Implementation Theory*

Eric Maskin and Tomas Sjöström

January 15, 2001

1 Introduction

1.1 The Implementation Problem

The implementation problem is the problem of designing a mechanism (game form) such that the equilibrium outcomes satisfy some criterion of social optimality. Early discussions focused on revelation mechanisms where each agent reports what he knows to a social planner who chooses an outcome. It was discovered that the Lindahl rule for allocation of public goods [Samuelson (1954, 1955)] and the Walrasian rule for allocation of private goods [Hurwicz (1972)] are not strategy-proof, i.e., all agents reporting the truth is not an equilibrium. For quasi-linear public goods environments, Groves (1970) and Clarke (1971) designed strategy-proof revelation mechanisms. But a strategy-proof revelation mechanism may, in addition to the truthful equilibrium, have undesirable untruthful equilibria which can only be eliminated by using more general message spaces. Groves and Ledyard (1977), Hurwicz and Schmeidler (1978) and Maskin (1999)[1] initiated the study of mechanisms with general message spaces. This line of research, which is known as implementation theory, provides an analytical framework for the design of institutions. It has been criticized for allowing mechanisms to be arbitrarily complicated, but much of the complexity is due to the fact that theorems are proved for very general environments. In many applications the optimal

*We are grateful to Luis Corchón and Sandeep Baliga for helpful comments.

[1] Maskin's paper was circulated as a working paper in 1977.

mechanisms have turned out to be quite simple.[2]

1.2 Definitions

The environment is represented by (A, N, Θ), where A is the set of feasible *alternatives* or *outcomes*, $N = \{1, 2, ..., n\}$ is the set of *agents*, and Θ is the set of possible *states of the world*. In some applications, the state $\theta \in \Theta$ specifies endowments and production technologies, so that the set of feasible alternatives depends on the state of the world. In fact, except for a brief discussion in Section 2.9, in this survey we will assume the feasible set A is the same in all states. The agents preferences over outcomes do, however, depend on the state.

Each agent $i \in N$ has a payoff function $u_i : A \times \Theta \to \mathbf{R}$. Thus, if the outcome is $a \in A$ and the state of the world is $\theta \in \Theta$, then agent i's payoff is $u_i(a, \theta)$. His weak preference in state θ is represented by $R_i = R_i(\theta)$, where for $x, y \in A$,

$$xR_iy \iff u_i(x, \theta) \geq u_i(y, \theta)$$

The strict part of his preference is denoted $P_i = P_i(\theta)$, and indifference is denoted $I_i = I_i(\theta)$:

$$xP_iy \iff u_i(x, \theta) > u_i(y, \theta)$$

$$xI_iy \iff u_i(x, \theta) = u_i(y, \theta)$$

The *preference profile* at state $\theta \in \Theta$ is denoted $R = R(\theta) = (R_1(\theta), ..., R_n(\theta))$. The *preference domain* is the set of a priori possible preference profiles:

$$\mathcal{R}(\Theta) \equiv \{R : R = R(\theta) \text{ for some } \theta \in \Theta\}.$$

Let $\mathcal{R}_A$ be the set which contains *all* profiles of complete and transitive preferences over A, the *unrestricted domain.* Let $\mathcal{P}_A$ be the set which contains all profiles of linear orderings over A (i.e., $\mathcal{P}_A$ is the subset of $\mathcal{R}_A$ which contains all *strict* preference profiles). The *preference domain for agent i* is the set

$$\mathcal{R}_i(\Theta) \equiv \{R_i : \text{there is } R_{-i} \text{ such that } (R_i, R_{-i}) \in \mathcal{R}(\Theta)\}.$$

[2]Earlier surveys of implementation theory include Moore (1992), Palfrey (1992), and Corchón (1996).

When Θ is fixed, we simplify by writing $\mathcal{R}$ and $\mathcal{R}_i$ instead of $\mathcal{R}(\Theta)$ and $\mathcal{R}_i(\Theta)$.

For any sets X and Y, let $X \backslash Y \equiv \{x : x \in X, x \notin Y\}$, let Y^X denote the set of all functions from X to Y, and let 2^X denote the set of all subsets of X. A *social choice rule* (SCR) is a function $F : \Theta \to 2^A \backslash \emptyset$ (i.e. a non-empty valued correspondence). The set $F(\theta)$ is the set of socially optimal (or more precisely *F-optimal*) alternatives in state $\theta \in \Theta$. The *image* or *range* of the SCR F is the set

$$F(\Theta) \equiv \{a \in A : a \in F(\theta) \text{ for some } \theta \in \Theta\}.$$

A *social choice function* (SCF) is a *single-valued* SCR, i.e., a function $f : \Theta \to A$.

Some important properties of SCRs are as follows. *Ordinality:* for all $\theta, \theta' \in \Theta$, if $R(\theta) = R(\theta')$ then $F(\theta) = F(\theta')$. *Weak Pareto optimality:* for all $\theta \in \Theta$ and all $a \in F(\theta)$, there is no $b \in A$ such that $u_i(b, \theta) > u_i(a, \theta)$ for all $i \in N$. *Pareto optimality:* for all $\theta \in \Theta$ and all $a \in F(\theta)$, there is no $b \in A$ such that $u_i(b, \theta) \geq u_i(a, \theta)$ for all $i \in N$ with strict inequality for some i. *Dictatorship:* there exists $i \in N$ such that for all $\theta \in \Theta$ and all $a \in F(\theta)$, $u_i(a, \theta) \geq u_i(b, \theta)$ for all $b \in A$. *Unanimity:* for all $\theta \in \Theta$ and all $a \in A$, if $u_i(a, \theta) \geq u_i(b, \theta)$ for all $i \in N$ and all $b \in A$ then $a \in F(\theta)$.

If F is an ordinal SCR, then there exists a unique correspondence $\bar{F} : \mathcal{R}(\Theta) \to 2^A \backslash \emptyset$ satisfying

$$\bar{F}(R(\theta)) = F(\theta) \tag{1}$$

for all $\theta \in \Theta$. The correspondence $\bar{F}$ is sometimes convenient to work with, since it is defined directly on preferences and avoids references to states of the world.

Given an SCR F, the *implementation problem* is the problem of finding a *mechanism* (or *game form*) such that the equilibrium outcomes are F-optimal in each state. A *normal form mechanism* is denoted $\Gamma = (\times_{i=1}^n M_i, h)$ and consists of a *message space* M_i for each agent $i \in N$, and an *outcome function* $h : \times_{i=1}^n M_i \to A$. Let $m_i \in M_i$ denote agent i's message. All messages are sent simultaneously, and the chosen outcome is $h(m_1, ..., m_n) \in A$. A *message profile* is denoted $m = (m_1, ..., m_n) \in M \equiv \times_{i=1}^n M_i$. An *extensive form mechanism* is a more complicated object since it allows agents to make choices sequentially; for a formal definition see Moore and Repullo (1988). The most common interpretation of the implementation problem is that the mechanism is designed by a *social planner* or *mechanism designer,* who cannot observe the true state of the world, but who wants the outcome to be F-optimal in each state.

Let $\mathcal{S}$-equilibrium be a game theoretic solution concept. For each mechanism Γ and each state $\theta \in \Theta$, the solution concept specifies a set of *$\mathcal{S}$-equilibrium outcomes* denoted $\mathcal{S}(\Gamma, \theta) \subseteq A$. Let F be an SCR. The mechanism Γ *implements F in $\mathcal{S}$-equilibria*, or simply *$\mathcal{S}$-implements F*, if and only if $\mathcal{S}(\Gamma, \theta) = F(\theta)$ for all $\theta \in \Theta$. Thus, the set of $\mathcal{S}$-equilibrium outcomes should coincide with the set of F-optimal outcomes in each state. If such a mechanism exists then F is *implementable in $\mathcal{S}$-equilibria* or simply *$\mathcal{S}$-implementable.* This notion is also referred to as *full implementation.* If $\mathcal{S}_1$ and $\mathcal{S}_2$ are two solution concepts, then Γ *doubly $\mathcal{S}_1$ and $\mathcal{S}_2$-implements F* if and only if $\mathcal{S}_1(\Gamma, \theta) = \mathcal{S}_2(\Gamma, \theta) = F(\theta)$ for all $\theta \in \Theta$.

The mechanism Γ *weakly $\mathcal{S}$-implements F* if and only if $\emptyset \neq \mathcal{S}(\Gamma, \theta) \subseteq F(\theta)$ for all $\theta \in \Theta$. That is, every $\mathcal{S}-$equilibrium outcome is F-optimal, but every F-optimal outcome need not be an equilibrium outcome. Weak implementation is actually subsumed by the theory of full implementation, since weak implementation of F is equivalent to full implementation of a sub-correspondence of F [Thomson (1996)].[3]

In general, whether or not an SCR F is $\mathcal{S}$-implementable depends on the solution concept $\mathcal{S}$. If solution concept $\mathcal{S}_2$ is a refinement of $\mathcal{S}_1$, in the sense that for any Γ we have $\mathcal{S}_2(\Gamma, \theta) \subseteq \mathcal{S}_1(\Gamma, \theta)$ for all $\theta \in \Theta$, then it is not a priori clear whether it will be easier to satisfy $\mathcal{S}_1(\Gamma, \theta) = F(\theta)$ or $\mathcal{S}_2(\Gamma, \theta) = F(\theta)$ for all $\theta \in \Theta$. However, the literature shows that refinements usually make things easier. More SCRs can be implemented in undominated Nash equilibria, or in trembling hand perfect equilibria, than in Nash equilibria. Harsanyi and Selten (1988) argue that game theoretic analysis should lead to an ideal solution concept which applies universally to all possible games, but experiments show that behavior in fact depends on the nature of the game (even on "irrelevant" aspects such as the labelling of strategies). Thus, for successful applications of implementation theory, the solution concept should be appropriate for the mechanism, but it is hard to make this criterion mathematically precise. For an insightful discussion, see Jackson (1992). From a theoretical point of view, Muench and Walker (1984), de Trenqualye (1998) and Cabrales (1996) have discussed the problem of how agents come to co-

[3]Sen (1995) discusses a notion of implementation which requires that $F(\theta) \subseteq \mathcal{S}(\Gamma, \theta)$ for all $\theta \in \Theta$. Thus, all F-optimal outcomes must be equilibrium outcomes, but all equilibrium outcomes need not be F-optimal. The problem is then to make the set $\mathcal{S}(\Gamma, \theta)$ as small as possible.

ordinate on a particular equilibrium. Cabrales and Ponti (1996) argued that in mechanisms that rely on the elimination of weakly dominated strategies, simple learning dynamics may lead to the "wrong" equilibrium. For experiments that test the performance of mechanisms for the provision of public goods, see Smith (1980), Banks, Ledyard and Porter (1988), Harstad and Marrese (1981, 1982), Chen and Plott (1996) and Chen and Tang (1997).

The notion of implementing an SCR discussed in this survey is *consequentialist*: the precise structure of the game form is unimportant as long as the equilibrium outcomes are F-optimal. However, game forms can be used use represent *rights* [Gärdenfors (1981), Gaertner, Pattanaik and Suzumura (1992), Deb (1994), Hammond (1997), Peleg (1998)]. Deb, Pattanaik and Razzolini (1997) introduce several properties of game forms that correspond to acceptable rights structures. For example, individual $i \in N$ has a *say* if there exists at least some circumstance where his message can influence the outcome. This requirement seems weak, yet there is nothing in the definition of implementation used in this survey that guarantees that each individual has a say.[4]

2 Nash Implementation

We start by assuming that the true state of the world $\theta \in \Theta$ is common knowledge among the agents. This is the case of complete information.

Given a normal form mechanism $\Gamma = (M, h)$, for any $m \in M$ and $i \in N$, let $m_{-i} = \{m_j\}_{j \neq i} \in M_{-i} \equiv \times_{j \neq i} M_j$ denote the messages sent by agents other than i. For message profile $m = (m_{-i}, m_i) \in M$, the set

$$h(m_{-i}, M_i) \equiv \{a \in A : a = h(m_{-i}, m_i') \text{ for some } m_i' \in M_i\}$$

is agent i's *attainable set at message profile* m. Agent i's *lower contour set* at $(a, \theta) \in A \times \Theta$ is $L_i(a, \theta) \equiv \{b \in A : u_i(a, \theta) \geq u_i(b, \theta)\}$. A message profile $m \in M$ is a (pure strategy) *Nash equilibrium* at state $\theta \in \Theta$ if and only if $h(m_{-i}, M_i) \subseteq L_i(h(m), \theta)$ for all $i \in N$. (For now we neglect mixed strategies: they are discussed in Section 3.7.) The set of Nash equilibria at state θ is denoted $N^\Gamma(\theta) \subseteq M$, and the set of *Nash equilibrium outcomes* at state θ is denoted $h(N^\Gamma(\theta)) = \{a \in A : a = h(m) \text{ for some } m \in N^\Gamma(\theta)\}$.

[4] Gaspart (1996, 1997) proposed a stronger notion of *equality (or symmetry) of attainable sets:* all agents, by unilaterally varying their strategies, should be able to attain identical (or symmetric) sets of outcomes, at least at equilibrium.

The mechanism Γ Nash-implements F if and only if $h(N^{\Gamma}(\theta)) = F(\theta)$ for all $\theta \in \Theta$.

2.1 Monotonicity

If $L_i(a, \theta) \subseteq L_i(a, \theta')$ then we say that $R_i(\theta')$ is a *monotonic transformation of* $R_i(\theta)$ *at alternative* a. The SCR F is *monotonic* if and only if the following is true for all $a \in A$ and all $\theta, \theta' \in \Theta$: if $a \in F(\theta)$ and $L_i(a, \theta) \subseteq L_i(a, \theta')$ for all $i \in N$, then $a \in F(\theta')$. Thus, if a is optimal in state θ, and when the state changes from θ to θ' outcome a does not fall in any agent's preference ordering relative to any other alternative, then monotonicity requires that a remains optimal in state θ'. Notice that if $R_i(\theta) = R_i(\theta')$ then $L_i(a, \theta) = L_i(a, \theta')$ for all $a \in A$. Therefore, if F is monotonic then F is ordinal. But many ordinal social choice rules are not monotonic.

Whether an SCR F is monotonic may depend on the preference domain $\mathcal{R}(\Theta)$. For example, in an exchange economy, the Walrasian correspondence is not monotonic in general. However, it is monotonic if the agents' preferences are restricted in such a way that Walrasian equilibria always occur in the interior of the feasible set [Hurwicz, Maskin and Postlewaite (1995)]. On the other hand, the weak Pareto correspondence, defined by $F(\theta) = \{a \in A :$ there is no $b \in A$ such that $u_i(b, \theta) > u_i(a, \theta)$ for all $i \in N\}$, is monotonic for any domain.

Since for any mechanism Γ, the Nash equilibrium outcome correspondence $h \circ N^{\Gamma} : \Theta \to A$ is monotonic, Maskin (1999) could obtain the following result.

Theorem 1 *[Maskin (1999)] If the SCR F is Nash implementable, then F is monotonic.*

Proof. Suppose $\Gamma = (M, h)$ Nash implements F. Then if $a \in F(\theta)$ there is $m \in N^{\Gamma}(\theta)$ such that $a = h(m)$. Suppose $L_i(a, \theta) \subseteq L_i(a, \theta')$ for all $i \in N$. Then, for all $i \in N$,

$$h_i(m_{-i}, M_i) \subseteq L_i(a, \theta) \subseteq L_i(a, \theta').$$

Therefore, $m \in N^{\Gamma}(\theta')$ and $a \in h(N^{\Gamma}(\theta')) = F(\theta')$. $\square$

Theorem 1 has a partial converse, originally stated by Maskin (1999). The SCR F satisfies *no veto power* if for all $j \in N$, all $\theta \in \Theta$ and all $a \in A$ the following is true: if $u_i(a, \theta) \geq u_i(b, \theta)$ for all $b \in A$ and all $i \neq j$ then $a \in F(\theta)$. In words, this condition says that if at least $n - 1$ agents prefer alternative a to all other alternatives, then a is F-optimal. In economic environments, no veto power is usually trivially satisfied. However, in other environments no veto power is not a trivial condition. For example, in voting over a finite set of alternatives with an unrestricted domain of preferences, the well-known Borda rule does not satisfy no veto power if the number of alternatives is greater than the number of voters.

Theorem 2 *[Maskin (1999), Repullo (1987), Saijo (1988)] Suppose $n \geq 3$. If the SCR F satisfies monotonicity and no veto power, then F is Nash implementable.*

Proof. The proof is constructive. Let each agent $i \in N$ announce an outcome, a state of the world, and an integer between 1 and n:

$$M_i = A \times \Theta \times \{1, 2, ..., n\}$$

A typical message for agent i is denoted $m_i = (a^i, \theta^i, z^i) \in M_i$. Let the outcome function be as follows.

Rule 1. If $(a^i, \theta^i) = (a, \theta)$ for all $i \in N$ and $a \in F(\theta)$, then $h(m) = a$.

