Discussion Paper No. 965

Implementing Social Choice Functions:
A New Look at Some Impossibility Results

by

Matthew O. Jackson
and
Sanjay Srivastava[*]

First Draft: September 1990

This Draft: May 1991

For some solution concepts, such as dominant strategies, Nash equilibrium, and undominated strategies, only dictatorial social choice functions are implementable on a full domain of preferences with at least three alternatives. For other solution concepts, such as the iterative removal of weakly dominated strategies, undominated Nash equilibrium, and maximin, it is possible to implement non-dictatorial social choice functions. We begin by offering simple proofs of several of the "impossibility" results. These proofs provide intuition into the properties of a solution concept which make it impossible to implement non-dictatorial social choice functions. This allows us to provide a characterization of solution concepts which lead to impossibility results, as well as two easily checked sufficient conditions.

[*]Kellogg Graduate School of Management, Northwestern University, Evanston, IL 60208-2001, and Graduate School of Industrial Administration, Carnegie Mellon University, Schenley Park, Pittsburgh, PA 15213-3890, respectively. We gratefully acknowledge financial support from the National Science Foundation.

Proposed Running Head: Implementing Social Choice Functions

Correspondence: Matthew Jackson, Kellogg Graduate School of Management, Northwestern University, Evanston, IL 60208-2001.

1. Introduction

The Gibbard (1973)-Satterthwaite (1975) Theorem states that a strategy–proof social choice function on an unrestricted domain of preferences must be dictatorial if it takes on at least three values. Equivalently, the result says that interesting social choice functions cannot be implemented in dominant strategies. This restriction to dictatorial social choice functions is often attributed to the strength of the requirement that there exist a dominant strategy for each agent and every preference profile.

It is interesting, however, that similar results obtain for much weaker solution concepts. Jackson (1989) shows that if a social choice function can be implemented in undominated strategies by a bounded mechanism[1] on a full domain of preferences, then it must be dictatorial. Undominated strategies is a very weak solution concept, quite the opposite of dominant strategies. Most games have undominated strategies: for instance, all mechanisms with finite action spaces are bounded and have undominated strategies for every preference profile. An impossibility result also holds for Nash equilibrium. The only social choice functions which take on at least three values and are Nash implementable are dictatorial, as shown Dasgupta, Hammond and Maskin (1979). These results indicate that it is not the strength of the solution concept which makes it impossible to implement interesting social choice functions .

In order to understand what makes it impossible to implement interesting social choice functions , it is important to recognize that there are solution concepts which avoid the negative results. Interesting social choice functions can be implemented on a full domain of preferences via undominated Nash equilibria, the iterated removal of weakly dominated strategies, maximin strategies, and other solutions. These observations lead to the following question: Is there some property of a solution concept which prevents it from implementing interesting social choice functions ?

The obvious answer to this question is that if a solution concept has a strategy–proof outcome function (for any mechanism for which it always predicts a single outcome), then it will be impossible implement non–dictatorial social choice functions via that solution

---

[1] A mechanism is bounded if, for each weakly dominated action, there exists and undominated action which dominates it. Definitions of various solution concepts and restrictions on mechanisms are provided in Section 2.

concept. We offer a direct proof of this fact (Theorem 2). Although this theorem is simply an extension of the Gibbard–Satterthwaite Theorem, we offer a proof for two reasons. First, the proof we offer is substantially simpler than existing proofs of the Gibbard–Satterthwaite theorem.[2] Second, the proof of Theorem 2 makes clear the critical role of the requirement that a solution lead to a single outcome for every preference profile; which helps us to identify the aspects of a solution concept which lead to an impossibility result.

Although Theorem 2 provides an obvious property common to solution concepts which are unable to implement interesting social choice functions , this property is not as useful as one might hope. For some solution concepts and mechanisms, it is difficult to verify that the outcome function is strategy–proof. With this in mind, we define conditions which are sufficient for an impossibility result (Theorem 3), and which are easily checked globally (without reference to a particular mechanism).

These conditions are called positive responsiveness and direct breaking. Roughly, positive responsiveness states that a solution accounts for improvements available to any agent. The direct breaking condition states that in certain circumstances, if a set of actions is not a solution then some agent can improve via some deviation. In addition to being easily verified, the conditions also provide intuition into the impossibility results. One may interpret the direct breaking condition as saying that actions to be ruled out are broken directly; or in a sense, "on the proposed equilibrium path". More precisely, if only one agent's preferences change and and the outcomes of the mechanism completely change, then for some set of actions which were a solution, some agent can do strictly better by deviating. The direct breaking condition is satisfied by undominated strategies (on bounded mechanisms), Nash equilibrium, and the iterative elimination of strictly dominated strategies. However, undominated Nash equilibrium, the iterated removal of weakly dominated strategies, and maximin strategies, which do not lead to impossibility results, do not satisfy the direct breaking condition, and in a sense incorporate "off the equilibrium path" information.

The paper is organized as follows. First, we provide definitions and simple proofs of impossibility theorems for several solution concepts. Next, we use the intuition from these impossibility theorems to prove a general impossibility theorem. The key steps in the proof

---

[2] An exception is a recent proof by Barbera and Peleg (1990). Our approach, however, offers different insight into the result. This is discussed more fully in Section 3.

3

of the general theorem lead us to define the positive responsiveness and direct breaking conditions and prove that they are sufficient for an impossibility result. Finally, we discuss potential extensions to allow for indifference and solutions which account for coalitional incentives. A table at the end of the paper summarizes the impossibility theorems.

2. Definitions

The finite set of alternatives is denoted A. It is assumed that $\#A \geq 3$.

The society is composed of a finite number, N, of individuals.

Individual preferences are represented by a binary relation which is complete, asymmetric, and transitive. [3] We use the notation P^i to represent such a binary relation for agent i, and for $a \neq b$ read $aP^i b$ to mean that i prefers a to b. Let $\mathcal{P}$ denote the set of all such strict preferences over A.

A *social choice function* is a map which associates an alternative to each preference profile. We use F to represent a social choice function, $F : \mathcal{P}^N \rightarrow A$. We assume that F is onto A.[4]

A social choice function is *strategy–proof* if for each i, P, and $\widehat{P}^i$ either $F(P) = F(P^{-i}, \widehat{P}^i)$ or $F(P)P^i F(P^{-i}, \widehat{P}^i)$.

A social choice function is *dictatorial* if there exists i such that $F(P)P^i a$ for all $P \in \mathcal{P}$ and $a \neq F(P)$ in the range of F.

A *mechanism* is a game form (M, g), where $M = M^1 \times \cdots \times M^N$ and $g : M \rightarrow A$. The set of mechanisms to be considered for the implementation problem is denoted $\mathcal{G}$.

A *solution* is a correspondence which indicates the set of actions which might be played for a given game form and profile of preferences. We denote solutions by S where S :

---

[3] The combination of completeness and asymmetry rules out indifference. As pointed out by previous authors, considering only strict preferences actually provides for stronger results than allowing for indifference since it is a more restricted domain. For the implementation problem, the consideration of indifference leads to some difficulties, which we discuss in our concluding remarks.

[4] This assumption is without loss of generality for our analysis. Any solution concept satisfying the conditions of the theorems we provide, depends only on the preferences of agents over alternatives available as outcomes of the game form being played. Furthermore, a solution which satisfies the conditions of any of our theorems also satisfies a unanimity condition: an alternative which is most preferred by all agents and is available via the game form will be an outcome. Taking these two observations together, we can let A be the range of F.

$\mathcal{G} \times \mathcal{P}^N \to 2^M$. Thus, $m \in S[(M,g),P]$ indicates that m is a solution under S to (M,g) at P.

The *outcome correspondence* associated with S is $O_S : \mathcal{G} \times \mathcal{P}^N \to 2^A$ is defined by

$$O_S[(M,g),P] = \{a \in A \mid \exists m \in S[(M,g),P] \text{ s.t. } g(m) = a\}.$$

A social choice correspondence F is implementable via the solution S if there exists a mechanism (M,g) such that $O_S[(M,g),P] = F(P)$ for all $P \in \mathcal{P}^N$.

