Envy-Freeness and Implementation in Large Economies *

Matthew O. Jackson† and Ilan Kremer‡

February 5, 2003
revised August 3, 2005
forthcoming: *The Review of Economic Design*

Abstract

We show that an asymptotic envy-freeness condition is necessary for a form of robust approximate implementation in large economies. In settings where allocations are excludable, asymptotic envy-freeness is also sufficient for implementation, while in non-excludable settings it is not sufficient.

*We are grateful for financial support from the National Science Foundation under grants SES-9986190 and SES-0316493. We thank Laurent Mathevet, Andy Postlewaite, Muhamet Yildiz, and an anonymous referee for helpful comments on earlier drafts. We dedicate this paper to Murat Sertel, with fond memories of his engaging personality, warm friendship, and wide-ranging contributions to economic theory.

†Division of Humanities and Social Sciences 228-77, California Institute of Technology, Pasadena, California 91125, USA. Email: jacksonm@hss.caltech.edu

‡Graduate School of Business, Stanford University, Stanford CA 94305, USA. Email: ikremer@stanford.edu

1 Introduction

The idea that competition leads to efficiency is central to economics, as is embodied in the first welfare theorem. Nevertheless, such conclusions clearly depend on the setting, as they need not hold in public goods settings. Moreover, even making precise the idea that approximately efficient allocations are implementable in private goods economies with many agents is much more subtle than one might guess.[1] There is still much that we do not understand about which allocations can be reached in large economies.

The purpose of this paper is to identify basic restrictions on implementability that any allocation rule that can be approximately implemented in large economies must satisfy. In particular, we show that in a wide variety of settings any social choice function that is approximately and robustly implementable in large economies must be envy-free. Foley's (1967) property of envy-freeness states that no agent should wish to swap allocations with another agent. Envy-freeness has a strong normative appeal and has served as the foundation of notions of "fairness."

The relationship between this fairness condition and incentives in large economies is embarrassingly simple. It comes in two parts. First, if a social choice function is approximately implemented in a large economy, then it must satisfy an incentive compatibility constraint so that an agent will not pretend to be of another type. Second, approximate implementation requires that in a large enough economy any single agent's announcement will not have much effect on the overall allocation. This means that an agent can obtain almost the same allocation as some other agent by pretending to be of the same type as that other agent. Putting these together we can conclude that an agent does not wish to swap allocations with any other agent in the limit and hence the limiting allocation must be envy-free.

This result seems to stand in conflict with a result by McLean and Postlewaite (2002) who show that in pure exchange economies if agents are all

[1] For instance, even in pure exchange economies this depends on the information structure, feasibility constraints, and notions of approximation. The seminal reference is Hurwicz (1972), and variations include: complete information implementation (e.g., Schmeidler (1981), Hurwicz, Maskin and Postlewaite (1995)), incentive properties of rational expectations equilibria (e.g., Blume and Easley (1983), Palfrey and Srivastava (1986)), incentive properties of the Walrasian correspondence (e.g., Roberts and Postlewaite (1976), Thomson (1979), Otani and Sicilian (1982, 1990), Jackson (1992), Gul and Postlewaite (1992), Jackson and Manelli (1996)), strategy-proofness in exchange settings (e.g., Barberà and Jackson (1995), Cordoba and Hammond (1998) and Kovalenkov (2002)).

"informationally small," then essentially any social choice function can be approximately implemented in the limit. This seems to conflict with our result, since envy-freeness is a significant restriction,[2] and yet never pops up in McLean and Postlewaite's analysis. Let us explain why this is the case. McLean and Postlewaite examine settings where uncertainty regards the "characteristics" of goods, and not individuals' preferences over the goods. While this seems to be a vague difference, the critical point is that in their setting under informational smallness if you know the information of all but one agent and the economy is large enough then you know with arbitrary accuracy what that remaining agent's preferences are. That definition of informational smallness is a strong one, as it precludes agents from having any private information about their own preferences. For instance, such a condition is violated in a classic Arrow-Debreu economy where an agent's preferences are his or her private information. In our setting, we allow agents to have private information that can concern their own preferences, and this is the key to implying envy-freeness as it allows them to mimic other agents, something which is not possible under informational smallness.

