The Theory of Implementation when the Planner is a Player

S. Baliga, L.C. Corchon and T. Sjöström

THE THEORY OF IMPLEMENTATION WHEN THE PLANNER IS A PLAYER[1]

Sandeep Baliga

King's College, Cambridge University, Cambridge, CB2 1ST, England.

Luis C. Corchon

Departamento de Fundamentos, Universidad de Alicante. Alicante, 03071 Spain.
Instituto Valenciano de Investigaciones Economicas.

Tomas Sjöström

Department of Economics, Harvard University, Cambridge MA 02138, USA.

ABSTRACT: *In this paper we study a situation were the planner cannot commit to a mechanism and the outcome function is substituted by the planner herself. We assume 1) agents have complete information and play simultaneously; and 2) given the messages announced by the agents, the planner reacts in an optimal way given her beliefs. This transforms the implementation problem into a signaling game. We derive necessary and sufficient conditions for* interactive implementation *under different restrictions on the planner's out-of-equilibrium beliefs. We compare our results to standard results on Nash implementation.*

---

[1] Many thanks to Eric Maskin for his comments. We are also grateful to seminar audiences at Harvard University, Yale University, Brown University, University of Alicante, University of Warwick, and University of Windsor. Any remaining errors are our responsibility. The financial support of EC-SPES grant CT910057 is gratefully acknowledged.

1. Introduction

Suppose there are a certain number of agents who share some information. This information is called the preference profile, type or state. An outside party, the Principal (also called designer and planner), wants to elicit the information from the agents in order to implement an outcome that is optimal for her in each possible state (the social choice rule). In the standard approach to this problem, known as implementation, the Principal can design a mechanism, i.e. a message space and an outcome function mapping messages profiles into allocations. Once such a task has been accomplished the implementation problem becomes completely mechanical. Agents learn the state of nature, send the corresponding equilibrium message and duly receive a certain allocation. In fact, the task of the mechanism can be performed by a machine or by a mindless servant.

With the development of the theory of implementation came an appreciation of certain unsatisfactory aspects on which the theory relied: in order to get rid of unwanted equilibria, certain message profiles led to allocations that, no matter what state occurred, were never optimal. Under the assumption that the designer can commit to the mechanism, and also prevent ex post renegotiation among agents, such allocations are credible. However, such assumptions are not universally regarded as satisfactory, so it is desirable to explore the consequences of assuming otherwise.

Maskin and Moore (1987) relaxed the assumption that agents cannot renegotiate the outcome given to them by the mechanism. They assumed the existence of an exogenously given renegotiation function which selected Pareto efficient and individually rational allocations. The question of whether the planner can know the renegotiation function is difficult. Baliga

(1994) has shown that if the planner does not know the renegotiation function, then implementation with renegotiation in the sense of Maskin and Moore is extremely difficult. In the exchange economy, it is possible to solve this problem if the planner can commit to destroying goods ex post (Sjöström (1994)). But again the assumption that the planner can commit to the outcome function is critical, since destroying goods will presumably not be optimal in any state.

In this paper our approach is in the spirit of Becker (1974). If the Principal is a (benevolent) player in the game, "incredible threats" are ruled out, as the Principal at each node of the game tree must maximize his expected payoff, given his beliefs. Becker considered a moral hazard model, however, where the Principal had perfect information (see also Ray (1993)). In our model the Principal is uninformed of the true state (adverse selection), so messages can convey information from the agents to the Principal.

Chakravorty, Corchon and Wilkie (1992) consider an adverse selection model where the mechanism is run by a real person who is in charge of delivering the allocation once the messages have been announced. They assume the person in charge is a benevolent (but mindless) "keeper" and not a player: she is neither allowed to figure out the equilibrium strategies of the agents nor to make inferences from the messages sent by the agents. Therefore, she is always uninformed. But she must keep in the spirit of the mechanism and therefore under no circumstances can she pick an allocation that is not in the range of the social choice rule.

In our paper, the Principal is a full-fledged player, and thus she stands on equal footing with respect to agents in terms of rationality (but, of course, not in terms of information). Thus, in equilibrium she holds correct beliefs about strategies (i.e. the mapping from types to messages). Contrary to the credible implementation approach, the outcome function is replaced by the Principal's optimal response to messages. The result is a cheap talk game where agents are the senders and the planner is the receiver. The notion of a mechanism (with its connotations of a mechanical interaction between agents and the planner) is substituted by a *communication network*. In the model we will present in this paper this does not differ from the standard message space but we prefer to call it a communication network as messages convey information to the planner and, in general, the planner can also send messages to the agents. Our theory is a theory of *interactive implementation* where the planner discovers the agent's equilibrium strategies, makes inferences from what agents say and might send them messages in an attempt to influence their behavior.

We take the simplest possible situation and consider a two stage game where agents play simultaneously in the first stage (i.e. they announce a message profile) and the planner reacts in the second (and last) stage. The planner is assumed to maximize her expected utility at the second stage, given her preferences and her beliefs.

With at least three agents, there always exists a truth-telling perfect Bayesian equilibrium (PBE). Since the planner is just one more player she knows the equilibrium strategies. So suppose everyone tells the truth in each state. Then, if an individual deviates, the planner can ignore his deviation and just implement the outcome that is optimal for the preference profile that is announced by the other two or more agents. Indeed, in any separating equilibrium where agents send a different message in each state, all the private information is released to the planner who can then implement an optimal outcome.

However, as this is a cheap talk game, there will exist "babbling" perfect Bayesian equilibria where messages do not convey any (or only partial) information. Since we insist on full implementation, i.e. all equilibria should be optimal for the planner, some refinement is needed. We use a version of Farrell's neologism proof equilibrium, which we call FGP (Farrell-Grossman-Perry) equilibrium (following Maskin and Tirole (1992)). The corresponding notion of implementation is *interactive implementation in FGP equilibrium.*

We define a necessary and sufficient condition for interactive implementation in FGP equilibrium. We relate this condition to the standard notion of implementation in the sense of Maskin (1977). In Maskin's model, the social optimum is given by a social choice rule. We interpret the social choice rule as representing the utility maximizing outcomes for the planner. Of course, there may be many preference orderings for the planner that are compatible with the same social choice rule, since the social choice rule is only concerned with the top-ranked elements in each state. We show that even if a social choice rule ϕ is Nash implementable in the usual sense,[2] there may not exist *any* preference ordering for the planner that is compatible with ϕ and which makes interactive implementation of ϕ possible. This should not be surprising, since in interactive implementation the planner cannot make incredible threats, and the occurrence of "bad" outcomes out of equilibrium can be crucial in Maskin's model. More surprisingly, if ϕ is *locally Maskin monotonic*, a condition slightly stronger than Maskin monotonicity, then there always exists *some* preference ordering for the

---

[2] In the exchange economy, this is equivalent to the well-known condition of Maskin monotonicity.

planner that is compatible with ϕ and which makes interactive implementation of ϕ possible.

The restriction of credibility in the planner's choice of outcome might seem to make implementation harder, but there are actually social choice rules that can be interactively implemented but cannot be Nash-implemented in the standard sense. This is because when the planner is a player, his response to a given set of messages can depend on the actual equilibrium being played. Let us give an intuition for this in terms of Maskin's (1977) model. In Maskin's mechanism, if the agents send a false message θ when the true state is θ', and outcome $a \in \phi(\theta)$ is about to be chosen, then agent i can "object" by demanding an outcome a' such that a' is (strictly) preferred to a when the true state is θ' but not when it is θ. (This is "rule 2" of Maskin's mechanism.) It is also necessary to stop the agent from objecting when the true state is really announced. Essentially, this implies that the θ and θ' indifference curves through a must cross. [3] With interactive implementation, on the other hand, the planner's reaction to the objection can depend on which equilibrium is actually played. If the equilibrium is truthful, objections by a single agent can be disregarded. Therefore there is no requirement that indifference curves must cross.