SOLUTION CONCEPTS.

Although most of the solution concepts we discuss are well–known, we provide definitions for them as used in an implementation context. An incomplete list of references for implementation via the solutions we discuss includes: Maskin (1977), (1985) for Nash implementation Palfrey and Srivastava (1986) and Jackson, Palfrey and Srivastava (1990) for undominated Nash implementation, Farquharson (1969), Moulin (1979), (1983), and Abreu and Matsushima (1990), for the iterated removal of weakly dominated strategies, Jackson (1989) and Börgers (1991) for undominated strategies, Thomson (1979) and Moulin (1982) for maximin strategies, Barbera and Dutta (1982) for protective equilibria, Herrero and Srivastava (1989) for implementation via backward induction, and Moore and Repullo (1988) for subgame perfect implementation. Moore (1991) provides an excellent survey of the (complete information) implementation literature.

DOMINANT STRATEGIES.

An action $m^i \in M^i$ is a *dominant strategy* for agent i at P^i if for each m^{-i} and $\widehat{m}^i$ either $g(m^i, m^{-i}) = g(\widehat{m}^i, m^{-i})$ or $g(m^i, m^{-i})P^i g(\widehat{m}^i, m^{-i})$.

UNDOMINATED STRATEGIES.

The action $\widehat{m}^i \in M^i$ *dominates* $m^i \in M^i$ at P^i if for each m^{-i} either $g(\widehat{m}^i, m^{-i}) = g(m^i, m^{-i})$ or $g(\widehat{m}^i, m^{-i})P^i g(m^i, m^{-i})$, with the preference being strict for some m^{-i}. The action m^i is *undominated* at P^i if it is not dominated by any other action.

STRICT DOMINATION.

5

The action $m^i \in M^i$ is *strictly dominated* by $\widehat{m}^i \in M^i$ at P^i if $g(\widehat{m}^i, m^{-i})P^i g(m^i, m^{-i})$ for each m^{-i}. The action m^i is *strictly undominated* if it is not strictly dominated by any $\widehat{m}^i$.

ITERATIVE REMOVAL OF DOMINATED STRATEGIES.

Given a mechanism (M, g) and sets $X^1 \succeq M^1, \ldots, X^N \succeq M^N$, an action $m^i \in X^i$ is dominated by $\widehat{m}^i \in X^i$ at P^i with respect to X if for each $m^{-i} \in X^{-i}$ either $g(\widehat{m}^i, m^{-i}) = g(m^i, m^{-i})$ or $g(\widehat{m}^i, m^{-i})P^i g(m^i, m^{-i})$, with the preference being strict for some $m^{-i} \in X^{-i}$. Let $D^i(X, P)$ be the set of actions which are not dominated for i at P^i with respect to X, and let $D(X, P) = D^1(X, P) \times \cdots \times D^N(X, P)$. Define a sequence $D_0(M, P), \ldots D_K(M, P) \ldots$ by $D_0(M, P) = D(M, P)$ and $D_K(M, P) = D(D_{K-1}(M, P), P)$. Finally, let $D^*(M, P) = \cap_K D_K(M, P)$. An action $m \in D^*(M, P)$ is said to be *iteratively undominated* at P.

Correspondingly, we can define *iteratively strictly undominated* by using strict domination instead of domination in the above definition.

NASH EQUILIBRIUM.

A (pure strategy) *Nash equilibrium* at P is a profile of actions $m \in M$ such that for each i and $\widehat{m}^i$ either $g(m) = g(\widehat{m}^i, m^{-i})$ or $g(m)P^i g(\widehat{m}^i, m^{-i})$.

UNDOMINATED NASH EQUILIBRIUM.

The actions $m \in M$ form an *undominated Nash equilibrium* at $P \in \mathcal{P}^N$ if m is a Nash equilibrium and each m^i is undominated at P^i.

MAXIMIN.

An action $m^i \in M^i$ is a *maximin* action for agent i at P^i, if for any $\widehat{m}^i \in M^i$ there exists $\widehat{m}^{-i}$ such that for each m^{-i} either $g(m^i, m^{-i}) = g(\widehat{m}^i, \widehat{m}^{-i})$ or $g(m^i, m^{-i})P^i g(\widehat{m}^i, \widehat{m}^{-i})$. Under maximin, agents 'rank' their strategies in terms of the worst outcomes they might lead to, and select from among those with the best worst outcome.

BOUNDED MECHANISMS.

A mechanism (M, g) is *bounded at P* if whenever m^i is dominated at P^i, there exists a undominated $\widehat{m}^i$ which dominates it. (M, g) is *bounded* if it is bounded at each $P \in \mathcal{P}$.

3. Implementing a Social Choice Function

Our starting point is the following version of the Gibbard–Satterthwaite Theorem.

THEOREM [GIBBARD (1973), SATTERTHWAITE (1975)]. *If a social choice function has at least three elements in its range, then it is implementable in dominant strategies if and only if it is dictatorial.*

Much of the intuition behind the Gibbard–Satterthwaite theorem, has been that the negative result derives from the strength of the requirement that a social choice function be strategy-proof on a full domain of preferences. Equivalently, from an implementation standpoint, the difficulty arises in attempting to construct a mechanism for which each agent always has a dominant strategy.

It is perhaps surprising that similar results occur for weaker solution concepts. In fact, the negative result holds for a wide variety of solution concepts. This includes implementation in undominated strategies by bounded mechanisms, Nash implementation, and implementation via the iterated elimination of strictly dominated strategies. We discuss these next, and in each case provide a particularly simple proof of the result. These proofs highlight the intuition that the inability to implement non–dictatorial social choice functions derives from the requirement that a solution always predict a single outcome.

THEOREM [JACKSON (1989)]. *If a social choice function has at least three elements in its range, then it is implementable in undominated strategies by a bounded mechanism if and only if it is dictatorial.*

The restriction to bounded mechanisms is critical to the theorem above. If we allow unbounded mechanisms, then the result is quite the opposite: *any* social choice function can be implemented in undominated strategies, as shown in Theorem 1 of Jackson (1989). The importance of boundedness is discussed in more detail in Example 3 below.

PROOF: Here, we offer a direct proof of the theorem for $N = 2$ and F has three elements in its range. This carries most of the intuition. Other cases are covered in the appendix. Let F be implemented by the bounded mechanism (M, g). The notation $P^i = (a, b, c)$ indicates that a is strictly preferred to b which is strictly preferred to c.

(1) There exist i, $m^i \in M^i$, and c such that $g(m^i, m^{-i}) = c$ for all $m^{-i} \in M^{-i}$.

Case 1. $g(m) = c$, for some P and m which is undominated at P, where c is j's worst outcome. In this case, since F is single valued, $g(m^i, \tilde{m}^j) = c$ for all $\tilde{m}^j$ which are undominated at P^j. Since (M, g) is bounded, it must be that $g(m^i, \hat{m}^j) = c$ for all $\hat{m}^j$.

Case 2. For any P and m which is undominated at P, $g(m)$ is not the worst outcome of either agent. In this case, if $\tilde{m}$ is undominated at $\tilde{P}^1 = (b, a, c), \tilde{P}^2 = (c, a, b)$, then $g(\tilde{m}) = a$. Similarly, if $\hat{m}$ is undominated at $\hat{P}^1 = (c, b, a), \hat{P}^2 = (a, b, c)$, then $g(\hat{m}) = b$. At $(\tilde{P}^1, \hat{P}^2)$, both $\tilde{m}^1$ and $\hat{m}^2$ are undominated. There is an undominated action $\overline{m}^2$ such that $g(\tilde{m}^1, \overline{m}^2) = a$ [since $g(\tilde{m}) = a$ makes a available to agent 2]. Likewise, there is an undominated action $\overline{m}^1$ such that $g(\overline{m}^1, \hat{m}^2) = b$. This implies that $F(\tilde{P}^1, \hat{P}^2) = \{a, b\}$, which contradicts the fact that F is a function.

