Abstract

We introduce the concept of potentials (formally analogous to Monderer and Shapley (1996)) for mechanism design problems with interdependent valuations, and relate potentials to implementation in ex-post equilibria. We go on to show that ordinal and cardinal potential coincide in settings with separable valuation functions.

1 Introduction

Aligning the interests of several 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) 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.

More generally, we say that 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 maximiz-
ing a single function, common to all agents. In this paper we introduce several notions of potentials for mechanism design problems with interdependent values, discuss their properties, and establish relations between these notions and implementability in ex-post equilibria. Ex-post implementation requires that, even after learning the information of others, agents are not willing to change their strategy. This notion has recently received a lot of attention because it ensures that agents need not know the distribution from which others’ signals are drawn in order to play.

The present notion of interest alignment bears a strong formal resemblance to the definition of potentials for normal form games, due to 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 in 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 Monderer and Shapley, we shall distinguish between ordinal and cardinal potentials. The former, weaker, concept roughly says that the potential function and each agent’s payoff function agree on the best alternative, whereas the latter stronger concept says that the quantitative differences between alternatives are equal for the potential function and for each agent’s payoff function.

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 in the

---

1The 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).


3See Bergemann and Morris (2002) for a formal treatment of this issue, and for the connection to “Wilson’s doctrine” about detail-free mechanisms.
potential function for a mechanism design problem. Also it is not true that a mechanism that admits a potential gives rise to a potential game\textsuperscript{4}.

Whereas in Monderer and Shapley’s general setting a potential need not have an intuitive ”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. Thus, it becomes possible to deduce results about implementation by studying the properties of potential functions, and vice-versa. In particular the important question ” Which choice rules are implementable in a given mechanism design setting?” can be translated into the language of potential functions.

We first show that ex-post implementability is related to the existence of a corresponding ordinal potential. While cardinal potentials are easily characterized via properties of second differences of valuation functions, ordinal potentials are much more fickle. Nevertheless, there are environments where every ordinal potential must correspond to a cardinal potential. For example, in a private values setting with multi-dimensional signals, this follows from a beautiful result by Roberts (1979) who has shown that all deterministic, dominant strategy, implementable social choice rules maximize affine functions of the agents’ signals. These maximizers are readily seen to be (weighted) cardinal potentials.

We next extend Roberts’ result to a setting with interdependent values (where the impact of an agent’s information on his own valuation can be separated from the impact of information which is private to others\textsuperscript{5}), and apply it to the study of potential in this setting.

The paper is organized as follows: In Section 2 we describe the mechanism design problem with interdependent values, and we define several notions of potentials for this setting. The focus of this section is on establishing links between potentials and implementation in ex-post equilibria. In Section 3 we consider settings with separable valuations functions, and show that cardinal and ordinal potentials coincide in this case. There are two appendices: Appendix A recalls the original definitions of potential games and the main

\textsuperscript{4}In contrast, Sandholm (2004) 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, where a dynamic learning process leads to an efficient outcome.

\textsuperscript{5}Note that in this setting the efficient allocation cannot be generically implemented. This holds even for the weaker concept of Bayes-Nash implementation, as shown by Jehiel and Moldovanu (2001). With one-dimensional signals efficient implementation is possible. A good survey of the recent literature on efficient auctions with interdependent valuations is Maskin (2001).
result of Monderer and Shapley (1996). Appendix B contains several proofs that would interrupt the flow of argument in the main text.

2 Potentials for Mechanism Design

Consider a setting with $N \in \mathbb{N}$ agents, who are affected by a decision among $K \in \mathbb{N}$ alternatives, $k \in \mathcal{K}$. Agent $i$'s utility $u^i = v_k^i + t^i$ 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 $k$, $v_k^i = v_k^i(s)$, depends on the state of the world $s \in S$.

Each agent gets a private signal $s^i \in S^i$ 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 combinations$^6$: $S = \prod_{i \in N} S^i$. We adopt the usual notation $s^{-i} = (s^j)_{j \in N, j \neq i}$ and $s = (s^i, s^{-i})$ when we focus on agent $i$.

We study choice rules $P$ that would interrupt the result of Monderer and Shapley (1996). Appendix B contains several proofs that would interrupt the flow of argument in the main text.

