Ex-post Implementation and Preference Aggregation via Potentials

Philippe Jehiel, Moritz Meyer-ter-Vehn and Benny Moldovanu*

February 26, 2007

Abstract

We introduce several notions of potentials for mechanism design problems with interdependent values, and relate them to implementation in ex-post equilibrium. Whereas ex-post implementation is closely linked to the ordinal concept of *best-alternative potentials*, the celebrated Vickrey-Clarke-Groves mechanism corresponds to the stronger notion of a *cardinal potential:* agents agree not only on the best alternative, but also on the quantitative differences between all alternatives. We characterize all valuations that allow for cardinal potentials, and use this characterization for: 1) Identifying valuations for which ex-post implementation is possible; 2) Identifying classes of valuations for which all ex-post implementable choice rules correspond to cardinal potentials. The latter allows us to extend to interdependent valuations a result for dominant strategy implementation in private values settings, due to Roberts (1979).

1 Introduction

We introduce several notions of potentials for mechanism design problems with interdependent values, discuss their properties, and establish relations between these notions and ex-post implementation. Roughly speaking, a mechanism design problem with given valuation functions admits a potential if there exist monetary transfers such that the maximization problem of each

---

agent coincides with the problem of maximizing a single "potential" function, common to all agents[1]. Thus, the existence of a potential suggests the possibility of aligning the interests of all agents.

Aligning the interests of heterogenous strategic agents that jointly control a decision is a central desideratum in mechanism design and implementation. By attaching different monetary transfers to different social alternatives, the designer can affect the agents' preferences over these alternatives so that, ultimately, all agents agree about the preferred alternative, and hence all agents find it in their own strategic interest to behave in a way that leads to the commonly preferred alternative.

The most famous example of successful alignment is offered by the Vickrey-Clarke-Groves mechanisms (see Vickrey (1961), Clarke (1971), and Groves (1973)) for private values environments with quasi-linear utility. There, an agent receives a transfer equal to the sum of valuations of the other agents in the chosen social alternative. With such transfers all individual payoff maximization decision problems coincide with the maximization of social surplus, yielding the well-known dominant strategy implementability of the efficient choice rule.

The notions of interest alignment analyzed in the present paper bear a strong formal resemblance to potentials for normal form games, defined by Monderer and Shapley (1996). Roughly speaking, a normal form game admits a potential if there exists a function (common to all players) from strategy profiles to the set of real numbers such that, for any player, changes in utility resulting from changes in own strategy (while keeping fixed others' strategies) are reflected in appropriate changes in the value of the common potential function. A main result is that a strategy profile is a Nash equilibrium of the original game if and only if it is a Nash equilibrium of the artificial game where each player's utility function is replaced by the common potential. Thus, the equilibria of strategic interaction in a potential game are mirrored in a much simpler game where all players' interests are identical.

Following the literature on potential games, we shall distinguish between *best-alternative, ordinal and cardinal potentials.* The first, weakest, concept says that the potential function and each agent's payoff function agree on the best alternative; the second requires that the potential function agrees with every agent's preference order over all alternatives; the third, strongest,

---

[1]The potentials introduced here should not be confused with the individual potential functions arising as expected equilibrium utility functions in Bayes-Nash implementation (see for example Jehiel, Moldovanu and Stacchetti, 1999). The common name reflects certain properties about path integrals first analyzed in the physical sciences (e.g., energy conservation).

concept requires that the potential function coincides, up to an affine transformation, with each agent's utility function (as is the case for social surplus in the Vickrey-Clarke-Groves mechanism).

Whereas in Monderer and Shapley's general setting a potential need not have an economic interpretation related to the game's features (see for example their derivation of a potential for a Cournot oligopoly), a potential in our mechanism design problem is closely related to the choice rule that is being implemented through the alignment of interests induced by that potential. In spite of the formal resemblance, the links between potentials in mechanism design and potential games are not immediate. Whereas an agent's preferences over her strategies are preserved by the potential function in a potential game, her preferences over alternatives are explicitly altered, via transfers, in the potential function for a mechanism design problem. It is also not true that a mechanism that admits a potential gives rise to a potential game in the corresponding revelation game[2].

Our present focus is on ex-post implementation - a weakening of dominant strategy implementation, appropriate also for settings with interdependent valuations. This notion requires that an agent is not willing to change her strategy for any type realization of the other agents. Ex-post implementation has recently received a lot of attention[3] because it ensures that neither the mechanism designer, nor the agents need to know the distribution from which signals are drawn in order to design the mechanism or to play optimally in the induced game[4].

Via a *taxation principle*, it can be easily shown, that best-alternative potentials are equivalent to ex-post implementable choice rules. Thus, statements over implementable choice rules can be translated into statements about best-alternative potentials[5]. Ordinal potentials go beyond ex-post implementation by requiring that, after eventual transfers, agents agree over the ranking of all alternatives (rather than just the best alternative), and

---

[2]In contrast, Sandholm (2005) shows how a price scheme administered by a designer can be used to augment an externality abatement game in order to yield a potential game a la Monderer-Shapley. A dynamic learning process leads there to an efficient outcome. Compare Example 6.4.

[3]See, among others, Dasgupta and Maskin (2000), Jehiel and Moldovanu (2001), Chung and Ely (2001), Bergemann and Välimäki (2002), Perry and Reny (2002), and Jehiel et al. (2006).

[4]See Bergemann and Morris (2005) for a formal treatment of this issue, and for the connection to "Wilson's doctrine" about detail-free mechanisms.

[5]Potentials should be seen as offering a different interpretation of the mechanism design problem. Whereas the latter studies joint decisions based on agents' signals (while providing incentives for truthful revelation), the former focuses on the aggregation of agents preferences.

cardinal potentials go even further by requiring that agents' utilities ultimately agree up to affine transformations. While implementation theory has essentially focused on best-alternative potentials, ordinal and cardinal potentials offer stronger notions of interest alignment, which should be attractive for a number of applications[6]. In particular, as mentioned above, the Vickrey-Clarke-Groves mechanisms correspond to cardinal potentials.

The paper is organized as follows: In Section 2 we describe the mechanism design problem with interdependent values, and state a well-known taxation principle. In Section 3 we define several notions of potentials, and link potentials to implementation in ex-post equilibria (Proposition 3.4).

In Section 4 we use existing results from the theory on potential games in order to provide conditions for the existence of cardinal potentials (Proposition 4.1). Roughly speaking, the existence of cardinal potentials requires, for each alternative k, the identity of the cross derivatives of agent i's and j's valuations with respect to the signals held by i and j. Equivalently, as shown by Ui (2000), this requirement can be expressed as a separability condition on the valuation functions.

It is a-priori not straightforward to determine whether ex-post implementation is possible for given valuation functions. But, the above sufficient conditions are readily checked. Furthermore, the conditions are constructive in the sense that, when satisfied, they indicate how to construct the cardinal potential, the implemented choice rule and the necessary transfers. If the conditions are not satisfied, however, ex-post implementation via weaker notions of potentials may still be possible.

After recalling that, with generic interdependent valuation functions, only constant choice rules can be ex-post implemented (see Jehiel et al., 2006), we next consider settings in which cardinal potentials exist. Our main question is whether and when all ex-post implementable choice rules can be represented by cardinal potentials in such cases.

Our main results can be summarized as follows:

1. When the valuation functions admit a cardinal potential and are generic within the class of settings admitting cardinal potentials and multidimensional type spaces, we show that every non-trivial implementable choice rule is represented by a translate of the cardinal potential (Proposition 4.4).

---

[6]For example, such stronger alignments may be desirable if there is some probability that the best-alternative will no longer be available at the time when the joint decision must be implemented. By contrast, in the literature on potential games, it is unclear what the stronger notions of potentials mean beyond providing technical simplifications to characterize the potential.

2. In a more special class of settings with "rich enough" multi-dimensional type spaces and with valuations that are additively separable in the agent's own signal and other agents' signals, we show that any ex-post implementable choice rule that is not trivial is an *affine maximizer:* any such choice rule can be represented by a maximization of a weighted sum of the parts of agents' utilities that depend on their own signals, augmented by signal-independent, alternative-specific weights (Proposition 4.7). Since almost all affine maximizers are cardinal potentials, it follows that in this case too almost all implementable non-trivial choice rules correspond to cardinal potentials.

3. In one-dimensional settings we observe that ex-post implementation is possible for an open set of valuations functions (thus implying that best-alternative potentials exist), and yet cardinal implementation is generically impossible. We also provide a one-dimensional example in which a cardinal potential exists and yet other (non-trivial) choice rules can be ex-post implemented. This example illustrates the significant difference between the two notions in such restricted domains.