(2) F is dictatorial.

Identify i, c and m^i from (1). We show that for any $a \neq c$, there exists $\tilde{m}^i$ such that $g(\tilde{m}^i, m^j) = a$ for all m^j. This [coupled with (1)] implies that i's undominated strategies are constant and provide i's most preferred alternative. Thus i dictates. Let $\overline{P}^i = (a, c, b)$, $\overline{P}^j = (b, a, c)$. Since $g(\hat{m}) = a$ for some $\hat{m}$, there exists an undominated $\overline{m}^i$ such that $g(\overline{m}^i, \hat{m}^j) = a$. If $\overline{m}^j$ is undominated at $\overline{P}^j$, it follows that $g(\overline{m}) = a$ or $g(\overline{m}) = b$. Since $g(m^i, \overline{m}^j) = c$ and $c\overline{P}^i b$, there exists an undominated m^i such that $g(\overline{m}) \neq b$. Thus, since F is single valued, $g(\overline{m}) = a$ for any $\overline{m}$ which is undominated at $\overline{P}$.

Let $\tilde{P}^i = (a, b, c)$. Either $\overline{m}^i$ is undominated at $\tilde{P}^i$, or it is dominated by an undominated action $\tilde{m}^i$ such that $g(\tilde{m}^i, \overline{m}^j) = a$. Since F is single valued, $g(\tilde{m}^i, \overline{m}^j) = a$ for any $\tilde{m}^i$ which is undominated at $\tilde{P}^i$.

Let $\tilde{P}^j = (b, c, a)$ and consider $\tilde{m}^j$ which is undominated at $\tilde{P}^j$. It follows that $g(\tilde{m}) \neq b$, since otherwise $g(\tilde{m}^i, \overline{m}^j) = b$. However, $g(m^i, \tilde{m}^j) = b$ for some m^i, since b is in the range of (M, g) and is most preferred by j. Since i prefers b to c, this implies that $g(\tilde{m}) \neq c$. It follows that $g(\tilde{m}) = a$.

Since a is least preferred by j at $\tilde{P}^j$ and F is single valued, $g(\tilde{m}^i, m^j) = a$ for any undominated m^j. Since (M, g) is bounded, $g(\tilde{m}^i, m^j) = a$ for all m^j. ∎

A similar result holds for Nash implementation.

THEOREM [DASGUPTA, HAMMOND, AND MASKIN (1979)]. *If a social choice function has at least three elements in its range, then it is Nash implementable if and only if it is dictatorial.*

The theorem stated above also holds for the case of $N = 2$ and $\#A = 2$, as shown by Maskin (1977) and Hurwicz and Schmeidler (1978). A simple proof for that case is presented in Jackson and Srivastava (1991).

PROOF: We provide a proof for $N = 2$ and $\#A = 3$. The extension to $N \geq 2$ and $\#A \geq 3$ is covered in the appendix. Consider F which is Nash implemented by the mechanism (M, g).

(1) There exist i, $m^i \in M^i$, and c such that $g(m^i, m^{-i}) = c$ for all $m^{-i} \in M^{-i}$.

Case 1. $g(m) = c$, for some P and m which is a Nash equilibrium at P, where c is j's worst outcome. Since m is a Nash equilibrium, it must be that $g(m^i, \tilde{m}^j) = c$ for all $\tilde{m}^j$.

Case 2. For any P and m which is a Nash equilibrium at P, $g(m)$ is not the worst outcome of either agent. In this case, if $\tilde{m}$ is a Nash equilibrium at $\tilde{P}^1 = (b, a, c), \tilde{P}^2 = (c, a, b)$, then $g(\tilde{m}) = a$. Similarly, if $\hat{m}$ is a Nash equilibrium at $\hat{P}^1 = (c, b, a)$, $\hat{P}^2 = (a, b, c)$, then $g(\hat{m}) = b$. It then follows that at $(\tilde{P}^1, \hat{P}^2)$ both $\tilde{m}$ and $\hat{m}$ are Nash equilibria, which contradicts the fact that F is single valued.

(2) F is dictatorial.

Identify i, c and m^i from (1). We show that for any $a \neq c$, there exists $\tilde{m}^i$ such that $g(\tilde{m}^i, m^j) = a$ for all m^j. This [coupled with (1)] implies that the Nash equilibria must result in i's i's most preferred alternative and so i dictates.

Consider a Nash equilibrium $\tilde{m}$ at preferences $\tilde{P}^i = (a, b, c)$ and $\tilde{P}^j = (c, b, a)$. It must be that $g(\tilde{m}) = a$ and so $g(\tilde{m}^i, m^j) = a$ for all m^j. To see this, suppose that $g(\tilde{m}) \neq a$. Then since m is a Nash equilibrium at $\tilde{P}$, $g(\hat{m}^i, \tilde{m}^j) \in \{b, c\}$ for all $\hat{m}^i$. It follows that $(m^i, \tilde{m}^j)$ (where m^i is identified in (1)) is a Nash equilibrium at $\overline{P}^i = (a, c, b)$, $\overline{P}^j = (a, c, b)$. There exists $\overline{m}$ such that $g(\overline{m}) = a$. It follows that $\overline{m}$ is also Nash equilibrium at $\overline{P}$, contradicting the fact that F is single valued. ∎

A similar result obtains for implementation by elimination of strictly dominated strategies. One way to prove this is to show that any social choice function implemented in this way is monotonic, and then to apply the Muller–Satterthwaite (1977) theorem. (The same is true for Nash implementation). Instead, we offer a direct proof which illustrates how properties of the implementing mechanism yield the result.

THEOREM 1. *If a social choice function has at least three elements in its range,[5] then it can be implemented by the iterated elimination of strictly dominated strategies if and only if it is dictatorial.*

We remark that while the result for undominated strategies relies on a restriction to bounded mechanisms, the above result does not require any limitation on mechanisms. The

---

[5] As with Nash implementation, the result holds for the case $N = 2$ and $\#A = 2$. This can be shown by a proof analagous to the one in Jackson and Srivastava (1991).

strict nature of the domination and the restriction to finite A places a natural bound on any string of dominating actions.

PROOF: Here we prove the theorem for $N = 2$ and $\#A = 3$. The case $N \geq 2$ and $\#A \geq 3$ appears in the appendix. Let F be implemented by the iterated elimination of strictly dominated strategies via the mechanism (M, g).

(1) There exist i, $m^i \in M^i$, and c such that $g(m^i, m^{-i}) = c$ for all $m^{-i} \in M^{-i}$.

Case 1. $g(m) = c$, for some P and m which is iteratively strictly undominated at P, where c is j's worst outcome. It follows that $F(P) = c$. Without loss of generality, let $P^j = (a, b, c)$. Any $\widetilde{m}^j$ such that $g(m^i, \widetilde{m}^j) = a$ would also be iteratively strictly undominated since a is maximal for j. Since F is single valued it must be that $g(m^i, \widetilde{m}^j) \neq a$ for all $\widetilde{m}^j$. Then by similar reasoning (given that a cannot be achieved against m^i) $g(m^i, \widetilde{m}^j) \neq b$ for all $\widetilde{m}^j$.

Case 2. For any P and m which is iteratively strictly undominated at P, $g(m)$ is not the worst outcome of either agent. In this case, if $\widetilde{m}$ is iteratively strictly undominated at $\widetilde{P}^1 = (b, a, c), \widetilde{P}^2 = (c, a, b)$, then $g(\widetilde{m}) = a$. Similarly, if $\widehat{m}$ is iteratively strictly undominated at $\widehat{P}^1 = (c, b, a), \widehat{P}^2 = (a, b, c)$, then $g(\widehat{m}) = b$. It then follows that at $(\widetilde{P}^1, \widehat{P}^2)$ both $\widetilde{m}$ and $\widehat{m}$ are iteratively strictly undominated, which contradicts the fact that F is single valued.