Rule 2. Suppose there exists $j \in N$ such that $(a^i, \theta^i) = (a, \theta)$ for all $i \neq j$ but $(a^j, \theta^j) \neq (a, \theta)$. Then $h(m) = a^j$ if $a^j \in L_j(a, \theta)$ and $h(m) = a$ otherwise.

Rule 3. In all other cases, let $h(m) = a^j$ where $j = (\sum_{i \in N} z^i) \pmod n$.[5]

We need to show that, for any $\theta^* \in \Theta$, $h(N^\Gamma(\theta^*)) = F(\theta^*)$.

Step 1: $h(N^\Gamma(\theta^*)) \subseteq F(\theta^*)$. Suppose $m \in N^\Gamma(\theta^*)$. If either rule 2 or rule 3 applies to m, then there is $j \in N$ such that any agent $k \neq j$ can get his top-ranked alternative, via rule 3, by announcing an integer z^k such that $k = (\sum z^i) \pmod n$. Therefore, we must have $u_k(h(m), \theta^*) \geq u_k(x, \theta^*)$ for all $k \neq j$ and all $x \in A$, and hence $h(m) \in F(\theta^*)$ by no veto power. If instead rule 1 applies, then there exists (a, θ) such that $(a^i, \theta^i) = (a, \theta)$ for all $i \in N$, and $a \in F(\theta)$. The attainable set for each agent j is $L_j(a, \theta)$, by rule 2. Since $m \in N^\Gamma(\theta^*)$, we have $L_j(a, \theta) \subseteq L_j(a, \theta^*)$. By monotonicity, $a \in F(\theta^*)$. Thus, $h(N^\Gamma(\theta^*)) \subseteq F(\theta^*)$.

Step 2: $F(\theta^*) \subseteq h(N^\Gamma(\theta^*))$. Suppose $a \in F(\theta^*)$. If $m_i = (a, \theta^*, 1)$ for all $i \in N$, then $m \in N^\Gamma(\theta^*)$ and $h(m) = a$. Thus, $F(\theta^*) \subseteq h(N^\Gamma(\theta^*))$. $\square$

The mechanism in the proof of Theorem 2 is the *canonical mechanism for Nash implementation*. Rule 3 is a *modulo game*.[6]

[5] $\alpha = \beta \pmod n$ denotes that integers α and β are conguent modulo n.

[6] It is sometimes replaced by an *integer game*, where each agent $i \in N$ announces a

2.2 Necessary and Sufficient Conditions

The no veto power condition is not necessary for Nash implementation. The necessary and sufficient condition for Nash implementation with $n \geq 3$ was given by Moore and Repullo (1990). It can be explained by introducing a modified version of the canonical mechanism. In fact, rules 1 and 3 will remain essentially unchanged, only rule 2 is modified.

Suppose we want to Nash implement a monotonic SCR F, and $a \in F(\theta)$. In the canonical mechanism of Section 2.1, there is a "consensus" Nash equilibrium $m \in N^{\Gamma}(\theta)$ where $(a^i, \theta^i) = (a, \theta)$ for each $i \in N$. Rule 1 applies to m, $h(m) = a$, and rule 2 specifies that agent j's attainable set at the consensus is $h(m_{-j}, M_j) = L_j(a, \theta)$. Now suppose at some state $\theta' \neq \theta$, there is $j \in N$ and an "awkward outcome" $c \in L_j(a, \theta)$ such that: (i) $L_j(a, \theta) \subseteq L_j(c, \theta')$; (ii) for each $i \neq j$, $L_i(c, \theta') = A$; (iii) $c \notin F(\theta')$. (Notice that (ii) and (iii) imply that F does not satisfy no veto power.) Since $c \in h(m_{-j}, M_j)$ there is $m_j' \in M_j$ such that $h(m_{-j}, m_j') = c$. Then, $(m_{-j}, m_j') \in N^{\Gamma}(\theta')$ since (i) c is the best outcome in agent j's attainable set $h(m_{-j}, M_j)$, and (ii) c is the best outcome in all of A for all other agents. By (iii), $c \notin F(\theta')$, so $h(N^{\Gamma}(\theta')) \neq F(\theta')$, contradicting the definition of implementation. We must modify rule 2 to make sure that the awkward outcome c is not in agent j's attainable set. Remove all awkward outcomes from $L_j(a, \theta)$; several iterations of this procedure may be necessary [Sjöström (1991)]. When there are no more iterations to be made, what remains is some set $C_j(a, \theta) \subseteq L_j(a, \theta)$.[7] In the modified mechanism, this set replaces $L_j(a, \theta)$ in rule 2, i.e., $C_j(a, \theta)$ is agent j's attainable set at the "consensus". By construction, the set $C_j(a, \theta)$ will have the property that $c \in F(\theta')$ whenever the following is true: $c \in C_j(a, \theta) \subseteq L_j(c, \theta')$ and $L_i(c, \theta') = A$ for each $i \neq j$. This removes the possibility of undesirable Nash equilibria. If $n \geq 3$, and if F satisfies *unanimity*, then F is Nash implementable if and only if it can be implemented by this modified canonical mechanism. This leads to the following *necessary and sufficient* condition for Nash implementation (assuming unanimity and $n \geq 3$): if $a \in F(\theta)$ and $C_i(a, \theta) \subseteq L_i(a, \theta')$ for all $i \in N$ then $a \in F(\theta')$.

Consider the following two examples of how to construct the C_i sets, due

positive integer $z^i \in \{1, 2, ...\}$ and the outcome is $h(m) = a^j$ if there exists $j \in N$ such that $z^j > z^i$ for all $i \neq j$, and otherwise $h(m)$ is arbitrary.

[7] If A is infinite, then it is possible (but unlikely) that the algorithm of sequentially eliminating outcomes in $L_j(a, \theta)$ never terminates. The set $C_j(a, \theta)$ can still be defined, but not by an algorithm [Sjöström (1991)].

to Maskin (1985). First suppose $N = \{1, 2, 3\}$, $A = \{a, b, c\}$, $\mathcal{R}(\Theta) = \mathcal{P}_A$. The SCR F is defined as follows. For any $\theta \in \Theta$, $a \in F(\theta)$ if and only if a majority prefers a to b, $b \in F(\theta)$ if and only if a majority prefers b to a, and $c \in F(\theta)$ if and only if c is top-ranked in A by all agents. This SCR is monotonic and satisfies unanimity but not no veto power. Fix $j \in N$ and suppose θ is such that $bP_j(\theta)aP_j(\theta)c$, and $aP_i(\theta)b$ for all $i \neq j$. Then $F(\theta) = \{a\}$. Now suppose θ' is such that $bP_j(\theta')cP_j(\theta')a$ and $L_i(c, \theta') = A$ for all $i \neq j$. Since $L_j(a, \theta) = L_j(c, \theta') = \{a, c\}$ but $c \notin F(\theta')$, c is awkward. Removing c, we obtain $C_j(a, \theta) = \{a\}$. (Since $a \in F(\theta)$ whenever $L_i(a, \theta) = A$ for all $i \neq j$, a is not awkward and there is need to iterate on the procedure.) By the symmetry of a and b, $C_j(b, \theta) = \{b\}$ whenever $aP_j(\theta)bP_j(\theta)c$ and $bP_i(\theta)a$ for all $i \neq j$. It can be verified that this takes care of all awkward outcomes, and that the necessary and sufficient condition for Nash implementation is satisfied. Thus, F is Nash implementable. For a second example, consider any environment (with $n \geq 3$) and let a_0 be a fixed "status quo" alternative in A. The *individually rational SCR*, which is defined by $F(\theta) = \{a \in A : aR_i(\theta)a_0 \text{ for all } i \in N\}$, satisfies monotonicity and unanimity but not no veto power. If $a \in F(\theta)$ then $a_0 \in L_j(a, \theta)$ for all $j \in N$. If $c \in L_j(a, \theta) \subseteq L_j(c, \theta')$ and $L_i(c, \theta') = A$ for each $i \neq j$, then $cR_i(\theta')a_0$ for all $i \in N$ so $c \in F(\theta')$. Therefore there are no awkward outcomes, and so the necessary and sufficient condition reduces to monotonicity. Thus, F is Nash implementable.

Danilov (1992) gave an elegant formula for the C_i sets for the case where $\mathcal{R}(\Theta) = \mathcal{P}_A$. Consider any set $X \subseteq A$. An alternative $x \in X$ is *essential for agent $i \in N$ in set X* if and only if $x \in F(\theta)$ for some θ such that $L_i(x, \theta) \subseteq X$. The set of all essential elements for agent i in set X is denoted $Ess(F; i, X)$. An SCR F is *essentially monotonic* if and only if for all $a \in A$ and all $\theta, \theta' \in \Theta$ the following is true: if $a \in F(\theta)$ and $Ess(F; i, L_i(a, \theta)) \subseteq L_i(a, \theta')$ for all $i \in N$, then $a \in F(\theta')$. If $\mathcal{R}(\Theta) = \mathcal{P}_A$, then it turns out that, for any $\theta \in \Theta$ and $a \in F(\theta)$, $C_i(a, \theta) = Ess(F; i, L_i(a, \theta))$.

Theorem 3 *[Danilov (1992)] Suppose A is a finite set, $n \geq 3$, and $\mathcal{R}(\Theta) = \mathcal{P}_A$. The SCR F is Nash implementable if and only if it is essentially monotonic.*

There is no need to make the extra assumption of unanimity in the statement of Theorem 3. Yamato (1992) showed that any SCR which satisfies essential monotonicity and a weak form of unanimity is Nash implementable, even if the set of alternatives is infinite and the preference domain $\mathcal{R}(\Theta)$.

is arbitrary. However, in such environments essential monotonicity is not a necessary condition for Nash implementation.

2.3 Simplifications of the Canonical Mechanism

The canonical mechanism establishes theoretical limits on what can be achieved. But it may be impossible in practise for agents to report a complete description of the state of the world [cf. Hayek (1945)], and the modulo game in rule 3 is artificial. It is of interest to see if simpler and more natural mechanisms for Nash implementation exist.

Since any Nash implementable F is ordinal, it clearly suffices to let the agents announce a preference profile $R \in \mathcal{R}(\Theta)$ rather than a state of the world $\theta \in \Theta$. In fact, it suffices if each agent $i \in N$ announces a preference ordering for himself and one for each of his two "neighbors" agents $i-1$ and $i+1$, where agents 1 and n are considered neighbors [Saijo (1988)]. Indeed, since the attainable sets at consensus outcomes are lower contour sets, it suffices to let each agent announce an *indifference curve* for himself and his two neighbors [McKelvey (1989)]. Given any message process which "computes" (or "realizes") an SCR, Williams (1986) considered the problem of embedding the message process into a mechanism which Nash implements the SCR. If the original message process encodes information in an efficient way, then the same will be true for the messages in Williams' mechanism for Nash implementation. Williams found necessary and sufficient conditions for such an embedding to be possible.

Hurwicz (1979a, 1979c) and Schmeidler (1980) found simple "market mechanisms" that implement the Walrasian correspondence in m-good exchange economies. In these mechanisms, each agent proposes an $(m-1)$-dimensional consumption vector and an $(m-1)$-dimensional price vector (the consumption of the mth "numeraire" good is determined by a budget constraint). However, these mechanisms did not satisfy the feasibility constraint $h(m) \in A$ for all $m \in M$. A simple, feasible and continuous mechanism for implementation of the Walrasian correspondence was found by Postlewaite and Wettstein (1989). Reichelstein and Reiter (1988) show (under certain smoothness conditions on the outcome function) that the minimal dimension of the message space M for any mechanism Nash implementing the Walrasian correspondence is (approximately) $n(m-1)+m/(n-1)$, and they exhibit a feasible mechanism with this dimensionality. In this mechanism each of the n agents proposes an $(m-1)$-dimensional consumption vector. The dimensional increase $m/(n-1)$ comes from the need to also allow announcements of price variables, however, it is not necessary for each agent to announce an $m-1$ dimensional price vector as in Hurwicz (1979a, 1979c) and Schmeidler (1980).

Hurwicz (1960, 1972) looked at "proposed outcome" mechanisms with even smaller message spaces: in an exchange economy, let agent i's message consist of a proposed net trade for himself, and require that in equilibrium $h(m) = m$. No preferences are announced. Such mechanisms were studied by Sjöström (1996a) and Saijo, Tatamitani and Yamato (1997), who named the approach "natural implementation".[8] Saijo, Tatamitani and Yamato (1997) showed that, under some regularity conditions, "proposed outcome" mechanisms are incompatible with Pareto efficiency. Dutta, Sen and Vohra (1995) demonstrated how efficiency can be attained if agents can announce both consumption bundles and prices (i.e., marginal rates of substitution), as long as the SCR satisfies "local independence". For natural implementation in different economic environments, see Shin and Suh (1997) and Yoshihara (1995).

For implementations of the Lindahl correspondence in economies with m public goods and one private good using simple mechanisms see Hurwicz (1979a) and Walker (1981). In Walker's mechanism each agent announces a real number for each public good, and the production of the public good is the sum of these numbers. The dimension of M is therefore nm, the minimal dimension of any smooth Pareto efficient mechanism in this environment [Sato (1981), Reichelstein and Reiter (1988)].

2.4 Weak Implementation

If $\tilde{F}(\theta) \subseteq F(\theta)$ for all $\theta \in \Theta$ then $\tilde{F}$ is a *subcorrespondence* of F, denoted $\tilde{F} \subseteq F$. To weakly implement the SCR F is equivalent to fully implementing a non-empty valued subcorrespondence of F. Fix an SCR F, and for all $\theta \in \Theta$ define

$$F^*(\theta) \equiv \{a \in A : a \in F(\tilde{\theta}) \text{ for all } \tilde{\theta} \in \Theta \text{ such that } L_i(a,\theta) \subseteq L_i(a,\tilde{\theta}) \text{ for all } i \in N\} \tag{2}$$

[8] If the requirement that $h(m) = m$ in equilibrium is dropped, then it becomes possible to "smuggle" information by coding preference information into outcome announcements. Information smuggling can, however, be ruled out by imposing smoothness conditions on the outcome function [Hurwicz (1972), Mount and Reiter (1974), Reichelstein and Reiter (1988)].

Theorem 4 *If $F^*(\theta) \neq \emptyset$ for all $\theta \in \Theta$ then F^* is a monotonic SCR.*

Proof. Suppose $a \in F^*(\theta)$ and $L_i(a, \theta) \subseteq L_i(a, \theta')$ for all $i \in N$. Suppose $\tilde{\theta} \in \Theta$ is such that $L_i(a, \theta') \subseteq L_i(a, \tilde{\theta})$ for all $i \in N$. Then $L_i(a, \theta) \subseteq L_i(a, \theta') \subseteq L_i(a, \tilde{\theta})$ for all i. Since $a \in F^*(\theta)$ we must have $a \in F(\tilde{\theta})$. Therefore, $a \in F^*(\theta')$. $\square$

It is clear that if $F^*(\theta) = \emptyset$ for some $\theta \in \Theta$ then there does not exist any monotonic subcorrespondence. Also, if $F^*(\theta) \neq \emptyset$ for all $\theta \in \Theta$ then F^* is the maximal monotonic subcorrespondence of F. Moreover, F is monotonic if and only if $F^* = F$. If $F^*(\theta) \neq \emptyset$ for all $\theta \in \Theta$ and F^* satisfies no veto power then Theorem 2 implies that F^* is Nash implementable, hence F is weakly implementable. Conversely, if F is weakly Nash-implementable, then Theorem 1 implies that F has a monotonic non-empty valued sub correspondence $\hat{F} \subseteq F$. Then $\hat{F} \subseteq F^*$ so $F^*(\theta) \neq \emptyset$ for all $\theta \in \Theta$. Thus, Theorems 1, 2 and 4 imply the following.

Theorem 5 *If F can be weakly Nash-implemented then $F^*(\theta) \neq \emptyset$ for all $\theta \in \Theta$. Conversely, if $F^*(\theta) \neq \emptyset$ for all $\theta \in \Theta$ and F^* satisfies no veto power then F can be weakly Nash-implemented (and F^* is the maximal Nash-implementable subcorrespondence of F).*

2.5 Rich Domains of Preferences

By definition, $a \in L_i(b, \theta') \backslash L_i(b, \theta)$ means that

$$u_i(a, \theta) > u_i(b, \theta) \quad \text{and} \quad u_i(a, \theta') \leq u_i(b, \theta') \tag{3}$$

In this case we say a *improves with respect to* b for agent i as the state changes from θ' to θ. Notice that monotonicity says that if b is F-optimal in state θ', and no alternative improves with respect to b for any agent as the state changes from θ' to θ, then b is also F-optimal in state θ. The following condition was introduced by Dasgupta, Hammond and Maskin (1979).

Definition *Rich domain.* Let a, b be any two alternatives in A and θ, θ' any two states in Θ. Suppose for all $i \in N$, $a \notin L_i(b, \theta') \backslash L_i(b, \theta)$ and $b \notin L_i(a, \theta) \backslash L_i(a, \theta')$. Then there exists $\theta'' \in \Theta$ such that for all $i \in N$, $L_i(a, \theta) \subseteq L_i(a, \theta'')$ and $L_i(b, \theta') \subseteq L_i(b, \theta'')$.