In view of this result, it may be asked what would happen if the planner can commit to the outcome function, but can participate as a player by sending messages. In this case the fact that the planner "knows" equilibrium strategies leads to fairly dramatic results: in contrast to the standard model, even cardinal rules such as the utilitarian criterion can be

[3] Note that it does not matter to this argument if there were other states except θ' where a' would be preferred to a. With interactive implementation, however, such considerations are crucial.

implemented ! (See Baliga, Palfrey and Sjöström (1995))

2. Two Examples

2.1 A didactic example

In order to show how our ideas work, we will consider a famous problem in implementation theory: King Solomon's dilemma (see Moore (1992)). Two perfectly informed agents (Anna and Betsy) lay claim to the motherhood of the same child. King Solomon (the planner) has to decide to whom to award the baby. There are two possible states of nature, denoted by α and β. In state α (resp. β) Anna (resp. Betsy) is the mother. There are four possible allocations: a (the baby is given to Anna), b (the baby is given to Betsy), c (the child is cut in half) and d (death all around). The ordinal preferences of Anna, Betsy and Solomon are assumed to be as follows:

	STATE α			STATE β		
Preferences of:	Solomon	Anna	Betsy	Solomon	Anna	Betsy
	a	a	b	b	a	b
	d	b	c	d	c	a
	c	c	a	c	b	c
	b	d	d	a	d	d

We will assume that the communication network consists of a message space for each woman: $M_i = \Theta \times \{L,H\}$ where $\Theta = \{\alpha, \beta\}$ and L(ow) or H(igh) indicates the tone of voice. Once the mothers have spoken, King Solomon (after all a feudal monarch at a time when democracy was unimaginable) takes an irrevocable decision on the allocation.

Solomon's preferences are such that in state α he prefers a to d to c to b. On a cardinal scale, a is only slightly better than d and c, but he cannot suffer b. This is explained by the fact that Solomon likes to be right but hates the possibility that the populace gossips[4] that Anna and Betsy made a fool out of him. So, if the messages do not reveal the true state, rather than risking handing over the baby to the wrong person, he prefers the harsh outcome *pour encourager les autres*. If the true state is β, the situation is similar and Solomon prefers b to d to c to a.

If Solomon could disentangle the messages sent by Anna and Betsy and to acquire knowledge about the true state he would like to implement the following social choice rule, $\phi()$: $\phi(\alpha) = a$, $\phi(\beta) = b$. However, if after the mothers have spoken, he is not able to discern the truth he orders the "compromise" allocation d that maximizes his expected utility conditional on the belief that the true state is either α or β.[5] Formally, the *extended social choice rule* F, which specifies the optimal outcome also in cases where Solomon does not know the true state, is defined by: $F(\alpha)=a$, $F(\beta)=b$, $F(\{\alpha,\beta\})=d$.

Let us consider the perfect Bayesian equilibria of this game. Consider a separating equilibrium first. If both Anna and Betsy adopt a truth-telling strategy, Solomon knows it[6]. Since he also knows the message sent by them,

---

[4] The British monarchy demonstrates the destructive nature of gossip on the authority of the head of state.

[5] Formally, Solomon has priors $p(\alpha)>0$ and $p(\beta)>0$ for the two states. The compromise d maximizes his expected utility given these priors.

[6] Of course, the reason why King Solomon knows the strategies is not his legendary wisdom, but the fact that he is assumed to be a player (both things may be equivalent).

to pick the right choice is a trivial matter for him. Any deviation from Anna or Betsy only can be meant to fool Solomon and we know what to expect from this! Indeed, if Solomon allocates the priors after any zero probability message and therefore implements d, truth-telling is an equilibrium. Of course, any other equilibrium where at least one agent separates also reveals the true state to Solomon.

However, there are also non-optimal "babbling" equilibria. In such equilibria, for any message, Solomon's posteriors are identical to his priors, so no deviation from Anna or Betsy is going to make any difference. The equilibrium outcome is d. For instance, suppose Anna and Betsy always announce (α,L) and (β,L), respectively. As Solomon gets no information from these strategies, his priors go through and he implements d. Suppose now Anna "objects" by announcing (α,H). This amounts to the speech: "Please implement a and not d as the state is truly α". Formally, an *objection* is a zero probability message under the equilibrium strategies. To make sure that objections are always available, we include auxiliary messages such as high or low tone of voice. Nevertheless, if babbling means sending each message with positive probability, objections are impossible. Therefore, we rule out mixed strategies.

Unfortunately, Anna's objection is not reliable. Solomon may retort as follows: "Since the outcome that awaits you is d, it is *always* in your interest to try to fool me!. Therefore I can not be sure that your speech does not convey more lies. I am sorry". Similarly, if Betsy objects to the above pooling strategies, this also is not a reliable objection: whether or not Betsy is the true mother, she would like a or b to be implemented instead of d. An FGP equilibrium is defined to be an equilibrium which is free from reliable objections (a precise definition will be presented later). Consequently, the babbling equilibrium is an FGP equilibrium, and Solomon's preferred outcome cannot be interactively implemented in FGP equilibrium.[7]

Let us now go forward in time to a kinder, fairer, post-revolutionary era. Democracy is now the form of government in the ancient world and Solomon though still wise and still rich is wheeled out for state ceremonies only. No longer can he order anyone's death: capital punishment has been abolished by the radical government and there is no question of letting Solomon have any such independent power. Anna and Betsy are arguing over the motherhood of the child and Carla, their confidant, also knows the truth but is indifferent over a and b. Solomon is still sensitive about maintaining his reputation and preventing gossip, and, as he cannot implement d or c, he now prefers to give the two women a lot of money, outcome m, and allocate the baby randomly between Anna and Betsy if he is not sure of the true state. While this makes all the women very happy, the true mother still prefers to have her child:

	STATE α				STATE β			
Preferences of:	Solomon	Anna	Betsy	Carla	Solomon	Anna	Betsy	Carla
	a	a	m	m	b	m	b	m
	m	m	b	a, b	m	a	m	a, b
	b	b	a		a	b	a	

The extended social choice rule in this case is $F(\alpha)=a$, $F(\beta)=b$, $F(\alpha,\beta)=m$. The social choice rule ϕ defined by $\phi(\alpha)=a$, $\phi(\beta)=b$ is

7 More precisely, we have shown that these message spaces do not work. It is clear that no other message space will work either.

Nash-implementable in the standard sense.

Again, all the women announce α or β simultaneously in a high or low tone of voice, and then Solomon implements an outcome. Truthtelling can be turned into an equilibrium by assuming that Solomon will disregard individual deviation. There are also "babbling" equilibria where m is implemented and Solomon believes the priors no matter what message is sent. However, there is a difference: reliable objections to the babbling equilibria are available to the women. For example, suppose Anna, Betsy and Carla announce $(\alpha,L),(\beta,L),(\alpha,L)$ respectively in all states. Then, Solomon's best response is to implement m. Suppose Betsy deviates and announces (β,H). If β is indeed the true state she definitely prefers the socially optimal outcome b to m. However, if Anna is the true mother then Betsy prefers m to b and would have no incentive to make the objection. Therefore, her objection is reliable. In such a way all babbling equilibria can be knocked out. Thus, Solomon can interactively implement the extended social choice rule F in FGP equilibria.

2.2 An economic example

There are several economic settings where our assumption that the planner cannot commit to an outcome function has force. Suppose there is an authority relationship between a Principal such as the head of a firm or the government and the set of agents such as workers or regulated firms respectively. The Principal then has the authority to choose any outcome she wants and the agents do not have the power of recourse to some outside authority if the Principal reneges on the mechanism. An alternative setting is one where although contracts can be written, the messages sent by the agents are unverifiable so that the agents have no legal recourse if the Principal reneges on the outcome function.

