Mathematical Social Sciences 11 (1986) 109–127
North-Holland

GENERAL, DIRECT AND SELF-IMPLEMENTATION OF SOCIAL CHOICE FUNCTIONS VIA PROTECTIVE EQUILIBRIA*

Salvador BARBERÁ
Departamento de Teoría Económica, Universidad del Pais Vasco, Av. Ejército 83, 48015 Bilbao, Spain

Bhaskar DUTTA
Indian Statistical Institute, 203 Barrackpore Trunk Road, Calcutta 700035, India

Communicated by K.H. Kim
Received 21 December 1984
Revised 6 May 1985

This paper investigates the possibility of implementing social choice functions via protective equilibria – a refinement of maximin behavior. We describe a necessary condition, as well as a set of sufficient conditions for a social choice function to be implementable. The set of implementable functions is certainly not empty, and it contains interesting procedures, allowing for some agents to have some veto power. However, positional or Condorcet-type methods cannot be implemented via protective equilibrium. The connections and the differences between different notions of implementation are also discussed.

Key words: Social choice functions; protective equilibria; implementable functions.

1. Introduction

In an earlier paper,[1] we had explored the possibility of implementing social choice functions under a solution concept which we had called protective equilibria – essentially a refinement of maximin behavior. We characterized the class of implementable social choice functions under the assumption that the aggregation mechanism has to be the social choice function itself; in other words, the agents are confronted with a game form in which strategy sets are the set of possible preference orderings and the outcome function is the social choice function to be implemented.

[1] See Barberá and Dutta (1982). See also Moulin (1981) for a closely related solution concept.

* This paper was the basis for part of one Schunbrunn Lecture, delivered by S. Barberá at the Institute for Advanced Studies, Hebrew University of Jerusalem, July 1983.

The literature[2] on implementation has considered a more general framework which allows for a much wider class of game forms – players' strategy sets are permitted to be arbitrary finite sets, while the only restriction placed on the outcome function is the obvious one that it always select a feasible alternative for each n-tuple of strategies. For most solution concepts, this richer framework allows for a wider class of implementable social choice functions, although the much-talked about 'relevation principle' shows that for many solution concepts, if a social function can be implemented by an arbitrary game form, then it can also be 'truthfully implemented' by a *direct* mechanism, i.e. a mechanism in which strategy sets are the set of possible characteristics (preference orderings) of agents.[3]

The purpose of this paper is to see whether the class of implementable social choice functions under protective equilibria is enlarged when the implementing mechanism is permitted to be any arbitrary game form. Indeed, we distinguish between three different notions of implementation. The first and most restrictive notion is *self-implementability*, in which strategies are preference orderings and the outcome function is nothing but the social choice function to be implemented. A second possibility, called *direct implementability*, is to restrict strategies to be preference orderings but allow the outcome function to be a social choice function different from the one being implemented. The most liberal notion is *general implementability* (henceforth referred to as simply implementation) in which no restrictions at all are placed on the game form.[4]

Our most important result is that the main necessary condition for self-implementability continues to be necessary for implementability in general. This necessary condition is not satisfied by many familiar social choice functins. We also show that if a social choice function is implementable but not self-implementable, then the strategy sets in the implementing game form have to be 'larger' (in a sense to be made precise later on) than the set of preference orderings. Finally, we also provide a set of sufficient conditions for implementability. These sufficient conditions are satisfied by the class of voting by veto social choice functions, and at least one element in this class is not self-implementable.

The paper is organized as follows. Section 2 contains notation and definitions and presents the main result in our earlier paper regarding self-implementability. Section 3 contains our results on general implementability. In Section 4 we present some examples to show that there is no general connection between the three notions of implementability that we have referred to earlier, except the obvious one that self-implementability implies direct implementability and the latter implies general im-

[2] Excellent surveys are Dasgupta et al. (1979), Laffont and Maskin (1983).

[3] See Dasgupta et al. (1979) for a discussion of the distinction between implementation and truthful implementation.

[4] In Barberá and Dutta (1982), self-implementability was actually called direct implementability. The present definitions conform to the conventional practice in the literature.

plementability. We conclude in Section 5. An Appendix elaborates on the restrictiveness of upper strong positive association, a necessary condition for SCFs to be implementable via protective equilibria.

2. Notation and definitions

Let $A=\{x, y, z, \ldots\}$ be a finite set of *alternatives*, with cardinality $m \geq 3$. (When convenient, alternatives will also be denoted, $a_i, a_j, \ldots$)

Let $I=\{1, 2, \ldots, n\}$ be an initial segment of the integers, whose elements are called *individuals*. $\mathscr{P}$ is the set of asymmetric orderings over A. Elements of $\mathscr{P}$ are represented by $P, P', P_i, P_j, \ldots$ and are called *preferences*.

The following definitions are useful to describe partial characteristics of a preference ordering and to compare it with others.

If $P \in \mathscr{P}$, $Y \subset A$, we say that Y *is bottom for* P *iff* $(\forall y \in Y)(\forall x \in A - Y) xPy$. For $k \in \{1, \ldots, m\}$, $P \in \mathscr{P}$, the k-bottom of P, denoted by $B(k, P)$, is the unique subset of A which is bottom for P and contains exactly k alternatives.

For $P, P' \in \mathscr{P}$, $Y \subset A$, we say that P *and* P' *agree on* Y iff $(\forall x, y \in Y)[xPy \leftrightarrow xP'y]$.

For any $P \in \mathscr{P}$, $r \in \{1, 2, \ldots, m\}$, the rth *ranking worst alternative in* P, denoted by $a_r(P)$, is defined by $a_r(P)=\{x \in A \mid$ there exist exactly $(r-1)$ alternatives $y \in A : xPy\}$.

Let $\mathscr{P}^n$ be the n-fold cartesian product of $\mathscr{P}$. Elements of $\mathscr{P}^n$ are denoted by $\boldsymbol{P}, \boldsymbol{P}', \ldots$, and are called *preference profiles*.

Let $i \in I$. Given a preference profile $\boldsymbol{P}=(P_1, P_2, \ldots, P_{i-1}, P_i, P_{i+1}, \ldots, P_n)$, we may denote it by $\boldsymbol{P}=(P_i, P_{-i})$, where $P_{-i}=(P_1, P_2, \ldots, P_{i-1}, P_{i+1}, \ldots, P_n)$.

Given $\boldsymbol{P} \in \mathscr{P}^n$ and $P_i' \in \mathscr{P}$, $\boldsymbol{P}/P_i' \overset{\text{def}}{=} (P_1, P_2, \ldots, P_{i-1}, P_i', P_{i+1}, \ldots, P_n)$, i.e. $\boldsymbol{P}/P_i'$ stands for a profile obtained from $\boldsymbol{P}$ by changing its ith component from P_i to P_i'.

2.1. Social choice functions

A social choice function (SCF) is a function $f: \mathscr{P}^n \to A$.

Individual $i \in I$ *has a veto under* f iff, for every $x \in A$, there exists $P_i(x)$ such that $f(P_i(x), P_{-i}) \neq x$, for all $P_{-i} \in \mathscr{P}^{n-1}$.