(2) F is dictatorial.

Identify i, c and m^i from (1). We show that for any $a \neq c$, there exists $\widetilde{m}^i$ such that $g(\widetilde{m}^i, m^j) = a$ for all m^j. This [coupled with (1)] implies that i dictates.

Suppose that there does not exist such a $\widetilde{m}^i$. Then at preferences $\widetilde{P}^i = \widetilde{P}^j = (a, c, b)$, the action m^i [defined in (1)] is not strictly dominated. Given the structure of m^i, none of agent j's actions are strictly dominated as long as m^i has not been removed by i. It follows that m^i, m^j is iteratively strictly undominated for any m^j at $\widetilde{P}$ and so $F(\widetilde{P}) = c$. It is also true that any $\widehat{m}$ such that $g(\widehat{m}) = a$ is iteratively strictly undominated at $\widetilde{P}$, contradicting the fact that F is single valued. ∎

The proofs given above illustrate that much of the force behind the negative results is the fact that F is single valued. For each of the previous theorems, the proofs follow similar reasoning: If F is single valued, then some agent must be able to enforce some outcome (part (1) of each proof). Next, we show that the same agent must be able to enforce every outcome and thus is a dictator (part (2) of each proof).

There are solution concepts for which this line of reasoning is not valid. Such solution concepts can implement non-dictatorial social choice functions . We now discuss two

10

such solution concepts. The first is undominated Nash equilibrium, which is stronger than either undominated strategies or Nash equilibrium, and weaker than dominant strategies. This indicates that the strength of the solution concept has little to do with producing an impossibility result.

EXAMPLE 1. *Undominated Nash implementation.*

The following mechanism allows both agents some say in the selection of an outcome, and yet it has a unique undominated Nash equilibrium for any preference profile.

	m^2	$\overline{m}^2$
m^1	a	a
$\overline{m}^1$	b	c

The mechanism represented above always has a unique undominated Nash equilibrium. The column player always has a unique undominated action, depending on the preference between b and c. The row player has a unique best response to this action, which completes the equilibrium. Notice that an iterated elimination of (weakly) dominated strategies will lead to the same predictions as the undominated Nash equilibria for this mechanism.[6]

If we examine other solution concepts applied to the above mechanism, such as undominated strategies, Nash equilibrium, or dominant strategies, they do not lead to a unique prediction for the above mechanism at some preference profiles. At some profiles there are more than one predicted outcomes for the undominated strategy or Nash solution concepts, while agent 1 has no dominant strategy.

Another solution concept which permits a positive implementation result is maximin strategies.[7]

---

[6] More discussion of interesting social choice functions which can be implemented by an iterated elimination of weakly dominated strategies on a full domain of preferences is given in Moulin (1982), (1983), and Herrero and Srivastava (1989).

[7] For more discussion of implementation via maximin see Thomson (1979). Example 2 also works for implementation via protective equilibria [Barbera and Dutta (1982)], which is a refinement of the set of maximin strategies [see Barbera and Jackson (1988)].

EXAMPLE 2. *Maximin Strategies.*

The following mechanism shows that the maximin solution is single valued on a mechanism which is not dictatorial. In fact, the maximin outcome function for the mechanism below is anonymous. One way to think of this mechanism is that each agent can veto a single outcome. The unique maximin solution is to veto your worst outcome.

	m^2	$\overline{m}^2$	$\tilde{m}^2$
m^1	a	a	b
$\overline{m}^1$	a	c	c
$\tilde{m}^1$	b	c	c

4. Comparing Solution Concepts.

The proceeding discussion leads to the following question: Can we characterize the solution concepts which lead to impossibility results? Such a characterization is provided by Theorem 2 below.

Before stating Theorem 2, we remark that the condition it identifies need only be satisfied on mechanisms for which a solution provides a single valued outcome function. This makes it easy to verify for those solutions which satisfy it everywhere, but much more difficult to verify for solution concepts which only satisfy it for some mechanisms. Thus following Theorem 2 we offer Theorem 3 which provides sufficient conditions for an impossibility result. The sufficient conditions turn out to be easier to verify for most solutions.

STRATEGY RESISTANCE.

A solution S is *strategy-resistant*[8] with respect to the mechanism (M, g) if for each i, $P \in \mathcal{P}^N$, and $\hat{P}^i \in \mathcal{P}$, there exists $m \in S[(M, g), P]$ and $\hat{m} \in S[(M, g), P^{-i}, \hat{P}^i]$ such that either $g(m) = g(\hat{m})$ or $g(m) P^i g(\hat{m})$.

---

[8] A condition of strategy-resistance is defined for social choice correspondences in Jackson (1989). The condition stated here requires that a solution lead to a strategy-resistant outcome function in the sense of Jackson (1989).

12

The strategy–resistance condition states that agents do not wish that they had preferences different from their true preferences, in the sense that those preferences lead to outcomes, *all of which* are better according to their true preferences, than the outcomes they get with their true preferences. The term strategy–resistance is a bit of a misnomer here, since agents do not have the ability to change their preferences (recall that this is a condition on a solution concept).

THEOREM 2. *The solution S is strategy–resistant with respect to a mechanism via which it implements a social choice function (with at least three outcomes in its range), if, and only if, the social choice function is dictatorial.*

The above theorem is easily proven using the Gibbard–Satterthwaite theorem. We offer a direct proof which provides insight to the impossibility results. The proof is similar in structure to those of the preceeding theorems. It is first shown that strategy–resistance and single valuedness imply that some agent can enforce an outcome. It is then shown that the same agent can enforce any outcome. The proof is easily modified to prove either the Gibbard–Satterthwaite theorem or Muller–Satterthwaite theorem. As mentioned in the introduction, the proof offered here is substantially simpler than the existing proofs of these theorems, with the exception of a proof of the Gibbard–Satterthwaite theorem by Barberà and Peleg (1990). Our proof is a nice complement to theirs, as it employs a different approach. The Barberá and Peleg (1990) proof uses a key step of showing that a strategy–proof social choice function on a full domain is "tops–only". That is, the social choice function depends only on the information of which element is most preferred by each agent. The key step in our proof involves showing that some agent can enforce an outcome.

The proof below is for $N = 2$ when F has a range of three elements. The proof is completed in the appendix.

PROOF: It is easily seen that if S implements a dictatorial social choice function via (M, g), then it is strategy–resistant with respect to (M, g). Thus we prove the converse. Let F be a social choice function which has at least three elements in its range and is implemented via the solution S by the mechanism (M, g), for which S is strategy–resistance. We show that F is dictatorial.

(1) For some i and $c \in A$, there exists $P^i \in \mathcal{P}$ such that $F(P^i, P^j) = c$ for all $P^j \in \mathcal{P}$.

13

Case 1. There exists P, j and c such that $O_S[(M,g),P] = c$ and c is worst for j at P^j. Strategy–resistance then implies that $O_S[(M,g),(P^i,\cdot)]$ gives c everywhere.

Case 2. Nobody ever gets their worst outcome. Consider $P^1 = (b,a,c)$ and $P^2 = (a,b,c)$. Then $O_S[(M,g),P] \in \{a,b\}$. Without loss of generality, suppose that $O_S[(M,g),P] = b$. Strategy–resistance then implies that $O_S[(M,g),P^1,\widehat{P}^2] \neq a$ for all $\widehat{P}^2$. Since $O_S[(M,g),\cdot]$ is neither agent's worst outcome, $O_S[(M,g),P^1,\widehat{P}^2] \neq c$ for any $\widehat{P}^2$. Therefore, $O_S[(M,g),P^1,\widehat{P}^2] = b$ for all $\widehat{P}^2$ and (1) is satisfied.

(2) F is dictatorial.

Identify i, c and P^i from (1). We show that i can also enforce any $a \neq c$.