To give a more explicit example, suppose there are $n \geq 3$ firms which are regulated by the government. Each firm i faces a demand curve $p_i = a - bq_i$ in its own market and has a constant marginal cost $c_i \in \{c_L, c_H\}$, where $a > c_H > c_L$. A state here is a list $\theta = \{c_1, c_2, \ldots, c_n\}$. Assume θ is common knowledge among the firms but the government does not know θ. The government's objective is to choose quantity q_i in each market i to maximize surplus

$$U(q_1, \ldots, q_n, c_1, \ldots, c_n) = \Sigma_i \ (aq_i - bq_i^2/2 - c_i q_i)$$

The prior probability that the state is c_L in market i is $r \in (0,1)$ and that it is c_H is $(1-r)$ and priors are i.i.d. in the other firms. Firm i's payoff is the profit $(p_i - c_i)q_i$. This model then fits ours framework if we assume the government cannot commit to a mechanism. Similar models have been studied in the regulation literature (Laffont and Tirole (1988)). We return to this example after we have stated our results.

3. The Model

There are $n \geq 3$ agents. Let I be the set of agents. The set of feasible outcomes is denoted by A. Let Θ be the set of possible states of the world, assumed to be finite. The prior probability of state θ occurring is p_θ, and we assume $p_\theta > 0$ for all $\theta \in \Theta$. Let Ω be the set of all subsets of Θ. Weak preferences of agent i in state θ are given by the ordering $R_i(\theta)$. Thus, for a, b $\in$ A, $aR_i(\theta)b$ means agent i (weakly) prefers outcome a to outcome b in state θ. Let $P_i(\theta)$ represent strict preferences and $I_i(\theta)$ indifference. The lower contour set for agent i at allocation a and state θ is $L_i(a, \theta) = \{b \in A$

/ $aR_i(\theta)b\}$. We assume throughout that the true state θ is common knowledge among the agents.

At this point, most of the literature on mechanism design defines a concept of social welfare, a social choice rule (SCR), a mapping from states to outcomes. In our setting, the planner is just another player. Accordingly, she must have an objective function, an action set, a strategy etc. just like a standard player in a game. The outcomes that are at the top of the planner's objective function for some particular state can be thought of as the optimal outcomes identified by a social choice rule. The planner differs from all the other agents in the game in one fundamental respect: the state is common knowledge to them but not to the planner. If the agents are not using strategies that release their private information in all states, the planner will have non-degenerate beliefs after observing some messages. She will have to choose a best response even though she is not sure of the state. In this respect, the usual concept of a social choice rule has to be "extended" to cover the cases where the planner has non-degenerate beliefs.

If allocation a is chosen in state θ, the payoff to the planner is $U(a, \theta)$. We suppose the planner behaves as an expected utility maximizer. Let $p(\theta \mid m)$ be the posterior probability conditional on the planner receiving some "message" m from the agents. Let $T \subseteq \Theta$ be the set of states where message m is sent. (We are ruling out mixed strategies.) Then, after message m the posterior probabilities are given by $p(\theta \mid m) = 0$ if $\theta \notin T$, and $p(\theta \mid m) = p_\theta / p(T)$, where $p(T) \equiv \Sigma_{\theta \in T} p_\theta$, if $\theta \in T$. Conditional on receiving this message m, the planner's expected utility from alternative $a \in A$ is

$$\sum_{\theta \in T} (p_\theta / p(T))\, U(a, \theta)$$

The extended social choice rule F is the "best response" correspondence for the planner, defined for each subset $T \subseteq \Theta$ as follows:

$$F(T) \equiv \text{argmax} \{ \sum_{\theta \in T} (p_\theta/p(T)) U(a, \theta) : a \in A \} \qquad (1)$$

Thus, after receiving the message m, the maximization of expected utility yields a set of optimal outcomes F(T), where T is the set of possible states given m. If $T=\{\theta\}$ is a singleton, then we write $F(\{\theta\})=F(\theta)$ for convenience. We can define a SCR ϕ_F by $\phi_F(\theta)=F(\theta)$ for all θ. Then ϕ_F is the *restriction* of F. Conversely, if F is an extended SCR and $\phi: \Theta \to A$ is a standard SCR such that $\phi(\theta)=F(\theta)$ for all $\theta \in \Theta$ then F is *compatible with* ϕ.

The planner's preference ordering is the basic data which induces the extended social choice rule. However, we find it convenient to couch the discussion directly in terms of the more general concept of implementation of extended social choice rules. The generalization is only marginal. In the appendix, we show that under some weak assumptions, for any extended social choice rule F defined on the subsets of Θ, there is some preference ordering for the planner that rationalizes it (i.e. such that (1) holds).

Finally we recall the following definition. If aR_ib implies aR'_ib, then R'_i is a monotonic transformation of R_i at a. The social choice rule ϕ is (Maskin) monotonic if, whenever $a \in \phi(R)$ and for all i R'_i is a monotonic transformation of R_i at a, then $a \in \phi(R')$.

4. Interactive Implementation

The *communication network* is $M = \times_{i \in I} M_i$, where $M_i = \Theta \times A \times Q_i$.[8] Thus, each

[8] We can consider more general message spaces but nothing is lost by focusing our attention on the ones we consider.

player reports a state $\theta^i \subseteq \Theta$, and outcome $a^i \in A$ and a "nuisance message" $q^i \in Q_i$. A generic message is denoted by $m_i = (\theta^i, a^i, q^i)$. Let $m_{-i}=(m_1,\dots m_{i-1},m_{i+1},\dots,m_n)$. A strategy for player i is a map $\mu_i: \Theta \longrightarrow M_i$, where $\mu_i(\theta)$ is the message sent in state θ. Let $\mu_{-i}(\theta)=(\mu_1(\theta),\dots,\mu_{i-1}(\theta),\mu_{i+1}(\theta),\dots,\mu_n(\theta))$. A strategy for the planner is a function α: $M \longrightarrow A$, where $\alpha(m)$ is the allocation chosen when the message is m.

Suppose the agents use strategies μ. The range of μ, i.e. the set of messages that are sent for some state, is denoted $\mu(\Theta) = \{m \in M : m = \mu(\theta)$ for some $\theta \in \Theta\}$. For any $m \in M$, $\mu^{-1}(m) \equiv \{\theta \in \Theta : \mu(\theta) = m\}$ is the set of states where agents send message m. Similarly, $\mu_{-i}^{-1}(m_{-i}) \equiv \{\theta \in \Theta : \mu_{-i}(\theta) = m_{-i}\}$. If $\mu(\theta)=m$ for all $\theta \in T \subseteq \Theta$, we write $m=\mu(T)$. Similarly if $\mu_{-i}(\theta)=m_{-i}$ for all $\theta \in T \subseteq \Theta$, then $m_{-i}=\mu_{-i}(T)$.

Definition 1. *(μ^*, α^*) is a perfect Bayesian equilibrium if*

(1) for each $\theta \in \Theta$ and each i, $\alpha^*(\mu^*(\theta))R_i(\theta)\alpha^*(\mu^*_{-i}(\theta), m_i)$ for all $m_i \in M_i$,

(2) for each $m \in \mu^*(\Theta)$, $\alpha^*(m) \in F((\mu^*)^{-1}(m))$,

(3) for each $m \in M \backslash \mu^*(\Theta)$, there exists $T \subseteq \Theta$ such that $\alpha^*(m) \in F(T)$.