Dasgupta, Hammond and Maskin (1979) found the following result.

Theorem 6 *Suppose f is a monotonic SCF, the domain is rich, and $a = f(\theta) \neq f(\theta') = b$. Then there is $i \in N$ such that either $a \in L_i(b, \theta') \backslash L_i(b, \theta)$ or $b \in L_i(a, \theta) \backslash L_i(a, \theta')$ (or both).*

Proof. If not, then there exists $\theta'' \in \Theta$ such that for all $i \in N$, $L_i(a, \theta) \subseteq L_i(a, \theta'')$ and $L_i(b, \theta') \subseteq L_i(b, \theta'')$. By monotonicity, $a = f(\theta'')$ and $b = f(\theta'')$ but $a \neq b$, a contradiction. □

If f is monotonic, then it is ordinal. Thus, the social objectives can be represented by a function $\bar{f} : \mathcal{R}(\Theta) \to A$, satisfying $\bar{f}(R(\theta)) = f(\theta)$ for all $\theta \in \Theta$. If the domain is rich, Theorem 6 implies that if $a = \bar{f}(R_{-i}, R_i) \neq \bar{f}(R_{-i}, R_i') = b$, then aR_ib and $bR_i'a$ (where at least one preference is strict). Thus, $\bar{f}$ is *strategy-proof.*

2.6 Unrestricted Domain of Strict Preferences

In models of voting over a finite set of alternatives A, it is often assumed that any strict ranking of the candidates is a priori possible: $\mathcal{R}(\Theta) = \mathcal{P}_A$. This domain is rich (as is the unrestricted domain $\mathcal{R}_A$). The SCR F is *dictatorial on its image* if and only if there exists $i \in N$ such that $a \in F(\theta)$ implies $u_i(a, \theta) \geq u_i(x, \theta)$ for all $x \in F(\Theta)$.

Theorem 7 *[Muller and Satterthwaite (1977), Dasgupta, Hammond and Maskin (1979), Roberts (1979)] Suppose the SCF f is Nash-implementable, A is a finite set, $f(\Theta)$ contains at least three alternatives, and $\mathcal{R}(\Theta) = \mathcal{P}_A$. Then f is dictatorial on its image.*

Proof. By Theorems 1 and 6, the function $\bar{f} : \mathcal{R}(\Theta) \to A$, defined by $\bar{f}(R(\theta)) = f(\theta)$ for all $\theta \in \Theta$, is strategy-proof. Now the result follows from the Gibbard-Satterthwaite theorem [Gibbard (1973), Satterthwaite (1975)]. □

Theorem 7 is false without the hypothesis of single-valuedness. For example, the weak Pareto correspondence is monotonic and satisfies no veto power in any environment, so it can be Nash implemented by Theorem 2 (when $n \geq 3$). Theorem 7 is also false without the hypothesis that the image contains at least three alternatives. Let

$$N(a, b, \theta) \equiv \#\{i \in N : u_i(a, \theta) > u_i(b, \theta)\}$$

denote the number of agents who strictly prefer a to b in state θ. Fix two alternatives $\{x, y\} \subseteq A$. The *method of majority rule* over x and y is defined by

$$F(\theta) = \begin{cases} x & \text{if } N(x,y,\theta) > N(y,x,\theta) \\ y & \text{if } N(x,y,\theta) < N(y,x,\theta) \\ \{x,y\} & \text{if } N(x,y,\theta) = N(y,x,\theta) \end{cases}$$

If $n \geq 3$ is odd and $\mathcal{R}(\Theta) = \mathcal{P}_A$ then F is single valued and monotonic. On the set $A' = \{x, y\}$ no veto power is satisfied so this SCF can be Nash implemented by Theorem 2. [9]

In general, however, the results when $\mathcal{R}(\Theta) = \mathcal{P}_A$ are negative. Most voting rules, such as the Borda and Copeland rules, are not monotonic and hence cannot be Nash implemented. Finally, if an SCR F satisfies *strong unanimity*, in the sense that $u_i(a, \theta) > u_i(b, \theta)$ for all $b \in A$ and all $i \in N$ implies $F(\theta) = \{a\}$, then monotonicity implies Pareto efficiency.[10] Sen (1970) showed that a condition of *minimal liberty* is inconsistent with Pareto efficiency in this environment, hence no strongly unanimous SCR which satisfies minimal liberty can be Nash implemented [Peleg (1998, Theorem 5.1)].

2.7 Economic Environments

If agents have selfish preferences and there exists at least one private good which is desired by everybody, then no veto power is automatically satisfied when $n \geq 3$, since $n-1$ agents can never agree on the best way to distribute the desirable good. Thus, monotonicity is necessary and sufficient for implementation when $n \geq 3$. When $n = 2$, monotonicity is necessary and "almost" sufficient [Moore and Repullo (1990), Dutta and Sen (1991b), Sjöström (1991)].

[9] It is not sufficent to simply ask each agent to vote for either x or y, for then everybody voting for x (or y) is always a Nash equilibrium when $n \geq 3$. Notice also that if indifference is allowed or if n is even, then F can still be Nash implemented (by the canonical mechanism).

[10] For suppose $u_i(a, \theta) > u_i(b, \theta)$ for all $i \in N$ but $b \in F(\theta)$. Consider the state θ' where preferences are as in state θ except that a has been moved to the top of everybody's preference. Then, $R_i(\theta')$ is a monotonic transformation of $R_i(\theta)$ at b for all i so $b \in F(\theta')$ by monotonicity, but $F(\theta') = \{a\}$ by unanimity, a contradiction.

2.7.1 Social choice functions in economic environments

Suppose agents are "selfish" and care only about their own consumption of private goods (and, in public goods economies, about the level of public goods). The preference domain is $\mathcal{R}(\Theta) = \mathcal{R}^E$, the set of all preference orderings such that each agent has a continuous, strictly monotonic and strictly convex preference ordering on the m-good commodity space $\mathbb{R}^m_+$.[11] This domain of is rich [Dasgupta, Hammond and Maskin (1979)]. From Section 2.5, if the SCF $f : \Theta \to A$ is Nash implementable, then the function $\bar{f} : \mathcal{R}^E \to A$ defined by $\bar{f}(R(\theta)) = f(\theta)$ is strategy-proof. Unfortunately, a number of papers have demonstrated the near-impossibility of strategy-proof social choice in environments with private and/or public goods.[12] For example, in exchange economies with $n = 2$, strategy-proofness plus Pareto efficiency implies dictatorship [Hurwicz (1972), Dasgupta, Hammond and Maskin (1979), Zhou (1991a)]. However, further restrictions on Θ lead to more positive results. For example, consider an "Edgeworth box" exchange economy ($n = m = 2$) where $\mathcal{R}(\Theta) = \mathcal{R}^E_0 \subset \mathcal{R}^E$, the subset of preferences in $\mathcal{R}^E$ such that both goods are normal for both agents. Let ℓ be a fixed "downward sloping line" that passes through the Edgeworth box. For each $\theta \in \Theta$ there is a unique Pareto efficient and feasible point on ℓ, which we define to be $f(\theta)$. Then $f : \Theta \to A$ is a Pareto efficient and non-dictatorial SCF which can be Nash implemented by the mechanism described in Section 2.8.

2.7.2 Social choice correspondences in economic environments

Let the preference domain $\mathcal{R}(\Theta) = \mathcal{R}^E$ be as defined in Section 2.7.1. Postlewaite showed that the Walrasian correspondence W is not monotonic, since a change in preferences over *non-feasible* consumption bundles can destroy a Walrasian equilibrium on the boundary of the feasible set [Hurwicz, Maskin and Postlewaite (1995)]. The minimal monotonic extension of W, in the sense of Sen (1995), is the *constrained Walrasian correspondence* W^c [Hur-

[11] Since agents are selfish, we can talk about preferences over $\mathbb{R}^m_+$ rather than over $A = \times_{i=1}^n \mathbb{R}^m_+$.

[12] A small sample includes Ledyard and Roberts (1974), Hurwicz (1972, 1975), Hurwicz and Walker (1990), Jackson and Barbera (1995), Green and Laffont (1979), Walker (1980), Dasgupta, Hammond and Maskin (1979), Zhou, (1991a, 1991b), Satterthwaite and Sonnenschein (1981).

wicz, Maskin and Postlewaite (1995)].[13] The *core correspondence* C and the *individually rational and Pareto efficient correspondence* PI are both monotonic. It is known that $W^c \subseteq C \subseteq PI$. In fact, if F is *any* monotonic, Pareto-optimal, individually rational and continuous SCR, and the space of preferences is sufficiently rich, then $W^c \subseteq F$ [Hurwicz (1979b), Hurwicz, Maskin and Postlewaite (1995)]. Partial converses to this result, using the additional hypothesis that attainable sets are convex or starlike, were obtained by Hurwicz (1979c) and Schmeidler (1982). Thomson (1979) showed that similar results can be obtained if the condition of individual rationality is replaced a condition of *fairness*. In the public goods economy, if F is any monotonic, Pareto-optimal, individually rational and continuous SCR, and the space of preferences is sufficiently rich, then $L^c \subseteq F$, where L^c is the *constrained Lindahl correspondence* [Hurwicz, Maskin and Postlewaite (1995)]. Non-monotonic SCRs in economic environments include the Shapley value correspondence and various bargaining solutions that use cardinal information about preferences.

2.7.3 Preferences that satisfy the single crossing property

In economic applications preference often satisfy a *single crossing condition.* In such an environment, monotonicity is rather easy to satisfy, even for single valued social choice functions, and even when $n = 2$. Suppose there is a seller and a buyer, a divisible good and "money". The feasible set is $A = \{(q, x) : x \geq 0\} \subset \mathbb{R}^2$, where q is a transfer of money from the buyer to the seller and x is the amount of the good delivered from the seller to the buyer. The state of the world is denoted $\theta = (\theta_s, \theta_b) \in [0, 1] \times [0, 1] \equiv \Theta$. The seller's utility function is $u(q, x, \theta_s)$, with $\partial u/\partial q > 0$, $\partial u/\partial x < 0$. The buyer's utility function is $v(q, x, \theta_b)$, with $\partial v/\partial q < 0$, $\partial v/\partial x > 0$. An increase in θ_s represents an increase in the sellers marginal production cost, and an increase in θ_b represents an increase in the buyers marginal valuation. More formally, the single crossing condition states that

$$\frac{\partial}{\partial \theta_s} \left| \frac{\partial u/\partial x}{\partial u/\partial q} \right| > 0 \quad \text{and} \quad \frac{\partial}{\partial \theta_b} \left| \frac{\partial v/\partial x}{\partial v/\partial q} \right| > 0$$

Under this assumption, a monotonic transformation can only take place at a boundary allocation of the form $(\bar{q}, 0)$, i.e., where there is no trade at all.

[13] The maximal monotonic sub correspondence F^* (as defined by (2)) contains all *interior* Walrasian equilibria, so $F^*(\theta) = \emptyset$ if no interior equilibrium exists.

An SCR F is monotonic if and only if the following is true: if $(\bar{q}, 0) \in F(\theta)$, and if the seller's production cost increases ($\theta_s' \geq \theta_s$) or the buyer's valuation decreases ($\theta_b' \leq \theta_b$), or both, then $(\bar{q}, 0) \in F(\theta')$.

2.8 Two Person Implementation

In the canonical mechanism for Nash implementation with $n \geq 3$, rule 2 singles out a unique deviator from the "consensus". This method does not work if $n = 2$. Suppose $n = 2$ and consider a message profile m where agent 1 claims the state is θ and agent 2 claims the state is $\theta' \neq \theta$. Since there is no way of identifying a unique deviator from a "consensus", the outcome $h(m)$ must simultaneously guarantee that agent 2 has no incentive to claim the state is θ' if the true state is θ *and* that agent 1 has no incentive to claim the state is θ if the true state is θ'. Implementability requires the existence of such an outcome for any pair of states (θ, θ').

In economic environments, the outcome "zero consumption for all" will generally satisfy the requirement, so a consensus can be supported by punishing both agents in case of disagreement. A consequence of this is that in two-person economic environments, monotonicity is a necessary and sufficient condition for Nash implementation if we restrict attention to SCRs that never recommend zero consumption to any agent. To illustrate, consider a two-person, m-good exchange economy. The feasible set is $A = \left\{a = (a_1, a_2) \in \mathbb{R}_+^{2m} : a_1 + a_2 \leq \omega\right\}$, where $a_i \in \mathbb{R}_+^m$ is agent i's consumption vector, and ω the aggregate endowment vector. Assume $\mathcal{R}(\Theta) = \mathcal{R}^E$ as described in Section 2.7.1. Suppose F is a monotonic SCR, with $F(\Theta) \subseteq A^0 \equiv \{a \in A : a_1 \neq 0,\ a_2 \neq 0\}$. Since F is ordinal, we may assume it is defined directly on $\mathcal{R}^E$. Consider the following mechanism. Each agent $i \in \{1, 2\}$ announces an outcome $a^i = (a_1^i, a_2^i) \in A^0$ and a preference profile $R^i = (R_i^i, R_j^i) \in \mathcal{R}^E$. Let $h_i(m)$ denote agent i's consumption. Set $h_i(m) = a_i^i$ if $m_1 = m_2$ and $a^i \in F(R^i)$, *or* if $R_j^i = R_j^j, R_i^j \neq R_i^i$ and $a^j R_i^j a^i$. Otherwise, set $h_i(m) = 0$. It is easy to check that this mechanism Nash implements F.

For necessary and sufficient conditions for two-person Nash implementation in general environments, see Moore and Repullo (1990), Dutta and Sen (1991b), and Sjöström (1991). For the *unrestricted domain* it is in general impossible to secure an outcome which unambiguously punishes both agents in case of disagreement, and the result is negative.

Theorem 8 (Maskin (1999), Hurwicz and Schmeidler (1978)) *Suppose*

F is a weakly Pareto optimal SCR, $n = 2$, and $\mathcal{R}(\Theta) = \mathcal{R}_A$. Then, F is dictatorial.

2.9 Unknown Feasible Set

Consider an m-good exchange economy. In state $\theta \in \Theta$, agent i's endowment vector is $\omega_i(\theta) \in \mathbb{R}^m_+$ and the aggregate endowment is $\omega(\theta) = \omega_1(\theta) + ... + \omega_n(\theta) \in \mathbb{R}^m_+$ and the feasible set is

$$A(\theta) = \left\{ x \in \mathbb{R}^{mn}_+ : \sum_{i=1}^{n} x_i \leq \omega(\theta) \right\}.$$

Here $x_i \in \mathbb{R}^m_+$ is agent i's consumption vector and $x = (x_1, ..., x_n)$. If the message space M and the outcome function h do not depend directly on the state of the world, then feasibility requires that $h(m) \in A(\theta)$ for all $m \in M$. This requirement is exceedingly strong when A depends on θ in a non-trivial way. Hurwicz, Maskin and Postlewaite (1995) solved this problem by assuming that agents can report endowment vectors as part of their messages, and that exaggeration of the endowment is impossible. Postlewaite and Wettstein (1989), Tian (1989) and Hong (1995) considered implementation of the Walrasian correspondence when endowments are unknown to the planner. Hong (1995) also allowed for unknown production sets.

3 Other Equilibrium Concepts

In this section, we maintain the assumption that the true state of the world is common knowledge among the agents, but we consider solution concepts other than Nash equilibrium.

3.1 Undominated Nash Equilibrium

As we have seen, in some environments Nash implementation is difficult, especially for social choice *functions*. However, under a number of *refinements* of Nash equilibrium, almost *any* ordinal SCR or SCF can be implemented.

Message $m_i \in M_i$ is a *weakly dominated strategy* in state $\theta \in \Theta$ for agent $i \in N$ if and only if there exists $m_i' \in M_i$ such that $u_i(h(m_{-i}, m_i'), \theta) \geq u_i(h(m_{-i}, m_i), \theta)$ for all $m_{-i} \in M_{-i}$, and $u_i(h(m_{-i}, m_i'), \theta) > u_i(h(m_{-i}, m_i), \theta)$

for some $m_{-i} \in M_{-i}$. A Nash equilibrium is an *undominated Nash equilibrium* if and only if no player uses a weakly dominated strategy.

An SCR F satisfies *preference reversal* if and only if the following is true for all ordered pairs $(\theta,\, \theta') \in \Theta \times \Theta :$ if $F(\theta') \neq F(\theta)$ then there exists an agent $i \in N$ and outcomes b and c in A such that $cR_i(\theta)b$ and $bP_i(\theta')c$. If F satisfies preference reversal then F is clearly ordinal.[14]

Theorem 9 *[Palfrey and Srivastava (1991)] Suppose $n \geq 3$. If the SCR F satisfies preference reversal and no veto power then F is implementable in undominated Nash equilibrium.*

Proof. Since F is ordinal, we may without loss of generality assume F is defined directly of the set of possible preference profiles; that is, $F : \times_{i=1}^{n} \mathcal{R}_i \rightarrow A$ where $\mathcal{R}_i$ is agent i's preference domain. We prove the result under a mild strenghtening of preference reversal, called *value distinction:* for all $i \in N$ and all ordered pairs $(R_i, R_i') \in \mathcal{R}_i \times \mathcal{R}_i$, if $R_i' \neq R_i$ then there exists outcomes b and c in A such that cR_ib and $bP_i'c$. For the proof without this condition, see Palfrey and Srivastava (1991).