The following properties of SCFs will be relevant in analyzing their strategic properties. Upper strong positive association is the most important and attractive. It is the conjunction of a monotonicity property (if $x=f(\boldsymbol{P})$ and x 'ranks better' in $\boldsymbol{P}'$, which is otherwise identical to $\boldsymbol{P}$, then $x=f(\boldsymbol{P}')$), and an invariance property (if $x=f(\boldsymbol{P})$ and the set of elements better than x is the same for all individuals in $\boldsymbol{P}$ and $\boldsymbol{P}'$, then $x=f(\boldsymbol{P}')$). This condition is similar to, though strictly weaker than, Muller and Satterthwaite's (1977) *Strong Positive Association* which was shown to be equivalent to strategyproofness.

Lower conditional independence and bottom equivalence are rather special 'regularity' conditions on the connections between the images of profiles having 'similar'

bottoms. Although they do not seem to have any particular normative appeal, they turn out to be crucial conditions for a SCF to be self-implementable via protective equilibria.

Given any social choice function f, for any $P_i \in \mathcal{P}$ and $x \in A$, we denote by $g_f(x, P_i)$ the set $\{P_{-i} \in \mathcal{P}^{n-1} / f(P_{-i}, P_i) = x\}$.

Let $i \in I$, $P_i, P_i' \in \mathcal{P}$ and $Y \subset A$. We say that P_i *and* P_i' *are Y-equivalent for i under* f iff $(\forall y \in Y)[g_f(y, P_i) = g_f(y, P_i')]$. P_i and P_i' are equivalent for i iff they are A-equivalent.

A social choice function f satisfies:

(i) *Upper strong positive association* (USPA) iff, for all $i \in I$, all $P \in \mathcal{P}^n$ and all $P_i' \in \mathcal{P}$, $[(f(P) = a_r(P_i)) \& (B(r-1, P_i) = B(r-1, P_i')) \& (P_i \text{ and } P_i' \text{ agree on } B(r-1, P_i))] \rightarrow [f(P/P_i') = f(P)]$.

(ii) *Bottom equivalence* (BE) iff, for all $i \in I$, for all $P_i, P_i' \in \mathcal{P}$, $[(P_i \text{ and } P_i' \text{ are } B(k, P_i) \text{ equivalent, but are not } B(k+1, P_i) \text{ equivalent}) \& (P_i \text{ and } P_i' \text{ agree on } A - B(k, P_i))] \rightarrow [B(k, P_i') = B(k, P_i)]$.

(iii) *Lower conditional independence* (LCI) iff, for all $i \in I$, all $P \in \mathcal{P}^n$, all $P_i' \in \mathcal{P}$, $[(f(P) = a_{r+1}(P_i)) \& (B(r, P_i) = B(r, P_i')) \& (P_i \text{ and } P_i' \text{ are } B(r, P_i)\text{-equivalent and agree on } A - B(r, P_i))] \rightarrow [f(P/P_i) = f(P)]$.

2.2. *Game forms, protective equilibria and the implementation of social choice functions*

A social choice function f is implementable if there exists a game form such that, for every preference profile, and thus for every specification of the payoffs of players, the equilibrium outcomes coincide with the function's image at this profile. This general definition should be qualified in two directions. Firstly, we need to specify what is the relevant concept of equilibrium. Secondly, we may or may not impose restrictions on the type of game form to be considered. We will comment briefly on each of these points as they appear.

2.2.1. *Game forms*

A *game form* for (I, A) is an $n+1$-tuple $F = (\mathcal{S}_1, \ldots, \mathcal{S}_n; \psi) \equiv (\mathcal{S}; \psi)$, where for all i, $\mathcal{S}_i$ is the *strategy set* of agent i, and $\psi : \mathcal{S} \rightarrow A$ is the *outcome function*.

A game form F for (I, A) is *direct* iff $\mathcal{S}_i \subseteq \mathcal{P}$ for all $i \in I$.

A social choice function f's *own induced game form* is the one defined as $F_f = [\mathcal{P}^n; f]$.

Direct game forms, and in particular those induced by a SCF, are specific types of game forms which arise naturally in connection with particular interpretations of the problem of implementation. The strategies of a direct game form are of the same nature as the inputs of a SCF. The outcome function for the game form induced by a SCF establishes the same relationship between strategies and outcomes as the SCF prescribes between preferences and outcomes.

Given a game form $F = [\mathcal{S}; \psi]$, elements of $\mathcal{S} = \mathcal{S}_1 \times \cdots \times \mathcal{S}_n$ are denoted by

$s=(s_1, s_2, \ldots, s_n)\equiv(s_{-i}, s_i)$, where $s_i\in\mathcal{S}_i, s_j\in\mathcal{S}_j, \ldots$, and $s_{-i}\in\mathcal{S}_1\times\cdots\times\mathcal{S}_{i-1}\times\mathcal{S}_{i+1}\times\cdots\times\mathcal{S}_n\equiv\mathcal{S}_{-i}$.

The next two definitions relate the attainability of an issue through the use of certain strategies.

Given $T\subset\mathcal{S}$, let $\psi(T)=\{a\in A \mid a=\psi(s) \text{ for some } s\in T\}$.

Given $F=[\mathcal{S};\psi]$, for any $a\in A$ and $s_i\in\mathcal{S}_i$, let $g_F(a,s_i)=\{s_{-i}\in\mathcal{S}_{-i} \mid \psi(s_i,s_{-i})=a\}$.

Some strategies in a game form systematically lead to the same outcome as others. The following definitions are useful in order to classify strategies on the basis of the outcomes they give rise to.

For any $B\subseteq A$, s_i and s_i' are *B-equivalent* iff for all $a\in B$, $g_F(a,s_i)=g_F(a,s_i')$. When two strategies are A-equivalent, we simply say that they are equivalent.

2.2.2. *Games and protective equilibria*

A *game* (in normal form) for (I,A) is a $(2n+1)$-tuple $G=[\mathcal{S}_1,\ldots,\mathcal{S}_n;\psi;P_1,\ldots,P_n]\equiv[F;P]$, where F is a game form for (I,A) and $P=(P_1,\ldots,P_n)\in\mathcal{P}^n$ is a preference profile on A. We denote by $G(F,P)$ the game associated with the game form F and profile P. When the game form we are referring to is unambiguous, we may simply write $G(P)$.

Given a game $G(F,P^*)=[\mathcal{S}_1,\ldots,\mathcal{S}_n;\psi;P_1^*,\ldots,P_n^*]$, for any $i\in I$, a strategy s_i *protectively dominates* s_i', denoted by $s_i d_F(P_i^*)s_i'$ iff there exists $k\in[1,m]$, such that:

(i) $g_F(a_k(P_i^*),s_i)\subset g_F(a_k(P_i^*),s_i')$, and

(ii) $g_F(a_r(P_i^*),s_i)=g_F(a_r(P_i^*),s_i')$ for all $r<k$.

We then say that $s_iP_i^*$ dominates s_i' at the k-level.

