Available online at www.sciencedirect.com
ScienceDirect
Games and Economic Behavior 61 (2007) 344–358
GAMES and Economic Behavior
www.elsevier.com/locate/geb

Weak monotonicity and Bayes–Nash incentive compatibility

Rudolf Müller *, Andrés Perea, Sascha Wolf

Department of Quantitative Economics, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands

Received 14 February 2006
Available online 23 April 2007

Abstract

An allocation rule is called Bayes–Nash incentive compatible, if there exists a payment rule, such that truthful reports of agents' types form a Bayes–Nash equilibrium in the direct revelation mechanism consisting of the allocation rule and the payment rule. This paper provides a characterization of Bayes–Nash incentive compatible allocation rules in social choice settings where agents have multi-dimensional types, quasi-linear utility functions and interdependent valuations. The characterization is derived by constructing complete directed graphs on agents' type spaces with cost of manipulation as lengths of edges. Weak monotonicity of the allocation rule corresponds to the condition that all 2-cycles in these graphs have non-negative length. For the case that type spaces are convex and the valuation for each outcome is a linear function in the agent's type, we show that weak monotonicity of the allocation rule together with an integrability condition is a necessary and sufficient condition for Bayes–Nash incentive compatibility.

JEL classification: C72

Keywords: Mechanism design; Incentive compatibility; Weak monotonicity

1. Introduction

This paper is concerned with the characterization of Bayes–Nash incentive compatible allocation rules in social choice settings where agents have independently distributed, multi-dimensional types and quasi-linear utility functions, that is, utility is the valuation of an allocation

* Corresponding author. Fax: +31 433884874.
E-mail addresses: r.muller@ke.unimaas.nl (R. Müller), a.perea@ke.unimaas.nl (A. Perea), s.wolf@ke.unimaas.nl (S. Wolf).

0899-8256/$ – see front matter
doi:10.1016/j.geb.2007.01.008

minus a payment. We allow for interdependent valuations across agents. The central task addressed in this paper is the following: given such type distributions and valuations, characterize precisely those allocation rules for which there exists a payment rule such that truthful reporting of agent's types forms a Bayes–Nash equilibrium in the direct revelation mechanism consisting of the allocation rule combined with the payment rule. In addition, we aim for a framework that lets us construct a payment rule, if any, which makes a particular allocation rule Bayes–Nash incentive compatible. For example, given an allocation rule which decides in a combinatorial auction for each set of bids for each agent which set of items he wins, we want to be able to decide whether there exists a pricing scheme for winning bids that makes truthful bidding a Bayes–Nash equilibrium. If the answer is yes, we would like to have means to construct such a pricing scheme.

1.1. Related work

An allocation rule is dominant strategy incentive compatible, if there exists a payment rule such that for any report of the other agents an agent maximizes his own utility by reporting truthfully his type. Roberts (1979) implicitly uses a monotonicity condition on the allocation rule in order to derive his characterization of dominant strategy incentive compatible mechanisms in terms of affine maximizers for unrestricted preference domains. For a selection of restricted preference domains, Bikhchandani et al. (2006) characterize dominant strategy incentive compatibility directly in terms of a monotonicity condition on the allocation rule. Gui et al. (2004) extend these results to larger classes of preference domains by making a link to network theory. The most general results are by Saks and Yu (2005), who show that previous results extend to any convex multi-dimensional type space.

The environment considered by Saks and Yu (2005) features quasi-linear utilities and multi-dimensional types. The allocation rule maps agents' type reports into a finite set of m possible outcomes. An agent's type is a vector in $\mathbb{R}^m$ reflecting his valuation of the different possible outcomes, that is, the agent's valuation of some outcome a is given by the ath element of his type vector. Agents' type spaces are assumed to be convex. Saks and Yu (2005) show that dominant strategy incentive compatible allocation rules in this setting can be characterized in terms of *weak monotonicity*, a term introduced by Bikhchandani et al. (2006). In order to derive this result they construct complete directed graphs in the following way: Take some agent and fix a profile of type reports for the others. Now, a directed graph is constructed by associating a node with each outcome and putting a directed edge between each ordered pair of nodes. Take two outcomes a and b. Consider the difference of the valuation of a and the valuation of b with respect to every type for which truthfully reporting this type yields outcome a. The length of the network edge from a to b is defined as the infimum of all these differences. In this fashion a graph is constructed for every agent and every possible report profile of the other agents. Weak monotonicity states that for any two different outcomes a and b, the sum of the two edge lengths from a to b and from b to a is non-negative.

Earlier, Rochet (1987) characterized dominant strategy implementation in cases where the set of outcomes is not necessarily finite; an assumption that is crucial to the work of Saks and Yu (2005). He considers a setting where agents have multi-dimensional, convex type spaces and valuation functions which are linear w.r.t. their own true types. Making some additional differentiability assumptions, Rochet (1987) shows that in this case dominant strategy incentive compatibility can be characterized in terms of a monotonicity condition on the allocation rule plus an integrability condition.

Monotonicity has also been used to characterize Bayes–Nash incentive compatible allocation rules. Jehiel et al. (1999) and Jehiel and Moldovanu (2001) develop characterizations for social choice settings where agents have multi-dimensional, convex type spaces and valuation functions which are linear w.r.t. their true types. Their characterizations of Bayes–Nash incentive compatibility include a monotonicity condition on the allocation rule as well as an integrability condition comparable to the one presented by Rochet (1987).

1.2. Our contribution

Similar to the network approach of Gui et al. (2004) and Saks and Yu (2005) we construct graphs. If an allocation rule is Bayes–Nash incentive compatible, then there exists a payment rule such that an agent's expected utility for truthfully reporting his type t is at least as high as his expected utility for misreporting some type s. Similarly, an agent's expected utility for truthfully reporting type s is at least as high as his expected utility for misreporting type t. From combining these two conditions we get a weak monotonicity condition on the allocation rule. This condition is the expected utility equivalent of the monotonicity condition mentioned in the context of dominant strategy incentive compatible allocation rules. Weak monotonicity is a necessary condition for Bayes–Nash incentive compatibility. It expresses that the expected gain in valuation for truthfully reporting t instead of misreporting s should be at least as big as the expected gain in valuation for misreporting t instead of truthfully reporting s.

