CREDIBLE IMPLEMENTATION*

Bhaskar Chakravorti, Luis C. Corchón and Simon Wilkie**

WP-AD 93-02

** B. Chakravorti & S. Wilkie: Economics Research Group, Bellcore; L. Corchón: Instituto Valenciano de Investigaciones Económicas and Universidad de Alicante.

CREDIBLE IMPLEMENTATION

Bhaskar Chakravorti, Luis C. Corchón and Simon Wilkie

ABSTRACT

The theory of mechanism design and implementation abounds with clever mechanisms whose equilibrium outcomes are optimal according to some social choice rule. However, the cleverness of these mechanisms relies on intricate systems of rewards and punishments off-the-equilibrium path. Generally, it is not in the designer's best interest to go through with the reward/punishment in the "subgame" arising from some disequilibrium play. This would make the mechanism's outcome function non-credible. In the context of exchange economies, we define an appropiate notion of "credible" implementation and show that (a) the non-dictatorial Pareto correspondence can be credibly implemented (b) there exists no credibly implementable Pareto-efficient and individually rational social choice rule and (c) there exists no credibly implementable fair social choice rules. We derive necessary and sufficient conditions for credible implementability of choice rules. The main implication is paradoxical: it is suboptimal for the designer to be endowed with "too much" information about the economy. Finally, we show that the negative results persist even under weaker credibility requirements.

1. Introduction

The theory of mechanism design and implementation abounds with clever mechanisms whose equilibrium outcomes are optimal according to some social choice rule. However, the cleverness of these mechanisms relies on intricate systems of rewards and punishments that result in case of behavior off-the-equilibrium path. The voluminous literature on the subject generally ignores the designer's incentives to operate the mechanism ex post. After all the designer is usually a principal (in a principal-agent setting) or a government or a body of representatives of the agents in society whose interests and objectives are embodied in the social choice rule being implemented.

Suppose a disequilibrium strategy profile is played by the agents; it is not always in the designer's best interest to go through with the reward/punishment in this "subgame". Consequently, the mechanism's outcome function would be non-credible. In the absence of indefinitely repeated mechanism design, given the designer's inability to pre-commit to the rules of a mechanism, credibility poses a critical constraint over and above the familiar one of "incentive compatibility". This is the motivation for our paper.

Extreme examples of non-credible mechanisms are ones that either promise a drastic outcome such as confiscating all resources from the agents, or from any given agent; and extravagant ones such as a promise to give an agent whatever outcome the agent desires in certain circumstances, regardless of the effect on the rest of the economy. In general, the lack of credibility appears in more subtle ways. In the context of exchange economies, we define a notion of "credible" implementation appropriate for the standard model of social choice rule implementation and show that (a) the non-dictatorial Pareto correspondence can be credibly implemented, (b) there exists no Pareto-efficient and individually rational social choice rule that can be implemented in a credible manner and (c) there exists no envy-free and efficient social choice rule that can be implemented in a credible manner. We derive necessary and sufficient conditions for credible implementability.

Next, we address the following question: are these negative results an outcome of an excessively strong credibility requirement? We explore this issue with respect to the result (b) above. To define weaker (and, perhaps, more acceptable) requirements, we need to impose more structure on the standard implementation problem. In particular, not only must the target social choice rule be common knowledge, but the entire social utility function that the designer is maximizing needs to be specified. Note that the social choice rule, which is a correspondence from the domain of preference profiles to allocations simply conveys information about the allocations that maximize social utility for each preference profile. We show that for any social utility function satisfying a weak property of monotonicity, the negative results persist: there exists no Pareto-efficient and individually rational social choice rule[1] that can be implemented in a *weakly* credible manner. A similar result is obtained if the social utility function were to satisfy a property of lexicographic preference for efficiency. A social utility function is lexicographic in efficiency if efficient allocations are preferred by the designer to inefficient allocations. The motivation for such a requirement is strong in environments with costless renegotiation opportunities among the agents. In such circumstances, a mechanism that assigns an inefficient allocation of resources will be undermined by the agents renegotiating via a Pareto-improving trade. This indicates that the negative implications of credibility are rather robust.

The questions associated with the inability to commit to a mechanism has recently been addressed in Maskin and Moore (1989), Rubinstein and Wolinsky (1989) and Aghion, Dewatripont and Rey (1989). The approach taken in these papers is as follows: it is presumed that there is some renegotiation or bargaining process after the outcome is chosen by the planner, which leads to efficiency ex post. Our approach is radically different. Instead of focusing on the issue of commitment on the part of the agents, we consider that of commitment on the part of the designer. We do not extend the original model to allow for ex post renegotiation possibilities. Instead, the designer's objectives are consistently

---

[1]With one exception -- a rule that is not very interesting and fails strict individual rationality.

6

associated with the social choice rule she is committed to implement.

In spirit, the notion of credible implementation introduced in this paper is akin to the idea of a sequential equilibrium of a game between economic agents as first movers and the game designer as a second mover. The game designer is modelled as a dummy player whose preferences are embodied in the social choice rule. In case, the entire social welfare function is known, these preferences are fully specified by such a function. The economic agents simultaneously transmit messages to the designer who chooses an outcome (i.e. allocation of resources) in response to the messages received. For a mechanism to be credible, we require the outcome chosen to be a "best response" (consistent with knowledge of the social choice rule) according to some posterior beliefs about the underlying economy. On the other hand, for a mechanism to be weakly credible, we require the outcome chosen to be a "best response" (consistent with knowledge of the social utility function) according to some posterior beliefs about the underlying economy.

We proceed with a note of caution. We cannot simply re-write the standard simultaneous-move game form as a sequential game with the designer as second-mover (as outlined in the previous paragraph) and examine its sequential equilibria. In an equilibrium, the strategies of all players are assumed to be common knowledge. If the designer were to be modelled as a full-fledged player, technically, we would have to admit the possibility that in equilibrium, she knows what strategies the agents in the economy are playing. This would endow the designer with too much knowledge and the revelation problem would disappear since knowledge of strategies would mean that in revelation mechanisms, the designer would be able to discern truth-telling from lying. Hence, our notion of implementation incorporates the idea of a sequential equilibrium, but is not one in the formal sense.

The basic model presumes Nash equilibrium as the fundamental solution concept employed by economic agents. This translates to an implicit assumption of complete information among the agents. In addition it ignores the many criteria for refining equilibria. Since we wish to keep the focus on the issue of

credibility, we maintain the assumption of Nash equilibrium -- which has always been the starting point for all analyses. Once our main point is made in this context, it is easy to see that it applies to all other environments (with asymmetric information, refined solution concepts, abstract social choice domains, etc.) as well. The extension to the latter cases is the subject of future papers.

The central implication of our paper is the following paradox: it is sub-optimal for the designer to have "too much" information about the economy. We shall argue that the designer's freedom to design credible mechanisms decreases as more information is observable regarding the economy. Our conditions show just how much absence of information is sufficient to ensure credibility of implementation of sub-correspondences of the social choice rule.

The following section presents the basic model and definitions. Section 3 contains the results on credible implementation and Section 4 contains those on weakly credible implementation. The final section concludes.

2. The Model

Exchange economies with m goods and n agents are considered. The set of agents is N. Each $i \in N$ is a triple $\langle Z_i, \succ_i, \omega_i \rangle$, where $Z_i \subseteq \mathbb{R}^M_+$ is i's consumption set, $\succ_i$ is i's (weak) binary preference relation defined on $Z_i \times Z_i$ and $\omega_i \in \mathbb{R}^M_+$ is i's initial endowment. We assume that, for all i, $\succ_i$ is complete, reflexive, transitive and continuous and monotone in z_i. Also, ω_i is fixed and is publicly observable and the profile $(\succ_i)_{i \in N}$ is observable to all $i \in N$ but not to a social planner.

For any set X_i and an element x_i with $i \in N$, the following notational convention will be used. We use x to denote $(x_i)_{i \in N}$, x_{-i} to denote $(x_j)_{j \in N \setminus \{i\}}$, x_i to denote the i-th component of x, and X to denote $\times_{i \in N} X_i$.

8

Let $\mathcal{R}_i$ denote the domain of admissible preference relations for i. $\mathcal{R}$ completely characterizes the class of economies under consideration, and we refer to $\succ \in \mathcal{R}$ as an *economy*. Let $A \equiv \{z \in Z: \sum_{i \in N} z_i = \sum_{i \in N} \omega_i\}$ denote the set of feasible allocations and let $L_i(z, \succ) \equiv \{z' \in A: z_i \succ_i z'_i\}$ denote the lower contour set for i defined at allocation z in the economy $\succ$.

A *social choice rule* is a non-empty valued correspondence $\varphi: \mathcal{R} \twoheadrightarrow A$.

The following are examples of important social choice rules:

The *Pareto-efficiency* social choice rule, *PE* is defined by

$$PE(\succ) = \{z \in A: \nexists z' \in A \text{ such that } \forall i \in N, z'_i \succ_i z_i \text{ and for some } i \in N, z' \notin L_i(z, \succ)\}.$$

The *non-dictatorial Pareto efficiency* social choice PE^* is defined by

$$PE^*(\succ) = \{z \in PE(\succ): z \gg 0\}.$$

The *individual rationality* social choice rule, *IR*, is defined by

$$IR(\succ) = \{z \in A: \forall i \in N, z_i \succ_i \omega_i\}.$$

Suppose that for all $i, j \in N$, $Z_i = Z_j$. Given $z \in A$, $i, j \in N$, define z^{ij} as the vector in A obtained by replacing z_i with z_j and vice versa.

The *envy-freeness* social choice rule *EF* is defined by

$$EF(\succ) = \{z \in A: \forall i \in N, \nexists j \in N\setminus\{i\} \text{ such that } z^{ij}_i \succ_i z_i\}.$$

Given Δ as the $(m - 1)$-dimensional unit simplex, and $p \in \Delta$, define $\bar{H}(z, p) = \{z' \in A: \forall i \in N, pz'_i = pz_i\}$ and $H_i(z, p) = \{z' \in A: pz'_i \leq pz_i\}$.