Let $D_{iF}(P_i^*)=\{s_i\in\mathcal{S}_i \mid \sim s_i' d_F(P_i^*)s_i \text{ for any } s_i'\in\mathcal{S}_i\}$, and $D_F(P^*)=\prod_{i\in I}D_{iF}(P_i^*)$. $D_F(P^*)$ is the set of all n-tuples of protectively undominated strategies for game G.

Whether a strategy is more protective than another depends, then, not only on what outcomes will obtain when the strategy is used, but also on the circumstances under which these outcomes will obtain. For s to protectively dominate s' the set of plays by other individuals which would lead to some outcome x when s is used should be a proper subset of the set of plays by others yielding x when s' is used, with x being the worst-ranked alternative for which s and s' are not equivalent. Protective strategies are those which are protectively undominated. This type of lexical maximin comparisons among strategies may be justified in contexts where the players lack all information on their opponents, both on their payoff and on the type of strategic considerations they will engage in. They are, of course, open to much of the general criticism that can be addressed to maximin solutions, but not to all. In many frameworks, maximin-type solution concepts are hardly discriminating, and most strategies can be admissible; in our particular context and given our notion of protective domination we find classes of functions under which the set of undominated strategies are systematically small.

2.2.3. General, direct and self-implementation

A social choice function f is *implementable via protective equilibria* (PI) iff there is a game form $F=[\mathcal{S};\psi]$ such that for every $\boldsymbol{P}^* \in \mathcal{P}^n$, $f(\boldsymbol{P}^*)=\psi[D_F(P^*)]$.

If f is implementable by a *direct* game form, we say that it is *directly implementable*. If it is implementable by its own induced game form $F=[\mathcal{P}^n;f]$, we then say that f is *self-implementable*.[5]

The distinction between general, direct and self-implementation is more than a technicality. Implementability of each type, when possible, corresponds to a different type of solution for the economic problem of mechanism design. If a SCF is implementable by a game form, this game form represents the 'institution' which will guarantee that the relationship between preferences and outcomes as prescribed by the function does hold. This 'institution' involves both the type of messages that agents must generate and the rules by which these messages will be processed. When direct implementation is possible, the messages can just be the preferences (or their relevant characteristics), and institutional design concentrates on the outcome function through which preference information will be processed. Finally, when a SCF is self-implementable, announcing the desired function itself would guarantee its realization.[6] The following theorem, which is proved in Barberá–Dutta (1982), characterizes the SCFs which are self-implementable via protective equilibria. In the next sections we turn to more general types of implementation.

Theorem 1. *A social choice function is self-implementable via protective equilibria iff it satisfies USPA, BE and LCI.*

3. Implementability of social choice functions

Our first result in this section (Theorem 2) establishes that in order to be implementable[7] at all, an SCF must satisfy one of the conditions that we already know to be necessary for self-implementation: USPA. This is a very restrictive condition, and many interesting SCFs do not satisfy it. We stress this point in the Appendix, where it is shown that almost all positional SCFs, as well as all rules selecting generalized Condorcet alternatives, violate USPA. Yet the condition is met by an important class of procedures, the class of voting by veto SCFs. Thus we are far from an impossibility result, as we can also prove that different functions in this

[5] See footnote 2.

[6] These informal comments assume, in particular, that the individual plays will be equilibrium plays (in this case, protective equilibria).

[7] From here on, we may omit to qualify our notion of implementability with reference to protective equilibria, since this is the only equilibrium concept to be used throughout. We only make explicit reference to it in formal statements, but not in discussion.

important class are indeed implementable.

The two following results (Lemma 1 and Theorem 4) refer to the relationship between the preferences which serve as the basic input of SCFs, and the strategies to be used in game forms which implement it. Essentially, Lemma 1 establishes that a strategy in the game form cannot be undominated under two nonequivalent preferences, and the other result indicates that if a game form can implement a SCF which is not self-implementable, then there must be more nonequivalent strategies in the game form than there are nonequivalent preferences under the SCF. Finally, Theorem 5 provides a sufficient condition for a SCF to be implementable: that at least two individuals have at least one veto each. This condition falls short of being necessary for implementation and we thus do not have a full characterization of implementable functions. Notice, however, that there exist SCFs that are even self-implementable.[8] Theorem 5 shows that there is additional gain from relaxing the notion of implementability since there are other SCFs which are not self-implementable, but are implementable. An important example is the rule called *voting by alternating veto* by Moulin (1982). Under this SCF, the individuals do not exercise their veto power in one shot, but are allowed to do so in stages. Of course, there can also be several other rules which satisfy the conditions of Theorem 5 but violate either LCI or BE, which are both necessary conditions for self-implementability. We now proceed to state and prove the announced results.

Theorem 2. *A social choice function f is implementable via protective equilibria only if it satisfies USPA.*

Proof. Let f be implementable by the game form $F=[\mathscr{S};\psi]$. Suppose f violates USPA. Then, there exists $i\in I$, $\boldsymbol{P}\in\mathscr{P}^n$, $P_i'\in\mathscr{P}$ such that:

(a) $f(\boldsymbol{P})=a_r(P_i)=x$,

(b) $B(r-1,P_i)=B(r-1,P_i')$,

(c) P_i and P_i' agree on $B(r-1,P_i)$,

(d) $f(\boldsymbol{P}/P_i')=y\neq x$.

Since f is implementable, there exists $s_j\in D_{jF}(P_j)$ for all $j\neq i$, and $s_i\in D_{iF}(P_i)$, $s_i'\in D_{iF}(P_i')$, such that:

$$\psi(s_i,s_{-i})=x, \tag{1}$$

$$\psi(s_i',s_{-i})=y. \tag{2}$$

This means that s_i and s_i' are not equivalent. Implementability then requires that $s_i d_F(P_i)s_i'$ and $s_i' d_F(P_i')s_i$. This is because if $\sim s_i d_F(P_i)s_i'$, then $s_i'\in D_{iF}(P_i)$, (s_i',s_{-i}) would also be protective under the profile $\boldsymbol{P}$, and $\psi(s_i',s_{-i})=y\neq f(\boldsymbol{P})$ would contradict the assumption that f is implementable by the game form $F\equiv[\mathscr{S};\psi]$. Similarly, assuming that $\sim s_i' d_{iF}(P_i')s_i$ leads to a contradiction. We can now proceed to com-

[8] See Barberá–Dutta (1982) for examples.

plete the proof.

(A) Suppose yP_ix. Since $s_id_{P_i}(P_i)s_i'$, then (1) and (2) imply that for some $k<r$, s_i and s_i' are (i) $B(k-1,P_i)$ equivalent and (ii) $g(a_k(P_i),s_i)\subset g(a_k(P_i),s_i')$. However, if (i) and (ii) hold, then (b) and (c) imply that $s_id_F(P_i')s_i'$. Hence, $s_id_F(P_i)s_i'\to s_id_F(P_i')s_i'$, a contradiction to $s_i'\in D_{iF}(P_i')$.

(B) Suppose xP_iy. Then, in an analogous manner, it can be shown that $s_i'd_F(P_i')s_i\to s_i'd_F(P_i)s$, a contradiction to $s_i\in D_{iF}(P_i)$. In either case, violation of USPA leads to a contradiction, so that USPA is necessary for implementability of f.