Recognizing that the constraints inherent in the definition of Bayes–Nash incentive compatibility have a natural network interpretation we build complete directed graphs for agents' type spaces. To do so we associate a node with each type and put a directed edge between each ordered pair of nodes. The length of the edge going from the node associated with type s to the node associated with type t is defined as the cost of manipulation, that is, the expected difference in an agent's valuation for truthfully reporting t instead of misreporting s. Note that unlike the network approach of Gui et al. (2004) and Saks and Yu (2005) (see description above) we construct only *one* graph for each agent since we work in terms of expectations and do not consider each possible type profile of the other agents separately. Furthermore, each of these graphs contains an infinite number of nodes as we associate a node with each possible type of the agent. One could also construct outcome based graphs (as done by Gui et al., 2004; Saks and Yu, 2005) by associating a node with each possible probability distribution over outcomes. However, these graphs also contain an infinite number of nodes whenever the different possible type reports of an agent induce an infinite number of probability distributions over outcomes.

The outline of the paper is as follows: In Section 2 we state some basic assumptions and definitions. Throughout the paper we assume that agents have quasi-linear utility functions and independently distributed, privately known, multi-dimensional types. Furthermore, we allow for interdependent valuations. We do not put any restrictions on the number of possible outcomes.

In Section 3 we show that an allocation rule is Bayes–Nash incentive compatible if and only if the graphs described above contain no finite, negative length cycles. Rochet (1987) shows that dominant strategy incentive compatibility can be characterized in terms of the absence of finite, negative length cycles in similar graphs. Our result is the Bayes–Nash equivalent for his finding.

In Section 4 agents' type spaces are assumed to be convex and their valuation functions are assumed to be linear w.r.t. to their own true types. Even under these restrictions, weak monotonicity alone is not sufficient for Bayes–Nash incentive compatibility, which is illustrated by an example. However, we show that weak monotonicity together with an integrability condition is both necessary and sufficient for Bayes–Nash incentive compatibility. Using examples it is illustrated that weak monotonicity and the integrability condition do not imply each other. The setting of a single-item auction with externalities considered in Jehiel et al. (1999) and the social choice setting considered in Jehiel and Moldovanu (2001) are special cases of the framework presented in this section. Compared to their settings, our multi-dimensional framework allows for a broader class of possible interdependencies between agents' valuations.

The main contribution of this paper is thus to derive for the setting described above a complete characterization of Bayes–Nash incentive compatibility in terms of weak monotonicity and an additional integrability condition. Thereby we achieve a characterization that depends purely on the valuations and the allocation rule. The characterization resembles the one derived by Rochet (1987) for dominant strategy incentive compatibility. However, our result does not follow from Rochet (1987) immediately, as we cover interdependent valuations.

2. The model and basic definitions

There is a set of agents $N = \{1, \dots, n\}$. Each agent i has a type $t^i \in T^i$ with $T^i \subseteq \mathbb{R}^k$. T denotes the set of all type profiles $t = (t^1, \dots, t^n)$, and T^{-i} denotes the set of all type profiles $t^{-i} = (t^1, \dots, t^{i-1}, t^{i+1}, \dots, t^n)$. A *payment rule* is a function

$$P : T \mapsto \mathbb{R}^n,$$

so given a report profile r^{-i} of the others, reporting a type r^i results in a payment $P_i(r^i, r^{-i})$ for agent i. Denoting the set of outcomes by Γ, an *allocation rule* is a function

$$f : T \mapsto \Gamma.$$

We allow for interdependent valuations across agents, that is, agents' valuations do not only depend on their own types but on the types of all agents. As an example one can think of an auction for a painting (see Klemperer, 1999) where agents' types reflect how much they like the painting. An agent's valuation for owning the painting depends on the types of the others as they affect the possible resale value of the painting and the owner's prestige. Take agent i having true type t^i and reporting r^i while the others have true types t^{-i} and report r^{-i}. The value that agent i assigns to the resulting allocation is denoted by $v^i(f(r^i, r^{-i}) \mid t^i, t^{-i})$. Utilities are quasi-linear, that is, an agent's utility is his valuation of an allocation minus his payment.

Agents' types are independently distributed. Let π^i denote the density on T^i. The joint density π^{-i} on T^{-i} is then given by

$$\pi^{-i}(t^{-i}) = \prod_{\substack{j \in N \\ j \neq i}} \pi^j(t^j).$$

Assume that agent i believes all other agents to report truthfully. If agent i has true type t^i, then his expected utility for making a report r^i is given by

$$\begin{aligned} U^i(r^i \mid t^i) &= \int_{T^{-i}} \big(v^i\big(f\big(r^i, t^{-i}\big) \mid t^i, t^{-i}\big) - P_i\big(r^i, t^{-i}\big)\big)\pi^{-i}\big(t^{-i}\big)\, \mathrm{d}t^{-i} \\ &= E_{-i}\big[v^i\big(f\big(r^i, t^{-i}\big) \mid t^i, t^{-i}\big) - P_i\big(r^i, t^{-i}\big)\big]. \end{aligned} \tag{1}$$

We assume $E_{-i}[v^i(f(r^i, t^{-i}) \mid t^i, t^{-i})]$ to be finite $\forall r^i, t^i \in T^i$.

An allocation rule f is *Bayes–Nash incentive compatible* if there exists a payment rule P such that $\forall i \in N$ and $\forall r^i, \tilde{r}^i \in T^i$:

$$E_{-i}\big[v^i\big(f\big(r^i, t^{-i}\big) \,\big|\, r^i, t^{-i}\big) - P_i\big(r^i, t^{-i}\big)\big] \geqslant E_{-i}\big[v^i\big(f\big(\tilde{r}^i, t^{-i}\big) \,\big|\, r^i, t^{-i}\big) - P_i\big(\tilde{r}^i, t^{-i}\big)\big]. \tag{2}$$

Symmetrically, we have also

$$E_{-i}\big[v^i\big(f\big(\tilde{r}^i, t^{-i}\big) \,\big|\, \tilde{r}^i, t^{-i}\big) - P_i\big(\tilde{r}^i, t^{-i}\big)\big] \geqslant E_{-i}\big[v^i\big(f\big(r^i, t^{-i}\big) \,\big|\, \tilde{r}^i, t^{-i}\big) - P_i\big(r^i, t^{-i}\big)\big]. \tag{3}$$