Consider the following mechanism. Agent i's message space is

$$M_i = A \times \mathcal{R}_1 \times \cdots \times \mathcal{R}_n \times \mathcal{R}_i \times Z \times Z \times Z$$

where Z is the set of all positive integers. A typical message for agent i is $m_i = (a^i, R^i, r^i, z^i, \zeta^i, \gamma^i) \in M_i$, where $a^i \in A$ is an outcome, $R^i = (R_1^i, R_2^i, ..., R_n^i) \in \times_{j=1}^{n} \mathcal{R}_j$ is a statement about the preference profile, $r^i \in \mathcal{R}_i$ is an "extra" statement about agent i's *own* preference, and (z^i, ζ^i, γ^i) are three integers. Let P_j^i denote the asymmetric part of the announced R_j^i and p^i the asymmetric part of the announced r^i. The outcome function is as follows.

Rule 1. If there exists $j \in N$ such that $(a^i, R^i) = (a, R)$ for all $i \neq j$, and $a \in F(R)$, then $h(m) = a$.

Rule 2. If Rule 1 does not apply then: (a) if there is $j \in N$ such that $j = (\sum_{k=1}^{n} z^k) \bmod (2n)$ set

$$h(m) = a^j$$

[14] But not every ordinal SCR satisfies preference reversal. For example, suppose F is ordinal, and Θ includes a state θ' where all agents are indifferent over all outcomes in A. Suppose $F(\theta') \neq F(\theta)$ for some $\theta \in \Theta$. Then, it is impossible that $bP_i(\theta')c$ for some $i \in N$ and some $b, c \in A$, so preference reversal is violated.

(b) if there is $j \in N$ such that $n + j = (\sum_{k=1}^{n} z^k) \mod(2n)$ and $\gamma^j > \zeta^{j-1}$ set

$$h(m) = \begin{cases} a^{j-1} & \text{if } a^{j-1} r^j a^{j+1} \\ a^{j+1} & \text{if } a^{j+1} p^j a^{j-1} \end{cases}$$

(c) if there is $j \in N$ such that $n + j = (\sum_{k=1}^{n} z^k) \mod(2n)$ and $\gamma^j \leq \zeta^{j-1}$ set

$$h(m) = \begin{cases} a^{j-1} & \text{if } a^{j-1} R_j^j a^{j+1} \\ a^{j+1} & \text{if } a^{j+1} P_j^j a^{j-1} \end{cases}$$

Notice that rule 1 includes the case of a consensus, $(a^i, R^i) = (a, R)$ for *all* i, as well as the case where a single agent j differs from the rest. Rule 2a is a modulo game similar to rule 3 of the canonical mechanism for Nash implementation. Rule 2b chooses agent j's most preferred outcome among a^{j-1} and a^{j+1} according to preferences r^j, and rule 2c chooses agent j's most preferred outcome among a^{j-1} and a^{j+1} according to preferences R_j^j.

Notice that references to agents $j-1$ and $j+1$ are always "modulo n". That is, if $j = 1$ then agent $j - 1$ is agent n; if $j = n$ then agent $j + 1$ is agent 1.

Let $R^* = (R_1^*, ..., R_n^*)$ denote the true preference profile. Let $U^\Gamma(R^*)$ denote the set of undominated Nash equilibria when the preference profile is R^*. The proof consists of several steps.

Step 1. If m_j is undominated for agent j then $r^j = R_j^*$. Indeed, r^j only appears in rule 2b, where "truthfully" announcing $r^j = R_j^*$ is always at least as good as any false announcement. By value distinction there exists a^{j-1} and a^{j+1} such that the preference is strict.

Step 2. If m_j is undominated for agent j then $R_j^j = R_j^*$. For, if $R_j^j \neq R_j^*$ then (since $r^j = R_j^*$ by step 1) if $n + j = (\sum_{k=1}^{n} z^k) \mod(2n)$, agent j always weakly prefers rule 2b to rule 2c, and by value distinction there exists a^{j-1} and a^{j+1} such that this preference is strict. But increasing γ^j increases the chance of rule 2b at the expense of rule 2c, without any other consequence, so m_j cannot be undominated.

Step 3. If m is a Nash equilibrium then either $(a^i, R^i) = (a, R)$ for all $i \in N$ and $a \in F(R)$, or there is j such that for all $i \neq j$, $h(m) R_i^* a$ for all $a \in A$. This follows from rule 2a (the same argument was used in the canonical mechanism for Nash implementation).

Step 4. $h(U^\Gamma(R^*)) \subseteq F(R^*)$. For, if $m \in U^\Gamma(R^*)$, then by steps 1 and 2, $R_j^j = r^j = R_j^*$ for all j. By step 3 , either rule 1 applies, in which case

$(a^i, R^i) = (a, R^*)$ for all $i \in N$ and $h(m) = a \in F(R^*)$, or else $h(m) \in F(R^*)$ by no veto power.

Step 5. $F(R^*) \subseteq h(U^\Gamma(R^*))$. Each agent j announcing $(R^j, r^j) = (R^*, R_j^*)$ "truthfully" and $a^j = a \in F(R^*)$ (and three arbitrary integers) is an undominated Nash equilibrium. (Notice that as long as agent j sets $R_j^j = r^j$, there is no possibility that γ^j can change the outcome).

Steps 4 and 5 imply $h(U^\Gamma(R^*)) = F(R^*)$. Since R^* is arbitrary, F is implemented. □

Theorem 9 is a general possibility result for implementation of ordinal SCRs. A similar possibility result was obtained for *trembling hand perfect Nash equilibria* by Sjöström (1991). If agents have strict preferences over an underlying finite set of basic alternatives, and lotteries over basic alternatives are feasible, then a sufficient condition for F to be implementable in trembling hand perfect equilibria is that F satisfies no veto power as well as its "converse": if all but one agent agrees on which alternative is the worst, then this alternative is not F-optimal. For example, if F is Condorcet consistent (in the sense that Condorcet winners are always F-optimal but Condorcet losers are never F-optimal) then F satisfies the sufficient condition.

A mechanism is *bounded* if and only if (i) each dominated strategy is dominated by some *undominated* strategy, and (ii) each agent $i \in N$ has a best response to any message profile $m_{-i} \in M_{-i}$ [Jackson (1992)]. The mechanism used by Sjöström (1991) for trembling hand perfect Nash implementation uses a finite message space, hence it is bounded. Palfrey and Srivastava's (1991) mechanism for undominated Nash implementation contains infinite sequences of strategies dominating each other, hence it is not bounded. This is illustrated by step 2 of the proof of Theorem 9. However, in economic environments any SCR which satisfies preference reversal can be implemented in undominated Nash equilibria by a bounded mechanism which does not use integer or modulo games [Jackson, Palfrey and Srivastava (1994), Sjöström (1994)]. Jackson, Palfrey and Srivastava (1994) showed that an SCR can be implemented in undominated Nash equilibria by a bounded mechanism if it is "chained".

Theorem 9 assumes $n \geq 3$. In economic environments, or more generally environments where there exists an outcome which is always the worst possible outcome for each agent, the case $n = 2$ is not more difficult than the case $n \geq 3$. We illustrate this with Jackson, Palfrey and Srivastava's (1994) simple mechanism for bounded implementation of an ordinal SCF f in a two-person, m-good exchange economy, assuming value distinction.

(Their bounded mechanism for the case $n \geq 3$ is similar.) Sjöström's (1994) mechanism is in the same spirit but works only for $n \geq 3$.

The feasible set is $A = \{a = (a_1, a_2) \in \mathbb{R}_+^{2m} : a_1 + a_2 \leq \omega\}$, where $a_i \in \mathbb{R}_+^m$ is agent i's consumption vector, and $\omega > 0$ the aggregate endowment vector. Assume $\mathcal{R}(\Theta) = \mathcal{R}^E$ as described in Section 2.7.1. Suppose f is an ordinal SCF with $f(\mathcal{R}^E) \subseteq \{a \in A : a_1 \neq 0, a_2 \neq 0\}$. Since f is ordinal, we may let $f_j(R)$ denote agent j's f-optimal consumption vector when the preference profile is $R \in \mathcal{R}^E$. Each agent $i \in \{1, 2\}$ announces *either* a preference profile $R^i = (R_1^i, R_2^i) \in \mathcal{R}^E$, *or* a pair of outcomes $(a^i, b^i) \in A \times A$ (where $a^i = (a_1^i, a_2^i)$ and $b^i = (b_1^i, b_2^i)$). Let $h_j(m)$ denote agent j's consumption.

Rule 1. Suppose both agents announce a preference profile. If $R_j^i \neq R_j^j$, then $h_i(m) = 0$. If $R_j^i = R_j^j$, then $h_i(m) = f_i(R^j)$.

Rule 2 Suppose agent i announces a preference profile R^i and agent j announces outcomes (a^j, b^j). Then, $h_j(m) = 0$. If $a^j P_i^i b^j$ then $h_i(m) = a_i^j$, otherwise $h_i(m) = b_i^j$.

Rule 3. In all other cases, $h_1(m) = h_2(m) = 0$.

Suppose the true preference profile is $R^* = (R_1^*, R_2^*)$. It is a dominated strategy to announce outcomes, since that always gives zero consumption. Announcing $R_i^i = R_i^*$ "truthfully" dominates lying, since the only effect lying can have is to give agent i an inferior allocation under rule 2. (As long as value distinction holds, there exists a^j and b^j such that the preference is strict). But, if agent j is announcing preferences, any best response for agent i must involve matching agent j's announcement R_j^j about agent j's preferences (to avoid getting zero). Thus, in the unique undominated Nash equilibrium both agents announce the true preference profile, so this mechanism implements f. The most disturbing property of this mechanism is that agent i's only reason to announce $R_i^i = R_i^*$ truthfully is that it will give him a preferred outcome in case agent $j \neq i$ plays the dominated strategy of announcing outcomes. At the undominated Nash equilibrium, if agent i were to change his report to $R_i^i \neq R_i^*$ this would not affect his own consumption at all, but he would reduce agent j's consumption to zero.

3.2 Iterated Elimination of Dominated Strategies

The *iterated* removal of weakly dominated strategies was considered by Farquharson (1969) and Moulin (1979) in their analyses of dominance solvable voting schemes. Abreu and Matsushima (1994) have shown that, in an environment where the feasible set consists of lotteries over a set of basic alternatives, any ordinal SCR can be implemented using the iterated elimination of weakly dominated strategies, if the social planner can use "small fines". Like Jackson, Palfrey and Srivastava (1994) and Sjöström (1994), Abreu and Matsushima avoid integer and modulo games. In the Abreu and Matsushima mechanism, it does not matter in which order dominated strategies are eliminated. The same is true for the dominance solvable mechanism in Sjöström (1994), where agents only report a preference ordering for themselves and two "neighbors". In Jackson, Palfrey and Srivastava's (1994) bounded mechanism described at the end of Section 3.1, once outcomes are eliminated, making a false claim about your own preference is no longer a weakly dominated strategy. Hence, the order of elimination matters. A drawback of the Abreu-Matsushima type mechanism is that it requires many rounds of elimination of dominated strategies [Glazer and Rosenthal (1992)]. Sjöström's (1994) mechanism only requires two rounds (first, elimination of false announcement of own preference, second, elimination of false announcement of neighbors' preferences). However, in contrast to Abreu and Matsushima (1994), Sjöström (1994) uses "large punishments".

3.3 Strong Nash Equilibrium

A Nash equilibrium is *strong* if and only if no group $S \subseteq N$ has a joint deviation which makes all agents in S better off. Formally, if $J \subseteq N$ is a non-empty subset of agents and m a message profile, then write $m = (m_J, m_{-J})$, where $m_J = \{m_j\}_{j \in J} \in \times_{j \in J} M_j$ and $m_{-J} = \{m_j\}_{j \notin J} \in \times_{j \notin J} M_j$. Message profile $m^* \in M$ is a *strong Nash equilibrium* for the mechanism Γ in state θ if and only if, for any subgroup $J \subseteq N$ and any joint deviation $m_J \in \times_{j \in J} M_j$, there is $i \in J$ such that $u_i(h(m^*), \theta) \geq u_i(h(m^*_{-J}, m_J), \theta)$. Maskin (1979, 1985) showed that monotonicity is a necessary condition for strong Nash implementation. Moulin and Peleg (1982) studied strong Nash implementation using the notion of effectivity functions. A necessary and sufficient condition for strong implementation, called condition γ, was given by Dutta and Sen (1991). Suh (1995) provides an algorithm for checking condition γ.[15] For economic environments, Dutta and Sen (1991) show that an SCR is implementable in strong Nash equilibrium if it satisfies Pareto optimality,

[15]Suh (1995) also corrects a mistake in Dutta and Sen's condition γ.

an interiority condition, and a version of monotonicity. For example, the Walrasian correspondence can be implemented in strong Nash equilibrium as long as Walrasian equilibria do not occur on the boundary of the feasible set.

3.4 Implementation using Extensive Form Mechanisms

An SCR F is *implementable in subgame perfect equilibrium* if and only if there exists an extensive form mechanism such that in each state $\theta \in \Theta$, the set of subgame perfect equilibrium outcomes equals $F(\theta)$. [See Moore and Repullo (1988) for formal definitions.]

To illustrate the power of subgame perfect implementation, we consider the following example due to Moore and Repullo (1988). As illustrated in the figure, there are two agents, two goods and two states of the world, $\Theta = \{C, L\}$. In state C both players have Cobb-Douglas preferences (solid indifference curves), while in state L both agents have Leontief preferences (dotted indifference curves). The f-optimal outcomes are $f(C)$ and $f(L)$. Observe that agent 2's Leontief preferences are a monotonic transformation of his Cobb-Douglas preferences at $f(C)$. But $f(C) \neq f(L)$, hence f is not monotonic. By Theorem 1, f is not Nash implementable. Moore and Repullo then considered the following stage mechanism.

Stage 1. Agent 1 announces L, in which case $f(L)$ is chosen and the game ends; or he announces C, in which case we go to stage two.

Stage 2. Agent 2 can "agree", in which case $f(C)$ is chosen and the game ends; or he can "challenge", in which case we go to stage three.

Stage 3. Agent 1 chooses between x and y.

Suppose the state is C. If we reach stage 3, then agent 1 will choose x. Thus, if we reach stage 2 agent 2 will in effect choose between $f(C)$ and x. He prefers $f(C)$. Thus, at stage 1 agent 1 will know that agent 2 will not challenge the announcement C. Since he prefers $f(C)$ to $f(L)$, player one will announce C. Thus, the unique subgame perfect equilibrium outcome in state C is $f(C)$. But suppose the state is L. In this case agent 1 will choose y if the game reaches stage 3. Since agent 2 prefers y to $f(C)$, he will challenge if the game reaches stage 2. Since agent 1 prefers $f(L)$ to y, he will announce L at stage 1. Thus, the unique subgame perfect equilibrium outcome in state L is $f(L)$. Thus, the mechanism implements f in subgame perfect equilibria. Note that stage three is never reached in equilibrium.

Moore and Repullo (1988) gave a partial characterization of subgame perfect implementable SCRs. Their result was improved on by Abreu and Sen (1990).

Definition *Property* α. There exists a set $B \subseteq A$, containing the range of F, such that for all $a \in A$ and all pairs $(\theta, \theta') \in \Theta \times \Theta$, if $a \in F(\theta)$ but $a \notin F(\theta')$ then there exists a sequence of outcomes $a_0 = a,\ a_1, ..., a_\ell, a_{\ell+1}$ in A and a sequence of agents $j(0), j(1), ..., j(\ell)$ in N such that: (i)

$$u_{j(\ell)}(a_{\ell+1}, \theta') > u_{j(\ell)}(a_\ell, \theta')$$

(ii) for $k = 0, 1, ..., \ell$,

$$u_{j(k)}(a_k, \theta) \geq u_{j(k)}(a_{k+1}, \theta)$$

(iii) for $k = 0, 1, ..., \ell$, in state θ' outcome a_k is not the top-ranked outcome in B for agent $j(k)$
(iv) if in state θ', $a_{\ell+1}$ is the top-ranked outcome in B for each agent $i \neq j(k)$, then either $\ell = 0$ or $j(\ell - 1) \neq j(\ell)$.

