mu) \leq u(\mu(v),v) - u(\mu(\hat{v}),\hat{v}) \leq \bar{U}(v;\hat{v},\mu)$$

could not be satisfied at every $v \in V$.

The converse of this statement is not true. That is, even for a mechanism μ which is not incentive compatible there might exist an anchor type $\hat{v}$ such that $\underline{U}(v;\hat{v},\mu) \leq \bar{U}(v;\hat{v},\mu)$ for all $v \in V$. However, the following proposition shows that μ is incentive compatible if there is *some* anchor type $\hat{v}$ such that either

$$\underline{U}(v;\hat{v},\mu) = u(\mu(v),v) - u(\mu(\hat{v}),\hat{v}) \qquad \text{for all } v \in V$$

or

$$\bar{U}(v;\hat{v},\mu) = u(\mu(v),v) - u(\mu(\hat{v}),\hat{v}) \qquad \text{for all } v \in V$$

is satisfied.

Proposition 5. *The following is true:*

i) If the mechanism $\mu : V \to Y$ is incentive compatible then for any fixed $\hat{v} \in V$,

$$\underline{U}(v;\hat{v},\mu) \leq \bar{U}(v;\hat{v},\mu)$$

for all $v \in V$.

ii) Let $\mu : V \to Y$ be such that for some $\hat{v} \in V$, $\underline{U}(v;\hat{v},\mu) \leq \bar{U}(v;\hat{v},\mu)$ for all $v \in V$. Moreover, assume that

$$\begin{aligned} u(\mu(v),v) - u(\mu(\hat{v}),\hat{v}) &= \underline{U}(v;\hat{v},\mu) \text{ for all } v \in V \text{ or} \\ u(\mu(v),v) - u(\mu(\hat{v}),\hat{v}) &= \bar{U}(v;\hat{v},\mu) \text{ for all } v \in V. \end{aligned}$$

Then, μ is incentive compatible.

Proof. See the Appendix. □

Notice that the bounds on the variation of the agent's indirect utility for a given mechanism, μ, are not fixed but are themselves functions of μ. The convenient feature of quasi-linear environments is that the bounds depend only on the non-monetary components of the mechanism. Therefore $\underline{U}$ and $\bar{U}$ can be translated into the *fixed* (i.e. transfer independent) bounds on the variation of the transfers $\underline{t}$ and $\bar{t}$. It is exactly due to this independence of $\underline{t}$ and $\bar{t}$ from the transfers that in Section 3 we have been able to obtain a stronger characterization result for the set of implementable allocation rules than the one which we have formulated in Proposition 5.

7 Conclusion

We characterize the boundaries of the set of transfers implementing a given allocation rule. This yields as an immediate consequence a characterization of the set of implementable allocation rules and the set of allocation rules that satisfy Revenue Equivalence, but also enables us to deal with mechanism design environments that do not satisfy Revenue Equivalence. In particular, we show that in many mechanism design problems it is without loss of generality for the designer to restrict his attention to extremal transfers. Consequently, such problems can be reduced to the choice of an allocation rule only, irrespective of whether Revenue Equivalence holds or not.

We go on to characterize the set of allocation rules that can be implemented in an incentive compatible, individually rational and ex post budget balanced mechanism.

An interesting open question is under what conditions do extremal transfers admit a simpler representation, say in integral form. The first step toward that was made by Carbajal and Ely (2010).