As an application of our results, it is easy to derive results on the limits of implementability in large public goods settings. For instance, as shown by Mailath and Postlewaite (1990) and Al-Najjar and Smorodinsky (2000), there are strong limitations on what can be implemented in large public goods economies. The central intuition is that in a large economy an agent's announcement has very little chance of influencing the public good decision, and thus agents worry mainly about how their announcement affects their share of the cost. Such large economy results on public goods can be obtained as an easy corollary of our envy-free result. Envy-freeness directly implies that all agents must pay the same cost,[3] and hence imposing an individual rationality constraint implies that a public project will not be efficiently undertaken in a large economy.

Beyond envy-freeness being a necessary condition for approximate implementability in large economies, it is also in some cases sufficient. We describe a simple variation on a "folk" mechanism that provides for approximate implementation of envy-free social choice functions in settings where goods are excludable. This applies to all private goods and some public goods settings.

[2]For instance, see Pazner and Schmeidler (1974) who provide examples in which no Pareto optimal allocation is envy free.

[3]Even if public goods are excludable, if we require efficiency then all agents with non-zero valuations must have access.

In the case of non-excludable goods, envy-freeness is no longer sufficient as we illustrate. We leave open the difficult question of a full characterization for the case of non-excludable goods.

Finally, let us point out two aspects of our definitions of social choice functions and implementability that play a role in our results. First, we only consider social choice functions satisfying the well-known "equal treatment of equals" property. This is built directly into our definition of social choice function and is the requirement that if agents differ only in their name and not in any observable or reported characteristics, then they should receive the same allocation.[4] It is a very minimal anonymity condition. Clearly, if a social choice function always gives everything to agent 1, just because she is named agent 1, then that will be implementable and envious. We rule such social choice functions out by fiat. Of course, we do allow the social choice function to depend on what agents announce, and also on other observable characteristics which might be things like the agent's endowment, age, whether they are buyers or sellers, or some pre-existing rights that they have.

The other condition that plays a role in our analysis is the "robustness" of the implementation. This requires that the implementation work even when agents happen to learn extra information about the common state beyond what they know from their type. The idea is that the mechanism is designed based on a given informational setting, but we would like the mechanism to still function if agents happen to learn more information.[5] This condition plays a subtle but important role in our results. In the absence of such a condition an implementing mechanism can take very strong advantage of small differences in agents' expectations about the state to get them to completely separate themselves.[6] This is the insight that underlies approaches such as that of Crémer and McLean (1988). Such mechanisms, however, are very sensitive to the exact specification of the prior and the agents' information

[4]In our case, this is only required from an interim perspective, and identical agents may receive random allocations whose ex post realizations differ.

[5]Especially, as otherwise agents might in fact end up having incentives to learn more.

[6]Our notion of robustness is weaker than that of Bergemann and Morris (2003), who allow for richer type spaces that encode agents' beliefs. We work with "traditional" type spaces, but add some minimal robustness to rule out Crémer and McLean (1988) type of constructions. Further robustness constraints impose further restrictions, as can be seen in Bergemann and Morris (2003). The envy-freeness conditions that we focus on come in addition to any stronger incentive compatibility conditions derived from the strengthening of robustness.

sets. Requiring robustness is a step towards the practical in that it requires a mechanism still perform in situations where agents, either by accident or by their choice, have more information than we presume. This robustness beyond being desirable and perhaps even necessary in practice, is important in deriving envy-freeness. The ability of one agent to pretend to be of another's type is needed to derive envy-freeness and can be prevented if one can rely on non-robust mechanisms where slight differences in information are exploited.

2 Definitions

Agents and Economies

There are countably many agents, who are indexed by $i \in \mathbb{N}$.

We index economies by n indicating the number of agents in the economy, where economy n consists of the first n agents, $\{1, \ldots, n\}$.

Uncertainty, Types, and Information

Agents have two different sorts of characteristics. Some are openly observable and others are private information. For example, it may be that an agent's age or endowment is observable and that the agent's preferences are private information.

In addition to the potential observed and unobserved heterogeneity of agents, we also wish to have the model allow for some commonality in preferences and payoffs. To this end, a specification of the economy includes a profile of the agents' public labels, their private types, and a state of nature.

- L is the set of labels, and $\ell_i \in L$ denotes agent i's label which consists of the publicly observable characteristics of agent i.[7]

- T is the set of types, and $t_i \in T$ denotes agent i's type. The type of agent i is his or her private information.