Lemma 1. *Suppose f is implementable via protective equilibria by the game form $F=[\mathcal{S};\psi]$. Then*

(a) *if P_i and P_i' are not equivalent under f, $D_F(P_i)\cap D_F(P_i')=\emptyset$;*

(b) *if $s_i\notin D_F(P_i)$, there must be an $s_i^*\in D_F(P_i)$ such that $s_i^*d_F(P_i)s_i$.*

Proof. (a) If P_i and P_i' are not equivalent under f, there is a $P_{-i}^*\in\mathcal{P}^{n-1}$ such that $f(P_i,P_{-i}^*)=x$, $f(P_i',P_{-i}^*)=y$, and $y\neq x$. Suppose $s_i\in D_F(P_i)\cap D_F(P_i')$. Let $s_{-i}^*\in\prod_{j\neq i}D_F(P_j^*)$. Then, since f is implementable,

$$\psi(s_i,s_{-i}^*)=x \quad\text{and}\quad \psi(s_i,s_{-i}^*)=y, \quad\text{which is impossible.}$$

(b) Follows directly from transitivity of the $d_f(P_i)$ *relation.*

Remark 1. Since $D_F(P_i)\cap D_F(P_i')=\emptyset$ whenever P_i and P_i' are nonequivalent under f, it follows that for all $i\in N$, we need at least as many nonequivalent strategies per player under ψ as there are nonequivalent preferences under f. Moreover, if for some individual $j\in I$ there is a strategy s_j^* which is always dominated, i.e. $\exists s_j^*\notin\bigcup_{P_j\in\mathcal{P}}D_F(P_j)$, then the number of ψ-nonequivalent strategies will be greater than the number of f-nonequivalent preferences. The following theorem tells us that this must indeed be the case when implementation is possible but self-implementation is not.

Theorem 3. *Suppose f is not self-implementable via protective equilibria, but that it is implementable (via p.e.) by the game form $F=[\mathcal{S};\psi]$. Then, for some $j\in I$, the number of ψ-nonequivalent strategies is greater than the number of f-nonequivalent preferences.*

Proof. From Theorem 2, f satisfies USPA. Since it is not self-implementable, it violates either LCI or BE, by Theorem 1. Suppose it violates LCI. Then, there exists $i\in N$, $\boldsymbol{P}\in\mathcal{P}^n$, $P_i'\in\mathcal{P}$ such that $f(\boldsymbol{P})=a_{r+1}(P_i)=x$, $f(\boldsymbol{P}/P_i')=y$, $B(r,P_i)=B(r,P_i')$, and P_i and P_i' are $B(r,P_i)$-equivalent under f.

Since f is implementable by F, for all $\hat{\boldsymbol{P}}\in\mathcal{P}^n$, $f(\hat{\boldsymbol{P}})=\psi(\boldsymbol{s})$, where $\boldsymbol{s}$ is any element in $\prod_{i\in I}D_{iF}(\hat{P}_i)$. Pick $s_j\in D_F(P_j)$ for all $j\neq i$, $s_i\in D_F(P_i)$ and $s_i'\in D_F(P_i')$ such that $s_id(P_i)s_i'$. Then $\psi(s_i,s_{-i})=x$ and $\psi(s_i',s_i)=y$. Since P_i and P_i' are $B(r,P_i)$-equivalent under f, we must have yP_ix. Hence, $s_id_F(P_i)s_i'$ implies that for some $k\leq r$,

$$g_F(a_l(P_i), s_i) = g_F(a_l(P_i), s_i'), \quad \forall l < k, \tag{3}$$

$$g_F(a_k(P_i), s_i) \subset g_F(a_k(P_i), s_i'). \tag{4}$$

(4) implies that there is $s_{-i}^* \in \mathscr{S}_{-i}$ such that

$$\psi(s_i, s_{-i}^*) = z, \quad \text{with } z P_i a_k(P_i) \tag{5}$$

and

$$\psi(s_i', s_{-i}^*) = a_k(P_i). \tag{6}$$

Suppose now that $s_{-i}^* \in \prod_{j \in I - \{i\}} D_F(P_j^*)$ for some $P_{-i}^* \in \mathscr{P}^{n-1}$. Since f is implementable by F, (5) and (6) imply that

$$f(P_i, P_{-i}^*) = z, \tag{7}$$

$$f(P_i', P_{-i}^*) = a_k(P_i). \tag{8}$$

Since $k \leq r$, (5) and (6) contradict the assumption that P_i and P_i' are $B(r, P_i)$-equivalent under f. Hence, $s_{-i}^* \notin \prod_{j \neq i} D_F(P_j^*)$ for any $P_{-i}^* \in \mathscr{P}^{n-1}$. Thus, for some j, there is an $s_j^* \notin \bigcup_{P_j \in \mathscr{P}} D_F(P_j)$. By Remark 1, this means that the number of ψ-nonequivalent strategies is greater than the number of f-nonequivalent preferences.

The proof for the case when f violated BE is exactly analogous.

Remark 2. Theorem 3 shows the need for 'large' strategy sets when implementing an SCF which is not self-implementable. However, notice that 'largeness' is relative to the number of nonequivalent strategies in f. One can have seemingly the same strategy sets – revealing preference orderings, but the number of nonequivalent strategies under $F = [\mathscr{P}^{n-1}; f]$ may not be the same as under $F' = [\mathscr{P}^{n-1}; f']$. This will be exploited later to construct SCFs which are not self-implementable but are nevertheless directly implementable.

Our next theorem provides sufficient conditions for an SCF to be implementable.

Theorem 4. *Let f satisfy USPA, and have at least two players with veto power. Then, f is implementable via protective equilibria.*

The proof of this result is quite involved. It is based on the construction of a game form guaranteeing that each individual finds his true preferences to be the unique protective strategy available, whatever these true preferences might be. Before proceeding to the proof itself, we introduce two useful lemmata and provide a sketch of the main idea involved in the construction.

Lemma 2. *Let f satisfy USPA, and $F = [\mathscr{P}^n; f]$ be its own game form. Given two preference orderings P_i, P_i^*, let $B(P_i, P_i^*)$ be the maximal set such that*

(a) *$B(P_i, P_i^*)$ is bottom for both P_i and P_i^*, and*

(b) *P_i, P_i^* agree on $B(P_i, P_i^*)$.*

Let q be the cardinality of $B(P_i, P_i^*)$. Then, if $P_i d_F(P_i^*) P_i^*$, the level at which P_i dominates P_i^* is at least $q+2$.

Proof of the lemma. By USPA, P_i and P_i^* are equivalent on $B(P_i, P_i^*)$ and cannot dominate each other at P_i^* levels $l \leq q$. Suppose that P_i dominates P_i^* at the $q+1$ level. It must be that $a_{q+1}(P_i) \neq a_{q+1}(P_i^*)$, by the definition of $B(P_i, P_i^*)$, and that for some $P_{-i} \in \mathscr{P}^{n-1}$, $f(P_i^*, P_{-i}) = a_{q+1}(P_i^*)$ while $f(P_i, P_{-i}) P_i^* a_{q+1}(P_i^*)$. But then $f(P_i, P_{-i}) \neq f(P_i^*, P_{-i}) = a_{q+1}(P_i^*)$ is again a violation of USPA.