8 Appendix

Proof of Lemma 3. $I(v;\hat{v},q) \neq \emptyset$ and $-\infty < \underline{t}(v;\hat{v},q)$ imply $-\infty < \bar{t}(v;\hat{v},q)$, for every $v \in V$. We will show that $\bar{t}$ implements q. The case of $\underline{t}$ is handled analogously. We have to verify that

$$u(q(v),v) - \bar{t}(v;\hat{v},q) \geq u(q(v'),v) - \bar{t}(v';\hat{v},q),$$

for every $v, v' \in V$. By the definition of $\bar{t}(v';\hat{v},q)$, for every $\epsilon > 0$ there exists a sequence $\{v^j\}_{j=0}^n \in S(v';\hat{v})$ such that

$$\sum_{j=1}^{n} \left[u(q(v^j),v^j) - u(q(v^{j-1}),v^j)\right] - \epsilon \leq \bar{t}(v';\hat{v},q). \tag{12}$$

Now,

$$\begin{aligned} \bar{t}(v;\hat{v},q) - \bar{t}(v';\hat{v},q) &\leq \bar{t}(v;\hat{v},q) - \sum_{j=1}^{n} \left[u(q(v^j),v^j) - u(q(v^{j-1}),v^j)\right] + \epsilon \\ &\leq u(q(v),v) - u(q(v'),v) + \epsilon, \end{aligned} \tag{13}$$

where the first inequality follows from (12) and the second from the fact that $\bar{t}(v;\hat{v},q) \leq u(q(v),v) - u(q(v'),v) + \sum_{j=1}^{n} \left[u(q(v^j),v^j) - u(q(v^{j-1}),v^j)\right]$. That is, being the infimum over all sums, $\bar{t}(v;\hat{v},q)$ can be no larger than the sum corresponding to the sequence which is obtained by adding v as the final element to the sequence $\{v^j\}_{j=0}^n$. Since the inequality (13) holds for any ϵ we are done. □

Proof of Theorem 1. Statement i). The fact that implementability of q implies that the correspondence $I(\cdot;\hat{v},q)$ is nonempty is a direct consequence of Lemma 2. The other direction instead follows from Lemma 3.

Statement ii). ($\Rightarrow$) Suppose q satisfies Revenue Equivalence. Since $\bar{t}(\cdot;\hat{v},q)$ and $\underline{t}(\cdot;\hat{v},q)$ are two transfer schemes which implement q, Revenue equivalence implies that they can differ only by a constant. But given that the two transfers coincide at the anchor type $\hat{v}$ this constant must be zero, meaning that they coincide everywhere.

($\Leftarrow$) Let q be such that $|I(v;\hat{v},q)| = 1$ for every $v \in V$. By the first statement of the theorem q is implementable. So let t, t' be two transfer rules which implement q. Since $|I(v;\hat{v},q)| = 1$ is equivalent to $\bar{t}(v;\hat{v},q) = \underline{t}(v;\hat{v},q)$ for every $v \in V$, we have

$$\begin{aligned} \underline{t}(v;\hat{v},q) &= t(v) - t(\hat{v}) = \bar{t}(v;\hat{v},q), \quad \text{and} \\ \underline{t}(v;\hat{v},q) &= t'(v) - t'(\hat{v}) = \bar{t}(v;\hat{v},q), \end{aligned}$$

for every $v \in V$. But then

$$t(v) - t(\hat{v}) = t'(v) - t'(\hat{v})$$

for every $v \in V$, which is a restatement of revenue equivalence.

Statement iii) is simply a restatement of Lemma 2. □

Proof of Proposition 1. We prove that for any given $\hat{v} \in V$, $\bar{t}(v; \hat{v}, q) = \int_{\hat{v}}^{v} x d\bar{q}(x; \hat{v})$ holds for all $v \in V$. The following arguments can easily be adapted to show that on v, $\underline{t}(\cdot; \hat{v}, q)$ coincides with $\int_{\hat{v}}^{v} x d\underline{q}(x; \hat{v})$. The reader should keep in mind that $\int_{\hat{v}}^{v} = -\int_{v}^{\hat{v}}$ and thus negative when $\hat{v} > v$.