By adding (2) and (3) we get the following monotonicity condition.[1]

Definition 1 *(Weak monotonicity).* An allocation rule f satisfies weak monotonicity if $\forall i \in N$ and $\forall r^i, \tilde{r}^i \in T^i$:

$$E_{-i}\big[v^i\big(f\big(r^i, t^{-i}\big) \,\big|\, r^i, t^{-i}\big) - v^i\big(f\big(\tilde{r}^i, t^{-i}\big) \,\big|\, r^i, t^{-i}\big)\big] \geqslant E_{-i}\big[v^i\big(f\big(r^i, t^{-i}\big) \,\big|\, \tilde{r}^i, t^{-i}\big) - v^i\big(f\big(\tilde{r}^i, t^{-i}\big) \,\big|\, \tilde{r}^i, t^{-i}\big)\big].$$

This condition is the expected utility equivalent to the weak monotonicity (W-MON) condition of Bikhchandani et al. (2006) and the 2-cycle inequality of Gui et al. (2004). The rationale for naming the above condition weak monotonicity becomes evident once we consider valuation functions that are linear with respect to agents' types in Section 4. Obviously, weak monotonicity is a necessary condition for Bayes–Nash incentive compatibility. In Section 4 we present a setting where weak monotonicity together with an integrability condition is also a sufficient condition.

3. A network interpretation

We begin this section by briefly reviewing a well-known result from the field of network flow theory.[2] Let $X = \{x_1, \ldots, x_k\}$ be a finite set of variables. Consider the following system of constraints:

$$x_i - x_j \leqslant w_{ij} \quad \forall i, j \in \{1, \ldots, k\}, \tag{4}$$

where w_{ij} is some constant specific to the ordered pair (i, j). The system can be associated with a network by constructing a directed, weighted graph whose nodes correspond to the variables. A directed edge is put between each ordered pair of nodes. The length of the edge from the node corresponding to x_i to the node corresponding to x_j is given by w_{ij}.

It is a well-known result (see, e.g., Shostak, 1981) that the system of linear inequalities in (4) is feasible, that is, there exists an assignment of real values to the variables such that the constraints in (4) are satisfied, if and only if there is no negative length cycle in the associated network. Furthermore, if the system is feasible then one feasible solution is to assign to each x_i the length of a shortest path from the node associated with x_i to some arbitrary source node.[3]

[1] Expected payments cancel since we work under the assumption of independently distributed types.

[2] A comprehensive introduction to network flows can be found in Ahuja et al. (1993).

[3] In order to be consistent with the existing literature we defined the system of constraints as in (4). However, in network theory the constraints are commonly defined as $x_j - x_i \leqslant w_{ij}$. In this case, if the system is feasible then one feasible solution is to assign to each x_i the length of a shortest path from some arbitrary source node to the node associated with x_i.

In order to see that the constraints in (2) have a natural network interpretation it is useful to rewrite (2) as follows

$$E_{-i}\left[P_i\left(r^i,t^{-i}\right)-P_i\left(\tilde{r}^i,t^{-i}\right)\right] \\ \leqslant E_{-i}\left[v^i\left(f\left(r^i,t^{-i}\right)\bigm| r^i,t^{-i}\right)-v^i\left(f\left(\tilde{r}^i,t^{-i}\right)\bigm| r^i,t^{-i}\right)\right]. \tag{5}$$

Considering a specific allocation rule, the right-hand side of (5) is a constant. Thus, we have a system of difference constraints as described in (4) (except that we are now dealing with a potentially infinite number of variables).

Given this observation, we associate the system of inequalities (5) with a network in the same way as is described above. For each agent we build a complete directed graph T^i_f. A node is associated with each type and a directed edge is put between each ordered pair of nodes. For agent i the length of an edge directed from r^i to $\tilde{r}^i$ is denoted $l^i(r^i,\tilde{r}^i)$ and is defined as the *cost of manipulation*:

$$l^i\left(r^i,\tilde{r}^i\right)=E_{-i}\left[v^i\left(f\left(r^i,t^{-i}\right)\bigm| r^i,t^{-i}\right)-v^i\left(f\left(\tilde{r}^i,t^{-i}\right)\bigm| r^i,t^{-i}\right)\right]. \tag{6}$$

Given our previous assumptions, the edge length is finite. For technical reasons we allow for loops. However, note that an edge directed from r^i to r^i has length $l^i(r^i,r^i)=0$.

Using this definition of the edge lengths, the weak monotonicity condition can be written as

$$l^i\left(r^i,\tilde{r}^i\right)+l^i\left(\tilde{r}^i,r^i\right)\geqslant 0 \quad \forall i\in N,\ \forall r^i,\tilde{r}^i\in T^i.$$

So weak monotonicity corresponds to the absence of negative length 2-cycles in the graphs described above.

Rochet (1987) observed that dominant strategy incentive compatibility can be characterized in terms of the absence of finite, negative length cycles in similar graphs. Using the same proof technique, we can derive such a characterization for Bayes–Nash incentive compatibility as well.

Theorem 1. *An allocation rule f is Bayes–Nash incentive compatible if and only if there is no finite, negative length cycle in T^i_f $\forall i\in N$.*

Proof. (Adapted from Rochet, 1987.) Take some agent i and let $C=(r^i_1,\ldots,r^i_m,r^i_{m+1}=r^i_1)$ denote a finite cycle in T^i_f. Let us assume that f is Bayes–Nash incentive compatible. This implies, using (5) and the edge length definition (6), that for every $j\in\{1,\ldots,m\}$,

$$E_{-i}\left[P_i\left(r^i_j,t^{-i}\right)-P_i\left(r^i_{j+1},t^{-i}\right)\right]\leqslant l^i\left(r^i_j,r^i_{j+1}\right).$$

Adding up these inequalities yields

$$0\leqslant\sum_{j=1}^m l^i\left(r^i_j,r^i_{j+1}\right),$$

so C has non-negative length.