Of course, the same questions about relations between best-alternative and cardinal potentials can be asked in private values settings. As mentioned above, the classic work of Vickrey, Clarke, and Groves implies that the maximization of social surplus is implementable and corresponds to a cardinal potential. But, their work leaves open the question whether some alternative choice rules can be implemented in dominant strategies without corresponding to cardinal potentials. In a remarkable paper, Roberts (1979) shows that, in a sufficiently rich, high-dimensional environment, deterministic implementation in dominant strategy boils down to affine maximization[7]. Since affine maximizers are cardinal potentials, cardinal and best-alternative potentials coincide in the environments studied by Roberts. In fact, we heavily use Roberts' result for our analysis of the interdependent but separable values case.

Bikhchandani et al. (2006) characterize dominant strategy implementation for private values settings in terms of a monotonicity condition[8]. The fundamental difference to Roberts' work is that these authors consider a restricted domain of preferences, better suited to some of the applications they

---

[7]Lavi et al. (2004) offer alternative proofs for Roberts' main result, and for another characterization result (that uses a condition of "player-decisiveness"), due to Meyer-ter-Vehn and Moldovanu (2002).

[8]Similar characterizations in terms of a "no cycle condition" are given by Gui et al. (2004). Their work builds on an earlier insight about monotonicity properties of subdifferentials of convex functions, due to Rochet (1987).

have in mind (e.g., combinatorial auctions without externalities). In particular, the characterization of dominant strategy implementable choice rules as affine maximizers does not hold anymore in their framework. Thus, in restricted, private-values domains, best-alternative and cardinal potentials need not coincide.

There are two appendices: Appendix A recalls the original definitions of potential games and the main result of Monderer and Shapley (1996), and it shows that mechanism design problems generally do not give rise to potential games. Appendix B contains several proofs that would interrupt the flow of argument in the main text.

2   The Mechanism Design Model

We consider a situation where $N \in \mathbb{N}$ agents $i \in \mathcal{N}$ are affected by a decision among $K \in \mathbb{N}$ alternatives, $k, l \in \mathcal{K}$. Agent i's utility $u^i = v^i_k + t^i$ is determined by a quasi-linear utility function that takes into account the chosen alternative k and a monetary transfer $t^i \in \mathbb{R}$. Her valuation for alternative k, $v^i_k = v^i_k(s)$, depends smoothly on the state of the world $s \in S$.

Each agent gets a private signal $s^i \in S^i \subseteq \mathbb{R}^{d^i}$ about the state of the world $s \in S$. The signal s^i results from an exogenous draw. Thus, we identify states of the world with signal profiles: $S = \prod_{i \in \mathcal{N}} S^i$. We adopt the usual notation $s^{-i} = (s^j)_{j \in \mathcal{N}, j \neq i}$, and more generally $s^{\mathcal{I}} = (s^j)_{j \in \mathcal{I}}$ and $s^{-\mathcal{I}} = (s^j)_{j \in \mathcal{N} \setminus \mathcal{I}}$ for a subset of agents $\mathcal{I} \subseteq \mathcal{N}$, and write $s = (s^i, s^{-i})$ when we focus on agent i.

We denote by ∇_{s^i} the d^i-dimensional vector of partial derivatives with respect to s^i and assume that the value difference between any two alternatives k, l is not satiated in one's own signal s^i:

$$\nabla_{s^i} \left(v^i_k - v^i_l \right)(s) \neq 0 \text{ for all } s \in S. \tag{1}$$

Given an infinitesimal change ds^i of i's signal, we denote by $dv^i_k = \nabla_{s^i} v^i_k(s) \bullet ds^i$ the corresponding infinitesimal effect on i's valuation of alternative k.

We study choice rules $\psi : S \to \mathcal{K}$ with the property that there are transfers functions $t = (t^i)_{i \in \mathcal{N}} : S \to \mathbb{R}^N$ such that truth-telling **almost everywhere** is an **ex-post equilibrium** in the incomplete information game that is induced by the direct revelation mechanism (ψ, t), i.e.

$$v^i_{\psi(s)}(s) + t^i(s) \geq v^i_{\psi(\tilde{s}^i, s^{-i})}(s) + t^i\left(\tilde{s}^i, s^{-i}\right) \tag{2}$$

for all agents i, almost all signal profiles $s = (s^i, s^{-i})$ and all possible misrepresentations of i's signal $\tilde{s}^i \in S^i$. We shall call such ψ **implementable** (a.e.) and (ψ, t) an incentive compatible mechanism (a.e.), and will suppress the

"(a.e.)" from here on[9]. Consider an incentive compatible mechanism (ψ, t). By Fubini's Theorem, for almost all $s^i \in S^i$, the set $S^{-i}(s^i) \subseteq S^{-i}$ of other agents' signals s^{-i} such that constraint (2) is violated for $s = (s^i, s^{-i})$ and some misrepresentation $\tilde{s}^i \in S^i$ is a null-set. Thus almost all types s^i find it optimal to truthfully report their type given truthful reporting by the other agents, and given any continuous type distribution over S^{-i}.

We call two choice rules ψ, ψ' **equivalent** if they agree on the chosen alternative $\psi(s) = \psi'(s)$ for almost all signal profiles $s \in S$; we call ψ **exhaustive** if every alternative k is chosen on some set $S(k)$ of states with positive measure; we call ψ **trivial** if it chooses the same alternative k in almost all states s.

We conclude this section by stating a "taxation principle" for ex-post implementation, first pointed out by Chung and Ely (2001). This is a multi-agent generalization of a well-known idea in the monopolistic screening literature: Instead of asking an individual for her information and deciding on an alternative and a transfer based on the report, the central authority can, equivalently, post prices for the different alternatives and let the individual choose among them. In our setting with multiple agents, these prices are personalized and depend on the signals of the other agents. In equilibrium all agents must agree on the best alternative.

Lemma 2.1 (Taxation Principle) *A choice rule ψ is implementable if and only if for every agent i there are transfers $(t_k^i(s^{-i}))_k \in (\mathbb{R} \cup \{-\infty\})^K \setminus (-\infty, ..., -\infty)$[10], such that for almost all s it holds that:*

$$\psi(s) \in \arg\max_{k \in \mathcal{K}} \left\{ v_k^i(s) + t_k^i(s^{-i}) \right\} \qquad (3)$$

Proof. "if": Given $t_k^i(s^{-i})$, such that condition (3) holds, define $t^i(s) := t_{\psi(s)}^i(s^{-i})$. Agent i's problem in the game induced by the mechanism (ψ, t)

---

[9]The reason for requiring optimality almost everywhere rather than everywhere is that we want to allow some leeway on zero-measure sets of signals where agents are indifferent between multiple alternatives. This avoids tedious technical details in the definition of best-alternative and ordinal potentials in equations (5) and (6), while allowing us to focus on the main economic insights implied by ex-post implementation.

Results on the restrictiveness of ex-post implementation, such as the those in Jehiel et al. (2006) and Proposition 4.4 below, generalize to choice rules that are implementable almost everywhere, as the geometric condition driving the result is determined by the incentive constraints in a whole neighborhood (with positive measure) of the indifference set.

[10]It is necessary to allow for $t_k^i(s^{-i}) = -\infty$ for some alternative k in order to ensure that $\psi(s^i, s^{-i}) \neq k$ for all $s^i \in S^i$. On the other hand one cannot allow for $t_k^i(s^{-i}) = -\infty$ for all alternatives k, as this would make agent i indifferent between all alternatives, and yield him infinite disutility.

is: $\max_{\widetilde{s}^i}\left\{ v^i_{\psi(\widetilde{s}^i, s^{-i})}\left(s^i, s^{-i}\right) + t^i_{\psi(\widetilde{s}^i, s^{-i})}\left(s^{-i}\right)\right\}$. By condition (3), it is optimal for her to report $\widetilde{s}^i = s^i$ since then the choice rule ψ picks, for almost all s, her most preferred alternative.

"only if": Let ψ be ex-post implementable via a mechanism (ψ, t). Let $S^i\left(s^{-i}\right)$ be the set of signals s^i such that constraint (2) holds for all $\widetilde{s}^i$. By Fubini's Theorem, for almost all $s^{-i} \in S^{-i}$, the set $S^i\left(s^{-i}\right)$ has full measure. We define

$$t^i_k\left(s^{-i}\right) = \begin{cases} t^i\left(s^i, s^{-i}\right) & \text{if } \psi\left(s^i, s^{-i}\right) = k \text{ for some } s^i \in S^i\left(s^{-i}\right) \\ -\infty & \text{if } \psi\left(s^i, s^{-i}\right) \neq k \text{ for all } s^i \in S^i\left(s^{-i}\right). \end{cases} \qquad (4)$$

Note that $t^i_k\left(s^{-i}\right)$ is well-defined since, by i's incentive constraint,

$$\psi\left(s^i, s^{-i}\right) = \psi\left(\widetilde{s}^i, s^{-i}\right) = k \text{ for } s^i, \widetilde{s}^i \in S^i\left(s^{-i}\right) \;\Rightarrow t^i\left(s^i, s^{-i}\right) = t^i\left(\widetilde{s}^i, s^{-i}\right).$$

