

Working Paper 97-84
Business Economics Series 15
November 1997

Departamento de Economía de la Empresa
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (341) 624-9608

UNIQUE IMPLEMENTATION OF ACTION PROFILES: NECESSARY AND SUFFICIENT
CONDITIONS

Sandro Brusco*

Abstract

We study the general problem of a principal who desires to implement a given vector of actions and pay the agents according to a given compensation scheme. Previous work has provided mechanisms for implementation in various special cases. In this paper we fully characterize the set of action profiles and compensation schemes implementable in subgame perfect equilibrium, providing necessary and sufficient conditions for implementation.

* Universidad Carlos III de Madrid, Departamento de Economía de la Empresa, C/ Madrid, 126, 28903 Getafe (Madrid) Spain. E-mail: brusco@emp.uc3m.es.

Unique Implementation of Action Profiles: Necessary and Sufficient Conditions

Sandro Brusco
Departamento de Economía de la Empresa
Universidad Carlos III de Madrid
Calle Madrid 126
28903 Getafe (Madrid)
SPAIN
E-mail: brusco@emp.uc3m.es

November 1997

Abstract

We study the general problem of a principal who desires to implement a given vector of actions and pay the agents according to a given compensation scheme. Previous work has provided mechanisms for implementation in various special cases. In this paper we fully characterize the set of action profiles and compensation schemes implementable in subgame perfect equilibrium, providing necessary and sufficient conditions for implementation.

1 Introduction

In the traditional implementation problem with complete information there is a set of possible states of the world S , with typical element s , a set of possible decisions D and a social choice function (or correspondence) $f : S \rightarrow D$ that the planner wants to implement. The state of the world s is observed by N agents, who have utilities defined over $S \times D$, but not by the planner. The problem is therefore to set up a mechanism making sure that the agents report truthfully the state of the world they observe. The presence of complete information makes incentive considerations trivial, since no agent has truly private information. It is therefore easy to set up a mechanism delivering truthtelling as an equilibrium (in fact, by the revelation principle the direct mechanism suffices). The difficult part is to make sure that truthtelling is the *unique* equilibrium, or more precisely to make sure that $f(s)$ is the unique equilibrium outcome when the state of the world is s .

Whether or not is actually possible to design a mechanism delivering unique implementation depends on the properties of f and on the equilibrium notion adopted to predict the outcome of the mechanism. The literature has provided complete characterizations of the set of implementable social choice functions for an array of solution concepts, such as Nash

equilibrium (Moore and Repullo (1990) and Dutta and Sen (1991)), subgame perfect equilibrium (Moore and Repullo (1988) and Abreu and Sen (1990)) and virtual implementation with iterative elimination of dominated strategies (Abreu and Matsushima (1992)).

The traditional problem described above can be seen as an ‘adverse selection’, or ‘hidden information’ problem. In this paper we look at the ‘moral hazard’ or ‘hidden action’ counterpart. The problem can be described as follows. Let I be a finite set of agents. Each agent i can take an action a_i in the finite set A_i . Let $A = \times_{i \in I} A_i$ with typical element $a = (a_1, \dots, a_I)$, and $A_{-i} = \times_{j \in I \setminus i} A_j$ with typical element $a_{-i} = (a_1, \dots, a_{i-1}, a_{i+1}, \dots, a_I)$. The action profile a determines a probability distribution over some real-valued variable (the typical example is the revenue of a firm) v . Let V^a be the finite set of values which can be taken by v when the true action profile is a , and let $\rho(\cdot|a)$ be the probability distribution of v when the true action profile is a . Define $V = \cup_{a \in A} V^a$.

An *individual compensation scheme* for agent i is a function $w_i : V \rightarrow \mathfrak{R}$, i.e. w_i specifies a monetary transfer for each possible realization of v . A *compensation scheme* w is just a collection of individual compensation schemes, i.e. $w = (w_1, \dots, w_I)$. Each agent has preferences over actions and individual compensation schemes represented by a utility function of the form:

$$U^i(a, w_i) = \sum_{v \in V} \rho(v|a) u^i[a, w_i(v)]$$

where the ‘basic’ utility function $u^i : A \times \mathfrak{R} \rightarrow \mathfrak{R}$, is defined over the whole action profile a and the monetary transfer, and the probability distribution $\rho(\cdot)$ is common to all players.

The goal of the planner is to make sure that each agent takes a given action a_i^* , so that the action profile $a^* = (a_1^*, \dots, a_I^*)$ is collectively taken, and reward each agent with compensation scheme w_i^* when a^* is the true action profile. Defining $w^* = (w_1^*, \dots, w_I^*)$, we will refer to this as ‘the problem of implementing the pair (a^*, w^*) ’. A well known example is the case of a risk neutral employer with risk averse agents who desires to implement the action profile a^* maximizing expected revenue and pay a constant wage to each agent¹ (i.e. for each i there is a real number w_i such that $w_i^*(v) = w_i$ for each v).

The ‘complete information’ assumption in this model takes the form that each agent can observe the whole action profile a . This again eliminates incentive considerations, since an agent can be forced to take a_i^* simply because any different action could be reported to the principal, and a stiff fine could follow. It leaves however open the multiplicity problem.