Theorem 10 *[Moore and Repullo (1988), Abreu and Sen (1990)] If the SCR F is implementable in subgame perfect equilibrium, then it satisfies property α. Conversely, if $n \geq 3$ and the SCR F satisfies property α and no veto power, then F is implementable in subgame perfect equilibrium.*

It is interesting to compare the sufficiency part of Theorem 10 with Theorems 2 and 9. An SCR F satisfies *monotonicity* if and only if, whenever $a \in F(\theta)$ but $a \notin F(\theta')$, there exists an agent $j \in N$ and an outcome $b \in A$ such that $aR_i(\theta)b$ and $bP_i(\theta')a$. That is, someone's preferences reverse over a and some other outcome b. An SCR F satisfies *preference reversal* if and only if, whenever $a \in F(\theta)$ but $a \notin F(\theta')$, there exists an agent $j \in N$ and two outcomes $b, c \in A$ such that $cR_i(\theta)b$ and $bP_i(\theta')c$. That is, someone's preferences reverse over two *arbitrary* outcomes b and c. An SCR F satisfies *condition* α if and only if, whenever $a \in F(\theta)$ but $a \notin F(\theta')$, there exists an agent $j = j(\ell) \in N$ and two outcomes $b, c \in A$ such that $cR_i(\theta)b$ and $bP_i(\theta')c$, and in addition b and c are connected to a by the sequence of outcomes described in the definition of property α, where $b = a_{\ell+1}$ and $c = a_\ell$. It is clear that condition α is (much) weaker than monotonicity but stronger than preference reversal.

Vartiainen (1999) found a condition, called condition β, which is both necessary and sufficient for subgame-perfect implementation. If agents have strict preferences over an underlying finite set of basic alternatives and lotteries are possible, then all Condorcet consistent voting rules satisfy condition β but all scoring rules (such as the Borda rule) violate it. Moore and Repullo (1988) showed that in an economic environment with quasi-linear preferences, Θ a finite set, and $n \geq 3$, any SCF can be implemented in subgame perfect equilibria using a finite game of perfect information (i.e., a sequential mechanism with no simultaneous moves). Finite games of perfect information can be solved using backward induction. Herrero and Srivastava (1990) derived a necessary and sufficient condition for an SCR to be implementable via backward induction using a finite game of perfect information.

Sequential mechanisms have been used by a number of authors in various social choice problems. See Farquharson (1969) and Moulin (1979) for applications to voting models, and Jackson and Moulin (1992) for public goods models.

3.5 Virtual Implementation

The problem of virtual implementation was first studied by Abreu and Sen (1991) and Matsushima (1988). Let B be a finite set of "basic alternatives", and let the set of feasible outcomes be $A = \Delta(B)$, the set of all probability distributions over B. The elements of $\Delta(B)$ are called *lotteries*. Let $d(a, b)$ denote the Euclidean distance between lotteries $a, b \in \Delta(B)$. If $F : \Theta \to \Delta(B)$ and $G : \Theta \to \Delta(B)$ are two SCRs, then F and G are *ε-close* if and only if for all $\theta \in \Theta$ there exists a bijection $\alpha_\theta : F(\theta) \to G(\theta)$ such that $d(a, \alpha_\theta(a)) \leq \varepsilon$ for all $a \in F(\theta)$. An SCR F is *virtually Nash implementable* if and only if for all $\varepsilon > 0$ there exists an SCR G which is Nash implementable and ε-close to F. If F is virtually implemented, then the social planner accepts a strictly positive probability that the equilibrium outcome is some undesirable element of B. However, this probability can be made arbitrarily small.

Suppose for all $\theta \in \Theta$, no agent is indifferent over all alternatives in B, and preferences over lotteries satisfy the von Neumann-Morgenstern axioms. Let $\Delta^0(B)$ denote the subset of $\Delta(B)$ which consists of all lotteries that give strictly positive probability to all elements of B. If $n \geq 3$, then any ordinal SCR G satisfying $G(\Theta) \subseteq \Delta^0(B)$ is Nash implementable. To see this, suppose for a moment the set of feasible outcomes is $A = \Delta^0(B)$. Then, G satisfies no veto power because no agent has a most preferred outcome in $\Delta^0(B)$. Second, G is monotonic. For suppose $a \in G(\theta)$ but $a \notin G(\theta')$. Since G is ordinal, there is $i \in N$ such that $R_i(\theta) \neq R_i(\theta')$. But then, since indifference surfaces of von Neumann-Morgenstern utility functions are hyperplanes, it is easy to see that $R_i(\theta')$ cannot be a monotonic transformation of $R_i(\theta)$ at $a \in \Delta^0(B)$. Thus, G is monotonic. By Theorem 2, G is Nash implementable when the feasible set is $\Delta^0(B)$. But then $G : \Theta \to \Delta^0(B)$ is also Nash implementable when the feasible set is extended to include all alternatives in $\Delta(B)$, since when designing a mechanism to implement G, we can always just disregard the elements of $\Delta(B)\backslash\Delta^0(B)$.

Now, suppose F is an ordinal SCR F which sometimes recommends lotteries that give zero probability to some outcome(s), so the image of F is not contained in $\Delta^0(B)$. Then F is not necessarily monotonic, because a monotonic transformation of preferences can take place on a face of $\Delta(B)$, i.e., at $a \in \Delta(B)\backslash\Delta^0(B)$. However, it is clear that F can be approximated arbitrarily closely by an ordinal SCR G such that $G(\Theta) \subseteq \Delta^0(B)$. Since any such G is Nash implementable by the above argument, we conclude that any ordinal $F : \Theta \to \Delta(B)$ is virtually implementable.

Theorem 11 *[Abreu and Sen (1991), Matsushima (1988)] Suppose $n \geq 3$, and let B be a finite set of "basic alternatives". Suppose for all $\theta \in \Theta$, no agent is indifferent over all alternatives in B, and preferences over $\Delta(B)$ satisfy the von Neumann-Morgenstern axioms. Then any ordinal SCR $F : \Theta \to \Delta(B)$ is virtually Nash implementable.*

Our derivation of Theorem 11 does not do justice to the work of Abreu and Sen (1991) and Matsushima (1988), since their mechanisms are better behaved than the canonical mechanism used to prove Theorem 2. Abreu and Sen (1991) show that their Nash equilibria are *strict*, so they satisfy stringent requirements such as trembling hand perfection. In addition, the Abreu and Sen (1991) and Matsushima (1988) mechanisms are bounded. Abreu and Matsushima (1992) showed that any ordinal SCR can be virtually implemented using iterated elimination of *strictly* dominated strategies, if the social planner can impose "small fines". However, the number of rounds of elimination of strategies may be quite large.

3.6 Double Implementation

3.6.1 Nash and undominated Nash equilibrium

In the canonical mechanism for Nash implementation, if a is the worst outcome in A for agent $i \in N$ then agent i has no good reason to participate in a consensus which leads to outcome a. Indeed, announcing $a^i = a$ is a weakly dominated strategy for agent i. Thus the canonical mechanism may have Nash equilibria where the agents use weakly dominated strategies. However, Yamato (1999) constructed a version of the modified canonical mechanism with the property that any Nash equilibrium is an undominated Nash equilibrium. Yamato's construction shows that if $n \geq 3$, then any Nash implementable SCR is doubly implementable in Nash and undominated Nash equilibrium. If $n = 2$ then the set of doubly implementable SCRs is strictly smaller than the set of Nash implementable SCRs. However, Yamato (1999) found conditions under which the two sets coincide even for $n = 2$.

3.6.2 Nash and strong Nash equilibrium

Double implementation in Nash and strong Nash equilibrium was considered by Maskin (1985), Schmeidler (1980), Corchón and Wilkie (1991), Shin and Suh (1997) and others. A necessary and sufficient condition for double implementation in Nash and strong Nash equilibrium was given by Suh (1997). He showed that the Walrasian SCR in an exchange economy, and the Lindahl SCR in a public goods economy, are doubly implementable. Double implementation in Nash and strong Nash equilibrium is the appropriate solution concept when the social planner is not sure about whether the agents will collude. If the planner has specific information about who can form a coalition and who cannot, a different notion of implementation is required. Suh (1996) gave a characterization result which applies to this situation.

3.7 Mixed Strategies

In most of the implementation literature, only the pure strategy equilibria of the mechanism are verified to be F-optimal, leaving open the possibility that there may be non-F-optimal mixed-strategy[16] equilibria.[17] In particular, in the proof of Theorem 2 we did not establish that all mixed strategy Nash equilibria are F-optimal. In fact they need not be. To see the problem, consider a mixed strategy Nash equilibrium $(\mu_1, ..., \mu_n)$ in state $\hat{\theta}$ for which one possible realization is m such that for some $j \in N$, $\theta \in \Theta$, and $a \in F(\theta)$,

$$(a^i, \theta^i) = (a, \theta) \quad \text{for all } i \neq j \tag{4}$$

but $(a^j, \theta^j) \neq (a, \theta)$. If m were the only realization of $(\mu_1, ..., \mu_n)$, then since at m each agent $i \neq j$ can induce his favorite alternative a^i he must already be getting his favorite alternative. No veto power would then guarantee $h(m) \in F(\hat{\theta})$. But if there are other possible realizations of μ_{-i}, then player i might suffer by trying to induce a^i. Suppose, for example, that m'_{-i} is another realization in which $m'_k = (\theta', a')$ for all $k \neq i$, where $a' \in F(\theta')$. Assume furthermore that

$$u_i(a^i, \theta') > u_i(a', \theta') > u_i(a, \theta') \tag{5}$$

Then, although agent i can induce a^i when others play m_{-i}, formula (5) and rule 2 of the canonical mechanism imply that he cannot induce a^i when others play m'_{-i}. Indeed, if he tries to do so, the outcome will be a', which may be a much worse outcome for him that a (the outcome that, from (5) and rule 2, he would get by sticking to m_i). Hence, if m is a realization of a mixed strategy equilibrium of the canonical mechanism, (4) does not suffice to imply that each player i is getting his favorite alternative. And so we cannot infer that $h(m)$ is F-optimal. Nevertheless, the canonical mechanism can be readily modified to take account of mixed strategies.

Suppose that $n \geq 3$. The *modified canonical mechanism* is defined as follows. Let each agent i announce an outcome a^i, a state θ^i, a mapping $\alpha^i : \Theta^n \times A^n \to A$ from profiles of announced states and outcomes to outcomes, and an integer z^i. That is, his message space is

$$M_i = A \times \Theta \times A^{\Theta^n \times A^n} \times Z$$

Let the outcome function be defined as follows:

[16] A mixed strategy for player i is a probability distribution μ_i over his set of pure strategies M_i.

[17] Exceptions include Abreu and Matsushima (1992), Jackson, Palfrey and Srivastava (1994), Sjöström (1994).

Rule 1. If $(a^i, \theta^i, \alpha^i) = (a, \theta, \alpha)$ for all $i \in N$ and $\alpha(\theta, ..., \theta, a, ..., a) = a \in F(\theta)$, then $h(m) = a$. In words, if players are unanimous in their proposal of an alternative, state, and mapping, and their proposed alternative a is prescribed by their proposed mapping α and F-optimal given their proposed state θ, then the outcome is a.

Rule 2. Suppose there exists $j \in N$ such that $(a^i, \theta^i, \alpha^i) = (a, \theta, \alpha)$ for all $i \neq j$, and $\alpha(\theta, ..., \theta, a, ..., a) = a \in F(\theta)$, but $(a^j, \theta^j, \alpha^j) \neq (a, \theta, \alpha)$. Then $h(m) = \alpha^j(\theta, ..., \theta^j, ..., \theta, a, ..., a^j, ..., a)$ if $\alpha^j(\theta, ..., \theta^j, ..., \theta, a, ..., a^j, ..., a) \in L_j(a, \theta)$, and $h(m) = a$ otherwise. That is, suppose that all agents but j propose the same alternative a, state θ, and mapping α and that a is prescribed by α and is F-optimal in state θ. Then agent j gets the alternative prescribed by his proposed mapping α^j (evaluated at $(\theta, ..., \theta^j, ..., \theta, a, ..., a^j, ..., a)$, where θ^j and a^j occur in the jth positions) provided that he does not prefer it to a at the state θ which the other agents have announced.

Rule 3. In all other cases let $h(m) = \alpha^j(\theta^1, ..., \theta^n, a^1, ..., a^n)$ where $j = \max\{i : z^i = \max_k z^k\}$. That is, the outcome is determined by the proposed mapping of the agent whose index is the highest among those announcing the maximal integer.

If F is monotonic and satisfies no veto power, then this modified mechanism implements F even when we take account of mixed strategies [Maskin (1999)]. The reason for having agent i report a mapping α^i rather than just a fixed outcome is to avoid the sort of difficulty noted above. After all, which outcome is best for an agent to propose will in general depend on the states and outcomes that the other agents propose. But if the other agents are using mixed strategies, then a player may not be able to forecast (except probabilistically) what these proposals will be. Allowing him to propose a function enables him, in effect, to propose a *contingent* outcome.

Maskin and Moore (1999) show that the extensive form mechanisms considered by Moore and Repullo (1987) and Abreu and Sen (1990) to attain subgame perfect implementation can also be suitably modified for mixed strategies. We conjecture that analogous modifications can be made for mechanisms corresponding to most of the other solution concepts that have been considered in the literature.

3.8 Renegotiation

So far we have been assuming implicitly that the mechanism Γ is immutable. In this section, however, we shall allow for the possibility that agents might choose to *renegotiate* it. Papers on implementation theory are often written as though there were an exogenous planner who simply imposes the mechanism on the agents. But this is not the only possible interpretation of the implementation setting. The agents might choose the mechanism *themselves*, in which case we can think of the mechanism as a "constitution", or a "contract" that the agents have signed. Suppose that when this contract is executed (i.e., when the mechanism is played) it results in a Pareto-inefficient outcome. Presumably, if the contract has been properly designed, this could not occur in equilibrium: agents would not deliberately design an inefficient contract. But inefficient outcomes might be incorporated in contracts as "punishments" for *deviations* from equilibrium. However, if a deviation from equilibrium has occurred, why should agents accept the corresponding outcome given that it is inefficient? Why can't they agree to "tear up" their contract (abandon the mechanism) and sign a new one resulting in a Pareto-superior outcome? In other words, why can't they *renegotiate*? Studying renegotiation is thus motivated by the idea that agents cannot commit themselves to choosing an outcome a when they could all do better with outcome b. Renegotiation acts as a constraint because if punishment is renegotiated, it may no longer serve as an effective deterrent to deviation from equilibrium.

Consider the following example [drawn from Maskin and Moore (1999)]. Let $N = \{1, 2\}$, $\Theta = \{\theta, \theta'\}$, and $A = \{a, b, c\}$. Agent 1 always prefers a to c to b. Agent 2 has preferences $cP_2(\theta)aP_2(\theta)b$ in state θ and $bP_2(\theta')aP_2(\theta')c$ in state θ'. Let f be the SCF such that $f(\theta) = a$ and $f(\theta') = b$. If we leave aside the issue of renegotiation for the moment, there is a simple mechanism that Nash implements f, namely, agent 2 chooses between a and b. He will have an incentive to choose a in state θ (since $aP_2(\theta)b$) and b in state θ' (since $bP_2(\theta')a$) and so f will be implemented. But what if he happened to choose b in state θ ? Since b is Pareto-dominated by a and c, agents will be motivated to renegotiate. If, in fact, b were renegotiated to a, there would be no problem since whether agent 2 chose a or b in state θ, the final outcome would be $a = f(\theta)$. However, if b were renegotiated to c in state θ, then.agent 2 would intentionally choose b in state θ, anticipating the renegotiation to c. Then b would not serve to punish agent 2 for deviating from the choice he is supposed to make in state θ, and the simple mechanism would no longer work. Moreover, from Theorem 12 below, no other mechanism can implement f either. Thus renegotiation can indeed constrain the SCRs that are implementable. But the example also makes clear that whether or not f is implementable depends on the precise nature of renegotiation (if b is renegotiated to a, implementation is possible; if b is renegotiated to c, it is not). Thus, rather than speaking merely of the "implementation of f" we should speak of the "implementation of f for a given renegotiation process".