Part (1) of Definition 1 states that, given the anticipated response from the planner, each agent sends a message that maximizes his payoff. Part (2) of Definition 1 requires that, for each message m that is sent in equilibrium, the planner chooses what is best for her, conditional on the *correct* belief that the true state must belong to $(\mu^*)^{-1}(m)$.

The PBE is *separating* if $(\mu^*)^{-1}(m)$ is a singleton for all $m \in \mu^*(\Theta)$. In this case the planner inverts μ^* and is fully informed in equilibrium. A PBE which is not separating is *pooling*. In a pooling PBE some "compromise" must be chosen by the planner whenever $m \in \mu^*(\Theta)$ is such that $(\mu^*)^{-1}(m)$ is not a singleton, for the planner will not be fully informed.

Part (3) of the Definition 1 requires that if m is a message which is *not* sent in equilibrium, then there exists *some* belief for the planner, say the state belongs to the set T, such that the planner's response is optimal conditional on this belief. However, as Farrell and others have argued, the planner might plausibly infer something from out-of-equilibrium messages. We introduce restrictions on the planner's off the equilibrium path beliefs that capture this idea. [9] It is important to note that we are concerned with the case where there are at least three senders of messages. Therefore, if one agent makes a surprise announcement, the planner may be able to infer some information from the other agents' messages.

Definition 2. *Let (μ^*,α^*) be a PBE. Suppose $\mu^*_{-i}(\theta') = m_{-i}$, $(\theta',a',q') \in M_i$ but $(m_{-i},(\theta',a',q')) \notin \mu^*(\Theta)$. Then (θ',a',q') is an* objection *to (μ^*,α^*,m_{-i}) by player i.*

Definition 3. *Let (μ^*,α^*) be a PBE, and $\mu^*_{-i}(\theta') = m_{-i}$. An objection (θ',a',q') to (μ^*,α^*,m_{-i}) is* reliable *for player i if there exists a set $T' \subset (\mu^*)^{-1}(\mu^*(\theta'))$ such that*

(1) $\theta' \in T'$,

(2) $\alpha^*(\mu^*(\theta')) \notin F(T')$,

---

[9] As in Farrell's model, we could allow players to send messages which are sets of types: "I am in the set $T \subseteq \Theta$". However, it will be clear that for our purposes the two formulations are equivalent.

(3) $a' \in F(T')$

(4) if $\theta \in T'$, *then* $a' P_i(\theta) \alpha^*(\mu^*(\theta'))$ *and*

(5) if $\theta \in (\mu^*)^{-1}(\mu^*(\theta')) \setminus T'$, *then* $\alpha^*(\mu^*(\theta')) R_i(\theta) a'$.

A PBE is an FGP-equilibrium *if no player has a reliable objection.*

<u>Remark</u> Requirement (2) of Definition 3 implies that reliable objections can only be made in states where some pooling occurs, for if $(\mu^*)^{-1}(\mu^*(\theta'))=\{\theta'\}$, then by definition of PBE, $\alpha^*(\mu^*(\theta')) \in F(\theta')$.

A reliable objection amounts to the following speech: "The other agents have told you $m_{-i}=\mu^*_{-i}(\theta')$, I'm supposed to say $\mu^*_i(\theta')$ and you are supposed to pick $\alpha^*(\mu^*(\theta'))$ but I object: the state is truly θ' and you should pick a'. Your knowledge of strategies allows you to infer that the true state is in $(\mu^*)^{-1}(\mu^*(\theta'))$. There exists a set $T' \subset (\mu^*)^{-1}(\mu^*(\theta'))$, with $\theta' \in T'$, $a' \in F(T')$ and $\alpha^*(\mu^*(\theta')) \notin F(T')$, and with $a' P_i(\theta)\ \alpha^*(\mu^*(\theta'))$ if and only if $\theta \in T'$. Since I'm trying to convince you to choose a' you can believe that the state is in T', so you should indeed pick a'."

We say that f is a *selection* from F, and write $f \in F$, if f is a single-valued function such that $f(T) \in F(T)$ for all $T \subseteq \Theta$.

Definition 4. *The extended social choice rule F is* (interactively) implemented in FGP-equilibrium, *if:*

(i) for each selection $f \in F$, *there exists an FGP equilibrium* (μ,α) *such that*

$$\alpha(\mu(\theta))=f(\theta) \text{ for all } \theta, \text{ and}$$

(ii) if (μ,α) *is an FGP-equilibrium, then for all* θ, $\alpha(\mu(\theta)) \in F(\theta)$.

Remark We assume that if an objection convinces the planner that the state is θ', and if $F(\theta')$ is not a singleton, then the objecting agent can tell the planner which outcome $a' \in F(\theta')$ to pick, perhaps as a reward for informing the planner of the true state. (The planner is indifferent among the outcomes in $F(\theta')$ by definition, and it is in her interest to encourage agents to object in order to knock out pooling equilibria.) A straightforward modification would be to define a "safe" reliable objection to be one where *all* elements in $F(\theta')$ make the objecting agent better off.

Similarly to Definition 4, one can define interactive implementation in PBE (with no restrictions on out-of-equilibrium beliefs). However, the following result shows that, due to the existence of "babbling" PBE, interactive implementation in PBE is (almost) impossible.

Theorem 1. *Let F be an extended social choice rule. If F is interactively implementable in PBE, there exists an outcome a such that* $a \in F(\theta)$, $\forall\theta \in \Theta$.

Proof: Let $\prod_{i \in n} M_i$ be a communication network that implements F interactively in PBE. Let all agents send the message profile m independent of the state of the world. For any message, the planner's prior goes through and she implements some $a \in F(\Theta)$. These strategies and posteriors form a pooling PBE, and since F is implemented, we must conclude that $a \in F(\theta)$, $\forall\theta \in \Theta$. QED

5. A Necessary and Sufficient Condition for Interactive Implementation in FGP Equilibria

Consider the following definition:

Definition 5. *An extended social choice rule F is* reliably monotonic *if the following holds. If for any $T \subseteq \Theta$, there is $b \in F(T)$ such that $b \notin \bigcap_{t \in T} F(t)$, then there exists $i \in I$, $T' \subset T$ and $a \in F(T')$ such that:*

(i) if $\theta \in T'$ then $aP_i(\theta)b$

(ii) if $\theta \in T \setminus T'$, then $bR_i(\theta)a$

Theorem 2. *An extended social choice rule F is implementable in FGP equilibrium if and only if it is reliably monotonic.*

Proof: Necessity Suppose F is interactively implementable in FGP equilibrium. Let $\mathop{X}_{i \in I} M_i$ be the message space of the mechanism that implements F, where $M_i = \Theta \times A \times Q_i$. Suppose there is a set $T \subseteq \Theta$, $b \in F(T)$ but $b \notin \bigcap_{t \in T} F(t)$. Let $f \in F$ be such that $f(T)=b$.

Consider the following strategies (μ^*, α^*). For each i, there is $\bar{m}_i$ such that $\mu_i^*(t) = \bar{m}_i$ if and only if $t \in T$. If $t \notin T$, then for all i, $\mu_i^*(t) = \{t,.,.\}$. That is, agent i reveals the state truthfully if it is not in T; if it is in T he always submits $\bar{m}_i$. If $m_i = \mu_i^*(t)$ for $t \notin T$ and all $i \neq j$, then $\alpha^*(m) = f(t)$. If $m_i = \bar{m}_i$ for all $i \neq j$, then $\alpha^*(m) = f(T) = b$. For all other m, $\alpha^*(m)$ is arbitrary.