- S is the set of states of the world, and the realized state is denoted by $s \in S$. The state can enter into all agents' preferences.

[7]This need not include all of the publicly observable characteristics, but must include all of those that a mechanism will condition on regardless of any announcement by i. A society may wish to treat certain publicly observable characteristics as if they are private information.

Depending on the setting and application, what falls into L and what falls into T may vary. For instance, in one application it may be that endowments are observable and preferences are not. In another it may be that neither endowments nor preferences are observable. In yet another application it may be some skill or income that is not observable, while preferences are known.

We make several assumptions on the spaces and the uncertainty.

1. To simplify the exposition we assume that L, T, and S are finite.[8]

2. A probability distribution p on S describes the uncertainty on s, where $p(s)$ is the probability of state $s \in S$. Types are i.i.d. conditional on the agent's label and the state according to a family of conditional probability measures $p(\cdot|s, \ell_i)$ defined on T. So, $p(t_i|s, \ell_i)$ is the probability that an agent is of type $t_i \in T$ conditional on the state being s and the agent's label being ℓ_i.

3. Finally, we assume that the proportion of agents having different labels is well-defined in the limiting economy. That is, for each $l \in L$ there exists $q(l) \in [0, 1]$ such that

$$\frac{|\{i \leq n|\ell_i = l\}|}{n} \to q\,(l)$$

Without loss of generality, let $q\,(l) > 0$ for all $l \in L$.

Note that we are treating labels as given and not uncertain. This is because we assume that they are observable, and so we can think of everything conditional on their realization. A trivial extension would treat them as uncertain without any changes to our results below.

Allocations and Preferences

$A^n \subset \mathbb{R}^{(n+1)k}$ is the set of feasible allocations for economy n, where k is some positive integer.

An allocation $a^n \in A^n$ contains both public and private parts. We write $a^n = (a_0^n, a_1^n, \ldots, a_n^n)$, with the interpretation that a_0^n is public and may enter

[8]As mentioned in footnote 1 above, distinctions between finite and infinite spaces in such analyses can sometimes be important. In this case it is fairly clear that our results extend to the infinite case, provided some appropriate compactness and continuity conditions are imposed.

any agent's preferences while a_i^n is something that is private to agent i. This is essentially without loss of generality.

Allocations may be quite general, including public goods, private goods, rules, etc.

Let $A_0 = \cup_n \{a_0^n \mid a^n \in A^n\}$ and $A_i = \cup_{n \geq i} \{a_i^n \mid a^n \in A^n\}$.

It is possible for the set of feasible allocations to depend on agents' labels. However, we require that $A_i = A_j$ whenever $\ell_i = \ell_j$.

Agents evaluate allocations according to a function $u : A_0 \times A_i \times S \times L \times T \to I\!R$ so that $u(a_0, a_i, s, \ell_i, t_i)$ denotes the utility of allocation a_0, a_i to an agent of label ℓ_i and type t_i in state s.

Having a common u across agents is without loss of generality, as any idiosyncratic aspect of preferences can be encoded in t_i and ℓ_i.

Given an allocation a, state s and a type t_i of agent i, we write $u_i(a, s, t_i)$ to denote $u(a_0, a_i, s, \ell_i, t_i)$.

We assume that u is bounded and continuous.

Note that most economies of interest fit into our setting, including exchange economies, production economies, economies with (excludable or non-excludable) public goods, economies with club goods, commons problems, etc.

Social Choice Function

A *social choice function* on economy n is a function $f : S \times T^n \to A^n$, such that that $\ell_i = \ell_j$ and $t_i = t_j$ implies that $a_i = a_j$.[9]

Note that a social choice function does not take labels as an argument. That is because we treat the labels as known and given, and so the social choice function already can be made to depend on labels.

We build into the definition of a social choice function that it must treat identical agents (in terms of both labels and types) equally.[10] This still allows agents with different labels, say buyers and sellers, or agents with different endowments, to be treated differently.

Convergent Sequences of Social Choice Functions

[9] We may wish to allow for random allocations. This is already admitted under our formulation of the allocation space which lies in a Euclidean space, and u can be an evaluation of expected (or non-expected) utility.