In this section the feasible set is $A = \Delta(B)$, the set of all probability distributions over a set of basic alternatives B. We identify degenerate probability distributions that assign probability one to some basic alternative b with the alternative b itself. The renegotiation process can be expressed as a function $r : B \times \Theta \to B$, where $r(b, \theta)$ is the (basic) alternative to which the agents would renegotiate in state $\theta \in \Theta$ if the fall-back outcome (i.e., the outcome prescribed by the mechanism) is $b \in B$. Following Maskin and Moore (1999), assume renegotiation is *efficient* (for all b and θ, $r(b, \theta)$ is Pareto efficient in state θ) and *individually rational* (for all b and θ, $r(b, \theta) R_i(\theta) b$ for all i).[18] For each $\theta \in \Theta$, define a function $r_\theta : B \to B$ by $r_\theta(b) \equiv r(\theta, b)$. Let $x \in A$, assume for the moment that B is a finite set, and let $x(b)$ denote the probability that the lottery x assigns to outcome $b \in B$. Extend r_θ to lotteries in the following way: let $r_\theta(x) \in A$ be the lottery which assigns probability $\sum x(a)$ to basic alternative $b \in B$, where the sum is over the set $\{a : r_\theta(a) = b\}$. For B an infinite set, define $r_\theta(x)$ in the obvious analogous way. Thus we now have $r_\theta : A \to A$, for all $\theta \in \Theta$. Finally, given a mechanism $\Gamma = (M, h)$ and a state $\theta \in \Theta$, let $r_\theta \circ h$ denote the composition of r_θ and h. That is, for any $m \in M$, $(r_\theta \circ h)(m) \equiv r_\theta(h(m))$. The composition $r_\theta \circ h : M \to A$ describes the *de facto* outcome function in state θ, since any basic outcome prescribed by the mechanism will be renegotiated according to r_θ. Notice that if the outcome $h(m)$ is a non-degenerate randomization over B, then renegotiation takes place *after* the uncertainty inherent in $h(m)$ has been resolved and the mechanism has prescribed a basic alternative in B. Let $\mathcal{S}((M, r_\theta \circ h), \theta)$ denote the set of $\mathcal{S}$-equilibrium outcomes in state θ, when the outcome function h has been replaced by $r_\theta \circ h$. The mechanism $\Gamma = (M, h)$ is said to *$\mathcal{S}$-implement the SCR F for renegotiation function r* if

[18] Other assumptions on the renegotiation process have been considered. Jackson and Palfrey (1998) assume that in each state θ any agent can veto the outcome of the mechanism and enforce a "status quo" alternative $a(\theta)$. Thus, $r(b, \theta) = b$ if $bR_i(\theta)a(\theta)$ for all $i \in N$, and $r(b, \theta) = a(\theta)$ otherwise. In an exchange economy, the status quo may be the endowment point: in this case the constrained Walrasian correspondence turns out not to be implementable. Sjöström (1999) assumes *disagreement point monotonicity* (each agent prefers to renegotiate from a fall-back outcome that is better for him), while Sjöström (1996b) considers a situation where commodities can be destroyed but (re-)trade of commodities that are not destroyed cannot be prevented.

and only if $\mathcal{S}((M, r_\theta \circ h), \theta) = F(\theta)$ for all $\theta \in \Theta$. In this section we restrict our attention to SCRs F that are "essentially single-valued" in the sense that for all θ, if $a \in F(\theta)$ then $F(\theta) = \{b \in A : bI_i(\theta)a \text{ for all } i \in N\}$.

Much of implementation theory with renegotiation has been developed for its application to bilateral contracts. With $n = 2$, a simple set of conditions are necessary for implementation *regardless* of the refinement of Nash equilibrium that is adopted as the solution concept.

Theorem 12 *[Maskin and Moore (1999)] The two-agent SCR F can be implemented in Nash equilibrium (or any refinement of Nash equilibrium) for renegotiation function r only if there exists a random function $\tilde{a} : \Theta \times \Theta \to A$ such that, for all $\theta \in \Theta$,*

$$r_\theta(\tilde{a}(\theta, \theta)) \in F(\theta) \tag{i}$$

and for all $\theta, \theta' \in \Theta$,

$$r_\theta(\tilde{a}(\theta, \theta))R_1(\theta)r_\theta(\tilde{a}(\theta', \theta)) \tag{ii}$$

and

$$r_\theta(\tilde{a}(\theta, \theta))R_2(\theta)r_\theta(\tilde{a}(\theta, \theta')) \tag{iii}$$

If $\tilde{a}(\theta, \theta)$ is the (random) equilibrium outcome of a mechanism in state θ, then condition (i) ensures that the renegotiated outcome is F-optimal; condition (ii) ensures that agent 1 will not wish to deviate and act as though the state were θ'; and (iii) ensures that agent 2 will not wish to act as though the state were θ'.

The reason for introducing randomizations over basic alternatives in Theorem 12 and the following results is to enhance the possibility of punishing agents for deviating from equilibrium. By assumption, agents will always renegotiate to a Pareto-efficient alternative. Thus, if agent 1 is to be punished for a deviation (i.e., if his utility is to be reduced below the equilibrium level), then agent 2 must, in effect, be rewarded for this deviation (i.e., his utility must be raised above the equilibrium), once renegotiation is taken into account. But as we noted in Section 2.8, in a two-agent setting determining which agent has deviated may not be possible, so it may be desirable to punish *both* agents. However, this cannot be done if one agent is always rewarded when the other is punished.[19] That is were randomization comes in.

[19]Introducing a third party might make it possible to *simultaneously* punish *both* original parties in an efficient way by transferring resources to the third party. But it is sometimes argued that the third party will always collude with one of the original parties, in effect bringing us back to the case of $n = 2$.

Although, for each realization $b \in B$ of the random variable $\tilde{a} \in A$, $r_\theta(b)$ is Pareto optimal, the random variable $r_\theta(\tilde{a})$ need not be Pareto optimal (if the Pareto frontier in utility space is not linear). Hence, deliberately introducing randomization is a way to create mutual punishments despite the constraint of renegotiation.

In the case of a linear Pareto frontier[20] randomization does not help. In that case, the conditions of Theorem 12 become *sufficient* for implementation.

Theorem 13 *[Maskin and Moore (1999)] Suppose that the Pareto frontier is linear for all $\theta \in \Theta$. Then the two-agent F can be implemented in Nash equilibrium (or any refinement thereof) for renegotiation function r if there exists a random function $\tilde{a} : \Theta \times \Theta \to A$ satisfying conditions (i), (ii) and (iii) of Theorem 12.*

The conclusion that F is implementable for any refinement of Nash equilibrium follows from the fact that, under the hypothesis of Theorem 13, a mechanism in effect induces a two-person zero sum game (renegotiation ensures that outcomes are Pareto efficient, and the linearity of the Pareto frontier means that payoffs sum to a constant). Hence all refinements of Nash equilibrium are equivalent.

With "quasi-linear preferences", the Pareto frontier is linear. In this case Segal and Whinston (1998) have shown that Theorem 13 can be reexpressed in terms of first-order conditions.[21]

Theorem 14 *[Segal and Whinston (1998)] Assume (i) $N = \{1, 2\}$; (ii) the set of alternatives is*

$$A = \{(x, y_1, y_2) : x \in X,\ (y_1, y_2) \in \mathbb{R}^2,\ \text{and } y_1 + y_2 = 0\}$$

where X is a compact interval in $\mathbb{R}$; (iii) $\Theta = [\underline{\theta}, \overline{\theta}]$ is a compact interval in $\mathbb{R}$; and (iv) in each state $\theta \in \Theta$, each agent i's post-renegotiation preferences take the form: for all $(x, y_1, y_2) \in A$,

$$u_i(r_\theta(x, y_1, y_2), \theta) = v_i(x, \theta) + y_i$$

[20] Formally, the frontier is linear in state θ if, for all $b, b' \in B$ that are both Pareto optimal in state θ, the lottery $\lambda b + (1-\lambda)b'$ is also Pareto optimal, where λ is the probability of b.

[21] Notice that their feasible set is different from what we otherwise assume in this section.

where v_i is C^1. If the SCR $F : \Theta \to A$ is implementable in Nash equilibrium (or any refinement) for renegotiation function r, then there exists $\hat{x} : \Theta \to X$ such that, for all $\theta \in \Theta$ and all $i \in N$,

$$u_i(F(\theta), \theta) = \int_{\underline{\theta}}^{\theta} \frac{\partial v_i}{\partial \theta} (\hat{x}(t), t) \; dt + u_i(F(\underline{\theta}), \underline{\theta}) \tag{6}$$

Furthermore, if there is $i \in N$ such that $\frac{\partial^2 v_i}{\partial \theta \partial x} (x, \theta) > 0$ for all $x \in X$ and all $\theta \in \Theta$, then the existence of $\hat{x}$ satisfying (6) is sufficient for F's implementability by a mechanism where only agent i sends a message.

Notice that as F is essentially single valued, we may abuse notation and write $u_i(F(\theta), \theta)$ in (6).

When the Pareto frontier is not linear, then it becomes possible to punish both agents for deviations from equilibrium. We obtain the following result for implementation in subgame-perfect equilibrium.

Theorem 15 (Maskin and Moore (1999)) *The two-agent SCR F can be implemented in subgame-perfect equilibrium with renegotiation function r if there exists a random function $\tilde{a} : \Theta \to A$ such that*
(i) for all $\theta \in \Theta$, $r(\tilde{a}(\theta), \theta) \in F(\theta)$;
(ii) for all $\theta, \theta' \in \Theta$ such that $r(\tilde{a}(\theta), \theta') \notin F(\theta')$ there exists an agent k and a pair of random alternatives $\tilde{b}(\theta, \theta')$, $\tilde{c}(\theta, \theta')$ in A such that

$$r(\tilde{b}(\theta, \theta'), \theta) R_k(\theta) r(\tilde{c}(\theta, \theta'), \theta)$$

and

$$r(\tilde{c}(\theta, \theta'), \theta') P_k(\theta') r(\tilde{b}(\theta, \theta'), \theta');$$

(iii) if $Z \subseteq A$ is the union of all $\tilde{a}(\theta)$ for $\theta \in \Theta$ together with all $\tilde{b}(\theta, \theta')$ and $\tilde{c}(\theta, \theta')$ for $\theta, \theta' \in \Theta$, then no alternative $z \in Z$ is maximal for any agent i in any state $\theta \in \Theta$ even after renegotiation (that is, there exists some $d^i(\theta) \in A$ such that $d^i(\theta) P_i(\theta) r(z, \theta)$); and
(iv) there exists some random alternative $\tilde{e} \in A$ such that, for any agent i and any state $\theta \in \Theta$, every alternative in Z is strictly preferred to $\tilde{e}$ after renegotiation (that is, $r(z, \theta) P_i(\theta) r(\tilde{e}, \theta)$ for all $z \in Z$).

The definition of implementation with renegotiation suggests that characterization results should be r-translations of those for implementation when renegotiation is ruled out. That is, for each result *without* renegotiation, we can apply r to obtain the corresponding result *with* renegotiation. This is particularly clear if Nash equilibrium is the solution concept. From Theorems 1 and 2 we know that monotonicity is the key to Nash implementation. By analogy, we would expect that some form of "renegotiation-monotonicity" should be the key when renegotiation is admitted. More precisely, we say that the SCR F is *renegotiation monotonic for renegotiation function* r provided that, for all $\theta \in \Theta$ and all $x \in F(\theta)$ there is $a \in A$ such that $r(a, \theta) = x$, and if $L_i(r(a, \theta), \theta) \subseteq L_i(r(a, \theta'), \theta')$ for all $i \in N$ then $r(a, \theta') \in F(\theta')$.

Theorem 16 (Maskin and Moore (1999)) *The SCR F can be implemented in Nash equilibrium with renegotiation function r only if F satisfies renegotiation monotonicity for r. Conversely, if $n \geq 3$ and no alternative is maximal in A for two or more agents, then F is implementable in Nash equilibrium with renegotiation function r if F satisfies renegotiation monotonicity for r.*

Sjöström (1999) shows that in economic environments with $n \geq 3$, *any* Pareto-optimal and ordinal SCR can be implemented in undominated Nash equilibrium for any renegotiation function satisfying some very weak properties (*the same* mechanism can be used for *any* renegotiation function satisfying these properties). Moreover, any undominated Nash equilibrium is *coalition-proof*, so that collusion is not necessarily a problem when $n \geq 3$ (cf. footnote 19). Baliga and Brusco (1996) demonstrate that a very wide class of n-agent ($n \geq 3$) SCRs can be implemented in strong subgame-perfect equilibrium for reasonable renegotiation functions. There is also a literature on renegotiation that takes the view that the renegotiation process itself can be regulated by the original mechanism or contract [see Aghion, Dewatripont and Rey (1994) and Rubinstein and Wolinsky (1992)].

3.9 The Planner as a Player

Suppose the mechanism Γ is designed by a social planner who cannot observe the true state of the world, but who would like the set of equilibrium outcomes to coincide with the set of F-optimal outcomes in each state. The canonical mechanism for Nash implementation can be given the following intuitive explanation. Rule 1 states that if there is a consensus among the agents on state θ and outcome $a \in F(\theta)$, then a is chosen by the planner. By Rule 2, agent j's attainable set at the consensus outcome is the lower contour set $L_j(a,\theta)$. Intuitively, agent j can "object" and attain any outcome $a^j \in L_j(a,\theta)$. Monotonicity is the condition that makes such objections effective. For if $\theta' \neq \theta$ is the true state and $a \notin F(\theta')$, then by monotonicity some agent j strictly prefers to deviate from the consensus with an objection $a^j \in L_j(a,\theta)\backslash L_j(a,\theta')$. Agent j would have no reason to propose a^j in state θ since $a^j \in L_j(a,\theta)$, but he does have such an incentive in state θ' since $a^j \notin L_j(a,\theta')$. Following the logic of Farrell (1993) and Grossman and Perry (1986), this objection may convince the social planner that θ is not the true state (and therefore that a is not the right outcome), although it may not convince the planner that the true state must be θ' (there may be some third state θ'' where the agent also would have an incentive to propose a^j). Worse, even if the objection should convince the planner that the state is θ', she does not actually want to choose a^j unless it should happen that $a^j \in F(\theta')$. Thus, there is a commitment problem for the planner in the sense that she may want to deviate *ex post* from the rules she herself has laid down, much like the agents renegotiated outcomes in Section 3.8.

Corchón, Wilkie and Chakravorty (1997) discuss the planner's commitment problem under the assumption that the mechanism is operated by a "mindless servant" who is not a player. Baliga, Corchón and Sjöström (1997) assume the planner herself operates the mechanism. She gets payoff $u_0(a,\theta)$ from alternative a in state θ, and the SCR F she wants to implement is given by

$$F(\theta) \equiv \arg\max_{a\in A} u_0(a,\theta) \tag{7}$$

for all $\theta \in \Theta$. Since the planner has no commitment power, after receiving the agents' messages she must choose an alternative a which maximizes $u_0(a,\theta)$ given her beliefs about θ. Baliga, Corchón and Sjöström (1997) found necessary and sufficient conditions for implementation, assuming the planner's beliefs must satisfy restrictions similar to those in Farrell (1993) and Grossman and Perry (1986). Although there is no simple relationship between these conditions and standard conditions such as monotonicity, in many cases removing the planner's commitment power makes the implementation problem much more difficult.

If the planner can commit, then explicitly allowing her to participate as a player in the game expands the set of implementable social choice rules.

Consider a utilitarian social planner, with

$$u_0(a,\theta) = \sum_{i=1}^{n} u_i(a,\theta)$$

The SCR F she wants to implement is the utilitarian SCR which is not ordinal, since multiplying some u_i by a scalar can change the utilitarian optimum. If the planner does not play, then F cannot be implemented using any non-cooperative solution concept (it cannot even be virtually implemented). But suppose we let the social planner (who has no information about θ) participate as a player by sending a message.[22] Let the solution concept be Bayesian-Nash equilibrium. In the expanded set of players, a preference reversal condition is trivially satisfied, since by definition the planner has different ordinal preferences in any two states that have different utilitarian optima. The utilitarian optimum can now be implemented for "generic" prior beliefs over Θ [Baliga and Sjöström (1999)].

4 Bayesian Implementation

Now we drop the assumption that each agent knows the true state of the world and consider the case of *incomplete information*. If the true state is θ, but agent j thinks it can be any of a number of states, then agent j will need to predict how the other agents would behave in all those states that he considers possible. This links the states together, and we can no longer consider each state separately. The basic solution concept for incomplete information environments is *Bayesian Nash equilibrium*. Bayesian Nash implementation in economic environments with *non-exclusive*[23] information was studied by Postlewaite and Schmeidler (1986) and Palfrey and Srivastava (1987). Palfrey and Srivastava (1989a), Mookherjee and Reichelstein (1990) and Jackson (1991) proved more general results for environments where the agents may have exclusive information. Blume and Easley (1990) showed that the Walrasian correspondence can be Bayesian Nash implemented if information is non-exclusive but not otherwise.

[22] Hurwicz (1979b) implemented the Walrasian and Lindahl SCRs under the assumption that there is an "auctioneer" whose payoff function agrees with the social choice rule as in equation (7). However, he assumed the auctioneer knew the state of the world.

[23] Information is non-exclusive if each agent's information can be inferred with certainty by pooling the other $n-1$ agents' information.