Define $z^c \in A$ by $z^c_i = z^c_j$, $\forall i, j \in N$.

The *(constrained) Walrasian* social choice rule, *W*, is defined by

$$W(\succ) = \{z \in A: \exists p \in \Delta \text{ such that } \forall i \in N, H_i(\omega, p) \subseteq L_i(z, \succ)\}.$$

9

The *(constrained) Walrasian from equal-division* social choice rule, W^e, is defined by

$$W^e(\succ) = \{z \in A: \exists p \in \Delta \text{ such that } \forall i \in N, H_i(z^e, p) \subseteq L_i(z, \succ)\}.$$

Let $\bar{\Delta}(\succ)$ and $\hat{\Delta}(\succ)$ denote the sets of Walrasian and Walrasian from equal-division prices for the economy $\succ$.

A *game form* Γ is a triple $<N, S, g>$, where, given that for all i, S_i is a message space, $S = \times_{i \in N} S_i$ and $g: S \to A$ is an outcome function. A strategy for i in Γ is a function $\sigma_i: \mathcal{R} \to S_i$, with Σ_i denoting the strategy space.

A *game designer* is a social planner who cannot observe the true economy $\succ$ and possesses a prior probability distribution defined on the domain of economies, denoted $\beta^o: \mathcal{R} \to [0, 1]$. Depending on the application, the designer could be a principal in a principal-(multi-)agent relationship; or a government; or a body of representatives of the agents. We assume that these priors are simple probability measures and use $\mathbb{B}$ to denote the class of admissible priors. The support of β^o is denoted *supp(β^o)*. The designer has preferences on A that are dependent on the economy $\succ$ and representable by a *social utility function* u: $A \times \mathcal{R} \times \mathbb{B} \to \mathbb{R}$. For each $\succ \in \mathcal{R}$, the set of allocations in A that maximize the designer's utility is $\varphi(\succ)$. The function u is normalized so that for all $\succ \in \mathcal{R}$, for all $z \in \varphi(\succ)$, for all $\beta^o \in \mathbb{B}$, $u(z, \succ, \beta^o) = u^{max}$. The class of *social utility functions consistent with* φ, $\mathcal{U}^\varphi$, is defined by $\mathcal{U}^\varphi = \{u: \forall \succ \in \mathcal{R} \ \forall z \in \varphi(\succ), u(z, \succ, \beta^o) = u^{max}\}$. Given $u \in \mathcal{U}^\varphi$, the designer's expected utility from an allocation $z \in A$ given a probability distribution $\hat{\beta}: \mathcal{R} \to [0, 1]$ is denoted $\tilde{u}(z, \hat{\beta})$.

The standard model of social choice rule implementation does not provide much information about the designer other than the fact that her objective is defined by φ; hence it is common knowledge that the underlying social utility function is <u>some</u> element of $\mathcal{U}^\varphi$. We have added some further structure by defining beliefs to formalize the credibility notion to be defined subsequently. To develop a weaker concept of credibility than the one given in the first definition below, we need to add to the minimal

10

structure available from the standard model; we must assume that a <u>particular</u> social utility function $u \in$ $\mathcal{U}^{\varphi}$ is known as well.

The agents and the designer participate in the following process. The designer chooses Γ, then the agents (simultaneously) submit their messages in Γ to the designer, who in turn decides on an allocation by employing the outcome function in Γ.

The set of *Nash equilibria of* Γ is

$$NE(\Gamma) \equiv \{\sigma \in \Sigma: \forall i \in N, \forall \sigma_i' \in S_i, \forall \succ \in \mathcal{R}, g(\sigma_i'(\succ), \sigma_{-i}(\succ)) \in L_i(g(\sigma(\succ)), \succ)\}.$$

The set of *Nash equilibrium allocations of* Γ *in* $\succ$ is

$$NE_A(\Gamma, \succ) \equiv \{z \in A: \exists \sigma \in NE(\Gamma) \text{ such that } g(\sigma(\succ)) = z\}.$$

Next, we introduce the criterion which is central to this paper.

Definition 1A: A social choice rule φ is *credibly implementable in* $\mathbb{B}$ if

$\forall \beta^\circ \in \mathbb{B}, \exists \Gamma, \exists \beta: \mathcal{R} \times S \to [0, 1]$ such that $\forall u \in \mathcal{U}^{\varphi}, \forall \succ \in \mathcal{R}$,

$$\left[\succ \in supp(\beta^\circ)\right] \Rightarrow \left[NE_A(\Gamma, \succ) = \varphi(\succ)\right] \quad \text{and}$$

$$\forall s \in S, \left[z = g(s)\right] \Rightarrow \left[\forall z' \in A, \tilde{u}(z, \beta(\bullet, s)) \geq \tilde{u}(z', \beta(\bullet, s))\right].$$

$$\left[\succ \notin supp(\beta^\circ)\right] \Rightarrow \left[\forall s \in S, \beta(\succ, s) = 0\right].$$

The credible implementation concept is an amalgam of the standard notions of Nash implementation and sequential equilibrium. Any outcome chosen by the designer should constitute a best response according to the designer's beliefs. Other than satisfaction of Bayes' Law, when applicable, these beliefs are unrestricted.

The standard objective of implementation theory is to check for the existence of a game form with certain desirable properties. Our concept checks for a game form (with the implementation properties) and

a set of beliefs (in terms of posterior probabilities) for the designer which make the game form credible.

Three important remarks are in order:

(i) The implementability concept defined above is a global one (as in Palfrey and Srivastava (1987), Chakravorti (1992a, 1992b)). A pair (Γ, β) with the appropriate properties must be shown to exist for every conceivable prior distribution in a class $\mathbb{B}$. This specification should be distinguished from one where such a pair must satisfy the appropriate properties in a given domain of economies. Hence, our notion covers both wide and narrow domains; we ensure the designer that no matter how good (or bad) her information is about the economy, she can indeed implement φ.

(ii) Also, the pair (Γ, β) must implement φ for every $u \in \mathcal{U}^{\varphi}$. This formulation is in keeping with the standard implementation model wherein the designer's preferences define a partial ordering on A: the only distinction is between φ-optimal allocations and sub-optimal allocations. This is consistent with every $u \in \mathcal{U}^{\varphi}$; hence, the mechanism must be credible with respect to any such u.

(iii) It should be observed that we have <u>not</u> defined implementation in sequential equilibrium. Also, we have not used the notion of sequential equilibrium in a formal sense.

The first of the three conditions in the definition above corresponds to the usual condition of Nash implementation. The second condition requires that any outcome recommended by the game form must be rationalizable as a best-response, via φ and the designer's beliefs, regardless of whether or not the outcome is expected to be realized in equilibrium. The third condition is a minimal requirement imposed by Bayes' Law, whenever it applies.

In environments in which there is more information available about the designer's objectives, and the underlying social utility function is common knowledge, we would have a weaker notion of credible implementability.

Definition 1B: Suppose $u \in \mathcal{U}^{\varphi}$ is known. A social choice rule φ is *weakly credibly implementable in* $\mathbb{B}$,

12

u) if

$$\forall \beta^{o} \in \mathbb{B}, \exists \Gamma, \exists \beta: \mathcal{R} \times S \to [0, 1] \text{ such that } \forall \succ \in \mathcal{R},$$

$$\left[\succ \in supp(\beta^{o}) \right] \Rightarrow \left[NE_{A}(\Gamma, \succ) = \varphi(\succ) \right] \quad \text{and}$$

$$\forall s \in S, \left[z = g(s) \right] \Rightarrow \left[\forall z' \in A, \tilde{u}(z, \beta(\bullet, s)) \geq \tilde{u}(z', \beta(\bullet, s)) \right].$$

$$\left[\succ \notin supp(\beta^{o}) \right] \Rightarrow \left[\forall s \in S, \beta(\succ, s) = 0 \right].$$

Next, we shall define some properties that are crucial to the analysis of the implications of credibility. Some additional notation is required.

Let $\tilde{\varphi}(\beta^{o}) \equiv \{ z \in \varphi(\succ): \succ \in supp(\beta^{o}) \}$.

This is the set of allocations that are φ-optimal in some economy assigned positive probability a priori by the designer.

Given $p \in \Delta$, let $\succ^{p} \in \mathcal{R}$ be defined by

$$\forall i \in N, \forall z \in A, \{ z' \in A: z'_{i} \succ_{i}^{p} z_{i} \text{ and } z_{i} \succ_{i}^{p} z'_{i} \} = \bar{H}(z, p).$$

These are the preferences with linear indifference surfaces with normal p.

Definition 2: A social choice rule φ satisfies *Property 1(B)* if

$$\forall \beta^{o} \in \mathbb{B}, \forall \succ, \succ' \in supp(\beta^{o}), \left[z \in \varphi(\succ) \text{ and } \{ L_{i}(z, \succ) \cap \tilde{\varphi}(\beta^{o}) \} \subseteq L_{i}(z, \succ') \right] \Rightarrow \left[z \in \varphi(\succ') \right].$$

In the sequel, since it will always be clear that the condition is stated relative to a class of priors B, we shall omit B from "1(B)" and simply refer to the condition above as Property 1. Property 1 is a strengthening of Maskin-monotonicity, due to Maskin (1977), which is given in the next definition.

Definition 3: A social choice rule φ satisfies *Maskin-monotonicity* if

$$\forall \succ, \succ' \in \mathcal{R}, \left[z \in \varphi(\succ) \text{ and } L_{i}(z, \succ) \subseteq L_{i}(z, \succ') \right] \Rightarrow \left[z \in \varphi(\succ') \right].$$

13

For $n > 2$, the property of Maskin-monotonicity is equivalent to that of *Nash implementability*, (see Maskin (1985)) i.e. φ is *Nash implementable* if

$$\exists \Gamma \text{ such that } \forall \succ \in \mathcal{R}, \, NE_A(\Gamma, \succ) \subseteq \varphi(\succ).$$

This concept ignores the issue of credibility of the game form Γ.