Let $\overline{P}^i = (a,c,b)$ and $\overline{P}^j = (b,a,c)$. Since $O_S[(M,g),P^i,\overline{P}^j] = c$, it follows from strategy–resistance that $O_S[(M,g),\overline{P}] \neq b$. Since a is in the range of F, $a = O_S[(M,g),\widehat{P}]$ for some $\widehat{P}$. Thus by strategy–resistance $a = O_S[(M,g),\overline{P}^i,\widehat{P}^j]$. Then since a is preferred by j to c at $\overline{P}^j$, $O_S[(M,g),\overline{P}] \neq c$. Therefore $O_S[(M,g),\overline{P}] = a$.

Let $\widetilde{P}^i = (a,b,c)$. By strategy–resistance $O_S[(M,g),\widetilde{P}^i,\overline{P}^j] = a$. Let $\widetilde{P}^j = (b,c,a)$. $O_S[(M,g),\widetilde{P}] \neq b$, since otherwise strategy–resistance implies that $O_S[(M,g),\widetilde{P}^i,\overline{P}^j] = b$. Since b is in the range of F, $b = O_S[(M,g),\widehat{P}]$ for some $\widehat{P}$. Thus, by strategy–resistance $b = O_S[(M,g),\widehat{P}^i,\widetilde{P}^j]$. Then since b is preferred by i to c at $\widetilde{P}^j$, $O_S[(M,g),\widetilde{P}] \neq c$. Thus $O_S[(M,g),\widetilde{P}] = a$. Since a is least preferred by j at $\widetilde{P}^j$, strategy–resistance implies that $O_S[(M,g),\widetilde{P}^i,P^j] = a$ for all P^j. ∎

Before proceding to Theorem 3, we examine how the strategy–resistance condition applies to several solution concepts.

EXAMPLE 3. *Undominated Strategies*

Implementation in undominated strategies shows that the possibility of non–trivial implementation depends critically on the domain of possible mechanisms $\mathcal{G}$. If we restrict attention to the class of bounded mechanisms, then the solution of undominated strategies is strategy–resistant, as shown in the theorem in section 3. Another way to see this is to apply Theorem 2. Consider a bounded mechanism (M,g) and any i, $P \in \mathcal{P}^N$, $\widehat{P}^i \in \mathcal{P}$, $\widehat{m} \in S[(M,g),P^{-i},\widehat{P}^i]$. If $\widehat{m}^i$ is undominated at P^i, then $\widehat{m} \in S[(M,g),P]$, and strategy–resistance is satisfied. If $\widehat{m}^i$ is dominated at P, then it is dominated by an undominated m^i. In this case $\widehat{m}^{-i},m^i \in O_S[(M,g),P]$ and either $g(\widehat{m}^{-i},m^i) = g(\widehat{m})$ or $g(\widehat{m}^{-i},m^i)P^i g(\widehat{m})$.

If $\mathcal{G}$ includes all mechanisms, then any social choice function is implementable in undominated strategies [Theorem 1 in Jackson (1989)]. The above argument breaks down in

14

trying to find the appropriate m^i in the argument above because for an unbounded mechanism, there exist infinite strings of strategies, with each strategy dominating the previous one, but none of which are undominated. For such mechanisms, an agent might find that a dominated strategy provides a better outcome than all of the undominated strategies, against a particular set of strategies of other agents [See example 1 in Jackson (1989)]. For such a mechanism, however, it seems unreasonable to argue that agents will only play undominated strategies.

EXAMPLE 4. *Undominated Nash implementation (part II)*.

Undominated Nash equilibrium does not satisfy strategy–resistance for the mechanism provided in Example 1. If $P^1 = (c, a, b)$ and $P^2 = (b, c, a)$, then the solution is m, with outcome a. If agent 2's preferences change to $\widehat{P}^2 = (c, b, a)$, then the solution is bottom right, with outcome c. This is not strategy–resistant, since agent 2 would rather have preferences $\widehat{P}^2$, when he or she has preferences P^2. Indeed, the social choice function implemented by the mechanism of Example 1 is not dictatorial.

Although strategy–resistance seems like a compelling condition for a solution to satisfy, we should be careful to consider its interpretation under different information structures. For solutions which operate in incomplete information settings, such as undominated strategies or dominant strategies, strategy–resistance seems natural since agents do not know the preferences of others and thus choose actions independent of the actions or preferences of others. The only change in actions from a change from P^i to $\widehat{P}^i$ is due to a change by agent i. The agent should not choose actions which do uniformly worse against the actions of the other agents.

However, when we move to a complete information setting, the preceeding argument can no longer be made. A solution such as undominated Nash equilibrium, looks for a stable point given that all agents know each others' preferences. In Example 1, m is ruled out at P^1, P^2 since agent 1 knows that it is a dominant strategy for agent 2 to play $\overline{m}^2$. Given this, agent 1 should play $\overline{m}^1$, even though the agent would prefer that both agents play m.

As mentioned previously, although Theorem 2 provides a characterization of solutions which lead to impossibility results, the strategy–resistance condition is not always easily

15

verified. It is difficult to verify, since it is only required to hold for mechanisms on which the solution concept has a single valued outcome correspondence. In the case of Nash equilibria, for instance, the strategy–resistance condition holds for 'single outcome' mechanisms, but not for other mechanisms. To overcome this problem, Theorem 3 provides conditions which are easy to check globally.

POSITIVE RESPONSIVENESS.

A solution S satisfies *positive responsiveness* with respect to the mechanism (M, g) if $g(m^{-i}, \widetilde{m}^i) P^i g(m)$ for some i, $m \in S[(M, g), P]$, and $\widetilde{m}^i \in M^i$, implies that there exists $\overline{m} \in S[(M, g), P]$ such that either $g(\overline{m}) P^i g(m^{-i}, \widetilde{m}^i)$ or $g(\overline{m}) = g(m^{-i}, \widetilde{m}^i)$.

The positive responsiveness condition essentially says that a solution is compatible with agents' preferences. If we consider any solution which is stable in a Nash equilibrium sense, then this condition is satisfied almost vacuously: there can exist no such improvement $\widetilde{m}^i$ for i. For solutions which work by means of domination arguments, the condition is also satisfied, but only when we restrict attention to bounded mechanisms. For example, if we consider undominated strategies, then such an action $\widetilde{m}^i$ is either undominated itself, or dominated by an undominated action which then must lead to at least as good an outcome for agent i as $\widetilde{m}^i$. If the mechanism is not bounded, then this is no longer true.[9] Two solutions which do not satisfy positive responsiveness are maximin and the protective criterion. Both solutions rely on information about the worst outcomes which an action may lead to, and do not account for the outcome of an action against particular actions of the other agents.

DIRECT BREAKING.

A solution S satisfies *direct breaking* with respect to the mechanism (M, g) if for each P, i, and $\widetilde{P}^i$ such that $O_S[(M, g), P^{-i}, \widetilde{P}^i] \cap O_S[(M, g), P] = \emptyset$, there exists j, $\widetilde{m} \in S[(M, g), P^{-i}, \widetilde{P}^i]$, and $\overline{m}^j$ such that $g(\widetilde{m}^{-j}, \overline{m}^j) P^j g(\widetilde{m})$.

The direct breaking condition may be interpreted as follows. Suppose that a change in one agent's preferences leads to a complete change in outcomes. Then it must be that the

---

[9] Example 1 in Jackson (1989) provides an illustration. In that example there are deviations $\widetilde{m}^i$ which are strict improvements, but they form a string each one dominating the previous one.

strategies leading to the alternative outcomes are not stable at the agents' original preferences. That is, *some* agent could benefit (in terms of original preferences) from deviating from at least one of the solutions associated with the one agent's alternative preference.