We now introduce potentials, the main concept of this paper:

Definition 2.1 1. A family $(P_k)_{k \in \mathcal{K}}$ of functions $P_k : S \rightarrow \mathbb{R}$ is an ordinal potential for valuations $v$ if there are payments $t^i_k(s^{-i}) \in \mathbb{R} \cup \{-\infty\}$ such that:

$$v^i_{\psi(s)}(s) + t^i_k(s^{-i}) \geq v^i_{\psi(s^i, s^{-i})}(s) + t^i_k(s^i, s^{-i})$$

(1)

for all $s^i, s^i, s^{-i} \in S^i$ and $s^{-i} \in S^{-i}$, where $s := (s^i, s^{-i})$. We shall call such $\psi$ implementable. Also, we call $\psi$ exhaustive if every alternative $k$ is chosen at some state $s$, and trivial if it chooses the same alternative $k$ in all states $s$.

We now introduce potentials, the main concept of this paper:

Definition 2.1 1. A family $(P_k)_{k \in \mathcal{K}}$ of functions $P_k : S \rightarrow \mathbb{R}$ is an ordinal potential for valuations $v$ if there are payments $t^i_k(s^{-i}) \in \mathbb{R} \cup \{-\infty\}$ such that:

$$\bigcap_i \arg \max_k \{v_k^i(s) + t^i_k(s^{-i})\} \cap \arg \max_k \{P_k(s)\} \neq \emptyset.$$  

(2)

2. Given agent specific weights $\alpha = (\alpha^i)_{i \in N} \geq 0$, a family $(P_k)_{k \in \mathcal{K}}$ of functions $P_k : S \rightarrow \mathbb{R}$ is an $\alpha$-potential for $v$ if there are payments $t^i_k(s^{-i}) \in \mathbb{R} \cup \{-\infty\}$ such that:

$$\forall i, k, s, \alpha^i v^i_k(s) + t^i_k(s^{-i}) = P_k(s)$$

(3)

$^6$There is no loss of generality in assuming that the agents’ joint information $(s^i)_{i \in N}$ completely determines the state of the world $s$.

$^7$To be precise, transfers $(t^i_k(s^{-i}))_{k \in (\mathbb{R} \cup \{-\infty\}) \setminus (-\infty, ..., -\infty)}$
We shall call any $\alpha$-potential a **cardinal potential** when we are not interested in the value of $\alpha$. If $\alpha \equiv 1$, we speak of an **exact cardinal potential**.

Remark 2.2 Some difficulties arise when ties have to be broken. In the generic case where the argmax sets are singletons, equation 2 boils down to

$$\arg \max_k \{ v_i^k (s) + t_i^k (s^{-i}) \} = \arg \max_k \{ P_k (s) \}$$

for each agent $i$.

These definitions closely parallel the ones of potential games, displayed in Appendix A. Whereas agents’ payoffs actually coincide in the presence of an (exact) cardinal potential, they only need to agree on a most preferred alternative in the presence of an ordinal potential. A cardinal potential for valuations $v$ is therefore a fortiori an ordinal potential.

The reason why the payments $t_i^k$ in the definition of potentials are only allowed to depend on $s^{-i}$ (and not on $s^i$) will become clear in the next section.

### 2.1 Potentials and Implementation

A potential represents preferences over alternatives, and thereby represents a set of choice rules. Let us formalize this connection: Let $S$ be the set of all choice rules $\psi : S \rightarrow K$, and let $P$ be the set of all functions $P = (P_k)_{k \in K} : S \rightarrow \mathbb{R}^K$. Define a function $\Phi : S \rightarrow P$ and a relation $\Xi : P \rightarrow S$ as follows:

1. The potential $\Phi (\psi)$ is given by

$$\Phi (\psi) (s) := \begin{cases} 1 & \text{if } \psi (s) = k \\ 0 & \text{else.} \end{cases}$$

2. The set of choice rules $\Xi (P)$ is the set of all $\psi \in S$ with

$$\psi (s) \in \arg \max_{k \in K} P_k (s) \text{ for all } s.$$ We say that a choice rule $\psi$ is represented by a potential $P$ if $\psi \in \Xi (P)$.