Definition 4: A social choice rule φ satisfies *Property 2 in* β^o if

$$\forall \succ \in supp(\beta^o), \left[\exists z \in \widetilde{\varphi}(\beta^o) \text{ and } i \in N \text{ such that } \forall z' \in \widetilde{\varphi}(\beta^o), \forall j \in N\backslash\{i\}, z \succ_j z'_j\right] \Rightarrow \left[z \in \varphi(\succ)\right].$$

Property 2 is a variant of the "No Veto Power" condition of Maskin (1977), in which the allocation z in the definition above is $\succ$-maximal in A for $n - 1$ agents. Our condition simply says that if in a given economy, $n - 1$ agents agree on an allocation as top ranked among all allocations in $\widetilde{\varphi}(\beta^o)$, then that allocation must be φ-optimal for the economy. Note that though Maskin's No Veto Power condition is trivially satisfied by any φ in the class of environments that we are studying, Property 2 is not necessarily met. If the prior distribution β^o has a sufficiently "rich" support, then most interesting social choice rules would satisfy Property 2 in β^o. A rich support ensures that no two agents would agree on a single allocation z as top ranked in $\widetilde{\varphi}(\beta^o)$, since $\widetilde{\varphi}(\beta^o)$ itself admits more allocations. This would lead to Property 2 being met trivially. Hence, Property 2 should be seen (with some exceptions) more as a condition on the richness of the support of β^o rather than on φ itself.

The following properties are direct conditions on the designer's beliefs.

Definition 5: A prior belief distribution β^o: $\mathcal{R} \to [0, 1]$ satisfies *Property 3* if

$$\exists p \in \Delta \text{ and } \succ^P \in \mathcal{R} \text{ such that } \succ^P \in supp(\beta^o).$$

Definition 6: A prior belief distribution β^o: $\mathcal{R} \to [0, 1]$ satisfies *Property 3A* if

$$\forall \succ \in \mathcal{R} \left[\succ \in supp(\beta^o)\right] \Rightarrow \left[\exists p \in \Delta(\succ) \text{ and } \succ^P \in \mathcal{R} \text{ such that } \succ^P \in supp(\beta^o)\right].$$

14

Definition 7: A prior belief distribution β°: $\mathcal{R} \to [0, 1]$ satisfies *Property 3B* if

$$\forall \succ \, \in \, \mathcal{R} \left[\succ \, \in \, supp(\beta^\circ) \right] \Rightarrow \left[\exists p \in \overset{\wedge}{\Delta}(\succ) \text{ and } \succ^P \, \in \, \mathcal{R} \text{ such that } \succ^P \, \in \, supp(\beta^\circ) \right].$$

The restrictions on the beliefs given in the definitions above have the following form: the designer's priors must assign positive probability on the existence of at least one "linear" economy in the domain of possible economies. Properties 3A and 3B distinguish two different criteria that such a linear economy must satisfy; if an economy $\succ$ is admissible then the priors must admit linear economies defined by a Walrasian price for $\succ$ (in the case of 3A) and a Walrasian price from equal equal-division for $\succ$ (in the case of 3B).

Next, we define a Pareto-indifference property from Gevers (1986), related to a regularity condition used in Thomson (1984, 1987).

Definition 8: A social choice rule φ is *Pareto-indifferent* if

$$\forall \succ \, \in \, \mathcal{R}, \, \forall z, z' \in A, \left[z \in \varphi(\succ) \text{ and } \forall i \in N, \, z_i \succ_i z_i' \text{ and } z_i' \succ_i z_i \right] \Rightarrow \left[z' \in \varphi(\succ) \right].$$

3. Results on Credible Implementation

In this section, we establish necessary and sufficient conditions for credible implementability of social choice correspondences. Furthermore, we identify correspondences that are credibly implementable and those that are not. These include ones that are Nash-implementable. Finally, we establish conditions on the designer's beliefs that ensure credible (partial) implementability of a large class of Nash-implementable correspondences.

Theorem 1: *If φ is credibly implementable in some $\mathbb{B}$, then it satisfies Property 1.*

Proof: Choose $\succ \, \in \, \mathcal{R}, \, \beta^\circ \in \mathbb{B}$ with $\succ \, \in \, supp(\beta^\circ)$ and $z \in \varphi(\succ)$. By the definition of credible implementation,

15

there exists $\Gamma = <N, S, g>$ and $\sigma \in NE(\Gamma)$ such that $g(\sigma(\succ)) = z \in \varphi(\succ)$. Let $\beta: \mathcal{R} \times S \to [0, 1]$ be the probability distribution representing the designer's beliefs. By the definition of a Nash equilibrium, for all $i \in N$, $\{z' \in A: \exists s'_i \in S_i \text{ such that } g(s'_i, \sigma_{-i}(\succ)) = z'\} \subseteq L_i(z, \succ)$.

Consider a second economy $\succ' \in supp(\beta^o)$ with $\{L_i(z, \succ) \cap \tilde{\varphi}(\beta^o)\} \subseteq L_i(z, \succ')$ for all $i \in N$. For any $i \in N$ and $\sigma'_i \in \Sigma_i$, by definition of implementability, in the case where $(\sigma'_i, \sigma_{-i}) \in NE(\Gamma)$, we know that $g(\sigma'_i(\succ), \sigma_{-i}(\succ)) = \varphi(\succ) \in \tilde{\varphi}(\beta^o)$. Hence, $\beta(\succ, \sigma'_i(\succ), \sigma_{-i}(\succ)) = 1$. In the case where $(\sigma'_i, \sigma_{-i}) \notin NE(\Gamma)$, by definition of credible implementation, there exists $\beta: \mathcal{R} \times S \to [0, 1]$ such that $\tilde{u}(g((\sigma'_i(\succ), \sigma_{-i}(\succ))), \beta) \geq \tilde{u}(\bar{z}, \beta)$ for all $\bar{z} \in A$ and all $u \in \mathcal{U}^\varphi$. Hence, given that (i) $\beta(\succ'', (\sigma'_i(\succ), \sigma_{-i}(\succ))) > 0$ implies $\beta^o(\succ'') > 0$ for all $\succ'' \in \mathcal{R}$ and (ii) $\varphi(\succ'')$ is the set of allocations that maximize $u(\bullet, \succ'')$ for all $\succ'' \in \mathcal{R}$ and all $u \in \mathcal{U}^\varphi$, we have $g(\sigma'_i(\succ), \sigma_{-i}(\succ)) \in \tilde{\varphi}(\beta^o)$. Otherwise, if $g(\sigma'_i(\succ), \sigma_{-i}(\succ)) \notin \tilde{\varphi}(\beta^o)$, then there exists $u \in \mathcal{U}^\varphi$ such that $u(g(\sigma'_i(\succ), \sigma_{-i}(\succ)), \succ'') < u(z'', \succ'')$ for all $z'' \in A$ and all $\succ'' \in \mathcal{R}$ -- in contradiction with the definition of credible implementation. Since $\{z' \in A: \exists s'_i \in S_i \text{ such that } g(s'_i, \sigma_{-i}(\succ)) = z'\} \subseteq \{L_i(z, \succ) \cap \tilde{\varphi}(\beta^o)\}$, and by hypothesis, $\{L_i(z, \succ) \cap \tilde{\varphi}(\beta^o)\} \subseteq L_i(z, \succ')$, it must be the case that $g(\sigma(\succ)) \in NE_A(\Gamma, \succ')$. By definition of credible implementability, $z = g(\sigma(\succ)) \in \varphi(\succ')$. ∎

Theorem 2: *Suppose* $n > 2$. *If* φ *satisfies Property 1 and Property 2 for all* $\beta^o \in \mathbb{B}$, *then* φ *is credibly implementable in* $\mathbb{B}$.

Proof: The proof is constructive and closely follows the standard proofs for establishing sufficient conditions for Nash-implementation in general social choice environments. Choose $\beta^o \in \mathbb{B}$. Consider the following procedure:

For all $i \in N$, let $S_i = \{(z(i), \succ(i), v(i)) \in A \times \mathcal{R} \times \{0, 1, 2,...\}: z(i) \in \varphi(\succ(i)); \succ(i) \in supp(\beta^o)\}$.

Definition 10: A message profile $s \in S$ satisfies *Condition A(i)* if

(I) $\forall j \in N\backslash\{i\}$, $z(j) = z^*$ for some $z^* \in A$,

(II) $\forall j \in N\backslash\{i\}$, $\succ(j) = \succ^*$ for some $\succ^* \in \mathscr{R}_i$

(III) $\forall j \in N\backslash\{i\}$, $v(j) = 0$.

Let $(z^*, \succ^*)$ be the allocation-preference pair for which parts (I) and (II) of Condition $A(i)$ are met and define $z(v)$ such that

$z(v) = z(i)$ with $\forall j \in N\backslash\{i\}$, $v(i) \geq v(j)$ and $[v(i) = v(j)] \Rightarrow [i > j]]$.

The outcome function $g: S \to A$ is defined as follows: $\forall s \in S$,

Rule 1: If Condition $A(i)$ is met for all $i \in N$, then $g(s) = z^*$.

Rule 2: If for some $i \in N$, Condition $A(i)$ is met for all $j \in N\backslash\{i\}$ and Rule 1 does not apply, then $g(s) = z(i)$ if $z(i) \in L_i(z^*, \succ^*)$. Otherwise, $g(s) = z^*$.

Rule 3: If neither Rule 1 nor Rule 2 applies, then $g(s) = z(v)$.

The resulting game form is referred to as $\Gamma(\varphi)$.