Conversely, let us assume that there exists no finite, negative length cycle in T^i_f $\forall i\in N$. For each agent i we pick an arbitrary source node $r^i_0\in T^i$ and define $\forall r^i\in T^i$

$$p^i\left(r^i\right)=\inf\sum_{j=1}^m l^i\left(r^i_j,r^i_{j+1}\right),$$

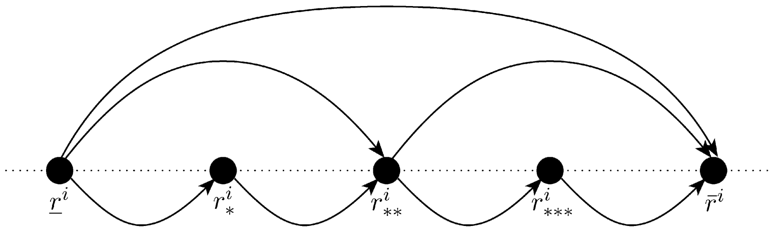

Fig. 1. Decomposition monotonicity.

where the infimum is taken over all finite paths $A = (r_1^i = r^i, \dots, r_{m+1}^i = r_0^i)$ in T_f^i, that is, all finite paths that start at r^i and end at r_0^i. Absence of finite, negative length cycles implies that $p^i(r_0^i) = 0$. Furthermore, $\forall r^i \in T^i$ we have

$$p^i(r_0^i) \leqslant p^i(r^i) + l^i(r_0^i, r^i)$$

which implies that $p^i(r^i)$ is finite. For every pair $r^i, \tilde{r}^i \in T^i$ we also have

$$p^i(r^i) \leqslant p^i(\tilde{r}^i) + l^i(r^i, \tilde{r}^i).$$

Thus, by setting[4] $P_i(r^i, t^{-i}) = p^i(r^i)\ \forall t^{-i} \in T^{-i}$, and using (6) we get

$$E_{-i}\big[P_i(r^i, t^{-i}) - P_i(\tilde{r}^i, t^{-i})\big] \leqslant E_{-i}\big[v^i\big(f(r^i, t^{-i}) \mid r^i, t^{-i}\big) - v^i\big(f(\tilde{r}^i, t^{-i}) \mid r^i, t^{-i}\big)\big].$$

Hence, the constraints in (5) are satisfied and f is Bayes–Nash incentive compatible. □

Let us conclude this section with a condition for the costs of manipulation that is used in the derivation of the characterization theorem presented in the following section.

Definition 2 *(Decomposition monotonicity).* The costs of manipulation are decomposition monotone if $\forall \underline{r}^i, \bar{r}^i \in T^i$ and $\forall r^i \in T^i$ s.t. $r^i = (1-\alpha)\underline{r}^i + \alpha \bar{r}^i$, $\alpha \in (0,1)$, we have

$$l^i(\underline{r}^i, \bar{r}^i) \geqslant l^i(\underline{r}^i, r^i) + l^i(r^i, \bar{r}^i).$$

So looking at a pair of nodes, if decomposition monotonicity holds then the direct edge between those nodes is at least as long as any path connecting the same two nodes via nodes lying on the line segment between them. Figure 1 gives an illustrative example. Decomposition monotonicity implies that the edge from $\underline{r}^i$ to $\bar{r}^i$ is at least as long as the path $A = (\underline{r}^i, r_{**}^i, \bar{r}^i)$ and that A is at least as long as the path $\tilde{A} = (\underline{r}^i, r_*^i, r_{**}^i, r_{***}^i, \bar{r}^i)$.

4. Weak monotonicity and path independence

In this section we restrict the rather general setting presented in Section 2. We assume that T^i is convex for each agent i. Furthermore, we now assume that an agent's valuation function is linear in his own true type. So if agent i has true type t^i and reports r^i while the others have true types t^{-i} and report r^{-i}, his valuation for the resulting allocation is

[4] Note that it is sufficient if P is set such that $E_{-i}[P_i(r^i, t^{-i})] = p^i(r^i) + c$. This allows for a variety of payment rules yielding the same expected payments up to an additive constant.

$$v^i\big(f\big(r^i,r^{-i}\big)\bigm| t^i,t^{-i}\big)=\alpha^i\big(f\big(r^i,r^{-i}\big)\bigm| t^{-i}\big)+\beta^i\big(f\big(r^i,r^{-i}\big)\bigm| t^{-i}\big)t^i. \tag{7}$$

Note that $\alpha^i:\Gamma\times T^{-i}\mapsto\mathbb{R}$ and $\beta^i:\Gamma\times T^{-i}\mapsto\mathbb{R}^k$, i.e. α^i assigns to every $(\gamma,t^{-i})\in\Gamma\times T^{-i}$ a value in $\mathbb{R}$, whereas β^i assigns to every $(\gamma,t^{-i})\in\Gamma\times T^{-i}$ a vector in $\mathbb{R}^k$. Similarly, assuming he believes all other agents to report truthfully, agent i's expected valuation for reporting r^i while having true type t^i is

$$\begin{aligned}&E_{-i}\big[v^i\big(f\big(r^i,t^{-i}\big)\bigm| t^i,t^{-i}\big)\big]\\&\quad=E_{-i}\big[\alpha^i\big(f\big(r^i,t^{-i}\big)\bigm| t^{-i}\big)\big]+E_{-i}\big[\beta^i\big(f\big(r^i,t^{-i}\big)\bigm| t^{-i}\big)\big]t^i.\end{aligned} \tag{8}$$

Using (8), the weak monotonicity condition becomes

$$\begin{aligned}&E_{-i}\big[\beta^i\big(f\big(r^i,t^{-i}\big)\bigm| t^{-i}\big)-\beta^i\big(f\big(\tilde r^i,t^{-i}\big)\bigm| t^{-i}\big)\big]\big(r^i-\tilde r^i\big)\geqslant 0\\&\quad\forall i\in N,\ \forall r^i,\tilde r^i\in T^i.\end{aligned} \tag{9}$$

In this restricted setting weak monotonicity implies that the costs of manipulation are decomposition monotone.