$\Xi$ is a left inverse of $\Phi$, i.e.: $\Xi (\Phi (\psi)) \equiv \{ \psi \}$. We cannot expect the converse to be true since the choice rules $\Xi (P)$ only contain the information about which of $K$ numbers is the greatest for any given $s$. The actual values of $P$ cannot be recovered from this information alone. We shall call two potentials $P, P'$ **equivalent**, if they represent the same choice rule, in the sense that $\Xi (P) \cap \Xi (P') \neq \emptyset$. Also, we shall call a potential $P$ **trivial** (resp. **exhaustive**), if $\Xi (P)$ contains a trivial (resp. exhaustive) choice rule.

---

8For $\alpha^i > 0$ this is achieved by dividing the payment from the definition of the cardinal potential $t_i^k (s^{-i})$ by $\alpha^i$. If $\alpha^i = 0$ for all $i$ in some subset of "irrelevant" agents $I \subseteq N$, choose some arbitrary alternative $k(s^{-I}) \in \arg \max_k \{ P_k (s^{-I}) \}$ for every signal combination of relevant agents $s^{-I}$ (by definition $P_k$ only is independent of $s^I$) and set $t_i^k (s^{-i}) = \begin{cases} 0 & \text{if } k = k(s^{-I}) \\ -\infty & \text{else.} \end{cases}$
A close link between potentials and implementable choice rules will be established below. We first establish a "Taxation principle" for ex-post implementation (this is a multi-agent generalization of the well-know idea from 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 equilibrium the agents agree on the best alternative. In our setting with multiple agents, these prices are personalized and depend on the signals of the other agents about implementable choice rules:

Lemma 2.3 (Taxation Principle) 9 A choice rule \( \psi \) is implementable, if and only if for every agent \( i \), every alternative \( k \) and others’ signals \( s^{-i} \), there are transfers \( t^i_k (s^{-i}) \in \mathbb{R} \cup \{-\infty\} \), such that

\[
\psi (s) \in \arg \max \{ v^i_k (s) + t^i_k (s^{-i}) \} \quad \text{for all agents } i. 
\]  

(4)

Proof. "if": Given \( t^i_k (s^{-i}) \), such that equation (4) holds, define \( t^i (s) := t^i_{\psi(s)} (s^{-i}) \). Agent \( i \) problem in the game induced by \( (\psi, t) \) is to maximize \( v^i_{\psi(s^{-i})} (s^i, s^{-i}) + t^i_{\psi(s^{-i})} (s^{-i}) \) by choice of \( s^i \). By equation (4), it is optimal for her to represent truthfully \( s^i = s^i \) and let the choice rule \( \psi \) pick her most preferred alternative.

"only if": Conversely let \( (\psi, t) \) be an ex-post mechanism. We define

\[
t^i_k (s^{-i}) := \begin{cases} 
  t^i (s^i, s^{-i}) & \text{if } \psi (s^i, s^{-i}) = k \\
  -\infty & \text{if } \psi (s^i, s^{-i}) \neq k \text{ for all } s^i \in S^i.
\end{cases}
\]  

(5)

Note that \( t^i_k (s^{-i}) \) is well-defined as by \( i \)'s incentive constraint

\[
\psi (s^i, s^{-i}) = \psi (s^n, s^{-i}) = k \text{ implies } t^i (s^i, s^{-i}) = t^i (s^n, s^{-i}).
\]

By \( i \)'s incentive constraint again, she will always report in a way such as to maximize her payoff. Thus, with \( t^i_k (s^{-i}) \) as defined in (5), Condition (4) is satisfied.

Proposition 2.4 \( P = (P_k)_{k \in K} \) is an ordinal potential for valuations \( v \) if and only if \( \Xi (P) \) contains an implementable choice rule.

Proof. If \( P \) is an ordinal potential, there is, by equation 2, a choice rule \( \psi \in \Xi (P) \) with \( \psi (s) \in \arg \max \{ v^i_k (s) + t^i_k (s^{-i}) \} \) for all agents \( i \). The Proposition follows from Lemma 2.3.

---

9The taxation principle for ex-post implementation has been first pointed out by Chung and Ely (2001).
Conversely, if some $\psi \in \Xi(P)$ is implementable, we have

$$\psi_i(s) \in \arg\max\{v_k^i(s) + t_k^i(s^{-i})\} \text{ for all agents } i \text{ by Lemma 2.3 and }$$

$$\psi(s) \in \arg\max_{k \in K} P_k(s) \text{ by the equation } 2.$$  

This proves that $\bigcap_i \arg\max_k \{v_k^i(s) + t_k^i(s^{-i})\} \cap \arg\max_k \{P_k(s)\} \neq \emptyset$ and we are done. □

Recall here Remark 2.2 about tie-breaking: In the generic case where the argmax sets in Definition 2.1 are singletons, the above defined correspondence $\Xi$ becomes a function, and Proposition 2.4 states that there is a one-to-one correspondence among ordinal potentials and implementable choice rules.