Remark 3. When $q=0$, this means that P_i and P_i^* have different bottom alternatives. Lemma 2 tells us that P_i cannot dominate P_i^* at level 1, and therefore that each P_i^* is at least $a_1(P_i^*)$-equivalent to any other strategy (or else dominates it), under f's satisfying USPA.

More in general, each P_i^* is at least $B(P_i, P_i^*) \cup a_{q+1}(P_i^*)$-equivalent to any other strategy P_i (or else dominates it).

Lemma 3. *If individual $i \in I$ has a veto under f and f satisfies USPA, then $f(P_i, P_{-i}) \neq a_1(P_i)$, for all P_i and all P_{-i}.*

The proof of this lemma is left to the reader.

A sketch of the proof, for the case with $\#A = 4$, $I = 2$. Given f, we shall keep all preferences as possible strategies for each individual. We will then add extra strategies to the strategy space of each of the two, and will specify the outcomes associated to the usage of such strategies in such a way that

(1) the added strategies guarantee that no sincere preference is ever dominated by any other preference, and

(2) none of the added strategies is ever protective.

Table 1 describes the outcomes associated with the use of each one of the six strategies added to player j, against each one of the 'original' strategies (preferences) of player i. We denote this (partial) outcome function by $\emptyset^{ij}$.

Consider now the game form defined by

strategy spaces:

$$S^i = \{\mathscr{P} \cup \sigma_1^i \cup \cdots \cup \sigma_6^i\} \equiv \{\mathscr{P} \cup \Sigma\} \quad (i = 1, 2)$$

outcome function:

$$\psi(s_1, s_2) = \begin{cases} f(s_1 s_2) & \text{if } s_1, s_2 \in \mathscr{P} \\ \emptyset^{ij}(s_1 s_2) & \text{if } s_i \in \mathscr{P} \text{ and } s_j \in \Sigma \\ \text{constant} & \text{if } s_1, s_2 \in \Sigma. \end{cases}$$

By Lemma 2, every preference which is not P_i^*-dominated by P_i must be equivalent to P_i^*, under f, for the set $B(P_i^*, P_i) \cup a_{q+1}(P_i^*)$. Our construct guarantees that all preferences P_i, P_i^* are still $B(P_i^*, P_i)$-equivalent under ψ when they are

Table 1

Preference strategies for player i (left is worse)	Added strategies for player j					
	σ_1^j	σ_2^j	σ_3^j	σ_4^j	σ_5^j	σ_6^j
$xyzw$	y	z	w	z	w	w
$xywz$	y	z	w	z	w	z
$xzyw$	y	z	w	w	y	w
$xzwy$	y	z	w	w	y	y
$xwyz$	y	z	w	y	z	z
$xwzy$	y	z	w	y	z	y
$yxzw$	z	w	x	z	w	w
$yxwz$	z	w	x	z	w	z
$yzxw$	z	w	x	w	x	w
$yzwx$	z	w	x	w	x	x
$ywxz$	z	w	x	x	z	z
$ywzx$	z	w	x	x	z	x
$zxyw$	w	x	y	y	w	w
$zxwy$	w	x	y	y	w	y
$zyxw$	w	x	y	w	x	w
$zywx$	w	x	y	w	x	x
$zwxy$	w	x	y	x	y	y
$zwyx$	w	x	y	x	y	x
$wxyz$	x	y	z	y	z	z
$wxzy$	x	y	z	y	z	y
$wyxz$	x	y	z	z	x	z
$wyzx$	x	y	z	z	x	x
$wzxy$	x	y	z	x	y	y
$wzyx$	x	y	z	x	y	x

under f. Given this equivalence, the added strategies guarantee that, for each P_i^*, P_i^* dom $(P_i^*)P_i$ at the $(q+1)$-th level, where $q = \#B(P_i^*, P_i)$.

For example, suppose that the preferences we use to evaluate strategies are $xyzw$. By USPA, strategies $xyzw$ and $xywz$ must be (x, y)-equivalent, since they agree on x and y as the bottom alternatives. By USPA, $xywz$ cannot dominate $xyzw$ at level z (relative to preferences $xyzw$, remember) when only preferences are strategies. Thus, either $xyzw$ already dominated $xywz$ in the direct game, or they were equivalent. In the latter case, addition of strategies σ_1^j to σ_6^j guarantees that $xyzw$ now dominates $xywz$, since the first five strategies still make them equivalent, and the last one has z as the outcome for $xywz$, while w (a better alternative) is the outcome for $xyzw$.

Adding new strategies is a way to create effective threats against deviations from truth-telling. But it would not be sufficient to guarantee that f is implementable if any of these added strategies could itself become at some point a protective equilibrium strategy. We use our assumption that some individuals have one veto in order to guarantee that these added strategies will always be dominated. By construction, we guarantee that if i has at least one veto under f, i still has one veto under ψ pro-

vided he uses preferences as strategies (the worst outcome for a given preference does not appear in the row corresponding to this preference). On the other hand, however, each nonpreference strategy allows for all outcome to arise (each column contains all possible outcomes). In particular, then, all nonpreference strategies allow for a player's worst alternative to be the outcome, and are thus dominated by all preference strategies where the worst alternative is placed last. This completes the (sketchy) proof that truth-telling is a unique protective strategy for our game form, and that f is implementable.

Proof of Theorem 4. Given f satisfying USPA, assume w.l.o.g. that the two individuals having veto power are 1 and 2. We shall construct a game form by enlarging the strategy spaces of these two individuals, while retaining all preferences as possible strategies for both, as well as for all other agents. Our outcome function ψ, when restricted to strategy n-tuples where the strategies of all individuals are preferences, will still be f. Thus, we shall describe ψ as an extension of f to those strategy n-tuples where some of the strategies are not preferences.

Let $\bar{P}$ be any ranking of all the alternatives in A. From now on we take this ranking as fixed. We can also represent it as $\bar{a}_1, \bar{a}_2, \ldots, \bar{a}_m$, where $\bar{a}_1$ is the worst alternative and $\bar{a}_m$ is the best according to $\bar{P}$. We want to represent the restrictions of $\bar{P}$ to subsets of A in a way that makes explicit what elements in A are not considered. We denote these restrictions as $\bar{P}(A-B)$; we also represent it as $\bar{a}_1^B \bar{a}_2^B \cdots \bar{a}_{m-|B|}^B$, since $m-|B|$ will be the number of elements to be ranked. To be consistent with notation, notice that $\bar{a}_t \equiv \bar{a}_t^{\emptyset}$.

We shall use these rankings in our description of an outcome function, and in this construction it will be useful to take the convention that $\bar{a}_t^B = \bar{a}_{t+m-|B|}^B$. Given m, it will also be convenient to partition the set of the first $m(m+1)/2$ integers. N_1 will stand for the first $m-1$ integers, N_2 will stand for the next $m-2$ (from m to $2m-3$), N_3 for the next $m-3$ (from $2m-2$ to $3m-6$), etc. We shall rename the integers accordingly, so that (i,j) will denote the ith ranking integer in N_j.