Lemma 1. *Suppose that every agent i has a valuation function which is linear in his true type: If f satisfies weak monotonicity then the costs of manipulation are decomposition monotone.*

Proof. Take some agent i and let $\underline{r}^i,\bar r^i\in T^i$. Let $r^i\in T^i$ such that $r^i=(1-\alpha)\underline{r}^i+\alpha\bar r^i$ for some $\alpha\in(0,1)$. Weak monotonicity implies that

$$E_{-i}\big[\beta^i\big(f\big(r^i,t^{-i}\big)\bigm| t^{-i}\big)-\beta^i\big(f\big(\bar r^i,t^{-i}\big)\bigm| t^{-i}\big)\big]\big(r^i-\bar r^i\big)\geqslant 0.$$

Note that $\underline{r}^i-r^i$ is proportional to $r^i-\bar r^i$, specifically $\underline{r}^i-r^i=\frac{\alpha}{1-\alpha}(r^i-\bar r^i)$. Since $\alpha\in(0,1)$, the above inequality implies that

$$E_{-i}\big[\beta^i\big(f\big(r^i,t^{-i}\big)\bigm| t^{-i}\big)-\beta^i\big(f\big(\bar r^i,t^{-i}\big)\bigm| t^{-i}\big)\big]\big(\underline{r}^i-r^i\big)\geqslant 0.$$

Adding $E_{-i}[\beta^i(f(\underline{r}^i,t^{-i})\mid t^{-i})-\beta^i(f(r^i,t^{-i})\mid t^{-i})]\underline{r}^i$ on both sides of the latter inequality and rearranging terms yields

$$\begin{aligned}&E_{-i}\big[\beta^i\big(f\big(r^i,t^{-i}\big)\bigm| t^{-i}\big)-\beta^i\big(f\big(\bar r^i,t^{-i}\big)\bigm| t^{-i}\big)\big]\underline{r}^i\\&\qquad+E_{-i}\big[\beta^i\big(f\big(\underline{r}^i,t^{-i}\big)\bigm| t^{-i}\big)-\beta^i\big(f\big(r^i,t^{-i}\big)\bigm| t^{-i}\big)\big]\underline{r}^i\\&\quad\geqslant E_{-i}\big[\beta^i\big(f\big(r^i,t^{-i}\big)\bigm| t^{-i}\big)-\beta^i\big(f\big(\bar r^i,t^{-i}\big)\bigm| t^{-i}\big)\big]r^i\\&\qquad+E_{-i}\big[\beta^i\big(f\big(\underline{r}^i,t^{-i}\big)\bigm| t^{-i}\big)-\beta^i\big(f\big(r^i,t^{-i}\big)\bigm| t^{-i}\big)\big]\underline{r}^i.\end{aligned}$$

Notice that the first and the last term on the left-hand side of the inequality cancel. Hence, using (6), the above can be written as

$$l^i\big(\underline{r}^i,\bar r^i\big)\geqslant l^i\big(\underline{r}^i,r^i\big)+l^i\big(r^i,\bar r^i\big),$$

so the costs of manipulation are decomposition monotone. □

It can be shown (Müller et al., 2005) that if agents' type spaces are one-dimensional then weak monotonicity is a sufficient condition for Bayes–Nash incentive compatibility. Unfortunately, if type spaces are multi-dimensional then weak monotonicity alone is not sufficient anymore, as is illustrated below in Example 1.

This example is constructed based on the following insight: Suppose that the allocation function f and the mapping β^i are such that we can write

$$E_{-i}\left[\beta^i\left(f\left(r^i, t^{-i}\right) \mid t^{-i}\right)\right] = r^i B_i,$$

where B_i is some agent specific $k \times k$ matrix. Weak monotonicity requires

$$\left(r^i - \tilde{r}^i\right) B_i \left(r^i - \tilde{r}^i\right)' \geqslant 0 \quad \forall r^i, \tilde{r}^i \in T^i,$$

where $'$ denotes "transposed." Note that

$$B_i = \frac{1}{2}(B_i + B_i') + \frac{1}{2}(B_i - B_i'),$$

that is, B_i can be decomposed into a symmetric part $\frac{1}{2}(B_i + B_i')$ and an anti-symmetric part $\frac{1}{2}(B_i - B_i')$. Weak monotonicity is already satisfied if the symmetric part of B_i is positive semi-definite. However, there are no finite, negative length cycles in T_f^i (and thus f is Bayes–Nash incentive compatible) if and only if B_i is symmetric and positive semi-definite (both results follow from Rockafellar, 1970, p. 240).

Example 1. For simplicity we assume that there exists only a single agent. Furthermore, we take the mapping β^i in (7) to be linear and the mapping α^i to be constant and equal to zero. The agent's type is one of three extreme types, denoted x, y and z, or any convex combination of these. His type space can be parameterized by a simplex with vertices $x = (1, 0, 0)$, $y = (0, 1, 0)$ and $z = (0, 0, 1)$. Thus, the agent's type space $T = \text{conv}\{x, y, z\}$ consists of the convex hull of the three unit vectors in $\mathbb{R}^3$. There are three elementary outcomes, denoted a, b and c. If the agent is of type x, his valuations for these outcomes are given by the first column of the following matrix:

$$V = \begin{pmatrix} 2 & 0 & 3 \\ 3 & 2 & 0 \\ 0 & 3 & 2 \end{pmatrix}.$$

The first element is his valuation for a, the second one for b and the third one for c. Similarly, if the agent is of type y or z, his valuations for the elementary outcomes are given by the second and the third column of V. The allocation rule f is a linear mapping associating each type report with a probability distribution over the three elementary outcomes. The outcome space Γ is the set of all possible probability distributions on $\{a, b, c\}$. Generic element $\gamma = (\gamma_a, \gamma_b, \gamma_c)$ indicates that a is achieved with probability γ_a, b with probability γ_b and c with probability γ_c. The allocation rule works as follows: If the agent reports x as his type then f awards him with the second-best outcome according to this type, that is $f(x) = (1, 0, 0)$. Similarly, $f(y) = (0, 1, 0)$ and $f(z) = (0, 0, 1)$. In general we have $f(r) = rI$, where I denotes the 3×3 identity matrix.

Using the above, the agent's valuation function becomes $v(f(r) \mid t) = rVt'$. As easily can be checked (by verifying that the symmetric part $\frac{1}{2}(V + V')$ of V is positive definite), weak monotonicity is satisfied, that is, $(r - \tilde{r})V(r - \tilde{r})' \geqslant 0$ $\forall r, \tilde{r} \in T$. Nevertheless, the 3-cycle $C = (x, y, z, x)$ has length $l(x, y) + l(y, z) + l(z, x) = -3$ (see also Fig. 2). The existence of such a negative length cycle implies that f is not Bayes–Nash incentive compatible (see Theorem 1).