### 2.2 Properties of Potentials

We next display necessary and sufficient conditions on the valuations $v$ for cardinal potentials to exist. Denote by $\partial(s^i,\bar{s}^i)\partial(s^j,\bar{s}^j)v_k^j(s^{-i,j})$ the second difference of $v_k^j$:

$$\partial(s^i,\bar{s}^i)\partial(s^j,\bar{s}^j)v_k^j(s^{-i,j}) = (v_k^j(s^i,\bar{s}^j, s^{-i,j}) - v_k^j(s^i, \bar{s}^j, s^{-i,j})) - (v_k^j(s^i, \bar{s}^j, s^{-i,j}) - v_k^j(s^i, \bar{s}^j, s^{-i,j}))$$

**Proposition 2.5** There exists an $\alpha$-potential for $v$, if and only if:

$$\alpha^i\partial(s^i,\bar{s}^i)\partial(s^j,\bar{s}^j)v_k^j(s^{-i,j}) = \alpha^j\partial(s^i,\bar{s}^i)\partial(s^j,\bar{s}^j)v_k^j(s^{-i,j})$$

(6)

for all $i, j \in \mathcal{N}, k \in \mathcal{K}, s^i, \bar{s}^j, s^{-i,j} \in S^{i,j}$ and $s^{-i,j} \in S^{-i,j}$. Thus, there is some cardinal potential for $v$ only if:

$$\left(\partial(s^i,\bar{s}^i)\partial(s^j,\bar{s}^j)v_k^j(s^{-i,j})\right) \left(\partial(s^i,\bar{s}^i)\partial(s^j,\bar{s}^j)v_k^j(s^{-i,j})\right)$$

(6)

for all $i, j \in \mathcal{N}, k, l \in \mathcal{K}, s^i, \bar{s}^j, s^{-i,j} \in S^{i,j}$ and $s^{-i,j} \in S^{-i,j}$.

**Proof.** See Appendix B. □

Assuming $v$ to be sufficiently smooth, equation 6 is equivalent to:

$$\alpha^i \frac{\partial^2}{\partial s^i \partial s^j} v_k^j(s) \equiv \alpha^j \frac{\partial^2}{\partial s^i \partial s^j} v_k^j(s).$$

where $\frac{\partial^2}{\partial s^i \partial s^j}$ denotes the $d^i \times d^j$-dimensional matrix of cross derivatives with respect to $i$ and $j$’s multi-dimensional signals $s^i$ and $s^j$.

\footnote{Note that Monderer and Shapley’s analogous sufficient conditions on cross derivatives use one dimensional strategy spaces and differentiability in order to apply Green’s Theorem. Our discrete version is more general and applies to their setting as well.}
3 Separable Valuations

In this section we focus on interdependent, yet separable valuation functions, i.e. on valuation functions \( v \) such that

\[
v^i_k(s) = f^i_k(s^i) + h^i_k(s^{-i}),
\]

for some functions \( f^i_k: S^i \to \mathbb{R} \) and \( h^i_k: S^{-i} \to \mathbb{R} \). Let \( f^i := (f^i_k)_{k \in K}: S^i \to \mathbb{R}^K \).

The following simple result displays the main role played by the separability assumption:

**Lemma 3.1** An SCR \( \psi: S \to K \) 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 \( \forall i, k, h^i_k \equiv 0 \).

**Proof.** See Appendix B. \( \blacksquare \)

Our main result in this section is:

**Theorem 3.2** Assume that \( f^i(S^i) = \mathbb{R}^K \), and that \( K \geq 3 \). Then every exhaustive ordinal potential is equivalent to a cardinal potential. In particular, every implementable exhaustive choice rule is represented by a cardinal potential.