The proof involves two steps. First, we show that $\varphi(\succ) \subseteq NE_A(\Gamma(\varphi), \succ) \neq \varnothing$ for all $\succ \in supp(\beta^0)$. Choose $\succ \in supp(\beta^0)$, $z \in \varphi(\succ)$ and $s_i = (z, \succ, 0)$ for all $i \in N$. By Rule 1, $g(s) = z$. Consider a unilateral deviation by i to s'_i. By construction, either Rule 1 or Rule 2 must apply to (s'_i, s_{-i}). In either case $g(s'_i, s_{-i}) \in L_i(z, \succ)$. Hence $z \in NE_A(\Gamma(\varphi), \succ)$.

The second step involves showing that $NE_A(\Gamma(\varphi), \succ) \subseteq \varphi(\succ)$ for all $\succ \in supp(\beta^0)$ and that there exists $\beta: \mathscr{R} \times S \to [0, 1]$ satisfying the credibility restrictions. Choose $\succ \in supp(\beta^0)$, $\sigma \in NE(\Gamma(\varphi))$ and $\sigma(\succ) = s$. Suppose $g(s) \notin \varphi(\succ)$ and s satisfies either Rule 2 or Rule 3. By Property 2, given $n > 2$, there exists $k \in N$ and $i, j \in N\backslash\{k\}$ such that i and j do not prefer $g(s)$ over every other allocation in $\tilde{\varphi}(\beta^0)$. Thus, one of them, say i, can deviate to $s'_i = (z'(i), \succ'(i), v'(i))$ so that $z'(i) \in L_i(g(s), \succ) \cap \tilde{\varphi}(\beta^0)$ and $v'(i) > v(\ell)$ for all $\ell \in N\backslash\{i\}$. Since Rule 3 applies to (s'_i, s_{-i}), $g(s'_i, s_{-i}) = z'$. Since $z'(i) \notin L_i(g(s), \succ)$, this is in contradiction with the assumption that $\sigma \in NE(\Gamma(\varphi))$.

17

Suppose Rule 1 applies to s. We know that for all $i \in N$, $s_i = (z^*, \succ^*, 0)$ and $g(s) = z^*$. Any unilateral deviation by i to $s_i' \in S_i$ must be such that either Rule 1 or Rule 2 applies to (s_i', s_{-i}). By construction, $g(s_i', s_{-i}) \in L_i(z^*, \succ^*)$. Moreover, by construction of the strategy space, $g(s_i', s_{-i}) \in \tilde{\varphi}(\beta^0)$. Also, $\{z \in g(s_i', s_{-i}): s_i' \in S_i\} = L_i(z^*, \succ^*) \cap \tilde{\varphi}(\beta^0)$. Since $\sigma \in NE(\Gamma(\varphi))$, $L_i(z^*, \succ^*) \cap \tilde{\varphi}(\beta^0) \subseteq L_i(z^*, \succ)$. By Property 1, $z \in \varphi(\succ)$. ∎

As expected, the additional requirement of credibility drives a wedge between necessary and sufficient conditions for implementability. The standard constructive proofs for sufficiency of Nash implementability rely on the ability of agents to obtain their "most desired outcome" in certain parts of the game (i.e. the one that corresponds to the chasing of the highest number) and thereby reducing the set of equilibrium outcomes. The credibility constraint restricts what is achievable in this part of the game; hence, an additional condition, i.e. Property 2, is needed.

As a second observation, we note that, contrary to what may appear to be the case at first glance, the beliefs in a credible implementation are not necessarily degenerate. For example, if $g(s) \in \varphi(\succ) \cap \varphi(\succ')$ for any $\succ, \succ' \in supp(\beta^0)$, then $\beta(\bullet, s)$ may be non-degenerate. The notion of credibility is such that it excludes allocations that are not in $\tilde{\varphi}(\beta^0)$ from the range of the outcome function. For weaker notions of credibility, we refer the reader to the next section.

The following theorems identify social choice rules that meet Property 1, and those that do not. The PE rule is trivially credibly implementable, in a partial sense, by a mechanism that allocates all resources to a single agent. The more interesting question relates to correspondences that exclude such trivial outcomes, i.e. the PE^*, $PE \cap IR$ and $PE \cap EF$ correspondences.

Theorem 3: *The correspondence PE^* satisfies Property 1 for any* $\mathbb{B}$.

<u>Proof:</u> Choose $\succ, \succ' \in \mathcal{R}$ and $\beta^0 \in \mathbb{B}$ with $\succ, \succ' \in supp(\beta^0)$. Suppose that there exists $z \in PE^*(\succ) \setminus PE^*(\succ')$.

Also suppose that the hypothesis of Property 1 is met, i.e. for all $i \in N$, $\{L_i(z, \succ) \cap \tilde{PE}^*(\beta^0)\} \subseteq L_i(z, \succ')$. Since $z \notin PE^*(\succ')$, there exists $z'' \in A$ such that z'' Pareto dominates z in the economy $\succ'$. By transitivity of the Pareto-dominance relation, and by closedness of preferences, there exists $z^* \in PE^*(\succ')$ such that z^* Pareto-dominates z in the economy $\succ'$. Hence, for all $i \in N$, $z^* \notin L_i(z, \succ')$. However, by assumption, for all $i \in N$, $\{L_i(z, \succ) \cap \tilde{PE}^*(\beta^0)\} \subseteq L_i(z, \succ')$, which in turn implies that $z^* \notin L_i(z, \succ')$ must mean that $z^* \in \{A \backslash L_i(z, \succ)\} \cup \{L_i(z, \succ) \backslash \tilde{PE}^*(\beta^0)\}$. Since $z^* \in PE^*(\succ')$, we have $z^* \in \tilde{PE}^*(\beta^0)$. Therefore, for all $i \in N$, $z^* \in \{A \backslash L_i(z, \succ)\}$. But this is in contradiction with the assumption that $z \in PE^*(\succ')$. ∎

In the two-person case the intuition underlying the argument above can be seen on a diagram. Refer to Figure 1. Suppose that the hypothesis of Property 1 is met and its conclusion is not. Without loss of generality, consider $z \in PE^*(\succ)$ as the point where the indifference curves for agent 1 in the two economies $\succ$ and $\succ'$ intersect. Such an intersection must occur to ensure that $z \notin PE^*(\succ')$. Next, consider the intersection of the set $PE^*(\succ')$ with the indifference curve for agent 1 through z in economy $\succ'$. This intersection must occur to the south-east of z, otherwise we would violate the requirement that for all $i \in N$, $\{L_i(z, \succ) \cap \tilde{PE}^*(\beta^0)\} \subseteq L_i(z, \succ')$. Let z^* be such a point of intersection. Now observe that $z^* \in \tilde{PE}^*(\beta^0)$ and is strictly preferred to z by agent 2 in economy $\succ'$. However, if $z \in PE^*(\succ)$, agent 2 must clearly prefer z to z^* in economy $\succ$. This violates the hypothesis that $\{L_2(z, \succ) \cap \tilde{PE}^*(\beta^0)\} \subseteq L_2(z, \succ')$.

Corollary to Theorem 3: *Suppose $n > 2$. The PE^* correspondence is credibly implementable in $\mathbb{B}$.*

<u>Proof:</u> Observe that the PE^* correspondence satisfies Property 2 trivially for every β^0. Each agent's most preferred outcome in A is the one which allocates all resources to him; however, such outcomes do not belong to $\tilde{PE}^*(\beta^0)$. Hence, each agent would like allocations arbitrarily close to these dictatorial outcomes. No agent has a $\succ$-maximal element in $\tilde{PE}^*(\beta^0)$ for any $\succ$. By Theorems 2 and 3, PE^* is credibly implementable in $\mathbb{B}$. ∎

Theorem 4: *Suppose $\varphi \subseteq PE \cap IR$ and $\mathbb{B}$ is unrestricted. Then φ fails Property 1.*

Proof: Consider β^{o}: $\mathscr{R} \to [0, 1]$ such that $\beta^{o}(\succ) = \beta^{o}(\succ') = 0.5$ and the $PE \cap IR$ correspondence defined for $\succ$ and $\succ'$ is as given in Figure 2. $PE \cap IR(\succ) = \overline{AB}$ and $PE \cap IR(\succ') = \overline{CD}$. The example is for a two-person, two-good economy. However, it easily generalizes.

Choose $z \in \varphi(\succ) \subseteq PE \cap IR(\succ)$. Observe that $L_1(z, \succ) \cap \tilde{\varphi}(\beta^{o}) \subseteq \overline{CD} \cup \overline{AE}$ and $L_2(z, \succ) \cap \tilde{\varphi}(\beta^{o}) \subseteq \overline{EB}$. By construction, $\overline{CD} \cup \overline{AE} \subseteq L_1(z, \succ')$ and $\overline{EB} \subseteq L_2(z, \succ')$. Clearly, $z \notin PE \cap IR(\succ')$. ∎

Theorem 5: *Suppose* $\varphi \subseteq PE \cap EF$ *and* $\mathbb{B}$ *is unrestricted. Then* φ *fails Property 1.*

Proof: The argument is identical to that of the previous proof. Consider β^{o}: $\mathscr{R} \to [0, 1]$ such that $\beta^{o}(\succ) = \beta^{o}(\succ') = 0.5$ and the $PE \cap EF$ correspondence defined for $\succ$ and $\succ'$ is as given in Figure 3. The figure is drawn so that the vectors $\overrightarrow{O_1X} = \overrightarrow{O_2Y}$ and $\overrightarrow{O_2X} = \overrightarrow{O_1Y}$. The locations of the points A, B, C and D are chosen so that $PE \cap EF(\succ) = \overline{AB}$ and $PE \cap EF(\succ') = \overline{CD}$. ∎

In the examples above we considered very narrow domains. However, if a mechanism fails to deliver in such domains, it fails to credibly implement. This is because we are interested in being able to state whether or not φ is implementable in every situation regardless of how good or bad the designer's prior information is. The examples above were designed to be as simple as possible and are not pathological cases. The same arguments can easily be shown to hold for economies with at least three people in which Property 2 is satisfied. Hence, the non-implementability of these crucial sub-correspondences of PE hinges entirely on the failure to satisfy Property 1.