From the above example it is evident that weak monotonicity alone is not enough to ensure Bayes–Nash incentive compatibility. However, in the following we are going to show that weak monotonicity together with an integrability condition is sufficient.

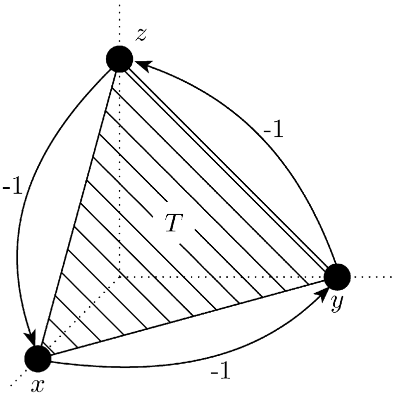

Fig. 2. The negative cycle C in Example 1.

Definition 3 *(Path independence).* Let $\psi : T^i \mapsto \mathbb{R}^k$ be a vector field. ψ is called path independent if for any two $\underline{r}^i, \bar{r}^i \in T^i$ the path integral of ψ from $\underline{r}^i$ to $\bar{r}^i$

$$\int_{\underline{r}^i, S}^{\bar{r}^i} \psi$$

is independent of the path of integration S.

Note that $E_{-i}[\beta^i(f(r^i, t^{-i}) \mid t^{-i})]$ is a vector field $T^i \mapsto \mathbb{R}^k$.

Theorem 2. *Suppose that every agent i has a convex type space and a valuation function which is linear in his true type. Then the following statements are equivalent*:

(1) *f is Bayes–Nash incentive compatible.*
(2) *f satisfies weak monotonicity and for every agent i, $E_{-i}[\beta^i(f(r^i, t^{-i}) \mid t^{-i})]$ is path independent.*

Proof. (1) $\Rightarrow$ (2). Let us assume that f is Bayes–Nash incentive compatible. As mentioned in Section 2, the necessity of weak monotonicity follows trivially. Furthermore, from Theorem 1 it follows that for every agent i the graph T^i_f has no finite, negative length cycles. Let $C = (r^i_1, \dots, r^i_m, r^i_{m+1} = r^i_1)$ denote a finite cycle in T^i_f. Absence of finite, negative length cycles implies that

$$\sum_{j=1}^{m} l^i\left(r^i_j, r^i_{j+1}\right) \geqslant 0$$

which can be rewritten using (6) and (8) as

$$\sum_{j=1}^{m} E_{-i}\left[\beta^i\left(f\left(r^i_j, t^{-i}\right) \mid t^{-i}\right) - \beta^i\left(f\left(r^i_{j+1}, t^{-i}\right) \mid t^{-i}\right)\right] r^i_j \geqslant 0.$$

This implies that

$$\sum_{j=1}^{m} E_{-i}\big[\beta^i\big(f\big(r^i_{j+1}, t^{-i}\big) \mid t^{-i}\big)\big]\big(r^i_{j+1} - r^i_j\big) \geqslant 0.$$

Thus, $E_{-i}[\beta^i(f(r^i, t^{-i}) \mid t^{-i})]$ is cyclically monotone.[5] From Rockafellar (1970, Theorem 24.8) it follows that there exists a convex function $\varphi : T^i \mapsto \mathbb{R}$ such that $E_{-i}[\beta^i(f(r^i, t^{-i}) \mid t^{-i})]$ is a selection from its subdifferential mapping, that is,

$$E_{-i}\big[\beta^i\big(f\big(r^i, t^{-i}\big) \mid t^{-i}\big)\big] \in \partial\varphi\big(r^i\big) \quad \forall r^i \in T^i.$$

This implies (see Krishna and Maenner, 2001, Theorem 1) that for any smooth path S in T^i joining $\underline{r}^i$ and $\bar{r}^i$ the following holds:

$$\int_{\underline{r}^i, S}^{\bar{r}^i} E_{-i}\big[\beta^i\big(f\big(r^i, t^{-i}\big) \mid t^{-i}\big)\big] = \varphi\big(\bar{r}^i\big) - \varphi\big(\underline{r}^i\big),$$

so $E_{-i}[\beta^i(f(r^i, t^{-i}) \mid t^{-i})]$ is path independent.

(2) $\Rightarrow$ (1). Let us assume that f satisfies weak monotonicity and that for every agent i, $E_{-i}[\beta^i(f(r^i, t^{-i}) \mid t^{-i})]$ is path independent. Take any edge from T^i_f and denote its starting node $\underline{r}^i$ and its ending node $\bar{r}^i$. Let L denote the line segment between $\underline{r}^i$ and $\bar{r}^i$, i.e. $L = \{r^i \in T^i \mid r^i = (1-\alpha)\underline{r}^i + \alpha\bar{r}^i,\ \alpha \in [0, 1]\}$. Now we pick any $r^i \in L$ and substitute the original edge with the path $A = (\underline{r}^i, r^i, \bar{r}^i)$ which has length $l^i(\underline{r}^i, r^i) + l^i(r^i, \bar{r}^i)$. By Lemma 1 we have

$$l^i\big(\underline{r}^i, \bar{r}^i\big) \geqslant l^i\big(\underline{r}^i, r^i\big) + l^i\big(r^i, \bar{r}^i\big), \tag{10}$$

that is, the original edge is at least as long as the path A. By repeated substitution we can generate a new path $\tilde{A} = (r^i_1 = \underline{r}^i, \ldots, r^i_m, r^i_{m+1} = \bar{r}^i)$, where $r^i_j \in L\ \forall j \in \{1, \ldots, m+1\}$. Then (10) implies that the original edge is at least as long as $\tilde{A}$, that is,

$$l^i\big(\underline{r}^i, \bar{r}^i\big) \geqslant \sum_{j=1}^{m} l^i\big(r^i_j, r^i_{j+1}\big)$$