[10] Note that this only needs to hold from the interim perspective. That is, if a_i is a random allocation, it simply requires that the distribution over allocations be the same for identical agents, while ex-post they may be treated differently.

A sequence of social choice functions $\{f^n\}$ is *convergent* if there exists $f : S \times L \times T \to A_0 \times \cup_i A_i$ (written $f = (f_0, f_1)$) such that

$$(f_0^n, f_i^n)[s, t^n] \to (f_0(s), f_1(s, \ell_i, t_i)) \text{ in probability}$$

for each i.

A convergent sequence of social choice functions is one that has a well-defined limiting allocation as a function of the state and agents' labels and types.

Envy-Freeness

A limit social choice function $f : S \times L \times T \to A_0 \times \cup_i A_i$ is *envy-free* if for each s, ℓ_i, t_i and t_i'

$$u(f_0(s), f_1(s, \ell_i, t_i), \ell_i, s, t_i) \geq u(f_0(s), f_1(s, \ell_i, t_i'), \ell_i, s, t_i).$$

Although our definition of envy-freeness only applies across agents with the same label, let us emphasize that this actually generalizes the usual notion of envy-freeness. This is easily seen as we can easily encode all information contained in both the type and the label directly into the type t_i and have ℓ_i simply be a singleton.

The definition of a limit social choice function being envy-free can equivalently be defined in terms of the sequence being asymptotically envy-free. Let us show this.

A social choice function f on economy n is *ε-envy-free* if for every i and j in $\{1, \ldots, n\}$ such that $\ell_i = \ell_j$

$$u(a_0^n, a_i^n, \ell_i, t_i, s) \geq u(a_0^n, a_j^n, \ell_i, t_i, s) - \varepsilon$$

for almost every s and t, where $a^n = f^n(s, t^n)$.

Epsilon envy-freeness implies that an agent does not envy the allocation of agents who have the same label by more than some minimal amount.

A sequence of social choice functions $\{f^n\}$ is *asymptotically envy-free* if for every $\varepsilon > 0$ there exists N such that f^n is ε envy-free for all $n > N$ and $i \leq n$.

Lemma 1 *Consider a convergent sequence $\{f^n\}$ of social choice functions. Its limit f is envy-free if and only if the sequence $\{f^n\}$ is asymptotically envy-free.*

Type Measurability and Direct Mechanisms

In our setting a *direct mechanism* is simply a social choice function that is *type-measurable*, that is f depends only on labels and types and is independent of s.[11]

Through the well-known revelation principle, direct mechanisms capture what is implemented through any more complex mechanism. This is seen by simply mapping the equilibrium outcomes as a function of agents' types.

Of course, this requires that the social choice function only involve information that can be obtained through the actions or reports of the agents as well as their observable labels - but does not require an omniscient social planner who knows the state. This is what is embodied in type-measurability.

The condition of type-measurability is usually vacuously satisfied, as most models let the state be the vector of types. For our purposes of defining a robust form of implementability, however it will be useful to generalize this to allow for a distinction between the two, and hence the necessity of the type-measurability condition in defining direct mechanisms.

We say that a limit social choice function $f : S \times L \times T \rightarrow A_0 \times \cup_i A_i$ is *type-measurable* if for every i and ℓ_i, and t_i, $f(s, \ell_i, t_i) = f(s', \ell_i, t_i)$ whenever s and s' lead to the same distribution over types (that is, such that $p(t|l, s) = p(t|l, s')$ for all $t \in T$ and $l \in L$).

The idea of type-measurability in the limit is that the social choice function can only use information from the state that could be deduced by observing the distribution of types.

Incentive Compatibility and Implementability

A direct mechanism f on economy n is *incentive compatible* if it satisfies

$$E\left[u_i(f(t_{-i}, t_i), t_i, s)|t_i\right] \geq E\left[u_i(f(t_{-i}, \widehat{t}_i), t_i, s)|t_i\right],$$

for all i, t_i, and $\widehat{t}_i$.

A direct mechanism f^n is *robustly incentive compatible* if it does not depend on s and satisfies

$$E\left[u_i(f^n(t_{-i}, t_i), t_i, s)|t_i, s\right] \geq E\left[u_i(f^n(t_{-i}, \widehat{t}_i), t_i, s)|t_i, s\right]$$

for all i, t_i, s, and $\widehat{t}_i$.