By i's incentive constraint we know that she reports in a way that maximizes her payoff. Thus, with $t^i_k\left(s^{-i}\right)$ as defined in equation (4), condition (3) is satisfied for $s = \left(s^i, s^{-i}\right)$ such that $s^i \in S^i\left(s^{-i}\right)$. ∎

3 Preference Aggregation via Potentials

Definition 3.1   *1. A family $P = (P_k)_{k \in \mathcal{K}}$ of functions $P_k : S \to \mathbb{R}$ is a **best-alternative potential** for valuations v if there are "transfers" $(t^i_k(s^{-i}))_k \in (\mathbb{R} \cup \{-\infty\})^K \setminus (-\infty, ..., -\infty)$ such that, at almost every state of the world s, the potential agrees with each agent i on the most favored alternative(s) k:*

$$\arg\max_k \left\{ v^i_k(s) + t^i_k\left(s^{-i}\right) \right\} = \arg\max_k \left\{ P_k(s) \right\} \qquad (5)$$

*2. A family $P = (P_k)_{k \in \mathcal{K}}$ of functions $P_k : S \to \mathbb{R}$ is an **ordinal potential** for valuations v if there are "transfers" $(t^i_k(s^{-i}))_k \in (\mathbb{R} \cup \{-\infty\})^K \setminus (-\infty, ..., -\infty)$ such that, at almost every state of the world s, the potential agrees with each agent i on the preference order of alternatives:*

$$\forall i, \forall k, l \in \mathcal{K}, \; v^i_k(s) + t^i_k\left(s^{-i}\right) \geq v^i_l(s) + t^i_l\left(s^{-i}\right) \Leftrightarrow P_k(s) \geq P_l(s) \quad (6)$$

*3. A family $P = (P_k)_{k \in \mathcal{K}}$ of functions $P_k : S \to \mathbb{R}$ is a **cardinal potential** for v if there exist (α^i, β^i), $\alpha^i > 0$ and "transfers" $(t^i_k(s^{-i}))_k \in (\mathbb{R} \cup \{-\infty\})^K \setminus (-\infty, ..., -\infty)$ such that, at almost every state of the world s, the utility of each agent i coincides with the potential up to an (α^i, β^i)-affine transformation:*

$$\forall i, k, \; \alpha^i \left[v^i_k(s) + t^i_k\left(s^{-i}\right) \right] + \beta^i = P_k(s) \qquad (7)$$

8

*A cardinal potential that is achieved through weights α^i will be called an α-**potential**.*

The above definitions closely parallel those for potential games (see Appendix A). Note that in the definition of a cardinal potential we can always choose $\beta^i = 0$ without loss of generality: if $P = (P_k)_{k \in \mathcal{K}}$ is a cardinal potential for some $\{(\alpha^i, \beta^i)\}_i$ it is also a cardinal potential for $\{(\alpha^i, 0)\}_i$[11]. This justifies the terminology for α-potentials.

The following result summarizes the simple relations among the above notions:

Proposition 3.2 *An ordinal potential for valuations v is a fortiori a best-alternative potential. A cardinal potential for valuations v is a fortiori an ordinal potential.*

Proof. If P is an ordinal potential, condition (6) obviously implies equation (5) and P is a best-alternative potential.

If P is an α-potential, $\alpha^i > 0$ ensures that condition (6) is satisfied. ∎

Recall that a choice rule identifies a unique "best" alternative for each profile of signals. Thus, it is not surprising that implementable choice rules are closely related to best-alternative potentials. We now formally express this connection.

Definition 3.3 *Two best-alternative potentials P, P' for valuations v are **equivalent** if $\arg\max_k \{P_k(s)\} = \arg\max_k \{P'_k(s)\}$ at almost every state of the world s[12].*

Proposition 3.4 *There is a one-to-one relation between equivalence classes of best-alternative potentials $\overline{P}$ and equivalence classes of ex-post implementable choice rules $\overline{\psi}$.*

Proof. We define for each class of best-alternative potentials $\overline{P}$ an equivalence class $\overline{\psi} = \Xi(\overline{P})$ of implementable choice rules, and we then show that Ξ is a bijection.

Given $\overline{P}$ with representative $P \in \overline{P}$, define $\overline{\psi} = \Xi(\overline{P})$ by defining a representative $\psi \in \overline{\psi}$ with $\psi(s) \in \arg\max_k \{P_k(s)\}$. By the taxation principle, the defined choice rule ψ is implementable. In order for Ξ to be well-defined

---

[11]For $\beta^i = 0$, consider the transfers $\widehat{t}_k^i(s^{-i}) = t_k^i(s^{-i}) + \frac{\beta^i}{\alpha^i}$ where $t_k^i(s^{-i})$ is the transfer associated with the (α^i, β^i)-potential.

[12]Note that by non-satiation, condition (1), and by the definition of best-alternative potentials, condition (5), the set $\arg\max_k \{P_k(s)\}$ is a singleton for almost all s.

we need to check that $\overline{\psi} = \Xi\left(\overline{P}\right)$ does not depend on the choice of $P \in \overline{P}$ in the construction of ψ. However, this is obvious from Definition 3.3.

To prove that Ξ is one-to-one, observe that $\Xi\left(\overline{P}\right) = \Xi\left(\overline{P}'\right)$ implies that $\arg\max_k \{P_k(s)\} = \arg\max_k \{P'_k(s)\}$ for all $P \in \overline{P}$, $P' \in \overline{P}'$ and for almost all $s \in S$. Thus we get that $\overline{P} = \overline{P}'$, proving that Ξ is injective. Finally, the taxation principle yields for a given implementable choice rule ψ that $P_k(s) = v_k^i(s) + t_k^i(s^{-i})$ defines a best-alternative potential with $\Xi\left(\overline{P}\right) = \overline{\psi}$, yielding surjectivity. ∎

4   Cardinal Potentials

It is obvious form the work of Vickrey, Clarke and Groves and from the above definitions that social surplus $P_k(s) = \sum_i v_k^i(s^i)$ is a cardinal potential for private values settings. Thus, if valuations are private, dominant-strategy implementation of the efficient choice rule can be achieved via the strong notion of cardinal potentials.

We next give a condition for the existence of cardinal potentials in interdependent values settings. This condition is not satisfied for generic valuations, and hence ex-post implementation via cardinal potentials is generically impossible. But, the non-generic set of valuation functions that allow for cardinal potentials goes far beyond private values, and these more general families of valuations play a role in various applications.

4.1   Existence of Cardinal Potentials

Denote by $\partial_{s^i,\widetilde{s}^i}\partial_{s^j,\widetilde{s}^j}v_k^i\left(s^{-\{i,j\}}\right)$ the second difference of v_k^i:

$$
\begin{aligned}
\partial_{s^i,\widetilde{s}^i}\partial_{s^j,\widetilde{s}^j}v_k^i\left(s^{-\{i,j\}}\right) \quad &:= \left(v_k^i\left(s^i,s^j,s^{-\{i,j\}}\right) - v_k^i\left(s^i,\widetilde{s}^j,s^{-\{i,j\}}\right)\right) - \\
&\quad - \left(v_k^i\left(\widetilde{s}^i,s^j,s^{-\{i,j\}}\right) - v_k^i\left(\widetilde{s}^i,\widetilde{s}^j,s^{-\{i,j\}}\right)\right)
\end{aligned}
$$

Proposition 4.1 *For $\alpha = (\alpha^i)_{i\in\mathcal{N}} \gg 0$ the following conditions on the valuation functions $v = (v_k^i)_{i,k}$ are equivalent:*

1. *There exists an α-potential for valuations v.*

2. *There exist functions $P_k : S \to \mathbb{R}$ and $Q_k^i : S^{-i} \to \mathbb{R}$ for all k and i, such that:*
$$\alpha^i v_k^i(s) = P_k(s) + Q_k^i(s^{-i}) \tag{8}$$

10

3. *For all agents i, j and alternatives k, the cross-differences in the valuations coincide:*

$$\alpha^i \partial_{s^i, \widetilde{s}^i} \partial_{s^j, \widetilde{s}^j} v_k^i \left(s^{-\{i,j\}} \right) = \alpha^j \partial_{s^i, \widetilde{s}^i} \partial_{s^j, \widetilde{s}^j} v_k^j \left(s^{-\{i,j\}} \right) \tag{9}$$

4. *There exists a family of functions $\Phi_k^{\mathcal{I}} : S^{\mathcal{I}} \to \mathbb{R}$ for each subset of agents $\mathcal{I} \subseteq \mathcal{N}$ such that:*

$$\alpha^i v_k^i (s) = \sum_{\mathcal{I} \text{ with } i \in \mathcal{I}} \Phi_k^{\mathcal{I}} \left(s^{\mathcal{I}} \right) \tag{10}$$

The family $\left(\Phi_k^{\mathcal{I}} \right)_{\mathcal{I}}$ is called an interaction potential and defines the potential function $P_k(s) = \sum_{\mathcal{I}} \Phi_k^{\mathcal{I}} \left(s^{\mathcal{I}} \right)$.