We shall now enlarge the strategy space of player 1 by adding to $\mathscr{P}$ another $(n-1)\times(m-1)!$ strategies. Each set of $(m-1)!$ strategies will be constructed in such a way that the threats to one particular agent (from 2 to n) make truth to be his unique protective strategy. This is why we denote the added strategies for player 1 with a superscript, indicating the player they are addressing. Let $\mathscr{S}^i$ be the set of all added strategies for individual 1. Its elements will be denoted by σ_k^i, with $i=2,3,\ldots,n$, and $k=1,2,\ldots,(m-1)!$. We shall also enlarge the strategy space of player 2, by adding to $\mathscr{P}$ another $(m-1)!$ strategies. These strategies are oriented to guarantee that player 1's truthful relevation of preference is always his unique protective strategy. Let $\mathscr{T}$ be the set of added strategies for individual 2. Its elements will be denoted by τ_h, with $h=1,2,\ldots,(m-1)!$.

The strategy spaces for the players will then the

$$\sum_i = \mathscr{P} \quad \text{for } i=3,\ldots,n,$$

$$\sum_1 = \mathscr{P} \cup \mathscr{S},$$

$$\sum_2 = \mathscr{P} \cup \mathscr{T}.$$

Let us now describe the outcome function ψ for our game form:

$$\psi(s_1, s_2, \ldots, s_n) = \begin{cases} f(s_1, s_2, \ldots, s_n) & \text{if } s_1, \ldots, s_n \in \mathscr{P} \\ \bar{a}_{t+r}^{B(l,P_i)} & \text{if } s_2, \ldots, s_n \in \mathscr{P}, \\ & \quad s_1 = \sigma_k^i, \\ & k = (r, l) \text{ and} \\ & \bar{a}_t \text{ is the } l\text{th} \\ & \text{ranking alternative} \\ & \text{for } s_i = P_i. \\ \bar{a}_{t+r}^{B(l,P_1)} & \text{if } s_1, s_3, \ldots, s_n \in \mathscr{P} \\ & \text{and } s_2 = \tau_k, \\ & k = (r, l) \text{ and} \\ & \bar{a}_t \text{ is the } l\text{th ranking} \\ & \text{alternative for } s_1 = P_1. \\ \text{constant} & \text{otherwise.} \end{cases}$$

This construct is a generalization of the one given in Table 1. Since it is quite involved, it may be helpful to have some specific examples spelled out. Suppose individual 1 uses strategy σ_1^i, and all other players use preferences as strategies. The outcome will depend on the least preferred alternative for player i, $B(1, s_i)$; if $B(1, s_i) = \bar{a}_t$, then $\psi(s_1, s_2, \ldots, s_n) = \psi(\sigma_1^i, P_2, \ldots, P_n) = \bar{a}_{t+1}$. Similarly, $\psi(\sigma_2^i, P_2, \ldots, P_n) = \bar{a}_{t+2}$, etc. and $\psi(s_{m-1}^i, P_2, \ldots, P_n) = \bar{a}_{t+m-1} = \bar{a}_{t-1}$.

If individual 1 uses strategy σ_k^i, where $k \in N_l$, $k = (r, l)$ and all other players use preferences as strategies, the outcome will depend on what alternatives are the worst l for individual i, and also on which of these alternatives is the lth worst ranking. If, for example, $l = 3$ and $k = (2, 3)$, the outcome will depend on what the worst three alternatives are for individual i, $B(3, P_i)$, and on what his third-worst ranking alternative is. Let this alternative, a_3, be equal to $\bar{a}_t^{B(3,P_i)}$. Then the outcome for $\psi(s_1, s_2, \ldots, s_n) = \psi(\sigma_k^i, P_2, \ldots, P_n) = \psi(\sigma_{(2,3)}^i, P_2, \ldots, P_n) = \bar{a}_{t+2}^{B(3,P_i)}$.

Now, if individual 2 uses strategy τ_k, where $k \in N_l$, $k = (r, l)$, and all other players use preferences as strategies, the outcome will be $\psi(P_1, \tau_k, P_2, \ldots, P_n) = \bar{a}_{t+r}^{B(l,P_1)}$, where t is such that $\bar{a}_t^{B(l,P_1)}$ is the lth ranking worst alternative for P_1.

Finally, notice that the function ψ is constant over all combinations of strategies where both agents 1 and 2 use nonpreference strategies.

We claim that the game form $F = [\Sigma_1 \times \Sigma_2 \times \cdots \times \Sigma_n; \psi]$ as defined implements f via protective strategies, because using their true preferences as strategies is always the unique protective equilibrium for all players. Let us first argue that this is true for players 3 to n. Suppose P_k^* are the true preferences of k ($k \in \{3 \ldots n\}$). Any other preference P_k is either dominated by P_k^* or $B(P_k, P_k^*) \cup a_{q+1}(P_k^*)$-equivalent under f. Suppose the latter is true. By adding strategies σ^k to 1's strategy space, the set of complementary profiles having $a_{q+1}(P_k^*)$ as an outcome when k uses strategy P_k^*

is strictly contained into its equivalent when P_k is used, i.e. $g_F(a_{q+1}(P_k^*), P_k^*) \subset g_F(a_{q+1}(P_k^*), P_k)$, and thus P_k^* dominates P_k at level $q+1$ in the extended game. On the other hand, if $P_k^* d_f(P_k^*) P_k$, it will still be the case that $P_k^* d_\psi(P_k^*) P_k$.

For players 1 and 2 we must add to the preceding argument one guaranteeing that no added strategy can ever be protective. This results from the fact that both 1 and 2 have a veto under f, that this veto is preserved by the definition of the new function (except for those cases where ψ is constant, which cannot make a difference) while the added strategies allow for all outcomes, and thus eventually for an agent's worse outcome, to arise.

4. Direct implementation

In this section we provide two examples, one showing that a SCF can be implementable but not directly implementable, and the other that an SCF can be directly implementable but not self-implementable. These examples thus show that each of these notions of implementability must be examined separately under the present solution concept, and it is up to each author to decide which of the three suits his particular concern.

Example 1. A social choice function which is directly implementable but not self-implementable.

Let $I=\{1,2\}$, $A=\{x,y,z,w\}$. For $C\subseteq A$, $P_i\in\mathscr{P}$, let

$$b(C,P_i)=\{a\in C \mid a'P_i a \ \forall a'\in C\}.$$

Now, define $f^1:\mathscr{P}^2\to A$ so that, for all $(P_1,P_2)\in\mathscr{P}^2$, $f^1(P_1,P_2)=A-\{a_3,a_2,a_1\}$, where

$$\begin{aligned} a_1&=b(A,P_1), & A_1&=A-\{a_1\},\\ a_2&=b(A_1,P_2), & A_2&=A_1-\{a_2\}.\\ a_3&=b(A_2,P_1), \end{aligned}$$

In words, f^1 assigns to each preference profile the alternative that would be left after a sequential elimination procedure in which 1, 2 and 1 again eliminate the alternatives they most dislike among those not yet eliminated.

f^1 is not self-implementable because it violates condition LCI. Indeed, for player 1, if P_i is such that $xP_iyP_izP_iw$ and P_i' is such that $xP_i'yP_i'wP_i'z$, then $D_F(P_i')\cap D_F(P_i)=\{P_i,P_i'\}$, although P_i and P_i' are not equivalent under f^1. (Here, $F=[\mathscr{P}^n;f^1]$, is f^1's own-induced game form.)