[11] As already pointed out, f can depend on the labels ℓ_i as those are given and known.

Robust incentive compatibility requires incentive compatibility even in the case where an agent happens to learn the common part of the state of the world s. Note that this is still substantially weaker than ex-post incentive compatibility, as this says nothing about what an agent knows regarding other agents' types - other than the distribution over them - and the social choice function depends only on the agents' types not the state of the world. For instance, in an independent private values world, robust incentive compatibility is exactly equivalent to incentive compatibility. It is a very minimal way of ensuring that a mechanism is not sensitive to extra information that agents might learn or might have an incentive to learn. The important aspect of this robust form of incentive compatibility is that it precludes the mechanism from taking advantage of very specific aspects of the correlation structure of uncertainty and hence rules out mechanisms such as those in Crémer and McLean (1988).

Implementability

A sequence of social choice functions $\{f^n\}$ is *robustly approximately implementable* if there exists a sequence of robustly incentive compatible direct mechanisms $\{\widehat{f}^n\}$ so that for each i

$$(\widehat{f}_0^n, \widehat{f}_i^n) - (f_0^n, f_i^n) \rightarrow 0 \text{ in probability}$$

We emphasize that we are using the term "implementable" in a weak sense, in that we are working only with incentive compatibility and not making claims about the full set of equilibria of a mechanism.

Note that the implemented social choice functions can depend on s, while the direct mechanisms which approximate them can only depend on information obtained from the agents. The idea is that through information obtained from the agents, the state can be estimated and so the implementing social choice functions can come to approximate ones which know the state. However, as we shall see, type-measurability will be a necessary condition for a sequence to be implementable.

3 Envy-Freeness as a Necessary Condition

We now show that envy-freeness is a necessary condition of robust approximate implementation.

Theorem 1 *A convergent sequence of social choice functions is robustly approximately implementable only if its limit is envy-free and type-measurable.*

As discussed in the introduction the proof is very simple. Still, this theorem provides us with an important conclusion as it provides a non-trivial condition that needs to be satisfied in order to implement a given social choice function. Moreover, as we shall see known results can be derived as a simple corollary.

Proof of Theorem 1: Given is a convergent sequence of social choice functions $\{f^n\}$ with limit f that is robustly approximately implemented by the sequence of direct mechanisms $\{\widehat{f}^n\}$. Given the type-measurability of the direct mechanisms, the type-measurability of f follows directly. Let us verify the envy-freeness of the limit of the sequence.

The theorem follows from Lemmas 2 and 3.

Lemma 2 *If a convergent sequence $\{f^n\}$ with limit f is robustly approximately implemented by the sequence $\{\widehat{f}^n\}$, then $\{\widehat{f}^n\}$ is also convergent with limit f.*

Proof of Lemma 2: This follows from the finiteness of the state and type spaces, which gives us uniform convergence of both sequences. ∎

Lemma 3 *If a convergent sequence $\{\widehat{f}^n\}$ is robustly incentive compatible then its limit f is envy-free.*

Proof of Lemma 3: Consider a convergent sequence $\{\widehat{f}^n\}$ that is robustly incentive compatible. It follows that

$$E\left[u(\widehat{f}_0^n(t_{-i},t_i),\widehat{f}_i^n(t_{-i},t_i),t_i,s)|s,t_i\right] \geq E\left[u(\widehat{f}_0^n(t_{-i},t_i'),\widehat{f}_i^n(t_{-i},t_i'),t_i,s)|s,t_i\right] \tag{1}$$

Observe that not only $\widehat{f}_0^n(t_{-i},t_i)$ but also $\widehat{f}_0^n(t_{-i},t_i')$ converges to $f_0(s)$ in probability, conditional on s and t_i. That is, in the limit if an agent misrepresents his type he does not change the public part of the allocation. To see this we consider $\widehat{f}_0^n(t_{-i},t^*)$ for some fixed type t^*. Conditional on the state being s there is a positive probability that $t_i = t^*$. Hence, the fact that $\widehat{f}_0^n(t_{-i},t_i)$ converges to $f_0(s)$ in probability implies that also $\widehat{f}_0^n(t_{-i},t^*)$ converges to $f_0(s)$ in probability.