Since F is implemented and $b \notin \bigcap_{t \in T} F(t)$, (μ^*, α^*) is not an FGP equilibrium. If $t \notin T$, then unilateral deviations are ignored by the planner, hence no deviation is profitable. Therefore, some agent i must have a reliable objection at some state $\theta' \in T$. That is, there exists a reliable objection (θ', a', q') to $(\mu^*, \alpha^*, \bar{m}_{-i})$. By definition, $\mu^*(\theta') = \bar{m}$. Then there exists $T' \subset (\mu^*)^{-1}(\bar{m}) = T$ such that $\theta' \in T'$, $b \notin F(T')$, $a' \in F(T')$ and the following holds: if $\theta \in T'$, then $a' P_i(\theta) b$, and if $\theta \in T \setminus T'$, then $bR_i(\theta)a'$. Thus, F is reliably monotonic. This proves necessity.

Sufficiency: Let the message space for player i be $M_i=\Theta\times A\times\{1,2,\ldots,|\Theta|+1\}$. Truthtelling can be supported as an FGP-equilibrium by letting the planner disregard unilateral deviations. Thus, we only need to show that there are no non-optimal equilibria.

Suppose there exists a non-optimal FGP equilibrium (μ,α) such that for some $\theta^*\in\Theta$, $\alpha(\mu(\theta^*))=b\notin F(\theta^*)$. Let $\bar{m}=\mu(\theta^*)$ and $T=\{\theta:\mu(\theta)=\bar{m}\}$. By definition of PBE, $b\in F(T)$. Also, $b\notin\bigcap_{t\in T}F(t)$ since $b\notin F(\theta^*)$. Since F is reliably monotonic, there exists $i\in I$, $T'\subset T$ and $a\in F(T')$ such that: if $\theta\in T'$ then $aP_i(\theta)b$, and if $\theta\in T\setminus T'$, then $bR_i(\theta)a$. Let $\theta\in T'$. Consider $m_i'=\{\theta,a,z\}\notin\mu^*(\Theta)$. Then, m_i' is a reliable objection to $(\mu,\alpha,\bar{m}_{-i})$. Then (μ,α) is not an FGP equilibrium, contradiction. This proves sufficiency. QED

Recall our economic example (B). If the government knows the cost in market i is c_i, it will choose a quantity $q_i=(a-c_i)/b$ such that price p_i equals marginal cost c_i. If it is unsure of the cost function of firm i, it will choose a quantity such that price equals the expected marginal cost $rc_H+(1-r)c_L$. We now show that the implied extended social choice rule F is reliably monotonic if and only if

$$r \geq (a-c_H)/(c_H-c_L) \tag{2}$$

Consider any set $T\subseteq\Theta$ which is not a singleton. Then, there exists $\theta,\theta'\in T$ and a firm i such that $c_i=c_L$ under θ and $c_i=c_H$ under θ'. If $(q_1,\ldots,q_n)\in F(T)$ then $q_i=rq_H+(1-r)q_L$ so that the low cost firm makes a profit and the high cost firm makes a loss. Clearly the only objection firm i has an incentive to make is that its cost is high. If it can convince the planner, it will be asked to produce q_H. This is always preferred to $rq_H+(1-r)q_L$ by the high cost firm. It can be checked that the low cost firm prefers $rq_H+(1-r)q_L$ to q_H if and only if (2) holds. Thus, if (2) holds the set T' as required by definition 5 can consist of all the states in T where the cost for firm i is high. Conversely, if (2) does not hold no such set T' exists. Thus, (2) is equivalent to reliable monotonicity.

6. Comparison of Interactive Implementation and Standard Nash Implementation

We first give two examples showing that interactive implementation in FGP equilibrium is in general neither easier nor more difficult than standard Nash implementation. We suppose there are three consumers and three commodities, and two states θ' and θ''. However, the third consumer is only interested in the consumption of the third good and no other consumer has endowments of this commodity nor do they derive any utility from consuming the third good. We assume that in any (extended) SCR under consideration the third consumer consumes just her initial endowments so we in effect have a two-good, two consumer world. Although the examples are concerned with extended social choice rules, it follows from Proposition 1 in the appendix that in each case there exists a nice utility function for the planner that rationalizes it, i.e. such that (1) holds.

Example 1. *An extended SCR F which is not interactively implementable in FGP equilibria, even though the restriction ϕ_F is Nash-implementable.*

Suppose ϕ_F is the Walrasian correspondence. In Figure 1 we have pictured the competitive equilibria for the states θ' and θ''. We know that in this case ϕ_F is Nash implementable. However because of the position of the "compromise" $\hat{a}=F(\{\theta',\theta''\})$, F is not reliably monotonic.

Example 2. *An extended SCR F which is interactively implementable in FGP equilibrium, but ϕ_F is not Nash implementable.*

Figure 2 shows the Walrasian correspondence for states θ' and θ''. In this case this correspondence does not satisfies monotonicity and thus it is not Nash implementable. However if the compromise $\hat{a}=F(\{\theta',\theta''\})$ is in the right place (as it happens in Figure 2) then F is reliably monotonic.

INSERT FIGURES 1 AND 2 ABOUT HERE

It should be clear that the reason why Maskin monotonicity and no veto power are neither necessary nor sufficient for implementation in FGP equilibria is that the compromise may be in the wrong place. Therefore, given a Maskin monotonic social choice function, we may wonder whether there exists *some* preferences for the planner which are compatible with the social choice function and allows interactive implementable in FGP equilibria. The answer is no, since our next example exhibits a Maskin monotonic social choice function such that no extended social rule compatible with it can be interactively implemented in FGP equilibria.

Example 3. *A Maskin-monotonic social choice rule ϕ such that, if F is any extended social choice rule which is compatible with ϕ, then F cannot be interactively implemented in FGP equilibria.*

Consider a three person exchange economy with two goods. The social endowment of good i is ω_i. There are four states, $\Theta=\{\theta_1,\theta_2,\theta_3,\theta_4\}$. The preferences of player 3 are fixed at $R_3(\theta)=R_3$ for all θ. The preferences of player 1 are $R_1(\theta_1)=R_1(\theta_2)=R_1$ and $R_1(\theta_3)=R_1(\theta_4)= R_1'$. The preferences of player 2 are $R_2(\theta_1)=R_2(\theta_3)=R_2$ and $R_2(\theta_2)=R_2(\theta_4)= R_2'$. Let ϕ be single valued and let $a=\phi(\theta_1), b=\phi(\theta_2), c=\phi(\theta_3)$, $d=\phi(\theta_4)$ be four distinct outcomes. Suppose in all four cases player 3 gets some small amount $\epsilon>0$ of each good. Let $x_i=(x_{i1},x_{i2})$ denote the amount of goods 1 and 2 consumed by agent i at allocation x. The preferences of player 1 are given in figure 3, where the dotted (resp. solid) line represents an R_1' (resp. R_1) indifference curve. The preferences of player 2 are given in figure 4. R_1' and R_2' are actually isomorphic, and also R_1 and R_2. [10]

Clearly ϕ is monotonic. Now let F be any extended social choice rule compatible with ϕ. We claim F cannot be interactively implemented in FGP equilibria.

Let $d_{11}=a_{21}$ be the greatest amount of good 1 consumed by any player at any of the outcomes a, b, c, d, and let $a_{12}=d_{22}$ denote the greatest amount of good 2 consumed by any player at any of the outcomes a, b, c, d. Let K_1 and K_2 be numbers such that $d_{11} < K_1 < \omega_1$ and $a_{12} < K_2 < \omega_2$. Let the shaded area $C_1=A_1\cup B_1$ in figure 3 be the union of the sets $A_1=\{x_1:\ x_{11}\le\omega_1-K_1$ and $x_{12}\le K_2\}$ and $B_1=\{x_1:\ x_{11}\le K_1$ and $x_{12}\le\omega_2-K_2\}$. Let $C_2=A_2\cup B_2$ in figure 4 be similarly defined.