For the proof of the Theorem, it suffices (by Proposition 2.4) to find a cardinal potential representation for any given implementable choice rule. This will be achieved by showing that all implementable social choice rules belong to the following class:

**Definition 3.3** A SCR \( \psi: S \to K \) is said to be an affine maximizer if and only if it is of the form:

\[
\psi(s) \in \arg \max_{k \in K} \left\{ \sum_{j=1}^{N} \alpha^j f^j_k(s^j) + \lambda_k \right\}
\]

for agent-specific weights \( \alpha^j \geq 0 \) and alternative-specific weights \( \lambda_k \in \mathbb{R} \).

For a given affine maximizer, the weight \( \alpha^j \) can be interpreted as the importance of agent \( j \)'s information to the social choice, and the weight \( \lambda_k \) as the designer’s preference for alternative \( k \). Note that, for private values, a Clarke-Groves-Vickrey mechanism is the affine maximizer with weights \( \alpha^j = 1, \lambda_k = 0 \).

The Theorem follows now from the next result:
Proposition 3.4 Assume that $f^i(S^i) = \mathbb{R}^K$ for all $i \in N$ and that $K \geq 3$. Then an exhaustive choice rule $\psi : S \to \mathcal{K}$ is ex-post implementable only if it is an affine maximizer.

Proof. See Appendix. □

The proof of Proposition 3.4 is subtle, and is based on a hyperplane separation argument due to Roberts (1979). Roberts proved a similar result for dominant strategy implementation with private values, i.e. for $S^i = \mathbb{R}^K$, $f^i = \text{id}$, $h^i = 0$. Our proof consists of adapting Robert’s insight by showing that there is no loss of generality in assuming that an ex-post implementable SCR takes only payoff relevant information into account. This will mean that $\psi$ factors through $f$, i.e. for $X = (\mathbb{R}^K)^N$ there exists a function $\phi : X \to K$ such that $\psi = \phi \circ f$.

Remark 3.5 Not every affine maximizer is implementable! Problems arise if the weight $\alpha_i$ of some agent $i$ is zero. In such a case, equation 8 imposes no restriction on $\psi$ with respect to agent $i$’s signal $s^i$: Fix a signal combination $s^{-i}$ for which the argmax set in equation 8 is not a singleton. Then, $\psi$ may choose an arbitrary alternative from the argmax set, depending on $s^i$. But this arbitrariness may violate a monotonicity condition that needs to be satisfied by implementable choice rules - see Lemma 5.3 and Example 5.2 in Appendix B. It is not difficult to see how the definition of an affine maximizer would have to be adapted, to ensure implementability of all affine maximizers: Denote the set of irrelevant agents, i.e. agents with weight 0, by $I \subset N$. For every $s^{-I}$ for which $\arg \max_{k \in \mathcal{K}} \left\{ \sum_{j \in I} \alpha_j f^j_k (s^j) + \lambda_k \right\}$ is not a singleton a new choice rule, mapping $s^I$ to elements of this argmax set, must be defined. The set of these rules is again characterized by Roberts’ theorem (or by the analogue theorem of Laffont and Maskin (1982) in the case that the argmax set contains only two elements). This tie-breaking problem is related to Remark 2.2, and the remark after Proposition 2.4.

We conclude this section by showing that the assumptions of Proposition 3.4 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. Thus, not every ordinal potential has a cardinal representation in this case.

\footnote{This nuance seems not to be appreciated by much of the literature (including Muller and Vohra (2003), Lavi et al. (2003) and an earlier version of this paper).}
In order to demonstrate the necessity of the assumption that \( f^i(S^i) \) has full dimensionality, consider the following setting, taken from Jehiel and Moldovanu (2001): In their Proposition 5.1, these authors study a model with one-dimensional signals \( s^i \) where agent \( i \)'s valuation of alternative \( k \) is given by \( v^i_k(s) = \sum_{j \in N} a^j_{ki}s^j \). Their work only deals with the possibility of implementing the efficient choice rule, but it can be extended to show that any functions \( \alpha \) can be equivalent to such an ordinal potential, because \( \alpha \) would have to simultaneously satisfy the \( K \) equations \( \alpha^j a^j_{kj} = A^j_k \).

For bounded valuations \( f^i(S^i) \subset \mathbb{R}^K \), an example of an implementable choice rule that is not an affine maximizer is available from the authors upon request.\(^{12}\)

Finally, it is important to note that separability is crucial in order to obtain positive results: in a social choice setting with multi-dimensional type spaces \( S^i \) and with generic\(^{13}\) valuation functions, Jehiel et al. (2004) show that only trivial choice rules are implementable. Thus, by Proposition 2.4, all ordinal potentials must be trivial too.