Thus, conditional on any s and ℓ_i, t_i,

$$(\widehat{f}_0^n, \widehat{f}_i^n)[s, t_{-i}^n, t_i] \to (f_0(s), f_1(s, \ell_i, t_i)) \quad \text{in} \quad \text{probability},$$

and also for any t_i'

$$(\widehat{f}_0^n, \widehat{f}_i^n)[s, t_{-i}^n, t_i'] \to (f_0(s), f_1(s, \ell_i, t_i')) \quad \text{in} \quad \text{probability}.$$

Since u is continuous and bounded, given the robust incentive compatibility of the sequence, the claim then follows from (1). ∎

This completes the proof of the theorem. ∎

A Corollary on Public Good Provision

A simple application of Theorem 1 is to a public good problem. Consider a society deciding on whether to undertake a costly public project. Agents differ only in the utility they receive from the project and this utility is private information. An immediate corollary of Theorem 1 is that any allocation for large economies which is robustly approximately implementable must have all agents pay the same cost, and so a convergent sequence of social choice functions is robustly approximately implementable only if the cost allocations of agents are identical. This implies that if individual rationality constraints are imposed, then the project is never built in the limit. Thus, Theorem 1 implies results of Mailath and Postlewaite (1990) and Al-Najjar and Smorodinsky (2000).

4 Envy-Freeness as a Sufficient Condition

Let us also examine the sufficiency of envy-freeness for implementability.

Excludable Goods and Private Values

An *excludable goods* setting is one where A_0 is a singleton. In addition to private goods settings, this admits the usual public goods where agents can be excluded from consuming the public good. A *private values* setting is one such that each agent i's preferences depend only on ℓ_i and t_i, and are independent of s. As we shall show below, both of the above conditions are important for obtaining a converse to Theorem 1.

We also add a technical condition that ensures that feasibility of allocations is not a roadblock for implementation. Let us also consider a case where the A_i's converge at some finite point on the sequence, so that we do not have to worry about what is feasible for large enough economies. That is, let us say that the allocation space is convergent if for every type $l \in L$, there exists i with $\ell_i = l$ and $N \geq i$ such that $A_i^n = A_i$ for all $n \geq N$. In such settings, envy-freeness and type-measurability are also sufficient for robust approximate implementation.

Theorem 2 *Consider a private values excludable goods setting with a convergent allocation space. A convergent sequence of social choice functions is robustly approximately implementable if and only if its limit is envy-free and type-measurable.*

The idea behind the proof builds on a class of "folk" mechanisms for implementation that adapt easily to excludable-good settings.[12]

Proof of Theorem 2: Given Theorem 1 we need only show that in an excludable goods setting, a convergent sequence of social choice functions is robustly approximately implementable whenever its limit is envy-free and type-measurable.

So, given a convergent sequence of social choice functions with a limit f that is envy-free and type-measurable, we describe the associated sequence of approximately implementing direct mechanisms, which is easily seen to be robustly incentive compatible. Let us only describe the mechanism for large enough n so that $A_i^n = A_i$ for any i (recall the finiteness of the label set). Each agent i announces his or her type $\widehat{t}_i$. Randomly identify $\sqrt{n}$ agents from the population. Given the excludability, we set their allocations to be constant across their announcements - and pick any constant allocation. Based on that sample, let us take a maximum likelihood estimate of s, denoted $\widehat{s}$. There may be several such estimates, and break ties in any way. For any of the remaining agents, the allocation is $f(\widehat{s}, \ell_i, \widehat{t}_i)$.

This mechanism is easily seen to be robustly incentive compatible. In the case where an agent is one of the sampled agents, his or her allocation is independent of the announcement. In the case where he or she is not sampled, then by envy-freeness and private values, the agent's allocation is as good as any other one they could obtain by lying.

Given type-measurability, it follows from a law of large numbers (e.g., the Glivenko-Cantelli Theorem) that for large enough n, with arbitrarily

[12] See Cordoba and Hammond (1998), Kovalenkov (2002), and Jackson (2003) for examples of such "folk" mechanisms.

high probability we learn all that is needed to estimate s to the extent that it determines the distribution of t's; and so the sequence approximates the desired social choice function in the limit. ∎

Non-excludable (Public) Goods and Beyond Private Values

In settings where we cannot exclude agents from enjoying some goods, or where agents' preferences are not private valued, the above result can fail.

Let us provide two examples that give an idea of each type of failure.