For the proof of Proposition 4.1 it is convenient to first establish a formal link between potentials for mechanism design problems and potentials for normal form games. This will allow us to use several known results from the latter theory.

Given valuations v and given an alternative $k \in \mathcal{K}$, consider an artificial game in normal form $\Gamma_k(v)$ where the set of players is $\mathcal{N}$, the set of strategies of player $i \in \mathcal{N}$ is S^i, and the payoff function of player i is $v_k^i : S \to \mathbb{R}$.

Lemma 4.2 *A family $P = (P_k)_{k \in \mathcal{K}}$ of functions $P_k : S \to \mathbb{R}$ is an α-potential for v if and only if, for each $k \in \mathcal{K}$, the function P_k is an α-potential for the game $\Gamma_k(v)$*[13].

Proof. By the definition of potentials for normal form games (see Appendix A), P_k is an α-potential for the game $\Gamma_k(v)$ if and only if

$$\alpha^i (v_k^i(s^i, s^{-i}) - v_k^i(t^i, s^{-i})) = P_k(s^i, s^{-i}) - P_k(t^i, s^{-i})$$

holds for all $s^i, t^i \in S^i$, $s^{-i} \in S^{-i}$ and $i \in \mathcal{N}$. This condition is equivalent to the existences of functions $Q_k^i : S^{-i} \to \mathbb{R}$ such that

$$\alpha^i v_k^i(s) = P_k(s) + Q_k^i(s^{-i})$$

for all $s \in S$ and $i \in \mathcal{N}$[14]. ∎

Armed with the above Lemma, we now prove **Proposition 4.1**:

Proof. "1 ⇔ 2" follows from an immediate transformation of equation (7) by setting $Q_k^i(s^{-i}) = -\alpha^i t_k^i \left(s^{-i} \right) - \beta^i$.

---

[13]We are very grateful to an anonymous referee who suggested this result.

[14]This has been pointed out by Slade (1994) and Facchini et al. (1997) for one-dimensional signals s^i, but is easily seen to generalize to arbitrary signals.

"1 $\Leftrightarrow$ 3" follows from a path-integral argument, standard in the physics literature, that has been applied to potential games by Monderer and Shapley (1996). By their Corollary 2.9, condition (9) is equivalent to the fact that $\Gamma_k(v)$ is a potential game. The result follows then by Lemma 4.2.

"1 $\Leftrightarrow$ 4" follows from Lemma 4.2 and from the analogous result on potentials for normal form games, due to Ui (2000). ∎

For given valuations v, identities (9) provide an easy way to check whether implementation via cardinal potentials is possible or not. Moreover, when identities (9) hold it is fairly easy to construct the associated α-potentials. For the sake of illustration, assume that there are two agents i, j, that each agent holds a one-dimensional signal $s^i \in \mathbb{R}^+$, that v_k^i are smooth functions of s^i and s^j, and that identities (9) hold for all k and $\alpha = (1,1)$. Then $P = (P_k)_{k \in \mathcal{K}}$ is an α-potential if and only if there exist λ_k such that[15]

$$
P_k(s^i, s^j) = \int_0^{s^i} \frac{\partial v_k^i(x, s^j)}{\partial s^i} dx + v_k^j(0, s^j) + \lambda_k.
$$

Note that the conditions imposed by identities (9) are non-generic in the sense that if the valuations are "drawn randomly", identities (9) will typically not be satisfied. But, as highlighted by the equivalent conditions (8) or (10), the set of valuation functions that allow for implementation via cardinal potentials goes far beyond the private valuations case. These equivalent conditions provide alternative simple representations of those systems of valuation functions that admit an α-potential, such as the settings with separable or semi-separable valuations analyzed below.

4.2 Implementation via Cardinal Potentials

Having identified the conditions for the existence of cardinal potentials, we now turn to a systematic study of the following question: Assuming the existence of cardinal potential, we ask oursleves whether and when it is the case that all best-alternative potentials (or equivalently, ex-post implementatable choice rules) are cardinal potentials? As we shall see, this study also reveals interesting classes of valuations for which ex-post implementation is possible.

To get a flavor of the difference between best-alternative and cardinal potentials, let us give an interpretation of cardinal potentials in terms of rates of information substitution. Given a cardinal potential P that represents an implementable choice rule ψ, and an "indifference point" s where $\arg\max_{k'} \{P_{k'}(s)\} = \{k, l\}$, let us consider an infinitesimal change (ds^i, ds^j)

---

[15]This parallels Lemma 2.7 in Monderer and Shapley (1996) showing that a cardinal potential is unique up to a constant.

in the signals of agents i and j that does not affect their preferences between alternatives k and l:

$$\nabla_{s^i} \left(P_k - P_l \right) (s) \bullet ds^i + \nabla_{s^j} \left(P_k - P_l \right) (s) \bullet ds^j = 0$$

By the definition of the cardinal potential we get that

$$\alpha^i \nabla_{s^i} \left(v_k^i - v_l^i \right) (s) \bullet ds^i + \alpha^j \nabla_{s^j} \left(v_k^j - v_l^j \right) (s) \bullet ds^j = 0$$

which can be expressed as

$$-\frac{dv_{k,l}^i (s)}{dv_{k,l}^j (s)} = \frac{\alpha^j}{\alpha^i} \tag{11}$$

where $dv_{k,l}^i (s) = \nabla_{s^i} \left(v_k^i - v_l^i \right) (s) \bullet ds^i$ and $dv_{k,l}^j (s) = \nabla_{s^i} \left(v_k^j - v_l^j \right) (s) \bullet ds^j$. The left hand side of the above equation can be interpreted as a rate of information substitution. It compares the change in i's preference of alternative k over l due to an infinitesimal change of her signal with the analogous term for agent j. A high value for this quotient signifies that a large change in i's preference is needed to offset a smaller change in agent j's valuation. Thus, agent j can be considered to be "more important" to the decision making. Equation (11) says that this rate needs to be independent of the chosen alternatives k, l and of the indifference signal s. But, there is no *a-priori* reason why this strong requirement should be satisfied by an arbitray implementable choice rule (or by an arbitrary best-alternative potential). Indeed, we find that this condition is not generally satisfied by arbitrary implementable choice rules in settings with one-dimensional signals (compare with Figure 1).

By contrast, in many settings with multi-dimensional signals admitting a cardinal potential, we find the surprising result that almost every best-alternative potentials is equivalent to some cardinal potential, thereby implying that the strong information substitution condition is satisfied by almost all implementable choice rules.

The next three subsections analyze settings with multi-dimensional signals and proceed from general valuations to more and more restricted ones. In particular, we get larger and larger families of choice rules that can be ex-post implemented. The last subsection considers settings with one-dimensional signals.

4.2.1 Generic Multi-dimensional Settings

Jehiel et al. (2006) have shown that for generic[16] valuation functions any ex-post implementable choice rule must be constant (i.e., it can make no

---

[16]Genericity pertains here to both topological and measure-theoretic notions of residual sets and finite prevalence, respectively. See Jehiel et al. (2006) for precise definitions.

use of private information). In other words, any best-alternative potential is trivial. Since cardinal potentials are a-fortiori best-alternative potentials, we obtain:

Proposition 4.3 *For multi-dimensional signal spaces and generic valuation functions v with non-zero gradient $\nabla_{s^i}\left(v_k^i - v_l^i\right)(s) \neq 0$, the only implementable rules that are not cardinal potentials are trivial choice rules.*

4.2.2 Semi-Separable Valuations

We now focus on the less general class of settings with valuation functions that allow for the existence of a cardinal potential P, characterized in Proposition 4.1.

By the definition of potentials, we know that the choice rule represented by the potential $\psi(s) \in \arg\max_k \{P_k(s)\}$ is implementable, and it is easy to see that the same holds for all rules represented by its translates: $\psi(s) \in \arg\max_k \{P_k(s) + \lambda_k\}$, for any vector of real numbers $(\lambda_k)_k$. The next result says that, beyond these choice rules, only trivial rules are implementable.

Proposition 4.4 *Consider a setting with multi-dimensional signals and with valuation functions that admit an $\alpha-$potential (e.g., the valuations have the form $v_k^i(s) = \frac{1}{\alpha^i}(P_k(s) + Q_k^i(s^{-i}))$ where $P_k : S \to \mathbb{R}$ is a common value part, $Q_k^i : S^{-i} \to \mathbb{R}$ are arbitrary, and $\alpha^i > 0$). Then, for generic families $\{P_k(s)\}_k$, the only non-trivial ex-post implementable choice rules ψ are of the form $\psi(s) \in \arg\max_k \{P_k(s) + \lambda_k\}$ for a vector of constants $(\lambda_k)_k$, and are thus represented by cardinal potentials.*

The proof of the above result closely follows that of Proposition 3.3 and Theorem 4.2 in Jehiel et al. (2006), and we ask the reader to consult that paper for details.