We start by showing that for all $v, \hat{v} \in V$

$$\int_{\hat{v}}^{v} x d\bar{q}(x; \hat{v}) \geq \bar{t}(v; \hat{v}, q).$$

For each pair $v, \hat{v} \in V$, define $S^*(v; \hat{v})$ as the set of all finite monotonic sequences in V which start at $\hat{v}$ and end at v. Observe that

$$\begin{aligned} \bar{t}(v; \hat{v}, q) &= \inf_{\{v^j\} \in S(v; \hat{v})} \sum_{j=0}^{n} v^j [q(v^j) - q(v^{j-1})] \\ &\leq \inf_{\{v^j\} \in S^*(v; \hat{v})} \sum_{j=0}^{n} v^j [q(v^j) - q(v^{j-1})] \\ &= \inf_{\{v^j\} \in S^*(v; \hat{v})} \sum_{j=0}^{n} v^j [\bar{q}(v^j; \hat{v}) - \bar{q}(v^{j-1}; \hat{v})]. \end{aligned} \tag{14}$$

The first equation is definitional and the inequality is implied by the fact that $S^*(v; \hat{v}) \subset S(v; \hat{v})$. Finally, the second equality follows from the fact that q and $\bar{q}(\cdot; \hat{v})$ coincide on V.

We next show that

$$\inf_{\{v^j\} \in S^*(v; \hat{v})} \sum_{j=0}^{n} v^j [\bar{q}(v^j; \hat{v}) - \bar{q}(v^{j-1}; \hat{v})] \leq \int_{\hat{v}}^{v} x d\bar{q}(x; \hat{v}).$$

By its definition the Riemann-Stieltjes integral on the right hand side of this inequality is the greatest lower bound of the set of numbers which can be generated as sums of the form

$$\sum_{j=1}^{n} v^j [\bar{q}_i(v^j; \hat{v}) - \bar{q}_i(v^{j-1}; \hat{v})]$$

when all sequences in the interval contained between the points $\hat{v}$ and v are considered which are monotonic and whose endpoints coincide with the endpoints of the interval. Thus the two sides of the inequality differ only in the set of sequences over which the respective infimum operations are performed.

Suppose that contrary to our hypothesis

$$\Delta \equiv \inf_{\{v^j\}\in S^*(v;\hat{v})} \sum_{j=0}^{n} v^j[\bar{q}(v^j;\hat{v}) - \bar{q}(v^{j-1};\hat{v})] - \int_{\hat{v}}^{v} xd\bar{q}(x;\hat{v}) > 0. \tag{15}$$

By the definition of the Riemann-Stieltjes integral there exists a monotonic sequence $\{v^j\}_{j=0}^{n}$ in the interval defined by the points $\hat{v}$ and v, with $v^0, v^n \in \{\hat{v}, v\}$ such that

$$0 \leq \sum_{j=1}^{n} v^j \left[\bar{q}(v^j;\hat{v}) - \bar{q}_i(v^{j-1};\hat{v})\right] - \int_{\hat{v}}^{v} xd\bar{q}(x;\hat{v}) < \Delta/2. \tag{16}$$

Next, consider the sequence $\{\tilde{v}^j\}_{j=0}^{n}$ defined as follows: $\tilde{v}^j = v^j$ for all j such that $v^j \in V$; otherwise set

$$\tilde{v}^j = \begin{cases} \max\{v \in v : v < v^j\} & \text{if } v^j \geq \hat{v} \\ \min\{v \in v : v > v^j\} & \text{if } v^j < \hat{v}. \end{cases}$$

Notice that $\{\tilde{v}^j\}_{j=0}^{n}$ is well defined (since v is closed) and entirely contained in V. This means that $\bar{q}(\tilde{v}^j;\hat{v}) = q(\tilde{v}^j)$ for all j. Furthermore, by the definition of $\bar{q}_i$, and the construction of $\{\tilde{v}^j\}_{j=0}^{n}$ we have $\bar{q}(v^j;\hat{v}) = \bar{q}(\tilde{v}^j;\hat{v})$ for all j. Finally, notice that $v^j \geq \tilde{v}^j$ for all j if $v < \hat{v}$ and $v^j \leq \tilde{v}^j$ for all j if $v > \hat{v}$. In addition, q being implementable implies that it is nondecreasing along the increasing sequence from $\hat{v}$ to v, when $\hat{v} < v$ and nonincreasing along the decreasing sequence from $\hat{v}$ to v, when $\hat{v} > v$. Combining these observations we get