The examples used to prove the last two theorems point to the following paradoxical phenomenon: when the designer's information about the economy is relatively good, it is more difficult for her to design a mechanism in a credible manner. This follows from the requirement that the range of the outcome function must be contained in the support of her priors. Good information means that the support is narrow; hence there is relatively little flexibility. This raises the obvious question: just how much incompleteness of information is "enough" to guarantee credible implementability of Nash implementable rules? The rest of

the section deals with this question.

The following result establishes that Property 3A, applicable everywhere in $\mathbb{B}$ together with the mild Pareto-indifference assumption guarantees that any Nash-implementable Pareto-efficient and individual rational rule can also be credibly implemented in a partial sense, provided that the (constrained) Walrasian correspondence, W, satisfies Property 2 in every β^o. The role of the W correspondence in the credible implementability of an arbitrary correspondence φ becomes evident upon examination of the proof of the theorem. By partial implementability, we mean that for any φ, there is a sub-correspondence $\varphi' \subseteq \varphi$ which is implementable. Given that the designer is, typically, indifferent over the allocations in $\varphi(\succ)$ for given $\succ$, partial implementation of φ is quite acceptable.

Theorem 6: *Suppose $n > 2$.*

> *If (i) $\varphi \subseteq PE \cap IR$,*
>
> *(ii) φ satisfies Pareto-indifference,*
>
> *(iii) φ satisfies Nash implementability,*
>
> *(iv) every $\beta^o \in \mathbb{B}$ satisfies Property 3A,*
>
> *(v) W satisfies Property 2 for every $\beta^o \in \mathbb{B}$,*

then there exists $\varphi' \subseteq \varphi$ such that φ' is credibly implementable in $\mathbb{B}$.

Proof: First, we shall show that if every $\beta^o \in \mathbb{B}$ satisfies Property 3A, then the (constrained) Walrasian correspondence W satisfies Property 1. By Theorem 2, W is credibly implementable in $\mathbb{B}$ if Property 1, together with Property 2 for every $\beta^o \in \mathbb{B}$, is satisfied. Finally we argue that φ is credibly partially implementable in $\mathbb{B}$ by combining these two results with a result in Thomson (1984) which states that $W \subseteq \varphi$ if $\varphi \subseteq PE \cup IR$, satisfies Maskin-monotonicity and the Pareto-indifference condition.

The following result is crucial.

Lemma 1: *If every $\beta^o \in \mathbb{B}$ satisfies Property 3A, then W satisfies Property 1.*

Proof of Lemma 1: Choose β^o: $\mathcal{R} \to [0, 1]$ such that Property 3A is met and choose $\succ, \succ' \in$ $supp(\beta^o)$ such that

$\{L_i(z, \succ) \cap \widetilde{W}(\beta^o)\} \subseteq L_i(z, \succ')$ for some $z \in W(\succ)$.

By Property 3A, for some price vector $p \in \Delta$ for which z is $\succ$-maximal on $H_i(z, p)$ for all $i \in N$, there

exists $\succ^P \in supp(\beta^o)$ such that $L_i(z, \succ^P) = H_i(z, p)$ for all $i \in N$. By definition of W, $\bar{H}(z, p) \subseteq W(\succ^P) \subseteq$

$\widetilde{W}(\beta^o)$. $\{L_i(z, \succ) \cap \widetilde{W}(\beta^o)\} \subseteq L_i(z, \succ')$ implies that $\bar{H}(z, p) \subseteq L_i(z, \succ')$. By definition of W, $z \in$

$W(\succ')$. ∎

The proof of the Theorem follows from the following result:

Lemma 2 *(Thomson (1984)): If* $\varphi \subseteq PE \cup IR$ *satisfies Maskin-monotonicity, and the Regularity condition, then*

$W \subseteq \varphi$.

The following result establishes a similar set of sufficiency conditions for credible

implementability of Nash-implementable rules satisfying Pareto-efficiency and envy-freeness.

Theorem 7: *Suppose* $n > 2$.

   If (i) $\varphi \subseteq PE \cap EF$,

   (ii) φ *satisfies Pareto-indifference,*

   (iii) φ *satisfies Nash implementability,*

   (iv) every $\beta^o \in \mathbb{B}$ *satisfies Property 3B,*

   (v) W^c *satisfies Property 2 for every* $\beta^o \in \mathbb{B}$,

then there exists $\varphi' \subseteq \varphi$ *such that* φ' *is credibly implementable in* $\mathbb{B}$.

Proof: Similar to proof of Theorem 6. A corresponding result, from Thomson (1987), stated as the

following lemma, is needed.

Lemma 3: *If* $\varphi \subseteq PE \cup EF$ *satisfies Maskin-monotonicity, and the Regularity condition, then* $W^c \subseteq \varphi$.

22

The results of this section are, in general, negative in nature. Most interesting social choice rules are not credibly implementable. To ensure credibility, the designer's beliefs must be sufficiently diffuse which in turn assures the designer greater flexibility in designing the range of the outcome function of the mechanism. The conditions on the beliefs given above identify sufficient conditions on the extent of information asymmetry that must exist between the agents and the designer to ensure credible implementability.

It may be checked that a weaker condition such as Property 3 cannot be used to replace Properties 3A or 3B in the theorems above. Examples can be easily constructed showing that Property 1 is violated by any $\varphi \subseteq \{PE \cap IR\} \cup \{PE \cap EF\}$ even if Property 3 is met.

4. Results on Weakly Credible Implementation

The negative results of the previous section are obtained under the assumption that the only information available about the designer's objective is that she is commited to implementing a social choice rule φ. This yields a rather strong credibility requirement. In many situations, of course, there is more information regarding the designer; for example, the social utility function may be known in addition to φ. This yields the weaker requirement given in Definition 1A.

In this section, we shall suppose that the social utility function is drawn from some natural sub-classes of social utility functions. We shall first define a class of functions characterized by a weak property. The property simply requires that the social utility be monotone (not to be confused with Maskin-monotonicity) in individual preference orderings.

Definition 11: Suppose $\beta^o \in \mathbb{B}$ is given. A social utility function u is *monotonic* if

$$\forall \succ \in supp(\beta^o), \left[\forall i \in N, z \succ_i z'; \exists j \in N \text{ such that } z' \notin L_j(z, \succ)\right] \Rightarrow \left[u(z, \succ, \beta^o) > u(z', \succ, \beta^o)\right].$$

Next, we define an alternative property of social utility functions which is natural in a context in which it is common knowledge that subsequent to any inefficient allocation of resources by the designer, the agents can costlessly renegotiate and arrive at an efficient allocation through a Pareto-improving trade. Hence, the designer's utility from assigning an inefficient allocation cannot exceed that from assigning an efficient one.

Definition 12: Suppose $\beta^o \in \mathbb{B}$ is given. A social utility function u satisfies *lexicographic preference for Pareto-efficiency (LPPE)* if

$$\forall \succ \in supp(\beta^o), \forall z \notin PE(\succ), \exists z' \in PE(\succ) \text{ such that } \forall i \in N, z' \succ_i z \text{ and } u(z', \succ, \beta^o) \geq u(z, \succ, \beta^o).$$

The following results show that, provided the social utility function is monotonic, then the negative result relating to implementability of individually rational and efficient social choice rules persists. The social choice rules must satisfy a mild requirement of containing some strictly individually rational allocations.

Theorem 8:   *If (i)* $\varphi \subseteq PE \cap IR$

   (ii) $\forall \succ \in \mathcal{R}, int[IR(\succ)] \neq \varnothing \Rightarrow \varphi(\succ) \cap int[IR(\succ)] \neq \varnothing,$

   (iii) $\mathbb{B}$ *is unrestricted, and*

   (iv) u *satisfies monotonicity,*

then φ *cannot be weakly credibly implemented in* $(\mathbb{B}, u)$.

<u>Proof:</u> Consider $\beta^o: \mathcal{R} \rightarrow [0, 1]$ such that $\beta^o(\succ) = \beta^o(\succ') = 0.5$ and the $PE \cap IR$ correspondence defined for $\succ$ and $\succ'$ is as given in Figures 4a and 4b. The broken-line indifference curves represent the economy $\succ$, whereas the unbroken-line curves represent $\succ'$. For each economy, the indifference curves are derived by horizontal translations of the curves shown. $PE \cap IR(\succ) = \overline{AB}$ and $PE \cap IR(\succ') = \overline{BC}$. The example is for a two-person, two-good economy. However, it easily generalizes.

We begin by supposing that φ is weakly credibly implementable in $(\mathbb{B}, u)$. We shall arrive at a contradiction in two steps. Let $\Gamma = \langle\{1, 2\}, S, g\rangle$ be the game-form that weakly credibly implements φ.

1.) First observe that, by construction, the monotonicity property on u implies that for any $z \in A$, if $z \in g(s)$ then $z \in PE(\beta^o) = \overline{DE} \cup \overline{DO}_2 \cup \overline{EO}_1$. Suppose otherwise, i.e. $z \notin \overline{DE} \cup \overline{DO}_2 \cup \overline{EO}_1$. In this case, for both economies $\succ$ and $\succ'$, there will exist $z' \in \overline{DE}$ such that for $i = 1, 2$, $z' \succ_i z$ and $z' \succ_i' z$ and for each $\succ^* \in \{\succ, \succ'\}$, for at least one agent i, $z' \notin L_i(z, \succ^*)$. This follows from the construction of the indifference curves in Figure 4a. Recall that, for each economy, the indifference curves are horizontal translations of the curves shown. Thus, by monotonicity of u, $\tilde{u}(z', \beta) > \tilde{u}(z, \beta)$ for all β. This is in contradiction with the assumption that Γ weakly credibly implements φ.