(see also the example given in Fig. 1). Note that

$$\begin{aligned}
\sum_{j=1}^{m} l^i\big(r^i_j, r^i_{j+1}\big) &= \sum_{j=1}^{m} E_{-i}\big[v^i\big(f\big(r^i_j, t^{-i}\big) \mid r^i_j, t^{-i}\big) - v^i\big(f\big(r^i_{j+1}, t^{-i}\big) \mid r^i_j, t^{-i}\big)\big] \\
&= E_{-i}\big[v^i\big(f\big(r^i_1, t^{-i}\big) \mid r^i_1, t^{-i}\big) - v^i\big(f\big(r^i_{m+1}, t^{-i}\big) \mid r^i_m, t^{-i}\big)\big] \\
&\quad + \sum_{j=1}^{m-1} E_{-i}\big[v^i\big(f\big(r^i_{j+1}, t^{-i}\big) \mid r^i_{j+1}, t^{-i}\big) - v^i\big(f\big(r^i_{j+1}, t^{-i}\big) \mid r^i_j, t^{-i}\big)\big] \\
&= E_{-i}\big[v^i\big(f\big(r^i_1, t^{-i}\big) \mid r^i_1, t^{-i}\big) - v^i\big(f\big(r^i_{m+1}, t^{-i}\big) \mid r^i_{m+1}, t^{-i}\big)\big] \\
&\quad + \sum_{j=1}^{m} E_{-i}\big[v^i\big(f\big(r^i_{j+1}, t^{-i}\big) \mid r^i_{j+1}, t^{-i}\big) - v^i\big(f\big(r^i_{j+1}, t^{-i}\big) \mid r^i_j, t^{-i}\big)\big]
\end{aligned}$$

[5] The notion of cyclical monotonicity was introduced by Rockafellar (1966).

$$= E_{-i}\big[v^i\big(f\big(\underline{r}^i, t^{-i}\big) \mid \underline{r}^i, t^{-i}\big) - v^i\big(f\big(\bar{r}^i, t^{-i}\big) \mid \bar{r}^i, t^{-i}\big)\big] + \sum_{j=1}^{m} E_{-i}\big[\beta^i\big(f\big(r^i_{j+1}, t^{-i}\big) \mid t^{-i}\big)\big]\big(r^i_{j+1} - r^i_j\big).$$

The first equality follows from the definition of the edge length given in (6). The second equality follows from rearranging the terms of the summation. The third equality is derived by adding and subtracting $E_{-i}[v^i(f(r^i_{m+1}, t^{-i}) \mid r^i_{m+1}, t^{-i})]$. To derive the last equality we use (8) and that $r^i_1 = \underline{r}^i$, $r^i_{m+1} = \bar{r}^i$. By repeated substitution we can generate paths with more and more edges. In the limit the distance between neighboring nodes goes to zero and

$$\sum_{j=1}^{m} E_{-i}\big[\beta^i\big(f\big(r^i_{j+1}, t^{-i}\big) \mid t^{-i}\big)\big]\big(r^i_{j+1} - r^i_j\big) \to \int_{\underline{r}^i, L}^{\bar{r}^i} E_{-i}\big[\beta^i\big(f\big(r^i, t^{-i}\big) \mid t^{-i}\big)\big].$$

Thus, the length of $\tilde{A}$ goes to

$$E_{-i}\big[v^i\big(f\big(\underline{r}^i, t^{-i}\big) \mid \underline{r}^i, t^{-i}\big) - v^i\big(f\big(\bar{r}^i, t^{-i}\big) \mid \bar{r}^i, t^{-i}\big)\big] + \int_{\underline{r}^i, L}^{\bar{r}^i} E_{-i}\big[\beta^i\big(f\big(r^i, t^{-i}\big) \mid t^{-i}\big)\big], \tag{11}$$

as $m \to \infty$. Now, let $C = (r^i_1, \dots, r^i_m, r^i_{m+1} = r^i_1)$ denote a finite cycle in T^i_f. Furthermore, let L_j denote the line segment between r^i_j and r^i_{j+1}. The result in (11) and the path independence of $E_{-i}[\beta^i(f(r^i, t^{-i}) \mid t^{-i})]$ imply for the length of C that

$$\begin{aligned}\sum_{j=1}^{m} l^i\big(r^i_j, r^i_{j+1}\big) &\geqslant \sum_{j=1}^{m} E_{-i}\big[v^i\big(f\big(r^i_j, t^{-i}\big) \mid r^i_j, t^{-i}\big) - v^i\big(f\big(r^i_{j+1}, t^{-i}\big) \mid r^i_{j+1}, t^{-i}\big)\big] \\ &\quad + \sum_{j=1}^{m} \int_{r^i_j, L_j}^{r^i_{j+1}} E_{-i}\big[\beta^i\big(f\big(r^i, t^{-i}\big) \mid t^{-i}\big)\big] \\ &= 0,\end{aligned}$$

that is, C has non-negative length. In order to see the equality relation, note the following: the terms of the first summation cancel each other out. Furthermore, the second summation describes an integral over a closed path in T^i which, due to path independence, equals zero. □

Weak monotonicity of f and path independence of $E_{-i}[\beta^i(f(r^i, t^{-i}) \mid t^{-i})]$ do not imply one another. That weak monotonicity does not imply path independence follows directly from Example 1 and Theorem 2. It can also be derived directly from Example 1. If we consider for example path A consisting of the line segment between x and y and path $\tilde{A}$ consisting of the line segment between x and z and the line segment between z and y, we find that

$$\int_{x, A}^{y} \beta\big(f(r)\big) = -\frac{3}{2} \quad \text{and} \quad \int_{x, \tilde{A}}^{y} \beta\big(f(r)\big) = 3.$$

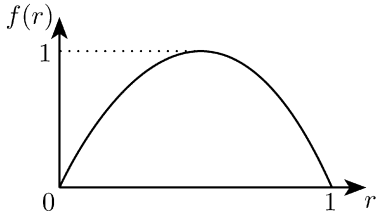

Fig. 3. The allocation function in Example 2.

So the path integral of $\beta(f(r))$ from x to y is not independent of the path of integration. That weak monotonicity of f does not imply path independence of $E_{-i}[\beta^i(f(r^i, t^{-i}) \mid t^{-i})]$ depends crucially upon the assumption of multi-dimensional type spaces. If we would consider one-dimensional type spaces instead, then weak monotonicity would indeed imply path independence.