$$\sum_{j=1}^{n} \tilde{v}_i^j \left[\bar{q}(\tilde{v}^j;\hat{v}) - \bar{q}(\tilde{v}^{j-1};\hat{v})\right] \leq \sum_{j=1}^{n} v^j \left[\bar{q}(v^j;\hat{v}) - \bar{q}(v^{j-1};\hat{v})\right]. \tag{17}$$

Using (15), (16) and (17) we obtain

$$\begin{aligned} \Delta &= \inf_{\{v^j\}\in S^*(v;\hat{v})} \sum_{j=0}^{n} v^j[\bar{q}(v^j;\hat{v}) - \bar{q}(v^{j-1};\hat{v})] - \int_{\hat{v}}^{v} xd\bar{q}(x;\hat{v}) && (18) \\ &\leq \sum_{j=1}^{n} \tilde{v}^j \left[\bar{q}(\tilde{v}^j;\hat{v}) - \bar{q}(\tilde{v}_i^{j-1};\hat{v})\right] - \int_{\hat{v}}^{v} xd\bar{q}(x;\hat{v}) && (19) \\ &\leq \sum_{j=1}^{n} v^j \left[q(v^j) - q(v^{j-1})\right] - \int_{\hat{v}}^{v} xd\bar{q}(x;\hat{v}) < \frac{\Delta}{2}, && (20) \end{aligned}$$

which is a contradiction. Thus we can conclude that

$$\int_{\hat{v}}^{v} xd\bar{q}(x;\hat{v}) \geq \inf_{\{v^j\}\in S^*(v;\hat{v})} \sum_{j=0}^{n} v^j[\bar{q}(v^j;\hat{v}) - \bar{q}(v^{j-1};\hat{v})]$$

must hold.

Next we show that for any given pair $v, \hat{v} \in V$,

$$\int_{\hat{v}}^{v} x d\bar{q}(x;\hat{v}) \leq \bar{t}(v;\hat{v},q).$$

Recall that $\bar{t}(\cdot;\hat{v},q)$ is the largest transfer schedule which i) implements q and ii) is equal to zero at $\hat{v}$. Thus we can verify the above inequality by showing that $t(v) = \int_{\hat{v}}^{v} x d\bar{q}(x;\hat{v})$ constitutes a transfer which implements q and vanishes at $\hat{v}$.

The condition $\int_{\hat{v}}^{\hat{v}} x d\bar{q}(x;\hat{v}) = 0$ is satisfied by construction. So it remains to be shown that (q,t) is incentive compatible.

Suppose that contrary to our hypothesis t fails to implement q. Then there must exist types $v', v'' \in V$ such that

$$v'q(v') - \int_{\hat{v}}^{v'} x d\bar{q}(x;\hat{v}) < v'q(v'') - \int_{\hat{v}}^{v''} x d\bar{q}(x;\hat{v}),$$

or equivalently

$$v'[q(v') - q(v'')] < \int_{\hat{v}}^{v'} x d\bar{q}(x;\hat{v}) - \int_{\hat{v}}^{v''} x d\bar{q}(x;\hat{v}) = \int_{v''}^{v'} x d\bar{q}(x;\hat{v}).$$

The integral on the right hand side of this inequality is bounded above by $v'[q(v') - q(v'')]$ which is incompatible with the strict inequality sign. This allows us to conclude that (q,t) is incentive compatible and therefore $\int_{\hat{v}}^{v} x d\bar{q}(x;\hat{v}) \leq \bar{t}(v;\hat{v},q)$ for all $v \in V$.