2.) Next, we shall argue that if Γ weakly credibly implements φ, then it must be the case that $\varphi(\succ) = \varphi(\succ')$. This is clearly, not the case in our example, except if $\tilde{\varphi}(\beta^o) = \{(\text{point } B)\}$. This is in contradiction with the assumption of the theorem that $\varphi(\succ) \cap int[IR(\succ)] \neq \varnothing$.

To show that $\varphi(\succ) = \varphi(\succ')$, choose $\sigma \in NE(\Gamma)$ such that $g(\sigma(\succ)) = z \in \varphi(\succ)$. By the definition of a Nash equilibrium, for all $i \in N$, $\{z' \in A: \exists s_i' \in S_i \text{ such that } g(s_i', \sigma_{-i}(\succ)) = z'\} \subseteq L_i(z, \succ)$.

Consider the second economy $\succ'$ and observe that, by construction, $\{L_i(z, \succ) \cap \tilde{PE}(\beta^o)\} \subseteq L_i(z, \succ')$ for all $i \in N$. By step 1 above, $\{z' \in A: \exists s_i' \in S_i \text{ such that } g(s_i', \sigma_{-i}(\succ)) = z'\} \subseteq \{L_i(z, \succ) \cap \tilde{PE}(\beta^o)\}$. Since $\{L_i(z, \succ) \cap \tilde{PE}(\beta^o)\} \subseteq L_i(z, \succ')$, it must be the case that $g(\sigma(\succ)) \in NE_A(\Gamma, \succ')$. By definition of weakly credible implementability, $z = g(\sigma(\succ)) \in \varphi(\succ')$. ∎

Next, suppose that the designer has a lexicographic preference for efficiency. Under this assumption, the negative conclusions of the previous theorem can be obtained as a corollary of the previous result.

Theorem 9:   *If (i) $\varphi \subseteq PE \cap IR$*

   (ii) $\forall \succ \in \mathscr{R}, int[IR(\succ)] \neq \varnothing \Rightarrow \varphi(\succ) \cap int[IR(\succ)] \neq \varnothing$,

   (iii) $\mathbb{B}$ is unrestricted, and

   (iv) u satisfies LPPE,

then φ cannot be weakly credibly implemented in $(\mathbb{B}, u)$.

<u>Proof:</u> Consider the example of given in the proof of the previous theorem. The argument is identical to that given in the proof above, with the exception that the *LPPE* property is used to prove step 1, instead of monotonicity.   ∎

The example given in this section is not pathological. Since it is designed to apply to every sub-correspondence of *PE∩IR*, it has certain special properties. For any particular φ (consider, as an excercise, the Walrasian correspondence) it can be seen that non-implementability argument given above is generic.

4. Conclusions

This paper introduces the notion of credible implementation. This concept ensures that it is rational for a designer to abide by the outcomes rules of the implementation mechanism even if the agents choose strategies with zero probability of occurrence in equilibrium. We explicitly model the designer's utility via the social choice correspondence and her beliefs about the true state of the economy.

It is shown that the positive findings of Nash implementation theory are severely affected once a credibility restriction is imposed. It is known that the Pareto correspondence and several sub-correspondences of the Pareto efficiency and individual rationality correspondence, such as the (constrained) Walrasian correspondence, are Nash implementable. We show that (a) the non-dictatorial Pareto correspondence can be credibly implemented (b) there exists no Pareto-efficient and individually

rational social choice rule that can be implemented in a credible manner and (c) there exists no Pareto-efficient and envy-free social choice rule that can be implemented in a credible manner.

The paper establishes necessary conditions for credible implementability and shows that it is possible to derive conditions on the designer' beliefs that are sufficient for credible implementability. An implication of these conditions is a paradoxical phenomenon: the less information a designer has, the easier it is to achieve her objective. The intuition behind this is clear from our conditions: a designer with relatively little information has a wider support for her priors -- which in turn gives her more room to design a mechanism, since the outcome function can map to a larger subset of allocations.

Finally, the paper explores the robustness of the negative conclusions by weakening the credibility requirement. Such a weaker definition would require additional information about the designer's preferences, in excess of what is normally available in a standard social choice problem. We find that there exists no Pareto-efficient and individually rational social choice rule that can be implemented in a weakly credible manner under natural assumptions about the designer's preferences.

The concerns raised here also arise in other contexts such as implementation in more general domains and implementation using solution concepts that refine Nash equilibria. In the case of the latter, the results of Moore and Repullo (1988), Palfrey and Srivastava (1991) and Palfrey, Srivastava and Jackson (1991) have shown that "almost" all social choice correspondences are implementable using alternative refinements. On the other hand, by weakening the requirement of exactness of implementation, Matsushima (1988) and Abreu and Sen (1991) have demonstrated similarly startling results. A natural line of inquiry is: how are these strong results limited by a restriction of credibility? We shall address this in future work.

In asymmetric information domains, the results have been largely negative (Palfrey and Srivastava (1987), Chakravorti (1992a, 1992b)) to begin with. However, given the satisfaction of an incentive

compatibility requirement, positive results emerge. These are, in fact, made stronger by the use of refinements (Palfrey and Srivastava (1989)). Here again, we must study the impact of credibility.

We close with the following observations:

(i) The credibility issue is especially severe for implementation theorems that require assumptions such as the existence of a universally worst element (e.g. see Palfrey, Srivastava and Jackson (1991) on bounded implementation or the papers on Bayesian implementation). It is hard to imagine that a designer (whose preferences are given by a reasonable social chocie rule) can rationalize an outcome in which she destroys all the goods in an economy. Even though in equilibrium, such a threat is never carried out, it is an incredible one and will make the rationale leading to an equilibrium unravel.

(ii) Our concern for out-of-equilibrium moves has been voiced in the past in the context of well-behavedness of the outcome function. The focus has been on continuity (Postlewaite and Wettstein (1990)) and feasibility (Hurwicz, Maskin and Postlewaite (1984)). The issue of credibility has not been addressed, though it may be tied into the concern for continuity and feasibility. Moreover, the credibility restrictions may be tied into the questions of renegotiation-proofness raised by Maskin and Moore (1989) and others. We have made a beginning in this paper with the *LPPE* property.

(iii) It may be noted that the implementation mechanism given in the paper is immune to signals being sent unilaterally by agents to the designer via disequilibrium moves and appealing to a Cho and Kreps (1987)-type argument. No agent has an incentive to send such a message and thereby hope to influence the designer's beliefs.

(iv) Note that a revelation principle can be stated where we replace "implementability" with "credible implementability". The principle essentially states that any social choice that can be implemented using an arbitrary game form can also be truthfully implemented using a direct revelation mechanism. But a direct mechanism, appropriately defined, is always credible provided its outcome function is a selection