For an intuition, assume that the signal spaces are d^i-dimensional cubes $S^i = [0;1]^{d^i}$, $d^i \geq 2$, that there are only 2 agents and 2 alternatives, k and l, assume that $\alpha^i = \alpha^j = 1$ and normalize $v_l^i = v_l^j = 0$. Let $v_k^i(s) = P_k(s) + Q_k^i(s^j)$, and let (ψ, t) be an ex-post incentive compatible mechanism with $t_l^i = t_l^j = 0$ and with smooth transfers t_k^i, t_k^j. By the taxation principle (Proposition 2.1) we have $v_k^i(s) + t_k^i(s^j) = 0 \Leftrightarrow v_k^j(s) + t_k^j(s^i) = 0$. Call signals for which these equalities hold "indifference signals", and denote the set of indifference signals by $I \subset S$. Differentiating the payoff functions, we obtain that $\nabla v_k^i + t_k^i$ and $\nabla v_k^j + t_k^j$ are both perpendicular to the indifference set I.

In other words, the vectors

$$\begin{pmatrix} \nabla_{s^i} P_k\left(s\right) \\ \nabla_{s^j} P_k\left(s\right) + \nabla_{s^j}\left(Q_k^i + t_k^i\right)\left(s^j\right) \end{pmatrix} \text{ and } \begin{pmatrix} \nabla_{s^i} P_k\left(s\right) + \nabla_{s^i}\left(Q_k^j + t_k^j\right)\left(s^i\right) \\ \nabla_{s^j} P_k\left(s\right) \end{pmatrix}$$

must be parallel for all $s \in I$.

Fix now an indifference signal $\widehat{s} \in I$. For multi-dimensional signals $(d^i, d^j \geq 2)$ the above implies that $\nabla_{s^i} P_k\left(s\right)$ is parallel to $\nabla_{s^i}\left(Q_k^j + t_k^j\right)\left(\widehat{s}^i\right)$ for all $s \in I$ with $\widehat{s}^i = s^i$. If the common value part P_k is a "randomly drawn" generic function, the direction of the gradient $\nabla_{s^i} P_k\left(s\right)$ will vary in s^j. The only way by which the term $\nabla_{s^i}\left(Q_k^j + t_k^j\right)\left(s^i\right)$, which cannot depend on s^j, can remain parallel to all these multiple directions is by having $\nabla_{s^i}\left(Q_k^j + t_k^j\right)\left(s^i\right) = 0$. This implies that $\left(Q_k^j + t_k^j\right)\left(s^i\right) = \lambda_k$, is a constant independent of s^i. The assertion follows then by the taxation principle.

The proof is completed by discarding the differentiability (and continuity) assumption on the transfer functions, and by showing that the set of functions $\{P_k\}_k$ for which the gradient $\nabla_{s^i} P_k\left(s\right)$ varies in s^j on every possible indifference set of P_k is residual and finitely prevalent in the Banach space of sufficiently smooth functions (see Jehiel et al., 2006).

Finally, note that the restriction on valuations that admit cardinal potentials is stronger than necessary in order to avoid the negative result of Proposition 4.3. Intuitively, this class of valuations is small within the class of valuations that allow for a non-trivial best-alternative potential: given valuations that admit a cardinal potential, only affine transformations of these valuations preserve the property that a cardinal potential exists, whereas a best-alternative potential will still exist for any monotone transformation; moreover, the set of affine transformations is small in the set of all monotone transformations. Thus, the present results do not imply that for all "reasonable" classes of valuations over multi-dimensional signals, best-alternative potentials and cardinal potentials coincide[17] [18].

4.2.3 Separable Valuations

In this subsection we restrict valuations even further, and we focus on interdependent, yet separable valuation functions of the form

$$v_k^i\left(s\right) = f_k^i\left(s^i\right) + h_k^i\left(s^{-i}\right), \tag{12}$$

---

[17]But, it is not hard to derive from the proofs in Jehiel et al. (2006) that non-trivial expost implementation does impose locally around the indifference set (instead of globally) the "cardinal conditions" on valuation functions.

[18]It also follows from the analysis in Jehiel et al. (2006) that, if there are three of more alternatives, the set of valuations allowing for an ordinal potential is small within the set of valuations allowing for ex-post implementation of an exhaustive choice rule.

for some functions $f_k^i : S^i \to \mathbb{R}$ and $h_k^i : S^{-i} \to \mathbb{R}$. Let $f^i = (f_k^i)_{k \in \mathcal{K}} : S^i \to \mathbb{R}^K$ and $f = (f^i)_{i \in \mathcal{N}} : S \to \mathbb{R}^{NK}$.

Definition 4.5 *A choice rule $\psi : S \to \mathcal{K}$ is said to be an **affine maximizer** if and only if it is of the form:*

$$\psi(s) \in \arg\max_{k \in \mathcal{K}} \left\{ \sum_{i=1}^{N} \alpha^i f_k^i(s^i) + \lambda_k \right\} \tag{13}$$

for agent-specific weights $\alpha^i \geq 0$ and alternative-specific weights $\lambda_k \in \mathbb{R}$.

For a given affine maximizer, the weight α^i can be interpreted as the importance of agent i's information to the social choice, and the weight λ_k as the designer's preference for alternative k. Note also that, for private values, a Vickrey-Clarke-Groves mechanism is simply the affine maximizer with weights $\alpha^i = 1, \lambda_k = 0$. Affine maximizers with weights $\alpha^i > 0$ can be implemented by transfers $t_k^i(s^{-i}) = -h_k^i(s^{-i}) + \sum_{j \neq i} \frac{\alpha^j}{\alpha^i} f_k^j(s^j)$.

Consider now a vector $\alpha \gg 0$. By defining

$$P_k(s) = \sum_{j=1}^{N} \alpha^j f_k^j(s^j) \text{ and } Q_k^i(s^{-i}) = -\sum_{j \neq i} \alpha^j f_k^j(s^j) + \alpha^i h_k^i(s^{-i})$$

we obtain:

$$v_k^i(s) = f_k^i(s^i) + h_k^i(s^{-i}) = \frac{1}{\alpha^i}[P_k(s) + Q_k^i(s^{-i})]$$

and

$$\psi(s) \in \arg\max_{k \in \mathcal{K}} \{P_k(s) + \lambda_k\}$$

for affine maximizers ψ.

Thus, separable valuations admit an $\alpha-$potential for any $\alpha >> 0$, and they constitute a (non-generic) subcase of the semi-separable valuation functions studied in Subsection 4.2.2 where, in contrast, the cardinal representation is generically unique (up to a multiplicative constant). The present non-genericity arises because $\forall i, k, s$, we have $\frac{\partial v_k^i(s)}{\partial s^i \partial s^j} = 0$ in the class of separable valuations, whereas, for semi-separable valuations, $\frac{\partial v_k^i(s)}{\partial s^i \partial s^j}$ is proportional to $\frac{\partial P_k(s)}{\partial s^i \partial s^j}$ which is allowed to vary in an arbitrary way.

Since $\nabla_{s^i} P_k(s)$ does not vary here in s^j, we cannot apply the method of proof illustrated in the above section. Quite surprisingly, by using a remarkable result about dominant strategy implementation for settings with private

16

values and multi-dimensional signals due to Roberts (1979), we are nevertheless able to prove a result similar to Proposition 4.4 (albeit under additional technical conditions): for separable valuations, the only implementable rules are affine maximizers. Thus, for all separable valuations, almost all implementable choice rules are represented by cardinal potentials. Exceptions are choice rules corresponding to affine maximizers where $\alpha^i = 0$ for some i. Such an affine maximizer is not a cardinal potential because Definition 3.1 insists that even an agent i with $\alpha^i = 0$ should have the same cardinal preference as the other agents. This requirement does not show up in the conditions for ex-post implementation since this agent's report can simply be ignored.

The following simple lemma explains why we can use here a result that was obtained for settings with private values:

Lemma 4.6 *A choice rule $\psi : S \to \mathcal{K}$ is ex-post implementable in the interdependent values model if and only if it is ex-post implementable in the associated private values model where $\forall i, k, s^{-i} : h_k^i (s^{-i}) = 0$.*

Proof. See Appendix B. ∎

Our main result in this subsection is:

Proposition 4.7 *Assume that valutions $v_k^i (s) = f_k^i (s^i) + h_k^i (s^{-i})$ are separable, that $f^i (S^i) = \mathbb{R}^K$ [19], and that $K \geq 3$. Then every implementable, exhaustive choice rule is an affine maximizer. Thus, almost all implementable choice rules are represented by cardinal potentials.*

The proof of Proposition 4.7 is based on a hyperplane separation argument due to Roberts (1979) who proved a similar result for dominant strategy implementation with private values, i.e. for $S^i = \mathbb{R}^K$, $f^i = id$, $h^i = 0$. Our proof adapts Roberts' insight by showing that there is no loss of generality in assuming that an ex-post implementable choice rule takes only payoff relevant information into account. This means that ψ *factors* through f, i.e. for $X = \left(\mathbb{R}^K\right)^N$ there exists a function $\phi : X \to \mathcal{K}$ such that $\psi = \phi \circ f$.