Now we draw the indifference curves for player 1 in such a way that if an indifference curve for preferences R_1 passes through the area C_1, then it coincides throughout the consumption set with an indifference curve for preferences R_1'. Similarly, if an indifference curves for preferences R_2

[10] Both player 1 and player 2 are always indifferent between a, b, c, and d but the example can be perturbed so that this indifference goes away.

passes through the area C_2, then it coincides throughout player 2's consumption set with an indifference curve for preferences R_2'.

Let G_1 and H_1 be the shaded areas in figure 3 given by:

$G_1=\{z:$ if $x \in A$ and $x_1=z$, then aP_1x and $xR_1'a\}$

$H_1=\{z:$ if $x \in A$ and $x_1=z$, then $aP_1'x$ and $xR_1a\}$

Let G_2 and H_2 be similar for player 2.

It is clear that we can draw the indifference curves in such a way that if $z=(z_1,z_2) \in G_1$ (where z_i is the consumption of good i) then $z_1>\omega_1-K_1$ and $z_2>K_2$. Similarly, if $z=(z_1,z_2) \in G_2$ then $z_1>\omega_1-K_1$ and $z_2>K_2$. Similar statements hold for H_1 and H_2.

Suppose F is reliably monotonic and let $e \in F(\Theta)$. Since $e \notin \bigcap_{t \in \Theta} F(t)$, there exists $i \in I$, $T' \subset \Theta$ and $g \in F(T')$ such that:

(i) if $\theta \in T'$ then $gP_i(\theta)e$

(ii) if $\theta \notin T'$, then $eR_i(\theta)g$.

There are four possibilities, call them I, II, III, IV. If i=1 then either (I) $T'=\{\theta_1,\theta_2\}$ so gP_1e and $eR_1'g$, or (II) $T'=\{\theta_3,\theta_4\}$ so $gP_1'e$ and eR_1g. Similarly, there are two possibilities (III and IV) for the case i=2.

Consider first possibility I, where $T'=\{\theta_1,\theta_2\}$. Since $g \notin \bigcap_{t \in T'} F(t)$, applying the definition of reliable monotonicity we find: there is $\theta' \in T'$ and $y \in F(\theta')$ such that:

(i) $yP_2(\theta')g$

(ii) if $\theta \in T' \setminus \{\theta'\}$, then $gR_2(\theta)y$

Again there are two possibilities to consider: (Ia) $\theta'=\theta_1$ or (Ib) $\theta'=\theta_2$.

(Ia) If $\theta'=\theta_1$ then $R_2(\theta')=R_2$ and $F(\theta')=a$. From (i) and (ii) it follows that aP_2g and $gR_2'a$. Thus, g_2 must be in area G_2 in figure 4. Then g_{21} >

ω_1-K_1 and $g_{22} > K_2$, so $g_{11} < K_1$ and $g_{12} < \omega_2$-K_2. Thus, g_1 belongs to the area B_1cC_1 of figure 3. By construction, if an indifference curve for preferences R_1 passes through this area, then it coincides throughout the consumption set with an indifference curve for preferences R_1'. However, this contradicts gP_1e and $eR_1'g$.

(Ib) This case is completely symmetric to (Ia).

Thus, possibility I leads to a contradiction. The remaining possibilities II, III, IV lead to similar contradictions. Thus, although ϕ is Maskin monotonic and hence Nash-implementable, if F is any extended social choice rule which is compatible with ϕ then F is not reliably monotonic, hence F cannot be interactively implemented in FGP equilibria.

However, a slightly stronger condition than Maskin monotonicity, *local Maskin monotonicity*, does guarantee that there exists *some* extended social choice rule F, which is compatible with ϕ, and which can be interactively implemented in FGP equilibria. The Walrasian correspondence is locally monotonic whenever the competitive equilibria occur in the interior of the feasible set. In Example 1, then, there must exist *some* c such that if $F(\{\theta',\theta''\})=c$, then F is interactively implementable in FGP equilibria.

We consider a special kind of environment. The feasible set A is a subset of Euclidean space with the usual Euclidean metric. Preferences are continuous and strictly convex. We also restrict attention to social choice *functions*.

Definition 6. (Shenker (1994)) A social choice rule ϕ is *locally Maskin monotonic* if for any θ and any set of open neighborhoods N_i around $a\in\phi(\theta)$, $a\in\phi(\theta')$ whenever $aR_i(\theta)x \Rightarrow aR_i(\theta')x$ for all allocations $x\in N_i$ and all i.

Notice that the SCR analyzed in Example 3 is not locally Maskin monotonic.

Theorem 3. *If a social choice function ϕ is locally Maskin monotonic, then there exists an extended social choice function F, which is compatible with ϕ, and which can be interactively implemented in FGP equilibria.*

Proof Since Θ is finite, we have $\Theta=\{\theta_1,\theta_2,\ldots,\theta_K\}$ for $K < \infty$. Let ϕ be locally Maskin monotonic. We need to define F(T) for all $T\subseteq\Theta$ in such a way that F is compatible with ϕ and reliably monotonic. Note that, trivially, F(T) can be said to be defined for all T such that $\theta_i\in T$ for some $i\leq 0$, since there is no such i.

Now suppose for some N such that $1\leq N\leq K$, we have defined F(T) for all T such that $\theta_i\in T$ for some $i\leq N-1$. If $N > 1$, define $T(N-1)\equiv\{\theta_1,\theta_2,\ldots,\theta_{N-1}\}$. If N=1 then $T(N-1)\equiv\emptyset$. We will proceed to define F(T) for all $T\subseteq\Theta\backslash T(N-1)$ such that $\theta_N\in T$. (After this has been done, we have defined F(T) for all T such that $\theta_i\in T$ for some $i\leq N$, and it is clear how this process can be continued.)

Consider θ_N and $a_N = \phi(\theta_N)$. Suppose, for some k such that $1\leq k\leq K-N+2$, F(T) has been defined for all $T\subseteq\Theta\backslash T(N-1)$ such that (a) $\theta_N\in T$, and (b) $|T| \geq k$. (This requirement is vacuously satisfied if $k=K-N+2$ since in this case there is no $T\subseteq\Theta\backslash T(N-1)$ that satisfies (a) and (b).) Also, suppose for all $T\subseteq\Theta\backslash T(N-1)$ satisfying (a) and (b), either $F(T)=\cap_{\theta\in T}F(\theta)$ or there exists $i(T)\in N$ and non-empty sets B(T) and N(T) satisfying:

(1) $T=B(T)\cup N(T)$

(2) $\theta_N\in B(T)$

(3) if $\theta\in B(T)$, then $a_N P_{i(T)}(\theta) F(T)$

(4) if $\theta \in N(T)$, then $F(T)P_{i(T)}(\theta)a_N$.

Now consider $T \subseteq \Theta \setminus T(N-1)$ such that $\theta_N \in T$ and $|T| = k-1$. We need to define F(T). If $a_N = \cap_{\theta \in T} F(\theta)$, then $F(T) \equiv a_N$. Suppose $a_N \neq F(\theta')$ for some $\theta' \in T$. Then there exists a player i(T) such that $R_{i(T)}(\theta')$ is not a locally Maskin monotonic transformation of $R_{i(T)}(\theta_N)$ at a_N. In this case (by continuity of preferences) there exists an outcome x arbitrarily close to a_N such that $a_N P_{i(T)}(\theta_N)x$, $xP_{i(T)}(\theta')a_N$, and there is no θ such that $a_N I_{i(T)}(\theta)x$. Then, set $F(T) \equiv x$.