from the correspondence φ. But that is exactly the manner in which a direct mechanism is constructed to prove the principle.
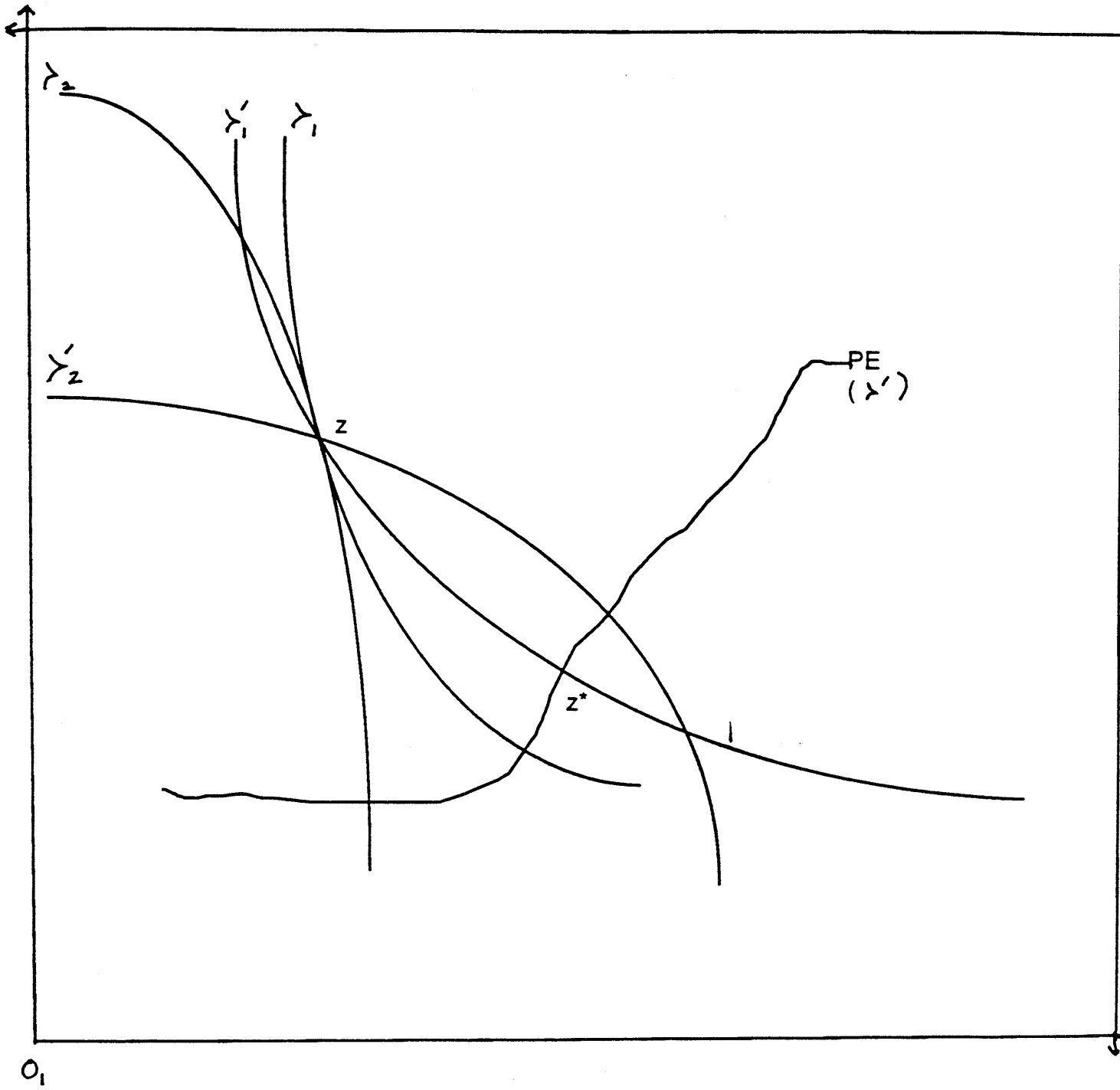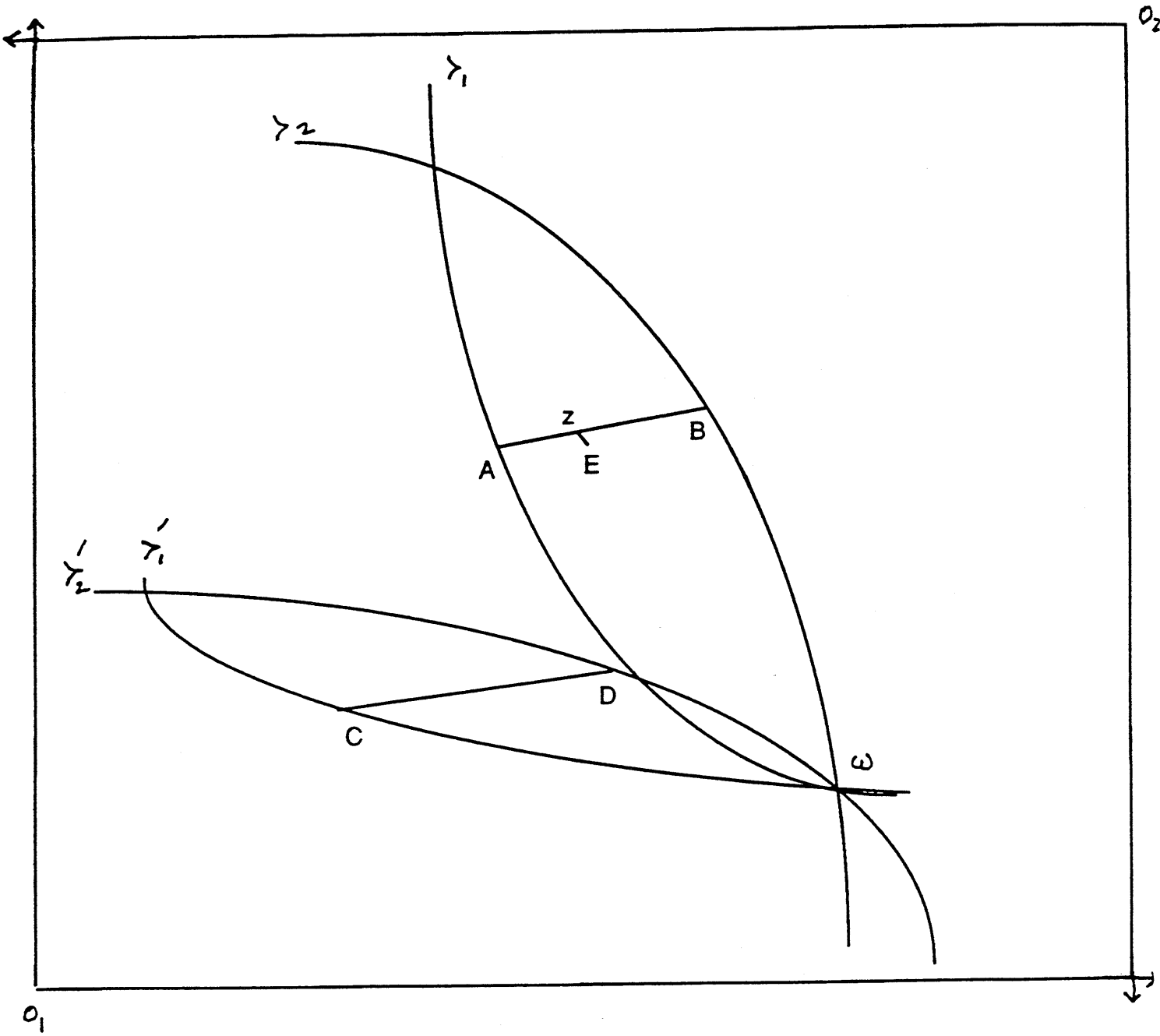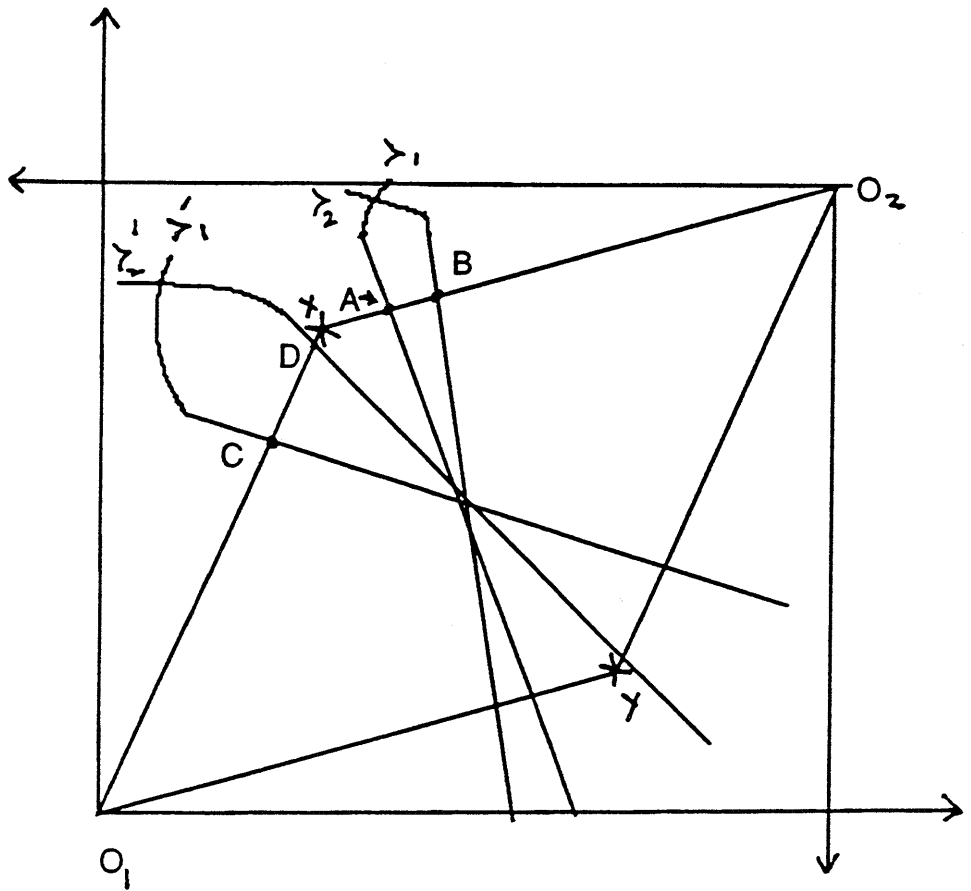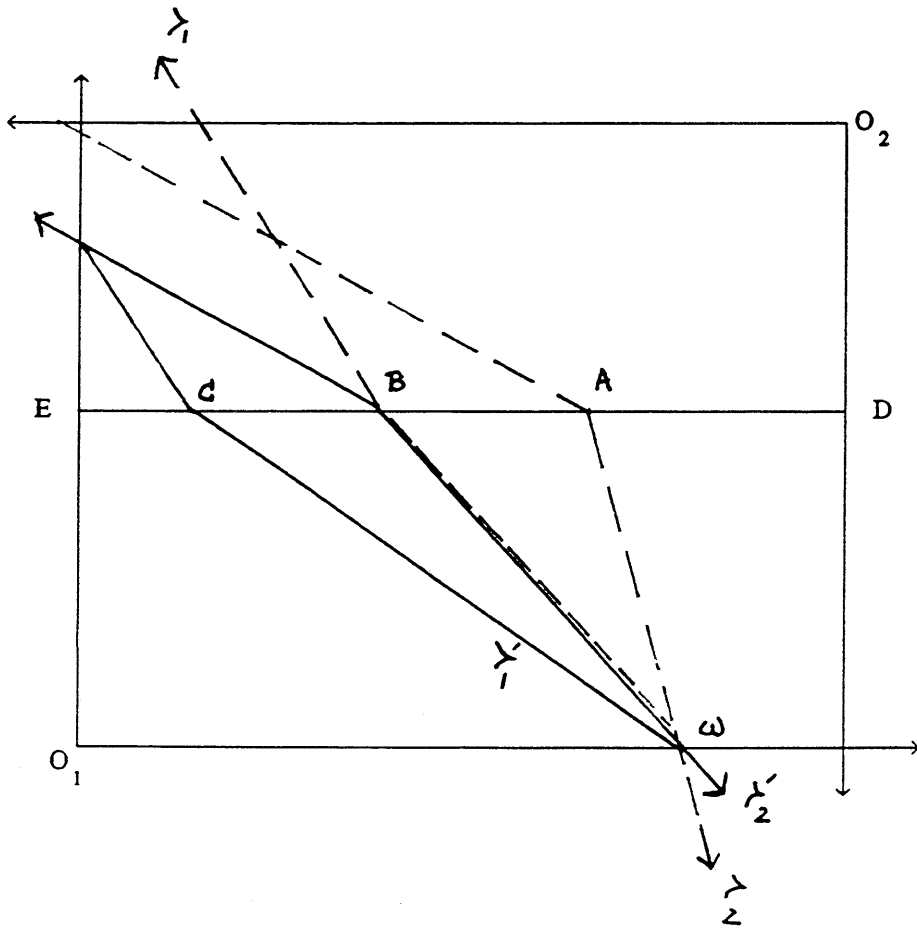
FIGURE 1

FIGURE 2

FIGURE 3

FIGURE 4

References

ABREU, D. AND A. SEN (1991): Virtual Implementation in Nash Equilibrium, *Econometrica*.