□

Proof of Proposition 3. As already mentioned in the main text the result follows from the following two facts:

i) By Proposition 2 we know that for any transfer scheme, T, which implements Q in an individually rational way we have $t_i(v_i) = E_{v_{-i}}[T_i(v_i, v_{-i})] \leq \bar{T}_i(v_i)$ for all i.[6] From this it immediately follows that in the set of all individually rational and incentive compatible transfer schemes for the implementation of Q, $\bar{T}$ is the one which delivers the largest ex ante surplus.

ii) The set of all transfers which implement Q in an individually rational way, contains an ex post budget balanced transfer scheme if and only if it contains a transfer scheme which generates an ex ante budget surplus.

[6] Observe that for each i the transfer scheme $\bar{T}_i$ only depends on v_i but not on v_{-i}. That is, under $\bar{T}$ each individual's ex post transfers coincide with its interim expected transfers.

In order to see that the second observation is true notice that ex post budget balance implies a zero ex ante budget surplus. Conversely, suppose that there exists a transfer scheme T such that (Q, T) satisfies IR and IC and generates a positive ex ante budget surplus, $\Pi \geq 0$. Denote the interim expected transfers associated with T by t and consider the transfer scheme $\tilde{T}$ defined as follows: For each $i < I$ set

$$\tilde{T}_i(v) = t_i(v_i) - \frac{\alpha_i}{1-\alpha_I}\left[\sum_{j \neq i} t_j(v_j) + E_{v_i}[t_i(v_i)]\right] + \frac{\alpha_i \alpha_I}{1-\alpha_I}\Pi,$$

where $0 \leq \alpha_i < 1$, such that $\sum_{i=1}^{I} \alpha_i = 1$. The transfer of the Ith agent instead is given by

$$\tilde{T}_I(v) = -\sum_{i \neq I} \tilde{T}_i(v).$$

Notice that $\tilde{T}$ has been constructed so that it satisfies ex post budget balance. We will now argue that $(Q, \tilde{T})$ also satisfies incentive compatibility and individual rationality. Both these properties depend only on the interim expected transfers. We will show now that the interim expected transfers are given by

$$\tilde{t}_i(v_i) = E_{v_{-i}}[\tilde{T}_i(v)] = t_i(v_i) - \alpha_i \Pi_i.$$

In the case $i < I$ this immediately follows from the fact that

$$E_{v_{-i}}\left[\sum_{j \neq i} t_j(v_j) + E_{v_i}[t_i(v_i)]\right] = \sum_{j \neq i} E_{v_j}[t_j(v_j)] + E_{v_i}[t_i(v_i)] = \Pi.$$

As for agent I, the interim expected transfer is given by

$$\tilde{t}_I(v_I) \quad = \quad -E_{v_{-I}}\left\{\sum_{i \neq I}\left(t_i(v_i) - \frac{\alpha_i}{1-\alpha}\left[\sum_{j \neq i} t_j(v_j) + E_{v_i}[t_i(v_i)]\right] + \frac{\alpha_i \alpha_I}{1-\alpha_I}\Pi\right)\right\}$$

Using $E_{v_{-I}}[t_i(v_i)] = E_{v_i}[t_i(v_i)]$ and $\sum_{i \neq I} E_{v_i}[t_i(v_i)] = \Pi - E_{v_I}[t_I(v_I)]$ this expression can be rewritten as follows

$$\tilde{t}_I(v_I) = -\left(\Pi - E_{v_I}[t(v_I)]\right) + \sum_{i \neq I} \frac{\alpha_i}{1-\alpha_I}\left\{t_I(v_I) + \left(\Pi - E_{v_I}[t_I(v_I)]\right)\right\} + \Pi \sum_{i \neq I} \frac{\alpha_i \alpha_I}{1-\alpha_I}.$$

Since $\sum_{i \neq I} \alpha_i = 1 - \alpha_I$ the first term of the sum cancels out with expression

$$\sum_{i \neq I} \frac{\alpha_i}{1-\alpha_I}\left(\Pi - E_{v_I}[t_I(v_I)]\right)$$

in the sum's second term. Moreover, the last term of the sum is equal to $\alpha_I \Pi$. We thus obtain the desired result

$$\tilde{t}_I(v_I) = t_I(v_I) - \alpha_I \Pi.$$

Observe that the interim expected transfer for player $i = 1, \ldots, I$, $\tilde{t}_i$, is a translation of t_i. Since we have assumed that T implements Q it therefore follows that so does $\tilde{T}$.