Example 1 *Non-implementability when Preferences are State-Dependent*

Consider a simple setting where $T = \{t, t'\}$ and $S = \{s, s'\}$, and where the probability of t is $3/4$ in state s and $1/4$ in state s'.

$A_i^n = \{a, a'\}$ for all i and n. The setting is excludable, and so let us ignore the public allocation. All agents have the same label so we also ignore the labels.

Let preferences be such that

$$u(a, s, t) = 1 = u(a', s, t')$$

$$u(a', s, t) = 0 = u(a, s, t')$$

$$u(a, s', t') = 1 = u(a', s', t)$$

$$u(a', s', t') = 0 = u(a, s', t)$$

This is a world where the state plays an important role in determining preferences. If the state is s, then a t type prefers a and a t' type prefers a'. If the state is s', then preferences are reversed.

Consider an f which gives an agent his or her preferred alternative in every situation. Such an f is clearly type-measurable and envy-free.[13] However, such an f is *not* robustly approximately implementable.

This is seen as follows. Consider any sequence of direct mechanisms $\{f^n\}$. In order to implement f in the limit, different types of agents must get different allocations (with some probability) far enough along the sequence. Given that f^n can depend only on type announcements, the allocations along the sequence must be independent of the state s, and depend only on the vector of announced t's. However, for any vector of t's either state could

[13] Given that the distribution of t depends on s, type measurability follows.

have generated them. Robustness requires that agents be happy with the allocation they get based on announcing their type rather than the other type regardless of what the state is, and even if they learn the state. However, here the agents' preferences reverse depending on the state and so this condition must be violated.

Example 2 *Non-implementability when goods are Non-Excludable*

Consider a pure public good setting with a binary public good decision, so that $A_0 = \{a, a'\}$. Let the private allocations be degenerate - so that we ignore the A_i's.

Again have two types $T = \{t, t'\}$ and two states $S = \{s, s'\}$, and where the probability of t is 3/4 in state s and 1/4 in state s'.

Preferences are such that all agents have the same label so we also ignore the labels.

Let preferences be state independent and such that

$$u(a, t) = 1 = u(a', t')$$

$$u(a', t) = -1 = u(a, t')$$

So, the t's are types who like a (say building the public project) and dislike a' (say not building the public project), while the t''s have the reverse preferences.

Consider f that minimizes the (expected) total sum of welfare of the agents: $f_0(s') = a$ and $f_0(s) = a'$.

This social choice function is envy-free and type-measurable. It is easy to check that this is not robustly incentive compatible, as agents have a clear incentive to pretend to be of the other type, as any approximating sequence must make the decision dependent on the announcements and in a way that goes against incentives far enough along the sequence.

These examples demonstrate that additional conditions will be needed to ensure implementability with non-excludable goods and/or interdependent valuations. We leave open the question of the full necessary and sufficient conditions for robust approximate implementation in the case where some goods are non-excludable and/or preferences may be state dependent.

5 References

Al-Najjar, N. and R. Smorodinsky (2000), "Pivotal Players and the Characterization of Influence," *Journal of Economic Theory,* Vol. 92, pp. 318-342.

Barberà, S. and M.O. Jackson (1995), "Strategy–proof Exchange," *Econometrica,* Vol. 63, pp. 51-88.

Bergemann, D. and S. Morris (2003), "Robust Mechanism Design," mimeo: Yale University.

Blume, L. and D. Easley (1983), "Implementation of Rational Expectations Equilibrium with Strategic Behavior," mimeo, Cornell University.

Cordoba, J. and P. Hammond, (1998), "Asymptotically Strategy-proof Walrasian Exchange," *Mathematical Social Sciences*, Vol. 36, pp 185-212.

Crémer, J. and R. McLean (1988), "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions," *Econometrica,* Vol. 56, pp. 1247–1257.

Foley, D. (1967), "Resource Allocation and the Public Sector," *Yale Economic Essays,* Vol. 7, pp. 45-98.

Gul, F. and A. Postlewaite (1992), "Asymptotic Efficiency in Large Exchange Economies with Asymmetric Information," *Econometrica,* Vol. 60, pp. 1273–1292.

Hurwicz, L. (1972), "On Informationally Decentralized Systems." In C.B. McGuire and R. Radner, *Decision and Organization*, North Holland, Amsterdam.