The direct breaking condition may seem somewhat similar to requiring that the outcome correspondence associated with a solution be monotonic.[10] There are important differences, however, and the direct breaking condition is much weaker. The direct breaking condition is binding only when all outcomes change due to a switch in preferences by some agent. In contrast, monotonicity is binding when any outcome changes due to a change in the preferences of an agent. Further, monotonicity then requires a preference switch between the outcome and some alternative for the agent whose preferences have changed. Direct breaking only requires that some agent have an improving deviation from some original solution. The agent does not necessarily have to have a preference switch or be the agent whose preferences have changed. These important differences are evident in the following example.

EXAMPLE 5. *The Iterated Removal of Strictly Dominated Strategies.*

As shown in Theorem 1, only dictatorial social choice functions are implementable via the iterated removal of strictly dominated strategies. Positive responsiveness and direct breaking are easily verified as follows:

Consider i, $m \in S[(M, g), P]$ and $\tilde{m}^i$ such that $g(m^{-i}, \tilde{m}^i) P^i g(m)$. Since $\#A$ is finite, there exists $\overline{m}^i$ such that for each $\hat{m}^i$ either $g(m^{-i}, \overline{m}^i) P^i g(m^{-i}, \hat{m}^i)$ or $g(m^{-i}, \overline{m}^i) = g(m^{-i}, \hat{m}^i)$. It follows that $m^{-i}, \overline{m}^i$ is iteratively undominated at P, and hence satisfies the requirement of positive responsiveness.

Checking the direct breaking condition is as straightforward. Consider m which is left after the iterated elimination of strictly dominated strategies at $P^{-i}, \tilde{P}^i$. If m is not a solution at P, Then there is a first stage such that m^j is strictly dominated by $\hat{m}^j$ for some j. This implies that $g(m^{-j}, \hat{m}^j) \overline{P}^j g(m)$.

---

[10] On the full domain we consider, monotonicity may be defined as follows. A social choice correspondence F is monotonic if for each a, P, i and $\overline{P}^i$ such that $a \in F(P)$ and $a \notin F(P^{-i}, \overline{P}^i)$, there exists b such that $aP^i b$ and $b\overline{P}^i a$. This condition has been called strong positive association by some authors.

Now we show that although the iterative elimination of strictly dominated strategies satisfies positive responsiveness and direct breaking, it does not always have a monotonic outcome correspondence. Consider the following mechanism.

	m^2	$\overline{m}^2$
m^1	a	a
$\overline{m}^1$	b	c

Let $P^1 = (b,a,c)$, $\overline{P}^1 = (b,c,a)$, and $P^2 = (c,b,a)$. At P, neither agent can remove a strategy and so the set of outcomes is $\{a,b,c\}$. At $\overline{P}^1, P^2$, agent 1 can remove m^1 since it is strictly dominated by $\overline{m}^1$. This then allows agent 2 to remove m^2. The solution is $\overline{m}$ with outcome c. This is inconsistent with monotonicity: the relative ordering of b remains unchanged, and yet it is dropped as an outcome. It is consistent, however, with direct breaking since $g(\overline{m})P^2 g(\overline{m}^1, m^2)$.

THEOREM 3. *If $\#A \geq 3$ and a solution satisfies positive responsiveness and direct breaking for a mechanism via which it implements a social choice function , then the social choice function is dictatorial.*

PROOF: Let F be a social choice function which has at least three elements in its range and is implemented via the solution S by the mechanism (M,g), for which S satisfies positive responsiveness and direct breaking. We show that S satisfies strategy–resistance with respect to (M,g). Thus by Theorem 2, F is dictatorial.

Suppose S is not strategy–resistant with respect to (M,g). It follows that for some P, i, and $\widehat{P}^i$, we have $O_S[(M,g), P^{-i}, \widehat{P}^i]P^i O_S[(M,g), P]$. Let $\widetilde{P}^i$ be the preference of i such that $O_S[(M,g), P^{-i}, \widetilde{P}^i]$ is most preferred and $O_S[(M,g), P]$ is second most preferred. [Recall that O_S is single valued.]

If $O_S[(M,g), P^{-i}, \widetilde{P}^i] \neq O_S[(M,g), P^{-i}, \widehat{P}^i]$, then by direct breaking, for some $\widehat{m} \in S[(M,g), P^{-i}, \widehat{P}^i]$ either (a) there exists $j \neq i$ and $\overline{m}^j$ such that $g(\widehat{m}^{-j}, \overline{m}^j)P^j g(\widehat{m})$, or (b) there exists $\overline{m}^i$ such that $g(\widehat{m}^{-i}, \overline{m}^i)\widetilde{P}^i g(\widehat{m})$. In case (a), positive responsiveness then implies that $O_S[(M,g), P^{-i}, \widehat{P}^i]$ is multi–valued, which is a contradiction. In case (b), the fact that $O_S[(M,g), P^{-i}, \widehat{P}^i]$ is most preferred at $\widetilde{P}^i$ is contradicted. Therefore, it must be that $O_S[(M,g), P^{-i}, \widetilde{P}^i] = O_S[(M,g), P^{-i}, \widehat{P}^i]$.

It follows that $O_S[(M,g), P^{-i}, \widetilde{P}^i] \neq O_S[(M,g), P]$. By direct breaking, for some $m \in S[(M,g), P]$ either (a) there exists $j \neq i$ and $\overline{m}^j$ such that $g(m^{-j}, \overline{m}^j)P^j g(m)$, or (b)

18

there exists $\overline{m}^i$ such that $g(m^{-i}, \overline{m}^i) \widetilde{P}^i g(m)$. there exists $\overline{m}^i$ such that $g(m^{-i}, \overline{m}^i) \widetilde{P}^i g(m)$. In case (a), positive responsiveness then implies that $O_S[(M,g), P]$ is multi–valued, which is a contradiction. In case (b), by the definition of $\widetilde{P}^i$, it follows that $g(m^{-i}, \overline{m}^i) = O_S[(M,g), P^{-i}, \widehat{P}^i]$. Therefore, $g(m^{-i}, \overline{m}^i) P^i g(m)$. Positive responsiveness then implies that $O_S[(M,g), P]$ is multi–valued which is a contradiction. Thus, our original supposition was wrong and S is strategy–resistant with respect to (M,g). ∎

EXAMPLE 6. *A Dictatorial Solution.*

The converse to Theorem 3 is not true. Consider S defined by

$$S[(M,g), P] = \{m \mid g(m) P^1 g(\overline{m}) \forall \overline{m} \in M\}.$$

S is the somewhat pathological solution concept which assumes that all agents choose actions which are best for the first agent. Clearly, the social choice functions which are implementable via S are dictatorial. Yet, S does not satisfy either positive responsiveness or direct breaking with respect to the following mechanism.

	m^2	$\overline{m}^2$
m^1	a	b
$\overline{m}^1$	c	d

Let $P^1 = (a, b, c, d)$, $\overline{P}^1 = (d, a, b, c)$, and $P^2(c, d, a, b)$. The solution at P is m and the solution at $\overline{P}^1, P^2$ is $\overline{m}$. Positive responsiveness is not satisfied since $g(\overline{m}^1, m^2) P^2 g(\overline{m})$. (This is part of what makes the solution so unappealing.) Direct breaking is not satisfied since neither agent has an improving deviation away from m.

EXAMPLE 7. *Nash implementation.*

Although it is difficult to check that the Nash equilibrium solution satisfies strategy–resistance for mechanisms on which it is single–valued, we can easily check that it satisfies both positive responsiveness and direct breaking for any mechanism. Positive responsiveness is satisfied since by the definition of Nash equilibrium there can never exist $\widehat{m}^i$ such that $g(m^{-i}, \widehat{m}^i) P^i g(m)$, when m is a Nash equilibrium at P. Direct breaking is satisfied since if m is a Nash equilibrium at P but not at $P^{-i}, \widetilde{P}^i$, then m^i must no longer be a best response for player i.