We will now show that f^1 is implemented by $F^1=[\mathscr{P}^n;\psi^1]$, where ψ^1 is an SCF whose construction is given below.

Let $P^{(1)}\overset{\text{def}}{=}(xyzw)$, $P^{(2)}\overset{\text{def}}{=}(xzyw)$, $P^{(3)}\overset{\text{def}}{=}(yzxw)$, $P^{(4)}\overset{\text{def}}{=}(ywxz)$, $P^{(5)}\overset{\text{def}}{=}(xwyz)$, $P^{(6)}\overset{\text{def}}{=}(zwxy)$. (Here, we have followed the earlier convention that alternatives to the left are worse.) Let $\bar{\mathscr{P}}=\{P^{(1)},P^{(2)},\dots,P^{(6)}\}$.

The SCF ψ^1 is now defined as follows:

(i) for all $P_2 \notin \bar{\mathscr{P}}$, $\psi^1(\boldsymbol{P}) = f^1(\boldsymbol{P})$,

(ii) for all $P_2 \in \bar{\mathscr{P}}$, $\psi^1(P_1, P_2) = \emptyset^{12}(P_1, \sigma_i^2)$ where $P_2 = P^{(i)}$ for some $i = 1, 2, \ldots, 6$ and $\emptyset^{12}$ is the partial outcome function defined by Table 1.

For exactly the same reasons as in the sketch of the proof of Theorem 4, $D_{F^1}(P_1) = \{P_1\}$ for all $P_1 \in \mathscr{P}$. Hence, 1 always reveals his true preference. If 2 reveals a preference $P_2' \in \bar{\mathscr{P}}$, then from Table 1 it is clear that 2 cannot veto any alternative. On the other hand, if 2 reveals a preference $P_2'' \notin \bar{\mathscr{P}}$, then he retains his power to veto one alternative. Suppose now 2's *true* preference is $P^{(1)}$. Clearly, 2's *unique* protective strategy is the strategy $(xywz) \notin \bar{\mathscr{P}}$. Since $(xywz)$ and $P^{(1)}$ are *equivalent* for 2 under f^1, F^1 implements f^1.

Note that this construction has been made possible by the fact that $N(2, \psi^1) > N(2, f^1)$.

Example 2. A social choice function which is implementable but not directly implementable.

Let $I = \{1, 2\}$, $A = \{x_1, x_2, x_3, x_4, x_5\}$.

Using the same notation as in Example 1, let $f^2 : \mathscr{P}^2 \to A$ be such that for all $(P_1, P_2) \in \mathscr{P}^2$,

$$f^2(P_1, P_2) = A - \{a_1, a_2, a_3, a_4\},$$

where

$$\begin{aligned} a_1 &= b(A, P_1), & A_1 &= A - \{a_1\}, \\ a_2 &= b(A_1, P_2), & A_2 &= A_1 - \{a_2\}, \\ a_3 &= b(A_2, P_1), & A_3 &= A_2 - \{a_3\}, \\ a_4 &= b(A_3, P_2), \end{aligned}$$

Here again the outcome is defined by elimination, with players 1 and 2 alternating in vetoing the outcome that is worst to them among those not previously eliminated. Again, f^2 violates LCI, and is thus not self-implementable. It cannot be directly implementable either, because $N(2, f^2) = |P|$, and from the proof of Theorem 3 it is clear that f^2 can be implemented only by a game form under which *both* 1 and 2 have a larger number of nonequivalent strategies than nonequivalent preferences.

However, f^2 satisfies USPA and has 2 vectors, and so by Theorem 4 it follows that f^2 is implementable.

5. Concluding remarks

We have proved that a large class of social choice functions is implementable via protective equilibria, thus strengthening the already known positive result that nondegenerate self-implementable social choice functions exist when the relevant solution concept is that of protective equilibria. We find this line of inquiry of some importance, since protective equilibria is one solution concept having the attractive feature that every player can compute his play while only knowing the game form

from which the actual game he is involved in was drawn from, and his own preferences. This property is shared by dominant strategy equilibria – the difference being that, while dominant strategy equilibria are more compelling than the ones we consider here, we are able to find reasonable functions which are implementable via protective equilibria, an impossible task with dominant strategies.

One possible objection to the use of our notion of protective equilibria could come from observation of our examples in Section 4 and of the construct used in order to prove Theorem 5. In all these cases some strategies of one individual, say i, are dominated by others only because some other player, say j, has some strategy s_j available, but this will never itself be protective, no matter what the preferences of j might be. Is it reasonable to discriminate between strategies of i on the basis of the expected consequences of playing them against a strategy that j will never use? The answer seems to be no, and therefore one may want to consider a different refinement of maximin, under which a player would first eliminate those strategies that will never be protective for the other players, regardless of what their preferences might be, and then compute protective equilibria relative to the uneliminated strategies. Such a concept would be consistent with a general assumption that each individual is ignorant of the preferences of others, and would lead to different results than the ones considered here.

However, we feel that the preceding criticism does not destroy the interest of protective equilibria – it just forces us to specify the type of information, or lack of it, characterizing the games we have in mind. Individuals may not only be unaware of the characteristics of their opponents, but they can well ignore the type of strategic considerations that will guide their opponents' actions. The preceding argument that a protective individual should disregard those threats coming from strategies which will never be protective for others assumes that protective individuals regard their opponents as being protective too. This symmetry assumption is unnecessary, and it may well be that the opponents are guided by different considerations – they may be trying to maximize their (expected) gain, say, rather than to maximize their minimum gain.

Thus, the notion of protective equilibria may apply to some attractive frameworks where information is lacking. The examples in Section 4 constitute, then, an interesting proof that the relationship between different types of implementation is not always as 'nice' as it is when the relevation principle holds.

Appendix

In order to bring out the restrictive nature of USPA, we show that this property is violated by all members of two wide classes of social choice functions.

A.1. Condorcet social choice functions (CSCF)

Let $D_f \subset 2^I$ be a set of *decisive* coalitions. For each preference profile $\boldsymbol{P}$, define

a binary relation $\boldsymbol{R}$ over A as follows:

$$\text{for all } x, y \in A,\ xRy \Leftrightarrow \{i \in I \mid yP_i x\} \notin D_f. \tag{3.1}$$

f is a *Condorcet social choice function* iff there is a set of decisive coalitions relative to which, $\forall \boldsymbol{P} \in \mathscr{P}^n$, $[CW(f,\boldsymbol{P}) \equiv \{x \in A \mid xRy\ \forall y \in A\} \neq \emptyset] \rightarrow [f(\boldsymbol{P}) \in CW(f;\boldsymbol{P})]$. Hence, R is the familiar social weak preference relation, and a CSCF selects a pairwise 'best' element if such an element exists. In order to economize on space, we appeal to the reader's familiarity with the several impossibility theorems in this context and merely assert that we consider Condorcet social choice functions for which $CW(f,\boldsymbol{P}) = \emptyset$ for some $\boldsymbol{P} \in \mathscr{P}^n$. Clearly, the class of such social choice functions is very large. For example, if $|A| \geq |I|$, every CSCF under which there is no vetoer belongs to this class. See Craven (1971) for the general restriction on the size of decisive coalitions and the relative cardinalities of I and A.