Finally, as for individual rationality observe that $\tilde{t}_i \leq t_i$. Thus, at the interim stage each type of each player must be at least as well of under the mechanism $(Q, \tilde{T})$ as under the mechanism (Q, T). Since the latter is individually rational so must be the first one.

□

Proof of Proposition 4. Let Q^* be an optimal allocation rule. Since the objective function is decreasing in transfers it follows that any transfer rule which implements Q^* in an optimal way must satisfy the resource constraint with equality. In order to see this, observe that whenever there is a strictly positive (ex ante) budget surplus then all transfers can be decreased by some (positive) constant without violating the ex ante budget constraint. Of course, a downward shift of all transfers by some constant also preserves incentive compatibility. At the same time, since all agents benefit from lower transfers, the shift cannot lead to a violation of individual rationality but implies an increase of the designer's objective function.

Exploiting the fact that at an optimum the budget constraint must hold with equality we can eliminate the transfer of agent 1 from the objective function in order to obtain

$$U(Q^*, T) = \sum_{i=1}^{I} \lambda_i E_v[u_i(Q^*(v), v)] + \sum_{i=2}^{I} (\lambda_1 - \lambda_i) E_{v_i}[t_i(v_i)].$$

Since $\lambda_1 \geq \lambda_i$ for all $i > 1$ this expression is non-decreasing in the expected transfer of each agent $i > 1$. By Proposition 2 we know that $T_i(v) = \bar{t}_i(v_i; v_i^*, Q^*) + c^*$ is the pointwise largest transfer scheme for agent i which is incentive compatible and satisfies individual rationality. This implies that there can be no other feasible transfer scheme which yields a larger expected payment from agent i and thus $T_i(v)$ must be an optimal transfer for the implementation of Q^*. □

Proof of Proposition 5. We have already seen in the discussion preceding this proposition that incentive compatibility of μ implies $\underline{U}(v; \hat{v}, \mu) \leq u(\mu(v), v) - u(\mu(\hat{v}), \hat{v}) \leq \bar{U}(v; \hat{v}, \mu)$. Hence, we only need to prove statement ii) of the proposition.

We will show that if for some $\hat{v} \in V$ we have $u(\mu(v), v) - u(\mu(\hat{v}), \hat{v}) = \bar{U}(v; \hat{v}, \mu)$ for all $v \in V$, then for every pair of types $v', v'' \in V$ we have $u(\mu(v'), v') \geq u(\mu(v''), v')$ (the case $\underline{U}(v; \hat{v}, \mu) = u(\mu(v), v)$ is treated in an analogous way and is therefore omitted). The condition $u(\mu(v'), v') \geq u(\mu(v''), v')$ is equivalent to

$$u(\mu(v'), v') - u(\mu(v''), v'') \geq u(\mu(v''), v') - u(\mu(v''), v'').$$

The left hand side of this inequality is by assumption equal to $\bar{U}(v'; \hat{v}, \mu) - \bar{U}(v''; \hat{v}, \mu)$. By the definition of $\bar{U}(v'; \hat{v}, \mu)$ we know that for every $\varepsilon > 0$ there exists a finite sequence $\{v^j\}_{j=1}^n \in S(v'; \hat{v})$ such that

$$\bar{U}(v'; \hat{v}, \mu) \geq \sum_{j=1}^{n} \left[u(\mu(v^j), v^j) - u(\mu(v^j), v^{j-1})\right] - \varepsilon.$$