EXAMPLE 8. *Undominated Strategies (part II).*

We close this section by verifying that the solution of undominated strategies satisfies both positive responsiveness and direct breaking for any bounded mechanism. Consider P, m which is undominated at P, and i and $\hat{m}^i$ such that $g(m^{-i}, \hat{m}^i)P^i g(m)$. Either $\hat{m}^i$ is undominated, or it is dominated by an undominated action $\overline{m}^i$. Positive responsiveness is thus satisfied by either $m^{-i}, \hat{m}^i$ or $m^{-i}, \overline{m}^i$, respectively. To verify direct breaking, consider P and $\overline{P}^i$ such that $O_S[(M,g), P] \cap O_S[(M,g), P^{-i}, \overline{P}^i] = \emptyset$. Let m be undominated at P. It follows that m^i is dominated at $\overline{P}^i$ by an action $\overline{m}^i$ which is undominated at $\overline{P}^i$. Thus, $g(m) \neq g(m^{-i}, \overline{m}^i)$, and so $g(m^{-i}, \overline{m}^i)\overline{P}^i g(m)$. Therefore, direct breaking is satisfied.

5. Concluding Remarks

In this paper we have examined properties of solution concepts which limit their ability to implement social choice functions on a full domain of preferences. The solutions with limited implementation results, had the common trait of "breaking" certain equilibria directly by requiring that some agent have an improving deviation against the actions of the other agents. In contrast, solutions which permit implementation of interesting social choice functions on a full domain of preferences incorporate information which permits them to break equilibria without requiring that any agent an improving deviation directly against the actions of the other agents. [Table 1 summarizes the results for various solutions.]

We have focussed attention on full preference domains and on the implementation of social choice functions . Comparisons across solution concepts might also prove useful in understanding implementation in more structured environments, where there are additional restrictions on the set of preferences considered, and where it is possible to implement correspondences instead of just functions.

Another extension would allow for the possibility of indifference in preferences. Considering a full domain of preferences with the possibility of indifference, produces difficulties for the implementation of social choice functions . For almost any solution we consider, there are no non–constant social choice functions which are implementable on a full domain of preferences where indifference is allowed. [Thus not even a dictatorial social choice function is implementable on such a domain.] This is easily seen by noting that when all agents are completely indifferent, then all actions will be possible under almost any

20

solution concept. An implemented social choice function must then take on all values at such a preference profile. Even if complete indifference is ruled out, allowing for some indifference will produce multiple outcomes for some preference profiles. Thus to extend the discussion of implementation to the domain of indifference, one has to consider social choice correspondences.

Finally, we have restricted our attention to non–cooperative solutions. For example, we have not discussed implementation in Strong Nash equilibria [see Maskin (1977), (1985), Dasgupta, Hammond, and Maskin (1979), and Dutta and Sen (1988)]. Strong Nash equilibrium leads to an impossibility result in the setting considered here, but does not satisfy the direct breaking condition. Coalitional arguments rely on slightly different intuition than that presented here. Direct breaking would be satisfied, however, if we changed it to allow a coalition of agents to take the place of agent j. In order to obtain an analog of Theorem 3, we then simply modify the positive responsiveness condition to account for coalitional deviations as well.

TABLE 1

D – Only dictatorial functions are implementable.

N – Non-dictatorial functions are implementable.

Solution	$N = 2$ $\#A = 2$	$N \geq 3$ $\#A = 2$	$N \geq 2$ $\#A \geq 3$
Dominant Strategies	N	N	D
Undominated Strategies (unbounded mechanisms)	N	N	N
Undominated Strategies (bounded mechanisms)	N	N	D
Iterated Elimination: Weakly Dominated	N	N	N
Iterated Elimination: Strictly Dominated	D	D	D
Nash Equilibria	D	N	D
Undominated Nash	N	N	N
Subgame Perfect	N	N	N
Maximin	D	N	N

References

ABREU, D. AND H. MATSUSHIMA (1990), "Exact Implementation," mimeo: Princeton University.

BARBERÀ, S. AND B. DUTTA (1982), "Implementation via Protective Equilibria," *Journal of Mathematical Economics*, **4**, pp. 49–65.

BARBERÀ, S. AND M. JACKSON (1988), "Maximin, Leximin, and the Protective Criterion: Characterizations and Comparisons," *Journal of Economic Theory*, **46**, pp. 34–44.

BARBERÀ, S. AND B. PELEG (1990), "Strategy–Proof Voting Schemes with Continuous Preferences," *Social Choice and Welfare*, **7**, pp. 31–38.

BÖRGERS, T. (1991), "Undominated Strategies and Coordination in Normalform games," *Social Choice and Welfare*, **8**, pp. 65–78.

DASGUPTA, P., P. HAMMOND, AND E. MASKIN (1979), "The Implementation of Social Choice Rules: some General Results on Incentive Compatibility," *Review of Economic Studies*, **46**, pp. 185–216.

DUTTA, B. AND A. SEN (1988), "Implementation Under Strong Equilibria: A Complete Characterization," mimeo: Indian Statistical Institute.

GIBBARD, A. (1973), "Manipulation of Voting Schemes: A General Result," *Econometrica*, **41**, pp. 587–601.

HERRERO, M. AND S. SRIVASTAVA (1989), "Implementation via Backward Induction," forthcoming: Journal of Economic Theory.

HURWICZ, L. AND D. SCHMEIDLER (1978), "Outcome Functions wich Guarantee the Existence and Optimality of Nash Equilibria," *Econometrica*, **46**, pp. 144–174.

JACKSON, M. (1989), "Implementation in Undominated Strategies: A Look at Bounded Mechanisms," mimeo: Northwestern University.

JACKSON, M., T. PALFREY, AND S. SRIVASTAVA (1990), "Undominated Nash Implementation in Bounded Mechanisms," mimeo..

JACKSON, M. AND S. SRIVASTAVA (1991), "On Two Person Nash Implementable Choice Functions," mimeo..

MASKIN, E. (1977), "Nash Equilibrium and Welfare Optimality," Department of Economics, MIT.

MASKIN, E. (1985), "The Theory of Implementation in Nash Equilibrium." In L. Hurwicz, D. Schmeidler, and H. Sonnenschein, *Social Goals and Social Organization: Essays in Honor of Elisha A. Pazner*, Cambridge: Cambridge University Press.

MOORE, J. (1991), "Implementation in Environments with Complete Information," mimeo London School of Economics.

MOORE, J. AND R. REPULLO (1988), "Subgame Perfect Implementation," *Econometrica*, **56**, pp. 1191–1220.

MOULIN, H. (1979), "Dominance Solvable Voting Schemes," *Econometrica*, **47**, pp. 1337–1351.

MOULIN, H. (1982), "Non–Cooperative Implementation: A Survey of Recent Results," *Mathematical Social Sciences*, **3**, pp. 243–257.

MOULIN, H. (1983), *The Strategy of Social Choice*, North Holland: Amsterdam.

MULLER, E. AND M. SATTERTHWAITE (1977), "On the Equivalence of Strong Positive Association and Strategy–Proofness," *Journal of Economic Theory*, **14**, pp. 412–418.

PALFREY, T., AND S. SRIVASTAVA (1986), "Nash Implementation Using Undominated Strategies," forthcoming *Econometrica*.

SATTERTHWAITE, M. (1975), "Strategy–proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Theorems," *Journal of Economic Theory*, **10**, pp. 187–217.

SCHMEIDLER, D. AND H. SONNENSCHEIN (1978), " Two Proofs of the Gibbard–Satterthwaite Theorem on the possibility of a Strategy–proof Social Choice Function." In Gottinger H. and W. Leinfellner, *Decision Theory and Social Ethics*, Reidel, Dordrecht.

THOMSON, W. (1979), "Maximin Strategies and the Elicitation of Preferences." In J.J. Laffont, *Aggreagation and the Revelation of Preferences*, North Holland, Amsterdam.

APENDIX

Proof of Theorem 2 when $N \geq 2$ and $\#A \geq 3$[11]

(1) F is monotonic.