The assumptions in Proposition 4.7 cannot be relaxed. If there are only two alternatives (i.e., $K = 2$), a characterization of dominant strategy implementable choice rules in a private values setting has been obtained by Laffont and Maskin (1982). Their characterization generalizes to separable, interdependent valuations, and yields a larger set than the set of affine maximizers. For bounded valuations $f^i (S^i) \subseteq \mathbb{R}^K$, an example of an implementable choice rule that is not an affine maximizer is available from the authors upon request.

---

[19] This implicitly assumes that the dimension of $S^i \subseteq \mathbb{R}^{d^i}$ is at least as large as K.
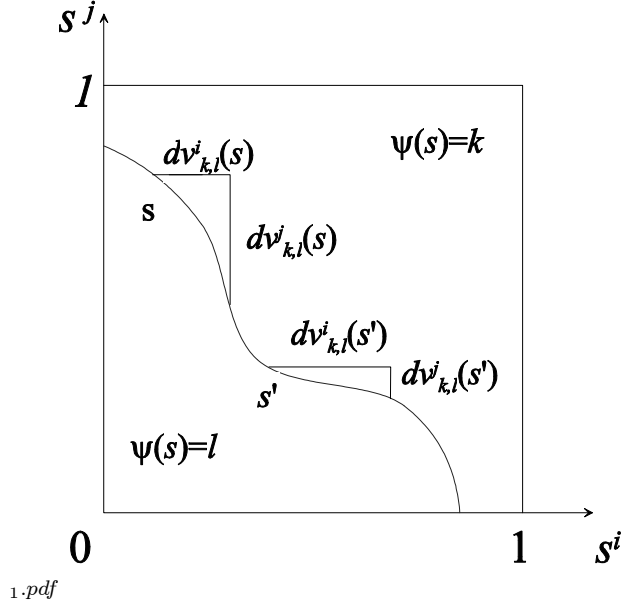
substitution1

Figure 1:

4.2.4 One-dimensional Settings

In one-dimensional settings, the conditions for the existence of cardinal potentials are still stringent and cannot be generically satisfied (see Proposition 4.1). Yet, as illustrated by Dasgupta and Maskin (2000), Jehiel and Moldovanu (2001), Bergemann and Välimäki (2002), Perry and Reny (2002), Chung and Ely (2001), and Jehiel et al. (2006), ex-post implementation of non-trivial choice rules can be achieved for open sets of valuation functions. Thus, in one-dimensional settings, the set of best-alternative potentials does generally not coincide with the set of cardinal potentials. Besides, even in situations in which a cradinal potential exists, many choice rules other than cardinal potentials can be implemented in one-dimensional settings.

To illustrate this phenomenon, consider a setting with two agents i, j with signals $s^i, s^j \in [0, 1]$, two alternatives k, l and private valuations $v_k^i = s^i, v_l^i = 0$ and $v_k^j = s^j, v_l^j = 0$. Clearly, total welfare is a cardinal potential in such a private values setting. Yet, many other choice rules that do not admit a cardinal representation can be implemented in this case. Indeed, standard arguments show that a choice rule ψ is implementable if and only if it is monotone in the sense that it satisfies $\psi(s) = k \Rightarrow \psi(\widetilde{s}) = k$ for any $s = (s^i, s^j)$ and $\widetilde{s} = (\widetilde{s}^i, \widetilde{s}^j)$ with $s^i \le \widetilde{s}^i$ and $s^j \le \widetilde{s}^j$. Figure 1 shows that,

18

for a general implementable ψ, the rate of information substitution $\frac{dv_{k,l}^i}{dv_{k,l}^j}(s)$ is not constant, and thus ψ cannot be represented by a cardinal potential.

5   Conclusion

Using ideas from the theory of potential games, we have introduced to mechanism design new notions of alignment between individual and social preferences, and we related these notions to ex-post implementation. We have characterized valuation functions that admit cardinal potentials. Since the required conditions are very easily checked, our results offer simple methods to construct the potentials and the corresponding transfers needed to implement them. Moreover, for settings with multi-dimensional type spaces that admit a cardinal potential and that are generic (within the class of such settings), we have established that almost all implementable choice rules maximize this potential or its translates. The class of separable valuations constitutes a non-generic sub-class, and the extra freedom leads there to the possibility of implementing the larger class of affine maximizers. Virtually all affine maximizers are cardinal potentials and we show that beyond them no additional choice rules are implementable.

To conclude, the focus on cardinal potentials and on the derived possibility results for ex-post implementation constitute, in our view, a significant complement to the earlier impossibility result for generic valuations with multi-dimensional signals (see Jehiel, Meyer-ter-Vehn, Moldovanu and Zame, 2006).

6   Appendix A: Potential Games

We briefly review the original definitions of potentials for normal form, complete information games due to Monderer and Shapley (1996). Let $\Gamma = \Gamma(u^1, u^2, ..., u^N)$ be a game in strategic form played by the agents in a finite set $\mathcal{N}$. The strategy set of player i is denoted by Y^i, and the payoff function of i is $u^i : Y \rightarrow \mathbb{R}$, where $Y = Y^1 \times Y^2 \times ...Y^N$.

Definition 6.1   *1. Given agent specific weights $\alpha = (\alpha^i)_{i \in \mathcal{N}} \gg 0$, a function $P : Y \rightarrow \mathbb{R}$ is an α-**potential** for Γ if for every $i \in \mathcal{N}$ and for every $y^{-i} \in Y^{-i}$*

$$\alpha^i(u^i(y^{-i}, x) - u^i(y^{-i}, z)) = (P(y^{-i}, x) - P(y^{-i}, z)) \qquad (14)$$

*for every $x, z \in Y^i$. If we are not interested in the value of α we simply speak of a **cardinal potential**.*

2. *A function* $P : Y \to \mathbb{R}$ *is an **ordinal potential** for* Γ *if for every* $i \in \mathcal{N}$ *and for every* $y^{-i} \in Y^{-i}$

$$u^i(y^{-i}, x) - u^i(y^{-i}, z) > 0 \Leftrightarrow P(y^{-i}, x) - P(y^{-i}, z) > 0 \qquad (15)$$

for every $x, z \in Y^i$.

An analogue for best-alternative potentials was defined by Dubey et al. (2006):

Definition 6.2 *A function* $P : Y \to \mathbb{R}$ *is a **pseudo-potential** for* Γ *if for every* $i \in \mathcal{N}$ *and for every* $y^{-i} \in Y^{-i}$

$$\arg\max_{x \in Y^i} \left\{ u^i \left(y^{-i}, x \right) \right\} \supseteq \arg\max_{x \in Y^i} \left\{ P \left(y^{-i}, x \right) \right\}. \qquad (16)$$

The following result shows how potentials align the agents' interests:

Proposition 6.3 (Monderer and Shapley (1996)) *Let* P *be an ordinal potential for* $\Gamma(u^1, u^2, ..., u^n)$. *Then a strategy profile is a Nash equilibrium of* $\Gamma(u^1, u^2, ..., u^n)$ *if and only if it is a Nash equilibrium of* $\Gamma(P, P, ..., P)$.

Proof. By definition, $\Gamma(P, P, ..., P)$ has the same best-response correspondence as $\Gamma(u^1, u^2, ..., u^n)$, which immediately implies the result. ∎

6.1 Potentials and Potential Games

From the above definitions it seems a-priori plausible to assume that a potential $P = (P_k)_{k \in \mathcal{K}}$ for valuations v gives rise to a potential for the revelation game that implements an associated choice rule $\psi \in \Xi \left(\overline{P} \right)$. In fact, Sandholm (2005) shows how an externality abatement mechanism gives rise to a potential game.

Example 6.4 *Consider a choice rule* $\psi = S \to \mathcal{K}$ *and valuations of the form* $v_k^i (s) = w_k^i + p_k^i (s^i)$ *with the property that* $p_k^i (s^i)$ *depends on the alternative* $k = \psi (s)$ *only through agent* i's *reported signal* $\widetilde{s}^i$, *i.e.* $p_{\psi(\widetilde{s}^i, s^{-i})}^i (s^i) = p_{\psi(\widetilde{s}^i, \widetilde{s}^{-i})}^i (s^i)$ *for all* $s^{-i}, \widetilde{s}^{-i} \in S^{-i}$. *Then, given true signals* s, *the revelation game* $(\psi, t) : S \to \mathcal{K} \times \mathbb{R}^N$ *with transfers* $t^i (\widetilde{s}) = \sum_{j \neq i} w_{\psi(\widetilde{s})}^j$ *admits the potential* $P_s (\widetilde{s}) = \sum_j \left(w_{\psi(\widetilde{s})}^j + p_{\psi(\widetilde{s})}^j (s^j) \right)$.