### 4 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 \( \Gamma = \Gamma(u^1, u^2, \ldots, u^N) \) be a game in strategic form played by the agents in a finite set \( N \). The strategy set of player \( i \) is denoted by \( Y^i \), and the payoff function of \( i \) is \( u^i : Y^i \rightarrow \mathbb{R} \), where \( Y = Y^1 \times Y^2 \times \ldots Y^N \).

**Definition 4.1**

1. Given agent specific weights \( \alpha = (\alpha^i)_{i \in N} \gg 0 \), a function \( P : Y \rightarrow \mathbb{R} \) is an \( \alpha \)-potential for \( \Gamma \) if for every \( i \in N \) and for every \( y^{-i} \in Y^{-i} \)
   
   \[ \alpha^i(u^i(y^{-i}, x) - u^i(y^{-i}, z)) = (P(y^{-i}, x) - P(y^{-i}, z)) \]
   
   for every \( x, z \in Y^i \). If we are not interested in the value of \( \alpha \) we simply speak of a cardinal potential. For \( \alpha \equiv 1 \) we shall speak of an exact cardinal potential.

---

\(^{12}\)Related characterizations of implementable choice rules for combinatorial auctions can be found in Lavi et al. (2003).

\(^{13}\)A topological definition of genericity is being used there.
2. A function \( P : Y \to \mathbb{R} \) is an ordinal potential for \( \Gamma \) if for every \( i \in \mathcal{N} \) and for every \( y^i \in Y^i \)

\[
u^i(y^i, x) - u^i(y^i, z) > 0 \iff P(y^i, x) - P(y^i, z) > 0
\]

for every \( x, z \in Y^i \).

The following result motivates the above definitions, and shows how potentials align the agents’ interests:

**Proposition 4.2 (Monderer and Shapley (1996))** Let \( P \) be an ordinal potential for \( \Gamma(u^1, u^2, \ldots, u^n) \). Then a strategy profile is a Nash equilibrium of \( \Gamma(u^1, u^2, \ldots, u^n) \) if and only if it is a Nash equilibrium of \( \Gamma(P, P, \ldots, P) \).

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

5 Appendix B: Proofs

For the proof of Proposition 2.5 we need the following:

**Lemma 5.1** Let \( f : X \times Y \times Z \to \mathbb{R} \) (where \( X, Y, Z \) are arbitrary sets) be any function that satisfies: \( \forall x, x', y, y', z, \)

\[
\frac{\partial}{\partial (x,x')} f (\cdot, \cdot, z) = \frac{\partial}{\partial (y,y')} f (\cdot, \cdot, z) = 0
\]

Then \( f \) must be additive in \( x \) and \( y \). That is, there are functions \( f_x(x, z) \) and \( f_y(y, z) \) such that \( f(x, y, z) = f_x(x, z) + f_y(y, z) \).

**Proof.** Fix \( x^*, y^* \) and define \( f_x(x, z) := f(x, y^*, z) \) and \( f_y(y, z) := f(x^*, y, z) - f(x^*, y^*, z). \) Solving \( \partial_x(x, y^*) \partial_y(y, y^*) f (\cdot, \cdot, z) = 0 \) for \( f(x, y, z) \) yields the result.

**Proof of Proposition 2.5.** As the alternative \( k \) does not matter for this proof, we simplify notation by leaving away the subscripts. We exhibit the proof for exact potentials. The weighted case is analogous.

The necessity of these conditions is immediate from the definitions. The sufficiency part follows:

**Lemma 5.1** gives the proof for the case of two agents \( i, j \): \( \partial_{(s^i, \tilde{s}^i)} \partial_{(y^j, \tilde{y}^j)} (v^i - v^j) \equiv 0 \) yields \( v^i(s) - u^i(s) = t^i(s^i) - t^i(s^i) \) and we can define the cardinal potential \( P(s) = v^i(s) + t^i(s^{-i}) \). We now use this as the initial step in an induction argument.
Assume that we have already defined a potential $P(s)$ and payments $v^j(s^{-j})$ such that $v^j(s) + t^j(s^{-j}) = P(s)$ for all agents $j$ in a subset $J \subset N$. We now construct a potential $P'(s)$ and payments $v'^j(s^{-j})$ such that $v^j(s) + t'^j(s^{-j}) = P'(s)$ for all agents $j$ in $J \cup \{i\}$ where $i \in N \setminus J$.