Consider $a = F(P)$ and preferences $\overline{P}$ such that $a \neq F(\overline{P})$. We need to show that there exists $b \in A$ and i such that $aP^i b$, while $b\overline{P}^i a$. Let i be the first agent such that $a = F(\overline{P}^1, \ldots, \overline{P}^{i-1}, P^i, \ldots, P^N)$, while $a \neq F(\overline{P}^1, \ldots, \overline{P}^i, P^{i+1}, \ldots, P^N)$. Let $b = F(\overline{P}^1, \ldots, \overline{P}^i, P^{i+1}, \ldots, P^N)$. By strategy–resistance, and the fact that F is single valued, $b\overline{P}^i a$ and $aP^i b$.

(2) If F is monotonic with range A, then F is Pareto efficient.

Let a Pareto dominate b at P and suppose that $F(P) = b$. Consider $\widetilde{P}$ where each agent has a most preferred and b second. By monotonicity $F(\widetilde{P}) = b$. Since F has range A, there exists $\widehat{P}$ such that $F(\widehat{P}) = a$. Hence, $F(\widetilde{P}) = a$ by monotonicity. This contradicts the fact that F is single valued.

A subset of agents S has *veto power* if $aP^i b$ for all $i \in S$ implies $F(R) \neq b$.

(3) If $N \geq 3$ and $\#A = 3$, then F is dictatorial.

(i) Partition agents into S and S^c. Either S or S^c has veto power.

Consider preference profiles for which all agents in S have identical preferences, and all agents in S^c have identical preferences. From the proof of Theorem 2 for $N = 2$ and the fact that strategy–proofness implies coalitional strategy–proofness (on a full domain of preferences), F gives the most preferred outcome of either S or S^c on this restricted domain. Say it is S. Suppose that for some P, $F(P) = b$ while $aP^i b$ for all $i \in S$. Consider $\overline{P}$ such that every agent in S has identical preferences with a most preferred and b second, and all agents in S^c have identical preferences with b most preferred. The outcome is the most preferred outcome of S, so $F(\overline{P}) = a$. Since $F(P) = b$ and F is monotonic, it follows that $F(\overline{P}) = b$. This is a contradiction.

(ii) If S has veto power and $j \in S$, then either $S - j$ or j has veto power.

---

[11] This portion of the proof is partly based on a proof in Schmeidler and Sonnenschein (1978).

Suppose not. Then by (i) it follows that $S^c \cup j$ and $S - j \cup S^c$ have veto power. Let $S - j$ have preferences (a, b, c), j have preferences (b, c, a), and S^c have preferences (c, a, b). $F(P) \neq b$, since $aP^i b$ for all $i \neq j$ $F(P) \neq c$, since $bP^i c$ for all $i \in S$ $F(P) \neq a$, since $cP^i a$ for all $i \in S^c \cup j$ This is a contradiction.

(iii) F is dictatorial.

Begin by (i) and then apply (ii). If j has veto power, then j is a dictator. If not, then pick $k \in S - j$ and apply (ii) again. Repeat until k is a dictator.

(4) If $N \geq 2$ and $A \geq 3$, then F is dictatorial.

Choose three alternatives in A, say a, b, and c. Fix an order for all other alternatives. Define $\widetilde{P}$ to be the subset of P such that a, b, and c are preferred to the other alternatives, which appear in the fixed order.

(i) F restricted to $\widetilde{P}^N$ is dictatorial.

By monotonicity, the range of F restricted to $\widetilde{P}^N$ is $\{a, b, c\}$. Apply previous steps.

(ii) If F restricted to $\widetilde{P}^N$ is dictatorial, then F is dictatorial.

First notice that by monotonicity, the fixed order for the alternatives other than a, b, and c is irrelevant. Next notice that if agent i is a dictator when a, b, and c are at the top, then agent i is a dictator when a, b, and d are at the top. To see this suppose the contrary. Then there exists some other agent j who is dictator. Thus b is chosen when it is at the top of agent j's preference, a is at the top of agent i's preference, and d ranks third for every agent. Now perform a monotonic change so that c replaces d in each agents' preferences. By monotonicity the alternative chosen is b, which contradicts the fact that i is a dictator when a, b, and c are on top.

On Two-Person Nash Implementable Choice Functions
Matthew O. Jackson and Sanjay Srivastava
April 1991

We provide an elementary proof of the following theorem: a two-person social choice function is Nash implementable if and only if it is dictatorial on its range. Maskin [1977] and Hurwicz and Schmeidler [1978] proved the result for Pareto optimal functions. Recently, Moore and Repullo [1990] provided a proof (given Pareto optimality) based on necessary and sufficient conditions for Nash implementation. In this note, we provide a very simple proof which does not require Pareto optimality of F.

Let A be a finite set and let P denote the set of linear orders on A. In a two agent setting, a *social choice function* is a map $F : P^2 \rightarrow A$. F is *dictatorial* if it always picks the best element of the same person. Let rng(F) denote the range of F, and define F to be *dictatorial on its range* if F always picks the best element of the same person within the range of F.

A *game form* is $G = (S^1, S^2, g)$ where $g : S^1 \times S^2 \rightarrow A$, and S^1 and S^2 are the strategy sets of the two agents and g assigns to each pair of (pure) strategies an element of A. Let $O_{NE}(G;P)$ be the set of pure strategy Nash equilibrium outcomes to G given $P \in P^2$. F is *Nash implementable* if there exists G such that $O_{NE}(G;P) = F(P)$ for all $P \in P$.

<u>Theorem</u> (Maskin [1977], Hurwicz and Schmeidler [1978]): If a two-person social choice function is Pareto optimal, then it is Nash implementable if and only if it is dictatorial.

We prove:

<u>Theorem:</u> A two-person social choice function is Nash implementable if and only if it is dictatorial on its range.

<u>Proof:</u> The "if" part is obvious. To prove the converse, note that the

1

implementing game form cannot involve any $a \in A \setminus rng(F)$, since any element available through the game form is a Nash equilibrium outcome when it is most preferred by both agents. If #rng(F)=1, the theorem is obvious. If #rng(F)≥2, consider any implementing game form in which player 1 chooses rows and 2 chooses columns.

Case 1: Every $a \in rng(F)$ appears in every row. Then, agent 2 is a dictator since every Nash equilibrium must lead to agent 2's best outcome in rng(F).

Case 2: Some $a \in rng(F)$ does not appear in some row, say row n. Then, a must appear in every column (if a does not appear in column m, then let c denote the entry in row n, column m. When both agents have preferences a preferred to c preferred to everything else, there are at least two Nash equilibrium outcomes, a and c, which contradicts the fact that the game form implements a function). Next suppose b≠a does not appear in some column. Then, as above, b must appear in every row. But then there is no Nash equilibrium when a is 1's best element and b is 2's best element. Thus b must appear in every column. Since b is arbitrary, this means that every element in rng(F) is in every column, which implies that agent 1 is a dictator over rng(F).

Remark: If rng(F) has three or more elements, then the theorem also holds for more than two players. This is shown in Dasgupta, Hammond, and Maskin [1979], and also follows from the Muller-Satterthwaite [1977] theorem. A simple proof for this case is given in Jackson and Srivastava [1991].

References

Dasgupta, P., P. Hammond, and E. Maskin, "The Implementation of Social Choice Rules: Some General Results on Incentive Compatibility," Review of Economic Studies, 46:185-216, 1979.

Hurwicz, L. and D. Schmeidler, "Outcome Functions which Guarantee the Existence and Optimality of Nash Equilibria," Econometrica 46:144-174, 1978.

Jackson, M. and S. Srivastava, "Implementing Social Choice Functions: A New Look at Some Impossibility Results," mimeo, 1991.

Maskin, E., "Nash Equilibrium and Welfare Optimality," mimeo, 1977.

Moore, J. and R. Repullo, "Nash Implementation: A Full Characterization," Econometrica 58:1083-1100, 1990.

Muller, E. and M. Satterthwaite, "The Equivalence of Strong Positive Association and Strategy-Proofness," Journal of Economic Theory, 14:412-418, 1977.