Let $B(T) = \{\theta \in T: a_N P_{i(T)}(\theta)x\} \neq \emptyset$ and $N(T) = T \setminus B(T) = \{\theta \in T: xP_{i(T)}(\theta)a_N\} \neq \emptyset$. If there exists one or several T' with $|T'| \geq k$ such that $B(T')=T$, then we choose $x \equiv F(T) \equiv F(B(T'))$ sufficiently close to a_N so that (using (3) and (4) above) the following holds:

(3') If $\theta \in B(T')$, then $xP_{i(T')}(\theta)F(T')$

(4') if $\theta \in N(T')$, then $F(T')P_{i(T')}(\theta)x$.

Choose F(T) this way for all $T \subseteq \Theta \setminus T(N-1)$ such that $\theta_N \in T$ and $|T| = k-1$. Then (1)-(4) will hold for all $T \subseteq \Theta \setminus T(N-1)$ such that $\theta_N \in T$ and $|T| \geq k-1$. Thus, we may continue this way to define F(T) for all $T \subseteq \Theta \setminus T(N-1)$ such that $\theta_N \in T$.

The following is clear from this procedure. Take any $T \subseteq \Theta \setminus T(N-1)$ containing θ_N. Then, either $F(T) = \cap_{\theta \in T} F(\theta) = a_N$. Or, there exists a player $i(T) \in I$ and non-empty sets B(T) and N(T) such that:

(1*) $T = B(T) \cup N(T)$

(2*) $\theta_N \in B(T)$

(3*) If $\theta \in B(T)$, then $F(B(T))P_{i(T)}(\theta)F(T)$

(4*) If $\theta \in N(T)$, then $F(T)P_{i(T)}(\theta)F(B(T))$.

After this procedure we have defined F(T) for all T such that $\theta_i \in T$ for some $i \leq N$. Thus, we may continue to define F(T) for all $T \subseteq \Theta$.

The result of this procedure is the following. Take any $T \subseteq \Theta$. Let N be the smallest integer such that $\theta_N \in T$. Then, either $F(T)=\cap_{\theta\in T}F(\theta)=a_N$. Or, there exists a player $i(T)\in I$ and non-empty sets $B(T)$ and $N(T)$ such that (1*)-(4*) hold.

We claim that F is reliably monotonic. Suppose there exists a set T and $\theta^*\in F(T)$ such that $F(T)\neq F(\theta^*)$. Let N be the smallest integer such that $\theta_N\in T$. Then, since $F(T)\neq\cap_{\theta\in T}F(\theta)$, there exists a player $i(T)\in I$ and non-empty sets $B(T)$ (containing θ_N) and $N(T)$ such that (1*)-(4*) hold. This is equivalent to F being reliably monotonic. Moreover, F is clearly compatible with ϕ. QED

Remark Notice from our construction that it is always possible to choose outcomes for all sets $T,T'\subseteq\Theta$ so that $F(T)=F(T')$ iff $F(T)=F(T')=\cap_{t\in T}F(t)=\cap_{t'\in T'}F(t')$. Therefore, by Proposition 2 in the appendix, there exists a utility function for the planner that implies such an extended social choice rule.

7. Conclusion

This paper has defined a new notion of interactive implementation and investigated the types of social choice rules that can be interactively implemented. Our analysis suggests that at least the following questions are of interest:

(1) There may be other restrictions on beliefs "off the equilibrium path" worth analyzing;

(2) Since messages in our model are cheap talk, it is necessary to postulate that the planner understands the "language" which the agents speak. On the other hand, if messages were costly to send, standard refinements such as

stability could be more powerful;

(3) Allowing for other types of interaction (i.e. having the planner move at the same time as the agents or many times) between the planner and agents may alter the set of extended social choice rules that can be interactively implemented in an interesting manner;

(4) The set of extended social choice rules that can be interactively implemented when there is incomplete information among the agents remains to be characterized; and

(5) The principal may be able to commit to an outcome function in some minimal way. For example, the principal may commit not to change the outcome from a to b if he is indifferent between a and b. Or, alternatively, he will not change from a to b if the expected gain is smaller than some $\varepsilon > 0$.

Appendix

In this appendix we show that extended social choice rules discussed in the paper can be derived from an underlying preference ordering for the planner.

First consider the exchange economy with two goods and two consumers, and no free disposal. An outcome is a pair $(a_1, a_2) \in \mathbb{R}^2_+$ indicating the consumption of consumer 1 of commodities 1 and 2. By no free disposal, the bundle allocated to consumer 2 can be inferred from (a_1, a_2). We will assume that there are two states of the world denoted by θ' and θ''. Let $F(\theta') \equiv a'$, $F(\theta'') \equiv a''$, $F(\theta', \theta'') \equiv \hat{a}$. Consider the following condition:

(R) $\exists q \in R^2$ such that $q\hat{a} > q\,a'$ and $q\hat{a} < qa''$

It is clear that (R) is satisfied unless $\hat{a}$, a' and a'' are lined up with $\hat{a}$ not in between a' and a''. In particular, (R) holds for Examples 1 and 2.

Proposition 1. *If a', a'' and $\hat{a}$ satisfy (R) then:*

a) There exists a utility function of the planner such that $U(\theta', a') > U(\theta', a)$ $\forall a \neq a'$ and $U(\theta'', a'') > U(\theta'', a)$ $\forall a \neq a''$. Moreover $U(\theta',.)$ and $U(\theta'',.)$ are strictly concave in a.

b) There is a prior p_θ such that

$$\hat{a} = argmax \{ p_{\theta'} U(\theta', a) + p_{\theta''} U(\theta'', a) : a \in A \}$$

i.e. $\hat{a}$ maximizes the planner expected utility when beliefs are the prior beliefs, $p_{\theta'} + p_{\theta''} = 1$.

<u>Proof</u> Construct U() such that $\hat{a}$ maximizes $U(\theta', a)$ subject to $q\,a \geq q\,\hat{a}$ and $a \in A$, and $\hat{a}$ maximizes $U(\theta'', a)$ subject to $q\,a \leq q\,\hat{a}$ and $a \in A$. It is clear that U() can be taken as required in a) above. Now notice that q is a separating hyperplane of the most preferred sets according to the preferences of the planner in states θ' and θ''. Therefore $\hat{a}$ is undominated in the sense that there is no $a \in A$ such that $U(\theta', a) \geq U(\theta', \hat{a})$ and $U(\theta'', a) \geq U(\theta'', \hat{a})$ with one inequality strict. By a standard result, the strict concavity of $U(\theta',.)$ and $U(\theta'',.)$ implies that there are non negative weights $(p, 1 - p)$ such that $\hat{a}$ maximizes $pU(\theta',a) + (1-p)U(\theta'',a)$. QED

Now consider the general environment and at least three agents. Clearly, if we are given some standard SCR ϕ, say the Walrasian correspondence, then we can easily find a utility function U for the planner that rationalizes it (set $U(a,\theta)=1$ if $a \in \phi(\theta)$, $U(a,\theta)=0$ otherwise.) Suppose we are given a (single valued) function F defined on the set of all subsets of Θ and taking values in A, and also we are given a prior probability distribution p. We show that there exists a utility function U for the planner such that (1) holds for the given priors if F satisfies: $a=F(T)=F(T')$ if and only if then $a=\cap_{t\in T}F(t)=\cap_{t'\in T'}F(t')$. This condition can clearly be satisfied by the construction in the proof of Theorem 3.

Proposition 2. *Let F be an extended social choice function and p_θ a prior probability distribution on Θ. Suppose for all sets $T,T'\subseteq\Theta$ and all $a\in A$, $a=F(T)=F(T')$ iff $a=\cap_{t\in T}F(t)=\cap_{t'\in T'}F(t')$. Then, there exists a utility function $U(a,\theta)$ such that for all $T\subseteq\Theta$:*

$$F(T) = \operatorname{argmax}\{ \sum_{\theta\in T} (p_\theta/p(T))\, U(a,\theta) : a\in A \} \qquad (1)$$

<u>Proof</u> We shall construct a utility function $U(a,\theta)$ such that (1) holds for all $T\subseteq\Theta$.