Proposition A1. *If f is a CSCF, and $CW(f,\boldsymbol{P}) = \emptyset$ for some profile $\boldsymbol{P}$, then f violates USPA.*

Proof. Consider $\boldsymbol{P} \in \mathscr{P}^n$ such that $CW(f,\boldsymbol{P}) = \emptyset$. Let $f(\boldsymbol{P}) = x$. Since $x \notin CW(f,\boldsymbol{P})$, there is y such that $\sim xRy$. Let $c = \{i \in I \mid yP_i x\}$. Then $c \in D_f$. Choose $\boldsymbol{P}'$ such that

(a) $\forall i \notin c$, $P_i = P_i'$;

(b) $\forall i \in c$, $[yP_i z$ for all $z \in A - \{y\}$ and P_i and P_i' agree on $\{z \in A \mid x = z$ or $xP_i z\}$.

By USPA, we should have $f(\boldsymbol{P}') = x$. But, since $c \in D_f$ and $yP_i z$ for all $i \in c$ and all $z \in A - \{y\}$, $CW(f,\boldsymbol{P}') = \{y\}$. Hence, $f(\boldsymbol{P}') = y$, which is the desired contradiction.

A.2. Positional social choice functions (PSCF)

Where m is the number of alternatives, an m-vector $e = (e_1, e_2, \ldots, e_m)$ is a positional scoring vector if $e_1 \geq e_2 \geq \cdots \geq e_m$ with $e_1 > e_m$.

For any P_i, for any $x \in A$,

$$w(x, P_i) \overset{\text{def}}{=} e_{m+1-k}, \quad \text{when } x = a_k(P_i),$$

and

$$w(x, \boldsymbol{P}) \overset{\text{def}}{=} \sum_{i \in I} w(x, P_i).$$

f is a positional SCF iff for all $\boldsymbol{P}$, $f(\boldsymbol{P}) \in \{a \in A \mid w(a,\boldsymbol{P}) \geq w(b,\boldsymbol{P})\ \forall b \in A\}$.

Call f a k-positional SCF if f is based on the positional scoring vector $\boldsymbol{e}$ such that $e_1 = e_2 = \cdots = e_{k-1} > e_k$. In other words, k is the *smallest* integer for which $e_1 > e_k$.

Proposition A2. *For all $k < m$, a k-positional SCF satisfies USPA only if $k \geq n$, where n is the number of individuals.*

Proof. Suppose $n>k$. Since $n>k$ and $k<m$, there are at least $(k+1)$ alternatives and $(k+1)$ individuals. Let the alternatives be $a_1, a_2, \ldots, a_m$.

Construct the following profile:

	Individuals							This set may be empty if $n=k+1$
Ranking	1	2	3	...	$k-1$	k	$k+1$	$k+2 \ldots n$
1	a_{k+1}	a_{k+1}	a_{k+1}		a_{k+1}	a_1	a_1	Every individual i in
2	a_k	a_k	a_k		a_k	a_2	a_2	this set ranks alternatives
3	a_2	a_3	a_4		a_1	a_3	a_3	in the order
⋮	⋮	⋮	⋮		⋮	⋮	⋮	$a_1, a_2, \ldots, a_k, a_{k+1}$
$k-1$	a_{k-2}	a_{k-1}	a_1		a_{k-3}	a_{k-1}	a_{k-1}	
k	a_{k-1}	a_1	a_2		a_{k-2}	a_k	a_{k+1}	
$k+1$	a_1	a_2	a_3		a_{k-1}	a_{k+1}	a_k	
$k+2$ ⋮ m	Any ordering of remaining alternatives, if any.							

Since $n>k$, there are two possibilities:
(1) $n=k+1$;
(2) $n>k+1$.

Case 1: $n=k+1$. Then, for each $i \in \{1, \ldots, k+1\}$,

$$w(a_i, \boldsymbol{P}) = (k-1)e_1 + e_k + e_{k+1}. \tag{A1}$$

Clearly, $f(P) \in \{a_1, \ldots, a_{k+1}\}$. Note that for each $i \in \{1, \ldots, k+1\}$, there is an individual j who ranks a_i in the $(k+1)$th position. So, without loss of generality, let $f(\boldsymbol{P}) = a_1$.

Consider P_1' such that a_{k-1} and a_{k-2} switch positions. Then,

$$w(a_{k-1}, (P_1', P_{-1})) = ke_1 + e_{k+1}. \tag{A2}$$

Since $e_1 > e_k$, (A1) and (A2) $\rightarrow f(\boldsymbol{P}) = a_{k-1}$.

However, by USPA, we should have $f(\boldsymbol{P}) = a_1$.

Case 2: $n>k+1$. Then, for each $i \in \{1, \ldots, k-1\}$,

$$\begin{aligned} & w(a_i, \boldsymbol{P}) = (n-2)e_1 + e_k + e_{k+1}, \\ \text{while} \quad & w(a_k, \boldsymbol{P}) = (k-1)e_1 + e_{k+1} + (n-k)e_k, \\ & w(a_{k+1}, \boldsymbol{P}) = (k-1)e_1 + e_k + (n-k)e_{k+1}. \end{aligned} \tag{A3}$$

Hence, $f(\boldsymbol{P}) \in \{a_1, \ldots, a_{k-1}\}$.

Let $f(\boldsymbol{P}) = a_i$. Note that for individual i, a_i is in the $(k+1)$th position, while the alternative in the kth position also belongs to the set $\{a_1, a_2, \ldots, a_{k-1}\}$. The rest of the proof is the same as in Case 1.

Corollary. *If $n \geq m$, and a positional f satisfies USPA, then f is m-positional.*

Remark. The plurality rule $(1, 0, \ldots, 0)$ and the Borda rule $(m, m-1, \ldots, 1)$ violate USPA.

References

S. Barberá and B. Dutta, Implementability via protective equilibria, Journal of Mathematical Economics 10 (1982) 49–65.

J. Craven, Majority voting and social choice, The Review of Economic Studies XXXVIII (2) (April 1971) 265–267.

J.J. Lafont and E. Maskin, The theory of incentives: an overview, in: Hildenbrand, ed., Advances in Economic Theory (Cambridge University Press, 1983).

H. Moulin, Prudence versus sophistication in voting strategy, Journal of Economic Theory 24 (1981) 318–412.

H. Moulin, Non-cooperative implementation: a survey of recent results, Mathematical Social Sciences 3 (1982) 243–257.

E. Muller and M.A. Satterthwaite, The equivalence of strong positive association and strategy-proofness, Journal of Economic Theory 14 (1977) 412–418.