Proof. It is easy to check that:

$$
\begin{aligned}
& u^i\left(\widetilde{s}^i, \widetilde{s}^{-i}\right) - u^i\left(\widetilde{s}'^i, \widetilde{s}^{-i}\right) \\
=\; & \left[\sum_j w^j_{\psi(\widetilde{s}^i, \widetilde{s}^{-i})} + p^i_{\psi(\widetilde{s}^i, \widetilde{s}^{-i})}\left(s^i\right)\right] - \left[\sum_j w^j_{\psi(\widetilde{s}'^i, \widetilde{s}^{-i})} + p^i_{\psi(\widetilde{s}'^i, \widetilde{s}^{-i})}\left(s^j\right)\right] \\
=\; & \left[\left(P_s\left(\widetilde{s}^i, \widetilde{s}^{-i}\right) - \sum_{j\neq i} p^j_{\psi\left(\widetilde{s}^j, \widetilde{s}^i, \widetilde{s}^{-\{i,j\}}\right)}\left(s^j\right)\right)\right] \\
& -\left[P_s\left(\widetilde{s}'^i, \widetilde{s}^{-i}\right) - \sum_{j\neq i} p^j_{\psi\left(\widetilde{s}^j, \widetilde{s}'^i, \widetilde{s}^{-\{i,j\}}\right)}\left(s^j\right)\right] \\
=\; & P_s\left(\widetilde{s}^i, \widetilde{s}^{-i}\right) - P_s\left(\widetilde{s}'^i, \widetilde{s}^{-i}\right)
\end{aligned}
$$

∎

The above example relies on the special circumstances that valuations are private and that the true information s^i and the stated information of the other agents $\widetilde{s}^{-i}$ enter the valuation function v^i in an additive separable way.

In general there is no reason to believe that revelation games are potential games. Ex-post incentive compatibility only requires truth-telling to be a best response to truthful revelation of other agents. It does not pose restrictions on the preference ordering of suboptimal announcements, and it is silent on best responses off the equilibrium path. To illustrate this phenomenon, we offer below a simple example of a second price auction for which we construct an improvement cycle in sub-optimal announcements, thereby proving that the auction is not a potential game.

Example 6.5 *The second-price auction with two bidders i, j, and private valuations v^i, v^j is not a potential game.*

Proof. The induced bidding game allows for the following improvement cycle. Let $\underline{b} = \left(\underline{b}^i, \underline{b}^j\right)$ and $\overline{b} = \left(\overline{b}^i, \overline{b}^j\right)$ be such that $\underline{b}^i < \underline{b}^j < \overline{b}^i < \overline{b}^j < v^i, v^j$ and consider the cycle $\left(\underline{b}^i, \underline{b}^j\right) \to \left(\overline{b}^i, \underline{b}^j\right) \to \left(\overline{b}^i, \overline{b}^j\right) \to \left(\underline{b}^i, \overline{b}^j\right) \to \left(\underline{b}^i, \underline{b}^j\right)$. Whereas the first change $\left(\underline{b}^i, \underline{b}^j\right) \to \left(\overline{b}^i, \underline{b}^j\right)$ makes agent i better off by $v^i - \underline{b}^j > 0$, the second change $\left(\overline{b}^i, \underline{b}^j\right) \to \left(\overline{b}^i, \overline{b}^j\right)$ makes agent j better off by $v^j - \overline{b}^j > 0$ and the last two changes leave the agents indifferent. It easily follows from condition (15) that an ordinal potential game does not allow for such improvement cycles and that this auction is therefore not a potential game. ∎

It is possible to amend the above example to construct a strong improvement cycle, i.e. a sequence of signal changes with the property that each

change makes the relevant agent strictly better off. Thus the revelation game does not even posses a generalized ordinal potential[20].

7 Appendix B: Proofs

Proof of Lemma 4.6. **"if":** Let ψ be ex-post implemented by a mechanism (ψ, t) under private values. Thus

$$f^i_{\psi(s)}\left(s^i\right) + t^i\left(s\right) = \max_{\widetilde{s}^i \in S^i}[f^i_{\psi(\widetilde{s}^i, s^{-i})}\left(s^i\right) + t^i\left(\widetilde{s}^i, s^{-i}\right)]$$

for all i, s^i, s^{-i}. For interdependent utilities of the form $f^i_k\left(s^i\right) + h^i_k\left(s^{-i}\right) + t^i$ we define monetary payments by $t'^i\left(s\right) = t^i\left(s\right) - h^i_{\psi(s)}\left(s^{-i}\right)$. One easily verifies that

$$f^i_{\psi(s)}\left(s^i\right) + h^i_{\psi(s)}\left(s^{-i}\right) + t'^i\left(s\right) = \max_{\widetilde{s}^i \in S^i}[f^i_{\psi(\widetilde{s}^i, s^{-i})}\left(s^i\right) + h^i_{\psi(\widetilde{s}^i, s^{-i})}\left(s^{-i}\right) + t'^i\left(\widetilde{s}^i, s^{-i}\right)]$$

for all i, s^i, s^{-i}. This shows that (ψ, t') ex-post implements ψ for the interdependent values case.

"only if": Analogously. ∎

The following result is a crucial ingredient for the proof of Proposition 4.7. It is also interesting in its own right, as it establishes a monotonicity property of implementable choice rules.

Lemma 7.1 (Monotonicity) *An implementable choice rule ψ is monotonic in the following sense: For every agent i, and for all signals $s = (s^i; s^{-i})$, $s' = (s'^i; s^{-i}) \in S$, such that $v^i_k\left(s'\right) - v^i_k\left(s\right) > v^i_l\left(s'\right) - v^i_l\left(s\right)$, $\psi\left(s\right) = k$ implies that $\psi\left(s'\right) \neq l$.*

Proof of Lemma 7.1. By the taxation principle, there are transfers $t^i_l\left(s^{-i}\right)$ such that $\psi\left(s\right) \in \arg\max_{k'}\left\{v^i_{k'}\left(s\right) + t^i_{k'}\left(s^{-i}\right)\right\}$ for all s. If alternative k is among i's favorite alternatives at signal s, we have $k \in \arg\max_{k'}\left\{v^i_{k'}\left(s\right) + t^i_{k'}\left(s^{-i}\right)\right\}$. If the change from s^i to s'^i makes alternative k strongly more preferable (for i) than l, $v^i_k\left(s'\right) - v^i_k\left(s\right) > v^i_l\left(s'\right) - v^i_l\left(s\right)$, it is immediate that l can not be preferred at signal s'^i. Thus, $l \notin \arg\max_{k'}\left\{v^i_{k'}\left(s'\right) + t^i_{k'}\left(s^{-i}\right)\right\}$. By the taxation principle, we can conclude that $\psi\left(s'\right) \neq l$. ∎

Proof of Proposition 4.7. We use an important result due to Roberts (1979) who studied deterministic choice rules that are implementable in dominant strategies in a private values setting. Roberts showed that such rules

---

[20]P is a generalized ordinal potential if it satisfies condition 15 with the " $\Longleftrightarrow$ " replaced by a " $\Rightarrow$ ".

must satisfy a monotonicity condition, called PAD. Using our notation, his proof relies on the following technical result[21]:

Theorem A (Roberts 1979): Let $X = \left(\mathbb{R}^K\right)^N$ and assume that $K \geq 3$. Then any function $\phi : X \to \mathcal{K}$ which satisfies PAD is an affine maximizer.

Here PAD means that for $x, x' \in X$ such that $x_k'^i - x_k^i > x_l'^i - x_l^i$ for all $i \in \mathcal{N}$ and all $l \neq k \in \mathcal{K}$, $\phi(x) = k$ implies $\phi(x') = k$. An affine maximizer ϕ satisfies $\phi(x) \in \arg\max_{k \in \mathcal{K}} \left\{ \sum_{j=1}^{N} x_k^j + \lambda_k \right\}$ for some $\{\alpha^j\}_{j \in \mathcal{N}}$ and some $\{\lambda_k\}_{k \in \mathcal{K}}$

Thanks to Lemma 4.6 we can assume that $h_k^i \equiv 0$ for all i, k. In order to apply Theorem A, assume first that $\psi : S \to \mathcal{K}$ factors through f, i.e. there is a function $\phi : X \to \mathcal{K}$ such that $\psi = \phi \circ f$. As ψ is implementable, we can recursively apply Lemma 7.1 to show that for all signals $s, s' \in S$ such that

$$f_k^i(s') - f_k^i(s) > f_l^i(s') - f_l^i(s) \text{ for all } i \in \mathcal{N} \text{ and } l \neq k \in \mathcal{K},$$

$\psi(s) = k$ implies $\psi(s') = k$. Consider the sequence of signals $s_{(0)} := s$, $s_{(i)} := \left(s_i', s_{(i-1)}\right)$ for all agents $i \leq n$ (this gives $s_{(n)} = s'$). The proof of Lemma 7.1 serves then as the induction step proving that, with $\psi\left(s_{(0)}\right) = k$, we have $\psi\left(s_{(i)}\right) = k$ for all i. This yields $\psi(s') = k$. Thus, we can apply Theorem A to ϕ as a function of the $f_k^i(s)$, and get $\psi(s) = \phi(f(s)) = \arg\max_{k \in \mathcal{K}} \left\{ \sum_{j=1}^{N} \alpha^j f_k^j(s^j) + \lambda_k \right\}$.