By adding v'' as final element to the sequence $\{v_\varepsilon^j\}$ we obtain an element of $S(v''; \hat{v})$ and hence

$$\bar{U}(v''; \hat{v}, \mu) \leq \left[u(\mu(v''), v'') - u(\mu(v''), v')\right] + \sum_{j=1}^{n} \left[u(\mu(v^j), v^j) - u(\mu(v^j), v^{j-1})\right]$$

Combining the last two inequalities delivers

$$\bar{U}(v'; \hat{v}, \mu) - \bar{U}(v''; \hat{v}, \mu) \geq u(\mu(v''), v') - u(\mu(v''), v'') - \varepsilon.$$

Since this inequality holds for every $\varepsilon > 0$ it follows that $\bar{U}(v'; \hat{v}, \mu) - \bar{U}(v''; \hat{v}, \mu) \geq u(\mu(v''), v') - u(\mu(v''), v'')$, which is what we had to show. $\square$

References

Carbajal, J. and J. Ely (2010). Mechanism design without revenue equivalence. Working Paper.

Chung, K. and W. Olszewski (2007). A non-differentiable approach to revenue equivalence. *Theoretical Economics 2*(4), 469–487.

Cramton, P., R. Gibbons, and P. Klemperer (1987). Dissolving a partnership efficiently. *Econometrica 55*(3), 615–632.

D'Aspremont, C. and L.-A. Gérard-Varet (1975). Individual incentives and collective efficiency for an externality game with incomplete information. CORE DP 7519.

D'Aspremont, C. and L.-A. Gérard-Varet (1979). Incentives and incomplete information. *Journal of Public Economics 11*, 25–45.

Hellwig, M. F. (2010). Incentive problems with unidimensional hidden characteristics: A unified approach. *Econometrica 78*, 1201 – 1237.

Heydenreich, B., R. Müller, M. Uetz, and R. Vohra (2009). Characterization of revenue equivalence. *Econometrica 77*(1), 307–316.

Kos, N. and M. Manea (2009). Efficient trade mechanisms with discrete values. Working Paper.

Kos, N. and M. Messner (2010a). Incentive compatible transfers in linear environments. Working Paper.

Kos, N. and M. Messner (2010b). Incentive compatible transfers in linear environments. Working Paper.

Krishna, V. and M. Perry (1998). Efficient mechanism design. Pennsylvania State University.

Makowski, L. and C. Mezzetti (1994). Bayesian and weakly robust first best mechanisms: Characterizations. *Journal of Economic Theory 64*, 500–519.

Matsuo, T. (1989). On incentive compatible, individually rational, and ex-post efficient mechanisms for bilateral trading. *Journal of Economic Theory 49(1)*, 189 – 194.

Milgrom, P. and I. Segal (2002). Envelope theorems for arbitrary choice sets. *Econometrica 70*(2), 583–601.

Myerson, R. (1981). Optimal auction design. *Mathematics of Operations Research 6*(1), 58–73.

Myerson, R. and M. Satterthwaite (1983). Efficient mechanisms for bilateral trading. *Journal of Economic Theory 29*, 265–281.

Rahman, D. (2010). Detecting profitable deviations. Working Paper.

Rochet, J. (1987). A necessary and sufficient condition for rationalizability in a quasi-linear context. *Journal of Mathematical Economics 16*, 191–200.

Segal, I. R. and M. D. Whinston (2011). A simple status quo that ensures participation (with application to efficient bargaining). *Theoretical Economics 6*, 109 – 125.

Skreta, V. (2006). Mechanism design for arbitrary type spaces. *Economic Letters 91*, 293–299.

Vickrey, W. (1961). Counterspeculation, auctions and competitive sealed tenders. *Journal of Finance 16*, 8–37.

Walker, M. (1980). On the existence of dominant strategy mechanisms for making optimal public decisions. *Econometrica 48*, 1521–1540.

Williams, S. (1999). A characterization of efficient, Bayesian incentive compatible mechanisms. *Economic Theory 14*, 155–180.