By the result in the two-agent case, we can construct a potential $P^j(s)$ and payments $t^j(s^{-j})$, $t^j_j(s^{-i})$ for each agent $j \in J$ such that $v^j(s) + t^j(s^{-j}) = P^j(s) = v^i(s) + t^i_j(s^{-i})$. Using the potentials $P^j(s)$ we now change $P(s)$ (which is a potential for $J$) into $P'(s)$, a potential for $J \cup \{i\}$. Fix an agent $h \in J$. By the properties of the potentials $P, P^j$, we have:

$$t^j(s^{-h}) - t^j_h(s^{-h}) = P^h(s) - P(s) = P^j(s) - P(s)$$

$$= t^j(s^{-j}) - t^j_j(s^{-i}) + t^i_j(s^{-j}) - t^j_i(s^{-i}), \forall j \in J$$

Since $t^j_h - t^j_j$ does not depend on $s^j$ and since $t^j - t^i_j$ does not depend on $s^j$, we get that $\partial_{(s^j, \tilde{s})} \partial_{(s^i, \tilde{s})} (t^j_h - t^j_i) \equiv 0$ for all $j \in J$. Lemma 5.1 yields now

$$t^j(s^{-h}) - t^j(s^{-h}) = t(s^i) - t(s^{-i,h})$$

(9)

for some functions $t$ and $t^i$.

For $j \in J$ set

$$P'(s) := P(s) + t(s^i) = P^h(s) + t^j(s^{-i,h}), t'(s^{-j}) := t^j(s^{-i}) + t(s^i)$$

(10)

Together with

$$t^j(s^{-i}) := t^j_h(s^{-i}) + t^j_i(s^{-i,h})$$

(11)

we get that $v^j(s) + t'^j(s^{-j}) = P'(s)$ for all agents $j$ in $J \cup \{i\}$.

**Proof of Lemma 3.1.** “if”: Let $\psi$ be ex-post implemented by a mechanism $(\psi, t)$ under private values. Thus

$$f^i_{\psi(s)}(s^i) + t^i(s) = \max_{s' \in S^i} \left( f^i_{\psi(s)}(s^i) + t^i(s) \right)$$

for all $i, s^i, s^{-i}$, where $s = (s^i, s^{-i}), \tilde{s} = (\tilde{s}^i, \tilde{s}^{-i})$. For interdependent utilities of the form $f^i_k(s^i) + h^i_k(s^{-i})$ we define monetary payments by $t^i_k(s) = t^i(s) - h^i_k(s^{-i})$. One easily verifies that

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

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

”only if”: Analogously. ■
Example 5.2 Consider a setting where an indivisible good has to be allocated to one of two agents $i, j$, who value the good between 0 and 1. Denote the alternative where $i$ (resp. $j$) receives the good by $i$ (resp. $j$), and consider the allocation rule that allocates the good to $i$ if $s_i > 0$ or $s_i = 0$ and $s_j > 0.5$. Denote the alternative where $i$ (resp. $j$) receives the good by $i$ (resp. $j$), and consider the allocation rule that allocates the good to $i$ if $s_i > 0$ or $s_i = 0$ and $s_j > 0.5$. This is an affine maximizer with weights $\lambda_i = \lambda_j = \alpha_j = 0$ and $\alpha_i = 1$.

Implementability of this allocation rule requires (by the taxation principle) the existence of a price $t_i(0)$ that would induce agent $j$ to buy the good when her value is smaller than 0.5, and not buy it when her value is greater than 0.5. This is clearly impossible.

The following result is a crucial ingredient for the proof of Proposition 3.4. It is also interesting in its own right, as it establishes a monotonicity property of implementable choice rules.

Lemma 5.3 (Monotonicity) An implementable choice rule $\psi$ 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

$$v_k^i(s') - v_k^i(s) > v_l^i(s') - v_l^i(s),$$

$\psi(s) = k$ implies that $\psi(s') \neq l$.