Hurwicz, L., E. Maskin, and A. Postlewaite (1995), "Feasible Nash Implementation of Social Choice Correspondences when the Designer does not Know Endowments or Production Sets." In J. Ledyard, *The Economics of Information Decentralization: Complexity, Efficiency, and Stability*, Kluwer, Amsterdam.

Jackson, M.O. (1992), "Incentive Compatibility and Competitive Allocations," *Economics Letters,* Vol. 40, pp. 299–302.

Jackson, M.O. (2003), "Efficiency and Information Aggregation in Auctions with Costly Information Acquisition," *Review of Economic Design*, Vol. 8, pp. - .

Jackson, M.O. and A. Manelli (1996), " Approximate Competitive Equilibria in Large Economies," *Journal of Economic Theory,* Vol. 77, pp. 354-376.

Kovalenkov, A(2002)"On a Folk Strategy-Proof Approximately Walrasian Mechanism,"*Journal of Economic Theory*, 103, pp 475-487.

Mailath, G. and A. Postlewaite (1990), "Asymmetric Information Bargaining Problems with Many Agents," *Review of Economic Studies,* Vol. 57, pp. 351–367.

McLean, R. and A. Postlewaite (2002), "Informational Size and Incentive Compatibility," *Econometrica,* Vol. 70, pp. 2421-2454.

Otani, Y. and J. Sicilian (1982), "Equilibrium Allocations of Walrasian Preference Games," *Journal of Economic Theory,* Vol. 27, pp. 47–68.

Otani, Y. and J. Sicilian (1990), "Limit Properties of Equilibrium Allocations of Walrasian Strategic Games," *Journal of Economic Theory,* Vol. 51, pp. 295–312.

Palfrey, T. and S. Srivastava (1987), "On Bayesian Implementable Allocations," *The Review of Economic Studies,* Vol. 54, pp. 193–208.

Pazner, E. and D. Schmeidler (1974), "A difficulty in the concept of fairness," *The Review of Economic Studies,* Vol. 41, pp. 441-443.

Roberts, D.J. and A. Postlewaite,(1976) "The Incentives for Price Taking Behavior in Large Exchange Economies,"

Thomson, W.*,(1979)"Equilibrium Allocations of Walras and Lindahl Manipulation Games,"DP 111, Center for Economic Research, University of Minnesota.*

6 *Appendix*

Proof of Lemma 1: First, suppose that $\{f^n\}$ is asymptotically envy-free but that f is not envy-free so that there is some s, ℓ_i, t_i and t_i', and ε^* such that

$$u(f_0(s), f_1(s, \ell_i, t_i), \ell_i, s, t_i) < u(f_0(s), f(s, \ell_i, t_i'), \ell_i, s, t_i) - \varepsilon^*.$$

Since $p(t_i' | s, \ell_i) > 0$ and since $q(\ell_i) > 0$ we can almost surely find a $j \neq i$ (for large enough n) so that $\ell_j = \ell_i$ and $t_j = t_i'$. This leads to a contradiction of asymptotic envy-freeness as $f_j^n(s, t_{-j}, t_j) \to f_1(s, \ell_i, t_i')$ (in probability, independently of t_i) and $f_i^n(s, t_{-i}, t_i) \to f_1(s, \ell_i, t_i)$ (in probability, independently of t_j), and so j will envy i by a fixed amount, even for large n.

Next consider the converse. Suppose that f is envy-free, but $\{f^n\}$ is not asymptotically envy-free. Since L, S and T are finite, we can find s, ℓ_i, t_i and an ε^* so that for arbitrary large n, some agent i who has label ℓ_i and type t_i envies another agent j of type t_j where $\ell_j = \ell_i$ by more than ε^*:

$$u(f_0^n(s,t), f_i^n(s,t), \ell_i, t_i, s) < u(f_0^n(s,t), f_j^n(s,t), \ell_i, t_i, s) - \varepsilon^*$$

f^n being convergent to f implies that:

$$\{f_0^n(s, t_{-i}, t_j), f_i^n(s, t_{-i}, t_j)\} \to (f_0(s), f_1(s, \ell_i, t_j)) \text{ in probability.}$$

Hence, a contradiction of envy-freeness of the limit f follows, noting that u is bounded and continuous. ▮