4.1 Definitions

A generic state of the world is denoted $\theta = (\theta_1, ..., \theta_n)$, where θ_i is agent i's *type*. Let Θ_i denote the finite set of possible types for agent i, and let $\Theta \equiv \Theta_1 \times ... \times \Theta_n$. Agent i knows his own type θ_i but may be unsure about $\theta_{-i} \equiv (\theta_1, ..., \theta_{i-1}, \theta_{i+1}, ..., \theta_n)$. Agent i's payoff depends only on his own type and the final outcome (*private values*). Thus, if the outcome is $a \in A$ and the state of the world is $\theta = (\theta_1, ..., \theta_n) \in \Theta$, then we will write agent i's payoff as $u_i(a, \theta_i)$ rather than $u_i(a, \theta)$. Suppose there exists a common prior distribution on Θ, denoted p. Conditional on knowing his own type θ_i, agent i's posterior distribution over $\Theta_{-i} \equiv \times_{j \neq i} \Theta_j$ is denoted $p(\cdot \mid \theta_i)$. It can be deduced from p using Bayes rule for any θ_i which occurs with positive probability. If $g : \Theta_{-i} \to A$ is any function, and $\theta_i \in \Theta_i$, then the expectation of $u_i(g(\theta_{-i}), \theta_i)$ conditional on θ_i is denoted

$$E\{u_i(g(\theta_{-i}), \theta_i) \mid \theta_i\} = \sum_{\theta_{-i} \in \Theta_{-i}} p(\theta_{-i} \mid \theta_i) u_i(g(\theta_{-i}), \theta_i)$$

A strategy profile in the mechanism $\Gamma = (M, h)$ is denoted $\sigma = (\sigma_1, ..., \sigma_n)$, where for each i, $\sigma_i : \Theta_i \to M_i$ is a function which specifies the messages sent by agent i's different types. The message profile sent at state θ is denoted $\sigma(\theta) = (\sigma_1(\theta_1), ..., \sigma_n(\theta_n))$, and the message profile sent by agents other than i in state $\theta = (\theta_{-i}, \theta_i)$ is denoted

$$\sigma_{-i}(\theta_{-i}) = (\sigma_1(\theta_1), ..., \sigma_{i-1}(\theta_{i-1}), \sigma_{i+1}(\theta_{i+1}), ..., \sigma_n(\theta_n)).$$

Let Σ denote the set of possible strategy profiles. Strategy profile $\sigma \in \Sigma$ is a *Bayesian Nash Equilibrium* if and only if for all $i \in N$ and all $\theta_i \in \Theta_i$,

$$E\{u_i(h(\sigma(\theta), \theta_i) \mid \theta_i\} \geq E\{u_i(h(\sigma_{-i}(\theta_{-i}), m_i'), \theta_i) \mid \theta_i\}$$

for all $m_i' \in M_i$. All expectations are with respect to θ_{-i} conditional on θ_i. Let $BNE^\Gamma \subseteq \Sigma$ denote the set of Bayesian Nash Equilibria for mechanism Γ.

A *social choice set* (SCS) is a collection $\hat{F} = \{f_1, f_2, ...\}$ of social choice functions, i.e. a subset of A^Θ. We identify the SCF f with the SCS $\hat{F} = \{f\}$. Define the composition $h \circ \sigma : \Theta \to A$ by $(h \circ \sigma)(\theta) = h(\sigma(\theta))$. The mechanism $\Gamma = (M, h)$ implements the SCS $\hat{F}$ in Bayesian Nash equilibrium if and only if (i) for all $f \in \hat{F}$, there is $\sigma \in BNE^\Gamma$ such that $h \circ \sigma = f$, and (ii) for all $\sigma \in BNE^\Gamma$ there is $f \in \hat{F}$ such that $h \circ \sigma = f$.

A set $\Theta' \subset \Theta$ is a *common knowledge event* if and only if $\theta' = (\theta'_{-i}, \theta'_i) \in \Theta'$ and $\theta = (\theta_{-i}, \theta_i) \notin \Theta'$ implies, for all $i \in N$, $p(\theta_{-i} \mid \theta'_i) = 0$. An SCS $\hat{F}$ satisfies *closure* if and only if the following is true: for any two common knowledge events Θ_1, Θ_2 that partition Θ and any pair $f_1, f_2 \in \hat{F}$, we have $f \in \hat{F}$ where f is defined by

$$f(\theta) = \begin{cases} f_1(\theta) & \text{if } \theta_1 \in \Theta_1 \\ f_2(\theta) & \text{if } \theta_2 \in \Theta_2 \end{cases}$$

Closure is a necessary property for Bayesian Nash implementation of an SCS.[24] In complete information environments, it is without loss of generality to consider implementation of an SCR rather than an SCS because Bayesian Nash implementation is equivalent to Nash implementation and any SCS which satisfies closure is equivalent to an SCR [Jackson (1991)]. This is not the case in incomplete information environments, however.

4.2 Incentive Compatibility

An SCR f is *incentive compatible* if and only if for all $i \in N$ and all $\theta_i, \theta'_i \in \Theta_i$,

$$E\left\{u_i(f(\theta_{-i}, \theta_i), \theta_i) \mid \theta_i\right\} \geq E\left\{u_i(f(\theta_{-i}, \theta'_i), \theta_i) \mid \theta_i\right\}$$

A SCS $\hat{F}$ is incentive compatible if and only if each $f \in \hat{F}$ is incentive compatible.

Theorem 17 *If the SCS $\hat{F}$ is implementable in Bayesian Nash equilibria, then $\hat{F}$ is incentive compatible.*

[24]Suppose $A = \{a, b\}$, where a is "a prize to Adam" and b is "a prize to Bob", and $\Theta = \{\theta, \theta'\}$. The social planner's preferences are not separable across states and are not compatible with any SCR, but are represented by the SCS $\hat{F} = \{f_1, f_2\}$, where $f_1(\theta) = f_2(\theta') = a$, $f_1(\theta') = f_2(\theta) = b$. That is, to give the prize to Adam in state θ and to Bob in state θ' is an optimal plan, and to give the prize to Bob in state θ and to Adam in state θ' is also optimal, but giving the prize to the same person in both states is not an optimal plan [cf. Diamond (1967)]. If the true state is common knowledge then $\hat{F}$ cannot be implemented since closure is violated. Intuitively, any mechanism that implements $\hat{F}$ would have both a and b as equilibrium outcomes in both states, but then there would be no way to guarantee that the outcomes in the two states would be different, as required by both f_1 and f_2.

Proof. To get a contradiction, suppose $\Gamma = (M, h)$ implements $\hat{F}$, but some $f \in \hat{F}$ is not incentive compatible. Then there is $i \in N$ and $\theta_i, \theta_i' \in \Theta_i$ such that

$$E\left\{u_i(f(\theta), \theta_i) \mid \theta_i\right\} < E\left\{u_i(f(\theta_{-i}, \theta_i'), \theta_i) \mid \theta_i\right\} \tag{8}$$

where $\theta = (\theta_{-i}, \theta_i)$. Suppose $\sigma \in BNE^\Gamma$ and $h \circ \sigma = f$. If agent i's type θ_i uses the equilibrium strategy $\sigma_i(\theta_i)$, his expected payoff is

$$E\left\{u_i(h(\sigma(\theta)), \theta_i) \mid \theta_i\right\} = E\left\{u_i(f(\theta), \theta_i) \mid \theta_i\right\} \tag{9}$$

If instead he were to send the message $m_i' = \sigma_i(\theta_i')$, he would get

$$E\left\{u_i(h(\sigma_{-i}(\theta_{-i}), \sigma_i(\theta_i')) \mid \theta_i)\right\} = E\left\{u_i(f(\theta_{-i}, \theta_i'), \theta_i) \mid \theta_i\right\} \tag{10}$$

But equations (8), (9) and (10) contradict the definition of Bayesian Nash equilibrium. □

The mechanism Γ is a *revelation mechanism* if and only if $M_i = \Theta_i$ for all $i \in N$. In a revelation mechanism, a message is simply an announcement of ones own type. Theorem 17 implies the *revelation principle*: if $\hat{F}$ is implementable, then for each $f \in \hat{F}$, truth telling is a Bayesian Nash equilibrium for the revelation mechanism (M, h) where $h = f$ and $M_i = \Theta_i$ for each $i \in N$ [Dasgupta, Hammond and Maskin (1979), Myerson (1979), Harris and Townsend (1981)]. However, the revelation mechanism will in general also have undesirable untruthful Bayesian Nash equilibria, in which case the revelation mechanism does not fully implement f [Postlewaite and Schmeidler (1986), Repullo (1986)]. We now consider the problem of (full) implementation.

4.3 Bayesian Monotonicity

A *deception for agent* i is a function $\alpha_i : \Theta_i \to \Theta_i$. A *deception* $\alpha = (\alpha_1, ..., \alpha_n)$ consists of a deception α_i for each agent i. Let $\alpha(\theta) \equiv (\alpha_1(\theta_1), ..., \alpha_n(\theta_n))$ and $\alpha_{-i}(\theta_{-i}) \equiv (\alpha_1(\theta_1), ..., \alpha_{i-1}(\theta_{i-1}), \alpha_{i+1}(\theta_{i+1}), ..., \alpha_n(\theta_n))$. A deception α is *compatible* if and only if $p(\alpha(\theta)) > 0$ for all θ such that $p(\theta) > 0$.

Definition *Bayesian monotonicity.* For all $f \in \hat{F}$, and all compatible deceptions α such that $f \circ \alpha \notin \hat{F}$, there exists $i \in N$, $\theta_i' \in \Theta_i$ and a function $y : \Theta \to A$ such that

$$E\left\{u_i(f(\theta), \theta_i) \mid \theta_i\right\} \geq E\left\{u_i(y(\theta), \theta_i) \mid \theta_i\right\} \tag{11}$$

for all $\theta_i \in \Theta_i$ and

$$E\left\{u_i(f(\alpha(\theta_{-i},\theta_i')),\theta_i') \mid \theta_i'\right\} < E\left\{u_i(y(\alpha(\theta_{-i},\theta_i')),\theta_i') \mid \theta_i'\right\} \tag{12}$$

for some $\theta_i' \in \Theta_i$.

This definition is due to Palfrey and Srivastava (1989a) and is weaker than the version given by Jackson (1991), who did not require deceptions to be compatible. A related condition called *selective elimination* was used by Mookherjee and Reichelstein (1990). They showed how mechanisms for full implementation can be built from incentive compatible revelation mechanisms by adding messages in order to eliminate undesirable equilibria.

The proof of Theorem 1 shows that with complete information, undesirable Nash equilibria always exist if the SCR is not monotonic. Bayesian monotonicity generalizes monotonicity to the case of incomplete information [Postlewaite and Schmeidler (1986), Palfrey and Srivastava (1989a), Jackson (1991)]. With incomplete information, undesirable Bayesian Nash equilibria always exist if the SCS is not Bayesian monotonic.

Theorem 18 *If the SCS $\hat{F}$ is implementable in Bayesian Nash equilibrium, then $\hat{F}$ is Bayesian monotonic.*

Proof. Suppose the mechanism $\Gamma = (M, h)$ implements $\hat{F}$ in Bayesian Nash equilibrium. For each $f \in \hat{F}$ there is $\sigma \in BNE^{\Gamma}$ such that $h \circ \sigma = f$. Let α be a deception such that $f \circ \alpha \notin \hat{F}$. Now, $\sigma \circ \alpha \in \Sigma$ is a strategy profile such that in state $\theta \in \Theta$ the agents behave as they would under σ if their types were $\alpha(\theta)$, i.e. they "deceptively" send message profile $(\sigma \circ \alpha)(\theta) = \sigma(\alpha(\theta))$. Since $h \circ (\sigma \circ \alpha) = f \circ \alpha \notin \hat{F}$, it follows that $\sigma \circ \alpha \notin BNE^{\Gamma}$. Therefore, some type $\theta_i' \in \Theta_i$ must have a message $m_i' \in M_i$ such that

$$E\left\{u_i(h(\sigma(\alpha(\theta))),\theta_i') \mid \theta_i'\right\} < E\left\{u_i(h(\sigma_{-i}(\alpha_{-i}(\theta_{-i})),m_i'),\theta_i') \mid \theta_i'\right\} \tag{13}$$

Let $y : \Theta \to A$ be defined by $y(\theta) = h(\sigma_{-i}(\theta_{-i}), m_i')$. Note that $y(\theta)$ is independent of θ_i, and

$$y(\alpha(\theta)) = h(\sigma_{-i}(\alpha_{-i}(\theta_{-i})), m_i')$$

Now (11) follows from the definition of Bayesian Nash equilibrium, and (12) follows from (13). $\square$

Thus, the three conditions of closure, Bayesian monotonicity and incentive compatibility are necessary for Bayesian Nash implementation. Jackson (1991) showed that in economic environments with at least three agents, these three conditions are also sufficient. Earlier, Palfrey and Srivastava (1989a) had shown that closure, Bayesian monotonicity, and a slightly stronger version of incentive compatibility are sufficient conditions in economic environments.[25] For general environments, Jackson (1991) shows that closure, Bayesian monotonicity and a condition called *monotonicity-no-veto* together are sufficient for implementation with $n \geq 3$.

As is the case in complete information environments, equilibrium refinements makes it possible to dispense with the monotonicity condition. Palfrey and Srivastava (1989b) showed that any incentive-compatible SCF can be implemented in undominated Bayesian Nash equilibrium if $n \geq 3$ and agents are never indifferent across all alternatives. Abreu and Matsushima considered an incomplete information version of the model described in Section 3.5, where A is the set of lotteries over a finite set of basic alternatives. They showed that an SCF can be virtually implemented using iterated elimination of undominated strategies if and only if it satisfies incentive compatibility and a weak measurability condition. Virtual Bayesian Nash implementation was studied by Duggan (1997) and Serrano and Vohra (1999). Baliga (1999), Bergin and Sen (1997) and Brusco (1995) studied Bayesian implementation using sequential mechanisms.

4.4 Non-Parametric, Robust and Fault Tolerant Implementation

In general, the Bayesian implementation literature implicitly assumes that the mechanism designer has knowledge about the common prior distribution p. As argued by Jackson and Moulin (1992) this requirement is very strong. The requirement is relaxed by Choi and Kim (1996), who construct a mechanism for *non-parametric* implementation in undominated Bayesian-Nash equilibrium in a public goods economy. They assume types are independent and agents share a common prior p but the mechanism designer does

[25] Postlewaite and Schmeidler (1986) showed that closure and Bayesian monotonicity are sufficient for implementation in economic environments with non-exclusive information if $n \geq 3$. This follows from the fact that in such environments, incentive compatibility is a vacuous condition.

not necessarily know p. Each agent is asked to *announce his own beliefs* as well as the beliefs of a "neighbor". The mechanism is designed in such a way that all agents announce their true beliefs at equilibrium. Duggan and Roberts (1997) introduced a notion of *robust* implementation, where the social planner is assumed to have a point estimate of the agents' prior p, but implementation is robust against small mis-specifications in this estimate.

A different kind of robustness was introduced by Corchón and Ortuño-Ortin (1995), who assumed agents are divided into local communities, each with at least three members. The social planner knows that information is complete within a community, but she does not necessarily know what agents in one community believe about members of other communities. Implementation should be robust against different possible inter-community information structures. Yamato (1994) showed that an SCR is robustly implementable in this sense if and only if it is Nash implementable.

Eliaz (2000) introduced *fault tolerant* implementation. The idea is that mechanisms ought not to break down if there are a few "faulty" agents who do not understand the rules of the game or make mistakes. Suppose neither the social planner nor the (non-faulty) agents know which agent (if any) is faulty, but all other aspects of the state are known to the (non-faulty) agents. Eliaz defines a Nash equilibrium to be k-fault tolerant if it is robust against deviations by at most k faulty players and gives necessary and sufficient conditions for implementation when $k < \frac{1}{2}n - 1$.

References

Abreu, D. and H. Matsushima (1990) "Virtual Implementation in Iteratively Undominated Strategies: Incomplete Information", Mimeo, Princeton University

Abreu, D. and H. Matsushima (1992) "Virtual Implementation in Iteratively Undominated Strategies: Complete Information" *Econometrica* **60**: 993-1008

Abreu, D. and H. Matsushima (1994) "Exact Implementation" *Journal of Economic Theory* **64:**1-19

Abreu, D. and A. Sen (1990) "Subgame Perfect Implementation: A Necessary and Sufficient Condition", *Journal of Economic Theory* **50**: 285-299

Abreu, D. and A. Sen (1991) “Virtual Implementation in Nash Equilibria”, *Econometrica* **59**: 997-1022

Aghion, P., M. Dewatripont and P. Rey (1994) “Renegotiation Design with Unverifiable Information”, *Econometrica* **62** (2): 257-282

Baliga, S. (1999) “Implementation in Incomplete Information Environments: the Use of Multi-Stage Games,” *Games and Economic Behavior* **27**: 173-183

Baliga S. and S. Brusco (1996) “Collusion, Renegotiation and Implementation,” to appear in *Social Choice and Welfare*

Baliga, S., L. Corchón and T. Sjöström (1997) “The Theory of Implementation when the Planner is a Player”, *Journal of Economic Theory* **77**:15-33

Baliga, S. and T. Sjöström (1999) “Interactive Implementation”, *Games and Economic Behavior* **27**: 38-63

Banks, J., J. Ledyard, and D. Porter (1988) “An Experimental Analysis of Unanimity in Public Goods Provision Mechanisms”, *Review of Economic Studies* **55**: 301-322

Bergin, J. and Sen, A. (1997) “Extensive Form Implementation in Incomplete Information Envirnoments”, Mimeo, Queens University

Blume, L. and D. Easley (1990) “Implementation of Walrasian Expectations Equilibria”, *Journal of Economic Theory* **51**: 207-227.