Proof of Lemma 5.3. By the taxation principle, there are transfers $t_i^i(s^{-i})$ such that $\psi(s) \in \arg\max_{k \in K} \{v_k^i(s) + t_k^i(s^{-i})\}$ for all $s$. If alternative $k$ is among $i$'s favorite alternatives at signal $s$, we have $k \in \arg\max_{k \in K} \{v_k^i(s) + t_k^i(s^{-i})\}$. If the change from $s'$ to $s'^i$ makes alternative $k$ strongly more preferable (for $i$) than $l$, $v_k^i(s') - v_k^i(s) > v_l^i(s') - v_l^i(s)$, it is immediate that $l$ cannot be preferred at signal $s'^i$. Thus, $l \notin \arg\max_{k \in K} \{v_k^i(s') + t_k^i(s'^{-i})\}$. By the taxation principle, we conclude that $\psi(s') \neq l$.

Proof of Proposition 3.4. We use an important result due to Roberts (1979) who studied deterministic SCRs that are implementable in dominant strategies in a private values setting. Roberts showed that such SCRs must satisfy a monotonicity condition, called PAD. Using our notation, his proof relies on the following technical result:

---

14 Mueller and Vohra (2003) characterize dominant strategy mechanisms with multidimensional types in terms of a monotonicity property of the underlying allocation rule.

15 Roberts' proof uses a hyperplane-separation argument which yields the weights in the affine representation.
Theorem A (Roberts 1979): Let $X = (\mathbb{R}^K)^N$ 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 $x^i_k - x^i_k > x'^i_k - x'^i_k$ for all $i \in \mathcal{N}$ and all $i \neq k \in \mathcal{K}$, $\phi(x) = k$ implies $\phi(x') = k$. An affine maximizer $\phi$ satisfies $\phi(x) \in \arg\max_{k \in \mathcal{K}} \{\sum_{j=1}^N x^j_k + \lambda_k\}$ for some $\{\alpha^j\}_{j \in \mathcal{N}}$ and some $\{\lambda_k\}_{k \in \mathcal{K}}$.

Thanks to Lemma 3.1 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 $\psi$ is implementable, we can recursively apply Lemma 5.3 to show that for all signals $s, s' \in S$ such that

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

$$\psi(s) = k \implies \psi(s') = k.$$ Consider the sequence of signals $s_{(i)} := s$, $s_{(i)} := (s', s_{(i-1)})$ for all agents $i \leq n$ (this gives $s_{(n)} = s'$). The proof of Lemma 5.3 serves then as the induction step proving that, with $\psi(s_{(0)}) = k$, we have $\psi(s_{(i)}) = k$ for all $i$. This yields $\psi(s') = k$.

Thus, we can apply Theorem A to $\phi$ as a function of the $f^i_k(s)$, and get $\psi(s) = \phi(f(s)) = \arg\max_{k \in \mathcal{K}} \{\sum_{j=1}^N \alpha^j f^j_k(s') + \lambda_k\}$.

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 $\psi$ to a function $\tilde{\psi}$ that factors $\tilde{\psi} = \phi \circ f$; b) Show that $\tilde{\psi}$ is ex-post implementable, and apply Theorem A to $\phi$ to show that $\tilde{\psi}$ is an affine maximizer; c) Show that $\tilde{\psi}$ is an affine maximizer if $\tilde{\psi}$ is one.

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

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

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

b) The SCR $\tilde{\psi}$ is ex-post implementable by the transfer rule $\tilde{t}((s^i)_{i \in \mathcal{N}}) := t((\tilde{s}^i)_{i \in \mathcal{N}})$, where $t : S \to \mathbb{R}^N$ are the transfers that implement $\psi$. Indeed, we readily check agent $i$’s incentive constraint:

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

14
for all \(i, s^i, s^{i-1}\). The first and third equality follow by the definitions of \(\hat{\psi}\) and \(\hat{t}\), and the inequality follows by the ex-post incentive compatibility of \((\hat{\psi}, \hat{t})\). By Lemma 5.3, \(\hat{\psi}\) satisfies monotonicity, which in turn means that \(\hat{\phi} : (\mathbb{R}^K)^N \rightarrow K\) satisfies PAD in the sense of Roberts’ Theorem A. Thus, there are constants \(\alpha^j \geq 0\) for \(j \in \mathcal{N}\) and \(\lambda_k\) for \(k \in K\) such that \(\hat{\phi}(f(s)) \in \arg\max_{k \in K} \left\{ \sum_{j=1}^N \alpha^j f^j_k(s^j) + \lambda_k \right\}\) for all \(s \in S\). This proves that \(\hat{\psi}\) is an affine maximizer.

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

References