Consider any (x,θ). There are two possibilities.

(a) There exists $T\subseteq\Theta$ such that $x=F(T)$ and $\theta\in T$. Let k be the smallest number such that there exists T' such that $x=F(T')$ and $\theta\in T'$ and $|T'|=k$. Then $U(x,\theta)=|\Theta|-k$.

(b) Any other case: set $U(x,\theta)=-K$, where $K>0$ will be determined later.

We claim that for K sufficiently large, (1) holds for all $T\subseteq\Theta$. This is clearly true if T is a singleton. Consider next $T\subseteq\Theta$ such that $|T|>1$. Then, if K is sufficiently large, the only candidate for the planners utility maximizing choice conditional on the state being in T is x such that $x=F(T')$ for some T' such that $T\subseteq T'$. To show this, we use the following result.

Claim. Let x be such that there does not exist T' such that $T\subseteq T'$ and $x=F(T')$. Then, there is $\theta\in T$ such that $U(x,\theta)=-K$.

Proof of claim. By the construction of U we need to show: there is $\theta\in T$ such that there does not exist T' such that $x=F(T')$ and $\theta\in T'$.

In order to obtain a contradiction, suppose for each $\theta\in T$ there is $T_\theta\subseteq\Theta$

such that $x=F(T_\theta)$ and $\theta\in T_\theta$. But $x=F(T_\theta)=F(T_{\theta'})$ for $\theta,\theta'\in T$ (where $\theta\neq\theta'$) implies by the hypothesis of the proposition that $\cap_{t\in T_\theta}F(t)=\cap_{t'\in T_{\theta'}}F(t')=x$.

Therefore, $x=F(\theta)$ for all $\theta\in T^*\equiv\cup_{\theta\in T}T_\theta$ and by definition $T\subseteq T^*$. Let $\theta\in T$. Then $x=\cap_{t\in T}F(t)=F(\theta)$ which, by the hypothesis of the proposition implies that $x=F(\theta)=F(T)$. This contradicts the assumption that there is no T', $T\subseteq T'$, such that $x=F(T')$. This proves the claim.

Since $p_\theta>0$ for all θ, the claim implies that for K large enough, it is not optimal to choose x when the state is in T and $\theta\in T$ is a possibility. Thus, the only candidate for the planners utility maximizing choice conditional on the state being in T is x such that $x=F(T')$ for some T' such that $T\subseteq T'$. It remains to show that actually $x=F(T)$.

Suppose $T\subset T'$ and $F(T')=x'\neq x= F(T)$. We claim $U(x,\theta) > U(x',\theta)$ for all $\theta\in T$. Suppose this is not the case for some $\theta\in T$. Then by our construction there must exist T'' such that $|T''| \leq|T|$ and $x'=F(T'')$ and $\theta\in T''$. Thus, $x'=F(T')=F(T'')$ which implies $x'=\cap_{t'\in T'}F(t')=\cap_{t''\in T''}F(t'')$. But $T\subset T'$. Therefore, $x'=\cap_{t\in T}F(t)=\cap_{t''\in T''}F(t'')$. By the hypothesis of the proposition this implies $x'=F(T)=F(T'')$, contradicting $F(T)=x\neq x'=F(T'')$. This contradiction establishes: $U(x,\theta) > U(x',\theta)$ for all $\theta\in T$. Then $x=F(T)$ is indeed the planner's (unique) utility maximizing choice conditional on the state being in T. QED

REFERENCES

Baliga, S. (1994). "Universal Collusion and Renegotiation, Dictators and Contracts". Mimeo.

Baliga, S., Palfrey, T. and Sjöström, T. (1995). "Interactive Implementation". Work in progress.

Becker, G. (1974). "A Theory of Social Interactions", Journal of Political Economy 82: 1063-1094.

Chakravorty, B., Corchon, L., Wilkie S. (1992). "Credible Implementation", Games and Economic Behavior, forthcoming.

Laffont, J-J, Tirole, J. (1988). "The Dynamics of Incentive Contracts", Econometrica, 56, 1153-1175.

Maskin, E. (1977). "Nash equilibrium and Welfare Optimality". Mimeo, MIT.

Maskin, E., Moore, J. (1987). "Implementation with Renegotiation", Mimeo, Harvard University.

Maskin, E., Tirole, J. (1992). "The Principal-Agent Relationship with an Informed Principal II: Common Values", Econometrica, 60, 1-42.

Moore, J. (1992)."Implementation, Contracts and Renegotiation in Environments with Complete Information" in J-J Laffont, *Advances in Economic Theory Sixth World Congress*. Cambridge University Press.

Ray, D. (1993). "Egalitarianism and Incentives in Teams", Working Paper, Boston University.

Shenker, S. (1994). "Some Technical Results on Continuity, Strategy-proofness and Related Strategic Concepts", Mimeo.

Sjöström, T. (1994) " Renegotiation and Credibility in Implementation", Mimeo Harvard.

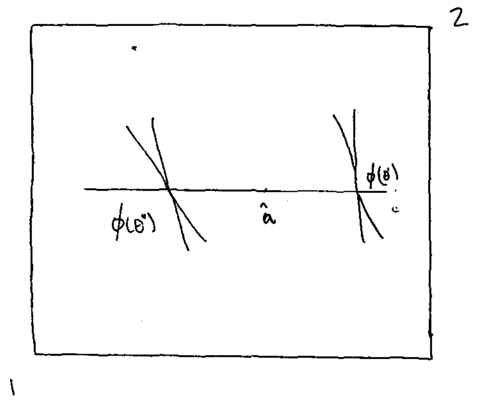

Figure 1

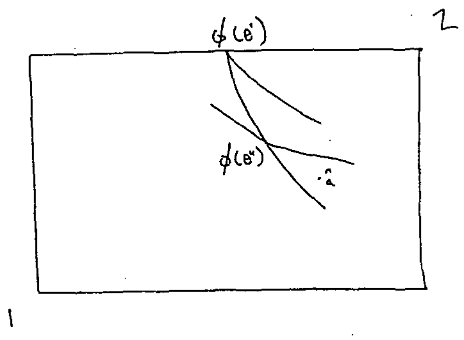

Figure 2

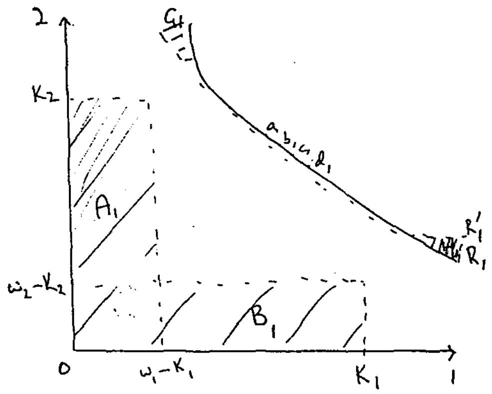

Figure 3

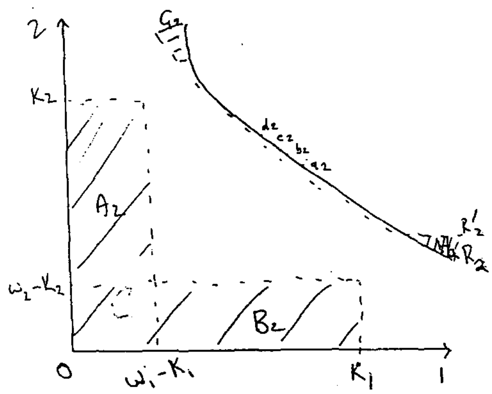

Figure 4