It remains to show that the above assertion holds also for the cases where $\psi : S \to \mathcal{K}$ does not factor through f. This proof can be broken down into three steps: **a)** Slightly change ψ to a function $\widetilde{\psi}$ that factors $\widetilde{\psi} = \phi \circ f$; **b.)** Show that $\widetilde{\psi}$ is ex-post implementable, and apply Theorem A to ϕ to show that $\widetilde{\psi}$ is an affine maximizer; **c)** Show that ψ is an affine maximizer if $\widetilde{\psi}$ is one.

a) Given functions $f^i = \left(f_k^i\right)_{k \in \mathcal{K}} : S^i \to \mathbb{R}^K$, denote $f^i(s^i) =: x^i$ and for each x^i fix $\widetilde{s}^i \in (f^i)^{-1}\{x^i\}$. We shall say that $\widetilde{s} = (\widetilde{s}^i)_{i \in \mathcal{N}}$ represents s. Given an ex-post implementable choice rule $\psi : S \to \mathcal{K}$ define $\widetilde{\psi} : S \to \mathcal{K}$ by setting

$$\widetilde{\psi}(s) := \psi(\widetilde{s})$$

where $\widetilde{s}$ represents s. Obviously, there is a function $\phi : \left(\mathbb{R}^K\right)^N \to \mathcal{K}$ such that $\widetilde{\psi} = \phi \circ f$.

b) The choice rule $\widetilde{\psi}$ is ex-post implementable by the transfer rule $\widetilde{t}(s) := t(\widetilde{s})$, where $t : S \to \mathbb{R}^N$ are the transfers that implement ψ. Indeed, we

---

[21]Roberts' proof uses a hyperplane-separation argument which yields the weights in the affine representation.

readily check agent i's incentive constraint:

$$f^i_{\widetilde{\psi}(s^i,s^{-i})}\left(s^i\right) + \widetilde{t}^i\left(s^i,s^{-i}\right) = f^i_{\psi(\widetilde{s}^i,\widetilde{s}^{-i})}\left(s^i\right) + t^i\left(\widetilde{s}^i,\widetilde{s}^{-i}\right) \geq$$

$$f^i_{\psi(\widetilde{s}'^i,\widetilde{s}^{-i})}\left(s^i\right) + t^i\left(\widetilde{s}'^i,\widetilde{s}^{-i}\right) = f^i_{\widetilde{\psi}(s'^i,s^{-i})}\left(s^i\right) + \widetilde{t}^i\left(s'^i,s^{-i}\right)$$

for all i, s^i, s'^i, s^{-i}. The first and third equality follow by the definitions of $\widetilde{\psi}$ and $\widetilde{t}$, and the inequality follows by the ex-post incentive compatibility of (ψ, t). By Lemma 7.1, $\widetilde{\psi}$ satisfies monotonicity, which in turn means that $\phi : \left(\mathbb{R}^K\right)^N \to \mathcal{K}$ satisfies PAD in the sense of Roberts' Theorem A. Thus, there are constants $\alpha^j \geq 0$ for $j \in \mathcal{N}$ and λ_k for $k \in \mathcal{K}$ such that $\phi\left(f\left(s\right)\right) \in \arg\max_{k \in \mathcal{K}}\left\{\sum_{j=1}^N \alpha^j f^j_k\left(s^j\right) + \lambda_k\right\}$ for all $s \in S$. This proves that $\widetilde{\psi}$ is an affine maximizer.

c) We now return to the original choice rule ψ. We will derive a contradiction by assuming that there exists $s \in S$ such that $\psi\left(s\right) = l \notin \arg\max_{k \in \mathcal{K}}$ $\left\{\sum_{j=1}^N \alpha^j f^j_k\left(s^j\right) + \lambda_k\right\}$. Consider $s' \in S$ such that $f^j_l\left(s'\right) = f^j_l\left(s\right) + \varepsilon$ for all j, $f^j_k\left(s'\right) = f^j_k\left(s\right)$ for all j and all $k \neq l$, where ε is sufficiently small so that $l \neq \widetilde{\psi}\left(s'\right) \in \arg\max_{k \in \mathcal{K}}\left\{\sum_{j=1}^N \alpha^j f^j_k\left(s'^j\right) + \lambda_k\right\}$. Let $\widetilde{s}'$ be the element representing s' in the definition of $\widetilde{\psi}$. By monotonicity, $\psi\left(s\right) = l$ implies $\psi\left(\widetilde{s}'\right) = l$, but, by the characterization of $\widetilde{\psi}$, we know that $l \neq \widetilde{\psi}\left(\widetilde{s}'\right)$ contradicting $\widetilde{\psi}\left(\widetilde{s}'\right) = \psi\left(\widetilde{s}'\right)$. This contradiction concludes the proof that $\psi\left(s\right) \in \arg\max_{k \in \mathcal{K}}\left\{\sum_{j=1}^N \alpha^j f^j_k\left(s^j\right) + \lambda_k\right\}$ for all $s \in S$. $\blacksquare$

References

[1] Bergemann, D. and S. Morris (2005), "Robust Mechanism Design," *Econometrica* **73**, 1771-1813.

[2] Bergemann, D. and J. Välimäki (2002): "Information Acquisition and Efficient Mechanism Design", *Econometrica* **70**, 1007-10035.

[3] Bikhchandani, S., Chatterji, S., Lavi, R., Mu'alem, A., Nisan, N., and Sen, A. (2006): "Weak Monotonicity Characterizes Deterministic Dominant-Strategy Implementation", *Econometrica* **74**, 1109-1132..

[4] Chung, K-S. and J. Ely (2003): "Ex-post Incentive Compatible Mechanism Design", discussion paper, Northwestern University.

[5] Clarke E. (1971): "Multipart Pricing of Public Goods", *Public Choice* **8**, 19-33.

[6] Dasgupta, P., and E. Maskin (2000): "Efficient Auctions", *Quarterly Journal of Economics* **115**, 341-388.

[7] Dubey, P., Haimanko, O., Zapechelnyuk, A. (2006): "Strategic Complements and Substitutes, and Potential Games", *Games and Economic Behavior* **54**, 77-94.

[8] Facchini, G., Van Meegen F., Borm, P. and S. Tijs (1997) "Congestion Models and Weighted Bayesian Potential Games", *Theory and Decision* **42**, 193-206.

[9] Groves T. (1973): "Incentives in Teams", *Econometrica* **41**, 617-631.

[10] Gui, H., Mueller, R. and Vohra, R. (2004): "Dominant Strategy Mechanisms with Multi-dimensional Types", discussion paper, Northwestern University.

[11] Jehiel, P., Moldovanu, B. and E. Stacchetti (1999) "Multi-dimensional Mechanism Design for Auctions with Externalities", *Journal of Economic Theory* **85**, 814-829.

[12] Jehiel, P., and B. Moldovanu (2001) "Efficient Design with Interdependent Valuations", *Econometrica* **69**, 1237-1259.

[13] Jehiel, P., Meyer-ter-Vehn, M., Moldovanu, B., and W. Zame (2006) "The Limits of Ex-post Implementation", *Econometrica* **74**, 585-611.

[14] Laffont, J.J. and E. Maskin (1982): "The Theory of Incentives: An Overview" in *Advances in Economic Theory*, ed. Hildenbrand W., Cambridge University Press 31-94.

[15] Lavi, R., Mu'alem, A. and Nisan, N. (2006): "Two Simplified Proofs for Roberts' Theorem", discussion paper, Hebrew University of Jerusalem

[16] Meyer-ter-Vehn, M. and Moldovanu, B. (2002): "Ex-Post Implementation with Interdependent Valuations", discussion paper, University of Bonn.

[17] Monderer D., and L. Shapley (1996): "Potential Games", *Games and Economic Behavior* **14**, 124-143.

[18] Perry, M. and P. Reny (2002): "An Efficient Auction", *Econometrica* **70**, 1199-1213.

[19] Roberts, K. (1979): "The Characterization of Implementable Choice Rules" in *Aggregation and Revelation of Preferences*, ed. Laffont J.J., North Holland.

[20] Rochet, JC. (1987): "A Condition for Rationalizability in a Quasi-Linear Context", *Journal of Mathematical Economics* **16**, 191-200.

[21] Sandholm, W. (2005): "Negative Externalities and Evolutionary Implementation" , *Review of Economic Studies* **72** , 885-915.

[22] Slade, M. (1994): "What Does an Oligopoly Maximize ?", *Journal of Industrial Economics* **17**, 45-61.

[23] Ui, T. (2000): "A Shapley Value Representation of Potential Games", *Games and Economic Behavior* **31**, 121-135.

[24] Vickrey, W. (1961): "Counterspeculation, Auctions and Competitive Sealed Tenders", *Journal of Finance* **16**, 8-37.