AGHION, P., M. DEWATRIPONT AND P. REY (1989): Renegotiation Design under Symmetric Information, mimeo.

CHAKRAVORTI, B. (1992a): Efficiency and Mechanisms with no Regret, *International Economic Review*, 33, 45-60.

CHAKRAVORTI, B. (1992b): Sequential Rationality, Implementation and Pre-play Communication, forthcoming *Journal of Mathematical Economics*.

CHO, I. AND D. KREPS (1987): Signaling Games and Stable Equilibria, *Quarterly Journal of Economics*, 102, 179-221.

GEVERS, L. (1986): Walrasian Social Choice: Some Simple Axiomatic Approaches, in W. Heller, R. Starr and D. Starrett (eds.) *Social Choice and Public Decision Making: Essays in Honor of Kenneth Arrow, Vol 1*, Cambridge University Press, Cambridge, UK.

HURWICZ L., E. MASKIN AND A. POSTLEWAITE (1984): Feasible Implementation of Social Choice Correspondences, mimeo, University of Minnesota.

MASKIN, E. (1977): Nash Equilibrium and Welfare Optimality, MIT mimeo.

MASKIN, E. (1985): The Theory of Implementation in Nash Equilibrium, in L. Hurwicz, D. Scmeidler and H. Sonnenschein (eds.) *Social Goals and Social Organizations: Essays in Memory of Elisha Pazner*, Cambridge University Press, Cambridge.

MASKIN, E. AND J. MOORE (1989): Implementation with Renegotiation, mimeo.

MATSUSHIMA, K. (1988): A New Approach to the Implementation Problem, *Journal of Economic Theory*, 45, 128-144.

MOORE, J. AND R. REPULLO (1988): Subgame Perfect Implementation, *Econometrica*, 56, 1191-1220.

PALFREY, T. AND S. SRIVASTAVA (1987): On Bayesian Implementable Allocations, *Review of Economic Studies*, LIV, 193-208.

PALFREY, T. AND S. SRIVASTAVA (1991): Nash Implementation Using Undominated Strategies, *Econometrica*,

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

mimeo, California Institute of Technology.

POSTLEWAITE, A. AND D. WETTSTEIN (1990): Implementing Constrained Walrasian Equilibria Continuously, *Review of Economic Studies.*

RUBINSTEIN, A. AND A. WOLINSKY (1989): Renegotiation-proof Implementation and Time Preferences, forthcoming *American Economic Review.*

THOMSON, W. (1984): The Manipulability of Resource Allocation Mechanisms, *Review of Economic Studies,* 51, 447-460.

THOMSON, W. (1987): The Vulnerability to Manipulative Behavior of Resource Allocation Mechanisms Designed to Select Equitable and Efficient Outcomes, in T. Groves, R. Radner and S. Reiter (eds.) *Information, Incentives and Economic Mechanisms: Essays in Honor of Leonid Hurwicz,* University of Minnesota Press, Minneapolis.

PUBLISHED ISSUES

16   "Equivalent Conditions on Solvability for Non-Linear Leontief Systems"
            J.A. Silva. 1986.

17   "A Weak Generalization of the Frobenius Theorem"
            J.A. Silva. 1986

18   "On the Fair Distribution of a Cake in Presence of Externalities"
            A. Villar. 1987.

19   "Reasonable Conjetures and the Kinked Demand Curve"
            L.C. Corchón. 1987.

20   "A Proof of the Frobenius Theorem by Using Game Theory"
            B. Subiza. 1987.

21   "On Distributing a Bundle of Goods Fairly"
            A. Villar. 1987.

22   "On the Solvability of Complementarity Problems Involving Vo-Mappings and its Applications
            to Some Economic Models"
            C. Herrero, A. Villar. 1987.

23   "Semipositive Inverse Matrices"
            J.E. Peris. 1987.

24   "Complementary Problems and Economic Analysis: Three Applications"
            C. Herrero, A. Villar. 1987.

25   "On the Solvability of Joint-Production Leontief Models"
            J.E. Peris, A. Villar. 1987.

26   "A Characterization of Weak-Monotone Matrices"
            J.E. Peris, B. Subiza. 1988.

27   "Intertemporal Rules with Variable Speed of Adjustment: An Application to U.K.
            Manufacturing Employment"
            M. Burgess, J. Dolado. 1988.

28   "Orthogonality Test with De-Trended Data's Interpreting Monte Carlo Results using Nager
            Expansions"
            A. Banerjee, J. Dolado, J.W. Galbraigth. 1988.

29   "On Lindhal Equilibria and Incentive Compatibility"
            L.C. Corchón. 1988.

30   "Exploiting some Properties of Continuous Mappings: Lindahl Equilibria and Welfare
            Egalitaria Allocations in Presence of Externalities"
            C. Herrero, A. Villar. 1988.

31   "Smoothness of Policy Function in Growth Models with Recursive Preferences"
            A.M. Gallego. 1990.

32   "On Natural Selection in Oligopolistic Markets"
            L.C. Corchón. 1990.

33   "Consequences of the Manipulation of Lindahl Correspondence: An Example"
          C. Bevía, J.V. LLinares, V. Romero, T. Rubio. 1990.

34   "Egalitarian Allocations in the Presence of Consumption Externalities"
          C. Herrero, A. Villar. 1990.

SECOND PERIOD

WP-AD 90-01   "Vector Mappings with Diagonal Images"
               C. Herrero, A.Villar. December 1990.

WP-AD 90-02   "Langrangean Conditions for General Optimization Problems with Applications to Consumer
               Problems"
               J.M. Gutierrez, C. Herrero. December 1990.

WP-AD 90-03   "Doubly Implementing the Ratio Correspondence with a 'Natural' Mechanism"
               L.C. Corchón, S. Wilkie. December 1990.

WP-AD 90-04   "Monopoly Experimentation"
               L. Samuelson, L.S. Mirman, A. Urbano. December 1990.

WP-AD 90-05   "Monopolistic Competition : Equilibrium and Optimality"
               L.C. Corchón. December 1990.

WP-AD 91-01   "A Characterization of Acyclic Preferences on Countable Sets"
               C. Herrero, B. Subiza. May 1991.

WP-AD 91-02   "First-Best, Second-Best and Principal-Agent Problems"
               J. Lopez-Cuñat, J.A. Silva. May 1991.

WP-AD 91-03   "Market Equilibrium with Nonconvex Technologies"
               A. Villar. May 1991.

WP-AD 91-04   "A Note on Tax Evasion"
               L.C. Corchón. June 1991.

WP-AD 91-05   "Oligopolistic Competition Among Groups"
               L.C. Corchón. June 1991.

WP-AD 91-06   "Mixed Pricing in Oligopoly with Consumer Switching Costs"
               A.J. Padilla. June 1991.

WP-AD 91-07   "Duopoly Experimentation: Cournot and Bertrand Competition"
               M.D. Alepuz, A. Urbano. December 1991.

WP-AD 91-08   "Competition and Culture in the Evolution of Economic Behavior: A Simple Example"
               F. Vega-Redondo. December 1991.

WP-AD 91-09   "Fixed Price and Quality Signals"
               L.C. Corchón. December 1991.

WP-AD 91-10   "Technological Change and Market Structure: An Evolutionary Approach"
               F. Vega-Redondo. December 1991.

WP-AD 91-11   "A 'Classical' General Equilibrium Model"
               A. Villar. December 1991.

WP-AD 91-12   "Robust Implementation under Alternative Information Structures"
               L.C. Corchón, I. Ortuño. December 1991.

WP-AD 92-01   "Inspections in Models of Adverse Selection"
               I. Ortuño. May 1992.

WP-AD 92-02   "A Note on the Equal-Loss Principle for Bargaining Problems"
               C. Herrero, M.C. Marco. May 1992.

WP-AD 92-03   "Numerical Representation of Partial Orderings"
               C. Herrero, B. Subiza. July 1992.

WP-AD 92-04   "Differentiability of the Value Function in Stochastic Models"
               A.M. Gallego. July 1992.

WP-AD 92-05   "Individually Rational Equal Loss Principle for Bargaining Problems"
               C. Herrero, M.C. Marco. November 1992.

WP-AD 92-06   "On the Non-Cooperative Foundations of Cooperative Bargaining"
               L.C. Corchón, K. Ritzberger. November 1992.

WP-AD 92-07   "Maximal Elements of Non Necessarily Acyclic Binary Relations"
               J.E. Peris, B. Subiza. December 1992.

WP-AD 92-08   "Non-Bayesian Learning Under Imprecise Perceptions"
               F. Vega-Redondo. December 1992.

WP-AD 92-09   "Distribution of Income and Aggregation of Demand"
               F. Marhuenda. December 1992.

WP-AD 92-10   "Multilevel Evolution in Games"
               J. Canals, F. Vega-Redondo. December 1992.

WP-AD 93-01   "Introspection and Equilibrium Selection in 2x2 Matrix Games"
               G. Olcina, A. Urbano. May 1993.

WP-AD 93-02   "Credible Implementation"
               B. Chakravorti, L. Corchón, S. Wilkie. May 1993.

WP-AD 93-03   "A Characterization of the Extended Claim-Egalitarian Solution"
               M.C. Marco. May 1993.