Brusco, S. (1995) “Perfect Bayesian Implementation”, *Economic Theory* **5**: 429-444

Cabrales, A. (1996) “Adaptive Dynamics and the Implementation Problem with Complete Information”, Mimeo, Universitat Pompeu Fabra

Cabrales, A. and G. Ponti (1996) “Implementation, Elimination of Weakly Dominated Strategies, and Evolutionary Dynamics”, Mimeo, Universitat Pompeu Fabra

Chakravorty, B. (1991) “Strategy Space Reduction for Feasible Implementation of Walrasian Performance,” *Social Choice and Welfare* **8**: 235-245

Chakravorty, B., L. Corchón and S. Wilkie (1997), "Credible Implementation", to appear in *Games and Economic Behavior*

Chen, Y. and C.R. Plott (1996), "The Groves-Ledyard Mechanism: An Experimental Study of Institutional Design" *Journal of Public Economics* **59**: 335-364

Chen, Y. and F.F. Tang (1997), "Learning and Incentive Compatible Mechanisms for Public Goods Provision: An Experimental Study" Mimeo, University of Michigan

Choi, J. and T. Kim (1996), "Nonparametric, Efficient Public Decision Mechanism: Undominated Bayesian Nash Implementation", Mimeo, Seoul National University

Clarke, E.H. (1971), "Multipart Pricing of Public Goods," *Public Choice* **2**:19-33

Corchón, L. (1996), The Theory of Implementation of Socially Optimal Decisions in Economics, New York: St. Martin's Press

Corchón, L. and I. Ortuño-Ortin (1995) "Robust Implementation under Alternative Information Structures", *Economic Design* **1**: 159-171

Corchón L. and S. Wilkie (1991), "Double Implementation of the Ratio Correspondence by a Market Mechanism," *Economic Design* **2**: 325-337

Crawford, V. (1979), "A Procedure for Generating Pareto-Efficient Egalitarian-Equivalent Allocations," *Econometrica* **47**: 49-60

Danilov, V. (1992), "Implementation Via Nash Equilibria", *Econometrica* **60**: 43-56.

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**: 185-216

Deb, R. (1994), "Waiver, Effectivity, and Rights as Game Forms", *Economica* **61**: 167-178

Deb, R., P. K. Pattanaik, and L. Razzolini (1997) "Game Forms, Rights and the Efficiency of Social Outcomes", *Journal of Economic Theory* **72**: 74–95.

Diamond, P. (1967), "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparison of Utility: Comment", *Journal of Political Economy* **75**: 765-6

Duggan, J. (1997) "Virtual Bayesian Implementation", *Econometrica* **67**: 1175-1199

Duggan, J. and J. Roberts (1997) "Robust Implementation", Mimeo, University of Rochester

Dutta, B. and A. Sen (1991a), "Implementation Under Strong Equilibria: a Complete Characterization," *Journal of Mathematical Economics* **20**: 49-67.

Dutta, B., A. Sen and R. Vohra (1995), "Nash Implementation through Elementary Mechanisms in Economic Environments," *Economic Design* **1**: 173-204

Dutta, B. and A. Sen (1991b), "A Necessary and Sufficient Condition for Two-Person Nash Implementation," *Review of Economic Studies* **58**: 121-128.

Eliaz, K. (2000), "Fault Tolerant Implementation", Mimeo, Tel Aviv University

Farquharson, R. (1969) *The Theory of Voting*, New Haven, Conn.: Yale University press

Farrell, J. (1993), "Meaning and Credibility in Cheap-Talk games", *Games and Economic Behavior* **5**: 514-531

Gaertner, W., P. Pattanaik and K. Suzumura (1992), "Individual Rights Revisited" *Economica* **59**: 161-177

Gärdenfors, P. (1981) "Rights, Games and Social Choice", *Nous* **15**: 341-356

Gaspart, F. (1996) "Fair Implementation in the Cooperative Production Problem: Two Properties of Normal Form Mechanisms", Mimeo, FUND Namur

Gaspart, F. (1997) "A General Concept of Procedural Fairness for One-Stage Implemenentation", Mimeo, FUND Namur

Gibbard, A. "Manipulation of Voting Schemes: A General Result," *Econometrica* **41**: 587-601

Glaser, J. and R. Rosenthal (1992) "A Note on Abreu-Matsushima Mechanisms", *Econometrica* **60**: 1435-1438

Grossman, S. and M. Perry (1986) "Perfect Sequential Equilibrium", *Journal of Economic Theory* **39**: 97-119

Groves, T. (1970) "The Allocation of Resources under Uncertainty", Ph.D. Dissertation. Berkeley: University of California

Groves, T. and J. Ledyard (1977), "Optimal Allocation of Public Goods: A Solution to the 'Free Rider' Dilemma", *Econometrica* **45**: 783-811

Hammond, P. (1997) "Game Forms Versus Social Choice Rules on Models of Rights", in: K.J. Arrow, A.K. Sen and K. Suzumura, eds, Social Choice Re-examined, Vol. II (IEA Conference Volume No. 117), London: Macmillan.

Harris, M. and R. Townsend (1981) "Resource Allocation with Asymmetric Information", *Econometrica* **49**: 33-64

Harsanyi J. and R. Selten (1988), A General Theory of Equilibrium Selection in Games, MIT Press

Harstad, R.M. and M. Marrese (1981), "Implementation of Mechanisms by Processes: Public Good Allocation Experiments," *Journal of Economic Behavior and Organization* **2**: 129-151

Harstad, R.M. and M. Marrese (1982), "Behavioral Explanations of Efficient Public Goods Allocations", *Journal of Public Economics* **19**: 367-383

Hayek, F. (1945), "The Use of Knowledge in Society," *American Economic Review* **35**: 519-530

Herrero, M. and S. Srivastava (1992) "Implementation via Backward Induction", *Journal of Economic Theory* **56**: 70-88

Hong, L. (1995) "Nash Implementation in Production Economies", *Economic Theory* **5**: 401-418

Hurwicz, L. (1960), "Optimality and Informational Efficiency in Resource Allocation Processes", in K.J. Arrow, S. Karlin, and P. Suppes, eds, Mathematical Methods in the Social Sciences, Stanford: Stanford University Press

Hurwicz, L. (1972),"On informationally decentralized systems," in R. Radner and C.B. McGuire, eds, Decision and Organization (Volume in honor of J. Marshak), pp. 297-336. Amsterdam: North-Holland.

Hurwicz, L. (1979a), "Outcome Functions Yielding Walrasian and Lindahl Allocations at Nash Equilibrium Points," *Review of Economic Studies* **46**: 217-225

Hurwicz, L. (1979b), "On Allocations Attainable through Nash Equilibria," *Journal of Economic Theory* **21**: 140-165

Hurwicz, L. (1979c), "Balanced Outcome Functions Yielding Walrasian and Lindahl Allocations at Nash Equilibrium Points for Two or More Agents," in: General Equilibrium, Growth and Trade, Ed. J. Green and J. Scheinkman, New York: Academic Press

Hurwicz, L., E. Maskin and A. Postlewaite (1995), "Feasible Nash Implementation of Social Choice Rules when the Designer does not know Endowments or Production Sets," in: The Economics of Informational Decentralization, Complexity, Efficiency and Stability, Ed. J. Ledyard, Kluwer Academic Publishers

Hurwicz L. and D. Schmeidler (1978) "Construction of Outcome Functions Guaranteeing Existence and Pareto-Optimality of Nash Equilibria," *Econometrica* **46**: 1447-1474.

Jackson M. (1991). "Bayesian Implementation," *Econometrica* **59**: 461-477.

Jackson M. (1992) "Implementation in Undominated Strategies: A Look at Bounded Mechanisms," *Review of Economic Studies* **59**: 757-775

Jackson, M. and H. Moulin (1992) “Implementating a Public Project and Distributing its Cost”, *Journal of Economic Theory* **57**: 125-140

Jackson, M. and T. Palfrey (1998) “Voluntary Implementation”, Mimeo, CALTECH

Jackson, M., T. Palfrey and S. Srivastava (1994) “Undominated Nash Implementation in Bounded Mechanisms,” *Games and Economic Behavior* **6**: 474-501

Ledyard J. and J. Roberts (1974) “On the Incentive Problems with Public Goods,” Discussion Paper No 116, CMSEMS, Northwestern University

Maskin, E. (1999), “Nash Equilibrium and Welfare Optimality,” *Review of Economic Studies* **66**: 23-38

Maskin, R. (1979), “Implementation and Strong Nash Equilibrium”, in: Aggregation and Revelation of Preferences, Ed. J.J. Laffont, Amsterdam: North-Holland

Maskin, E. (1985), “The Theory of Implementation in Nash Equilibrium: a Survey,” in: Social Goals and Social Organization, Ed. L. Hurwicz, D. Schmeidler and H. Sonnenschein, Cambridge: Cambridge University Press.

Maskin, E. and J. Moore (1999), “Implementation and Renegotiation,” *Review of Economic Studies* **66**: 39-56

Matsushima, H. (1988), “A New Approach to the Implementation Problem,” *Journal of Economic Theory* **45**: 128-144

Matsushima, H (1993), “Bayesian Monotonicity with Side Payments”, *Journal of Economic Theory* **39**:107-121

McKelvey, R. (1989), “Game Forms for Nash Implementation of General Social Choice Correspondences,” *Social Choice and Welfare* **6**: 139-156

Mookherjee, D. and S. Reichelstein (1990), “Implementation via Augmented Revelation Mechanisms,” *Review of Economic Studies* **57**: 453-476

Moore, J. (1992), "Implementation, Contracts and Renegotiation in Environments with Complete Information," in: Advances in Economic Theory (vol. I), Ed. J.J. Laffont, Cambridge: Cambridge University Press.

Moore J. and R. Repullo (1988), "Subgame Perfect Implementation," *Econometrica* **56**: 1191-1220

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

Moulin, H. (1979), "Dominance Solvable Voting Schemes," *Econometrica* **47**: 1337-1352

Moulin H. and B. Peleg (1982), "Cores of Effectivity Functions and Implementation Theory", *Journal of Mathematical Economics* **10**: 115-145

Mount, K. and S. Reiter (1974) "The Informational Size of Message Spaces," *Journal of Economic Theory* **8**: 161-192

Muench, T. and M. Walker (1984) "Are Groves-Ledyard Equilibria Attainable?," *Review of Economic Studies* **50**: 393-396

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

Myerson, R. (1979), "Incentive Compatibility and the Bargaining Problem", *Econometrica* **47**: 61-74

Palfrey, T. (1992): "Implementation in Bayesian Equilibrium: the Multiple Equilibrium Problem in Mechanism Design", in: Advances in Economic Theory: sixth world congress (vol. I), Ed. J.J. Laffont, Cambridge: Cambridge University Press.

Palfrey. T. and S. Srivastava (1987): "On Bayesian Implementable Allocations," *Review of Economic Studies* **54**: 193-208

Palfrey, T. and S. Srivastava (1989a): "Implementation with Incomplete Information in Exchange Economies," *Econometrica* **57**: 115-134

Palfrey, T. and S. Srivastava (1989b) "Mechanism Design with Incomplete Information: a Solution to the Implementation Problem," *Journal of Political Economy* **97**: 668-691

Pattanaik, P.K. and K. Suzumura (1994), "Rights, Welfarism and Social Choice", *American Economic Review* (Papers and Proceedings) **84**: 435–439.

Peleg, B. (1998), "Effectivity Functions, Game Forms, Games and Rights", *Social Choice and Welfare* **15:** 67-80

Postlewaite A. and D. Schmeidler (1986). "Implementation in Differential Information Economies," *Journal of Economic Theory* **39**: 14-33.

Postlewaite A. and D. Wettstein (1989), "Continuous and Feasible Implementation," *Review of Economic Studies* **56**: 603-611

Reichelstein, S. and S. Reiter (1988) "Game Forms with Minimal Message Spaces," *Econometrica* **56**: 661-692

Repullo, R. (1986) "The Revelation Principle under Complete and Incomplete Information," in: K. Binmore and P. Dasgupta, eds., Economic Organization as Games, Oxford: Basil Blackwell

Repullo R. (1987) "A Simple Proof of Maskin's Theorem on Nash Implementation," *Social Choice and Welfare* **4**: 39-41

Roberts, K. (1979), "The Characterization of Implementable Choice Rules," in: Aggregation and Revelation of Preferences, Ed. J.J. Laffont, Amsterdam: North-Holland

Rubinstein, A. and A. Wolinsky (1992): "Renegotiation-Proof Implementation and Time Preferences," *American Economic Review* **82**: 600-614.

Saijo, T. (1988): "Strategy Space Reduction in Maskin's Theorem," *Econometrica* **56**: 693-700.

Saijo, T. Y. Tatamitani and T. Yamato (1993): "Toward Natural Implementation," *International Economic Review* **37***: 949-980*

Samuelson, P. (1954) "The pure theory of public expenditure," *Review of Economics and Statistics* **36**: 387-389.

Samuelson, P. (1955), "Diagrammatic exposition of a theory of public expenditure", *Review of Economics and Statistics* **37**: 350-356.

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

Sato, F. (1981) "On the Informational Size of Message Spaces for Resource Allocation Processes in Economies with Public Goods", *Journal of Economic Theory* **24**: 48-69

Schmeidler, D. (1980) "Walrasian Analysis via Strategic Outcome Functions", *Econometrica*, **48**: 1585-1594

Schmeidler, D. (1982) "A Condition Guaranteeing that the Nash Allocation is Walrasian", *Journal of Economic Theory* **28**: 376-378

Segal, I. and M. Whinston (1998), "The Mirrlees Approach to Mechanism Design with Renegotiation", Mimeo, Stanford and Northwestern University

Sen, A. (1995), "The implementation of social choice functions via social choice correspondences: a general formulation and a limit result", *Social Choice and Welfare* **12**: 277-292

Serrano, R. and R. Vohra (1999), "Some Limitations of Virtual Bayesian Implementation", to appear in *Econometrica*

Shin, S. and S.S. Suh (1997), "Double Implementation by a Simple Game Form in the Commons Problem," *Journal of Economic Theory* **77**: 205-213

Sjöström, T. (1991): "On the Necessary and Sufficient Conditions for Nash Implementation", *Social Choice and Welfare* **8**: 333-340.

Sjöström, T. (1993): "Implementation in Perfect Equilibria," *Social Choice and Welfare* **10**: 97-106

Sjöström, T. (1994): "Implementation in Undominated Nash Equilibria Without Using Integer Games," *Games and Economic Behavior* **6**: 502-511

Sjöström, T. (1996a): "Implementation by Demand Mechanisms," *Economic Design* **1**: 343-354

Sjöström, T. (1996b): "Credibility and Renegotiation of Outcome Functions in Economics," *Japanese Economic Review* **47**:157-169

Sjöström, T. (1999): "Undominated Nash Implementation with Collusion and Renegotiation," *Games and Economic Behavior* **26**: 337-352

Smith, V. (1980), "Experiments with a Decentralized Mechanism for Public Goods Decisions", *American Economic Review* **70**: 584-599

Suh, S. C. (1995), "An Algorithm for Checking Strong Nash Implementability", *Journal of Mathematical Economics* ***25**: 109-122*

Suh, S. C. (1996) "Implementation with Coalition Formation: A Complete Characterization", *Journal of Mathematical Economics* ***26**: 409-428*

Suh, S. C. (1997), "Double Implementation in Nash and Strong Nash Equilibria", *Social Choice and Welfare* **14**: 439-447.

Thomson, W. (1979) "Comment on Hurwicz: On Allocations Attainable Through Nash Equilibria", in: Aggregation and Revelation of Preferences, Ed. J.J. Laffont, Amsterdam: North-Holland

Thomson, W. (1995) "Concepts of Implementation", *Japanese Economic Review* **47**:133-143

Tian, G. (1989) "Implementation of the Lindahl Correspondence by a Single-Valued, Feasible, and Continuous Mechanism," *Review of Economic Studies* **56**: 613-621

de Trenqualye, P. (1988) "Stability of the Groves and Ledyard Mechanism", *Journal of Economic Theory* **46**: 164-171

Vartiainen, H. (1999) "Subgame Perfect Implementation: A Full Characterization", Mimeo, University of Helsinki

Walker, M. (1980) "On the nonexistence of a dominant strategy mechanism for making optimal public decisions", *Econometrica* **48**: 1521-1540.

Walker, M. (1981) “A Simple Incentive Compatible Mechanism for Attaining Lindahl Alloications”, *Econometrica* **49**: 65-73

Williams, S. (1986) “Realization and Nash Implementation: Two Aspects of Mechanism Design”, *Econometrica* **54**: 139-151

Yamato, T. (1992) “On Nash Implementation of Social Choice Correspondences”, *Games and Economic Behavior* **4**: 484-492

Yamato, T. (1993) “Double Implementation in Nash and Undominated Nash Equilibrium”, *Journal of Economic Theory* **59**: 311-323

Yamato, T. (1994) “Equivalence of Nash Implementability and Robust Implementability with Incomplete Information”, *Social Choice and Welfare* **11**: 289-303

Yamato, T. (1999) “Nash Implementation and Double Implementation: Equivalence Theorems,” *Journal of Mathematical Economics* **31**: 215-238.

Yoshihara, N. (1995) “Natural and Double Implementation of Public Ownership Solutions in Differentiable Production Economies”, Mimeo, Hitotsubashi University

Zhou, L. (1991a), “Inefficiency of Strategy-Proof Allocation Mechanisms in Pure Exchange Economies”, S*ocial Choice and Welfare* **8**: 247-254

Zhou, L. (1991b), “Impossibility of Strategy-Proof Mechanisms in Economies with Pure Public Goods”, *Review of Economic Studies* **58**: 107-119