That path independence does not imply weak monotonicity is illustrated by the following example.

Example 2. Let us consider the allocation of a single, indivisible object. For simplicity we assume that there exists only a single agent to possibly allocate to. He has a type $t \in T = [0, 1]$ which reflects the value of the object for him. Given a report r of the agent, the allocation rule $f : T \mapsto [0, 1]$ assigns to him a probability for getting the object. The agent's valuation for the resulting allocation is $v(f(r) \mid t) = f(r)t$. Specifically, we set $f(r) = -(2r - 1)^2 + 1$ (see Fig. 3). Clearly, f is path independent but not weakly monotone.

If f is Bayes–Nash incentive compatible, the corresponding payments can be constructed by using shortest path lengths (as described in the proof of Theorem 1). For each $i \in N$, let us pick some a^i as the source node in T^i_f. Thus, if agent i reports t^i, he has to make a payment

$$P_i(t^i) = \inf \sum_{j=1}^{m} l^i(r^i_j, r^i_{j+1}), \tag{12}$$

where the infimum is taken over all finite paths from t^i to a^i. Take any finite path $A = (r^i_1 = t^i, \ldots, r^i_{m+1} = a^i)$ in T^i_f. Let L_j denote the line segment between r^i_j and r^i_{j+1}, whereas L_t denotes the line segment between the source and t^i. Following the repeated substitution approach presented in the second part of the proof of Theorem 2, we can construct paths that are shorter (or as long) by letting them visit the same nodes as A and also additional nodes along the line segments in between. In the limit, as the number of nodes goes to infinity, the distance between neighboring nodes goes to zero and the length of the paths goes to

$$\sum_{j=1}^{m} \Bigg(E_{-i}\big[v^i\big(f\big(r^i_j, t^{-i}\big) \mid r^i_j, t^{-i}\big) - v^i\big(f^i\big(r^i_{j+1}, t^{-i}\big) \mid r^i_{j+1}, t^{-i}\big)\big] + \int_{r^i_j, L_j}^{r^i_{j+1}} E_{-i}\big[\beta^i\big(f\big(r^i, t^{-i}\big) \mid t^{-i}\big)\big] \Bigg). \tag{13}$$

Using path independence in (13) we have that[6]

$$\sum_{j=1}^{m} \int_{r_j^i, L_j}^{r_{j+1}^i} E_{-i}\big[\beta^i\big(f\big(r^i, t^{-i}\big) \mid t^{-i}\big)\big] = \int_{t^i, L_t}^{a^i} E_{-i}\big[\beta^i\big(f\big(r^i, t^{-i}\big) \mid t^{-i}\big)\big].$$

Applying the above to (12) yields

$$P_i\big(t^i\big) = E_{-i}\big[v^i\big(f\big(t^i, t^{-i}\big) \mid t^i, t^{-i}\big) - v^i\big(f^i\big(a^i, t^{-i}\big) \mid a^i, t^{-i}\big)\big] - \int_{a^i, L_t}^{t^i} E_{-i}\big[\beta^i\big(f\big(r^i, t^{-i}\big) \mid t^{-i}\big)\big], \tag{14}$$

implying that the expected utility (see (1) for definition) for truthfully reporting t^i is[7]

$$U^i\big(t^i \mid t^i\big) = U^i\big(a^i \mid a^i\big) + \int_{a^i, L_t}^{t^i} E_{-i}\big[\beta^i\big(f\big(r^i, t^{-i}\big) \mid t^{-i}\big)\big]. \tag{15}$$

Acknowledgments

The authors are grateful to the participants of the Second World Congress of the Game Theory Society (2004) and the First Spain Italy Netherlands Meeting on Game Theory (2005) for helpful discussions. We especially thank Philip J. Reny, Rakesh V. Vohra and an anonymous associate editor of Games and Economic Behavior for their useful comments.

References

Ahuja, R.K., Magnanti, T.L., Orlin, J.B., 1993. Network Flows—Theory, Algorithms and Applications. Prentice-Hall, New Jersey.

Bikhchandani, S., Chatterji, S., Lavi, R., Mu'alem, A., Nisan, N., Sen, A., 2006. Weak monotonicity characterizes deterministic dominant-strategy implementation. Econometrica 74, 1109–1132.

Gui, H., Müller, R., Vohra, R.V., 2004. Dominant strategy mechanisms with multidimensional types. METEOR Research Memorandum 04/046.

Jehiel, P., Moldovanu, B., 2001. Efficient design with interdependent valuations. Econometrica 69, 1237–1259.

Jehiel, P., Moldovanu, B., Stacchetti, E., 1999. Multidimensional mechanism design for auctions with externalities. J. Econ. Theory 85, 258–293.

Klemperer, P., 1999. Auction theory: A guide to the literature. J. Econ. Surveys 13, 227–286.

Krishna, V., Maenner, E., 2001. Convex potentials with an application to mechanism design. Econometrica 69, 1113–1119.

Müller, R., Perea, A., Wolf, S., 2005. Weak monotonicity and Bayes–Nash incentive compatibility. METEOR Research Memorandum 05/040.

Roberts, K., 1979. The characterization of implementable choice rules. In: Laffont, J.-J. (Ed.), Aggregation and Revelation of Preferences. North-Holland, Amsterdam, pp. 321–348.

[6] The line segment L_t for the path of integration is picked for convenience. Due to path independence, it can be replaced with any other path connecting the source and t^i.

[7] In order to derive (15) one can use that by construction $P_i(a^i) = 0$ and thus add this term to the right-hand side of (14).

Rochet, J.-C., 1987. A necessary and sufficient condition for rationalizability in a quasi-linear context. J. Math. Econ. 16, 191–200.
Rockafellar, R.T., 1966. Characterization of the subdifferentials of convex functions. Pacific J. Math. 17, 487–510.
Rockafellar, R.T., 1970. Convex Analysis. Princeton Univ. Press, Princeton.
Saks, M., Yu, L., 2005. Weak monotonicity suffices for truthfulness on convex domains. In: Proceedings of the 6th ACM Conference on Electronic Commerce (EC05). ACM Press, New York, pp. 286–293.
Shostak, P., 1981. Deciding linear inequalities by computing loop residues. J. ACM 28, 769–779.