Such a problem was first considered by Demski and Sappington (1984) and Mookherjee (1984). They observed that standard ‘tournament’ models usually have multiple equilibria, and some of the ‘bad’ equilibria may be Pareto-superior from agents’ point of view. In particular, in ‘direct revelation’ types of mechanisms it is often an equilibrium for all the agents to take the lowest level of effort and report unanimously that the required effort was actually undertaken. Ma, Moore and Turnbull (1988) have shown how to get unique implementation in the Demski-Sappington model, and Ma (1988) has shown that unique implementation in subgame perfect equilibrium is achievable in the Mookherjee model. A number of recent contributions, among them Arya and Glover (1995), Sjöström (1996), Arya, Glover and Hughes (1997) and Brusco

¹See Ma (1988) for an analysis of this case.

(1997) have all proposed mechanism for implementation in some variant of the general model. The main goal of these papers is to provide mechanisms to solve a given implementation problem. Obviously, the simpler the mechanism the better it is. In other words, the above cited papers only look at sufficient conditions for implementation.

In this paper we take a different perspective, which is more related to the classical implementation perspective. The question we want to answer is: Can we identify necessary and sufficient conditions that a pair (a^*, w^*) has to satisfy in order to be implementable? We consider subgame perfect equilibrium as the solution concept, and provide such conditions. Since the goal is only to provide a full characterization of implementable pairs (a^*, w^*) , identifying exactly the class of implementable pairs, we do not try to use simple mechanisms in our proof. While the mechanism adopted for the general proof is complicated and incorporates the undesirable ‘integer game’ feature, it is clear that in many (not so special) cases simpler mechanism are available, as shown by the papers cited above.

The rest of the paper is organized as follows. In section 2 we provide a discussion of the problem of implementing action profiles and state the result about necessary conditions. In section 3 we show that, when there are at least three agents, the necessary conditions are also sufficient, and describe a general mechanism for implementation. Section 4 contains concluding remarks. The proofs are collected in the appendix.

2 Necessary Conditions

There are two main differences between the traditional implementation problem and the problem of implementing action profiles. First, agents will report on an endogenous variable (the action profile) rather than on some exogenously given state of the world.

Second, ‘action taking’ is given by some technological process and it cannot be manipulated by the principal. This imposes some physiological constraints on the form that the mechanism can take. In particular, any extensive game which is adopted to implement the pair (a^*, w^*) must be made up of 3 parts:

1. A ‘pre-action part’, where agents send messages. This has to be intended as an extensive form at which messages are exchanged, with the final vector of messages observed by the principal and the agents. Depending on the message profile, either the game ends and payments are determined, or the game proceeds to the action stage.
2. An ‘action part’ where each agent chooses an action, possibly as a function of the outcome of the pre-action stage. Simultaneously, the agent can also issue some message². The set of messages available to the agents may depend on the message profile announced at the pre-action stage. The action profile is observed by the agents but not by the principal, while the message profile issued at this stage is also observed by the principal.

²Here ‘simultaneously’ should be interpreted with respect to the information structure, that is messages are sent before actions are observed and actions are taken before messages are observed.

3. A ‘post-action part’ where messages are again issued. The particular extensive form used may depend on the messages sent at the pre-action stage and at the action stage. However, it cannot depend on the action profile a , since this is not observable by the principal.

The final vector of messages (i.e. messages issued at each stage) determines the compensation scheme selected by the principal, while the action profile implemented is simply the one chosen by the agents at the action stage.

The search for necessary conditions for implementation has to take into account this natural structure of any mechanism. In order to understand better what type of conditions we may need, it is useful to think about the necessary conditions in the ‘standard’ (i.e., hidden information) implementation problem. The basic idea is that if the social choice function prescribes two different allocations at state s and at state s' we want to avoid that the strategy profile adopted at s be an equilibrium when the state of the world is s' . When implementing action profiles, the main issue becomes to make sure that an action profile $a \neq a^*$ is not taken in equilibrium. In particular, we want to avoid equilibria in which $a \neq a^*$ is the action profile taken and the agents issue the same messages as in the equilibrium in which a^* is taken, thus leading to the selection of the compensation scheme w^* . In other words, whenever an action profile $a \neq a^*$ is going to be taken there must be a profitable deviation for some agent at some stage of the mechanism. As observed before, all mechanisms for implementation of action profiles can be partitioned in three parts, and we are going now to provide three conditions, each one corresponding to the case in which a profitable deviation can be found at a given part.

First, we have a condition making possible a deviation in the post-action part.

Condition 1 *An action profile a satisfies Condition 1 with respect to a^* if it is possible to find an agent $j(a, a^*)$ (shortly, j) and a pair of individual compensation schemes $w_j^x(a, a^*)$, $w_j^y(a, a^*)$ (shortly, w_j^x, w_j^y) such that:*

$$U^j(a^*, w_j^x) \geq U^j(a^*, w_j^y) \quad U^j(a, w_j^x) < U^j(a, w_j^y)$$

Condition 1 permits to ‘test’ independently a claim made by an agent that the true action profile is a rather than a^* . This structure is used for example by Ma (1988). Brusco (1997) discusses the relationship between condition 1 and Abreu and Sen’s condition α , a necessary condition for subgame perfect implementation in the ‘adverse selection’ framework.

Second, we have the condition making possible a deviation at the action part.

Condition 2 *An action profile a satisfies Condition 2 with respect to (a^*, w^*) if it is possible to find an agent $i(a, a^*)$ (shortly, i), an action $\tilde{a}_i(a, a^*)$ (shortly, \tilde{a}_i) and an individual compensation scheme $w_i^y(a, a^*)$ (shortly, w_i^y) such that:*

1. $U^i((\tilde{a}_i, a_{-i}), w_i^y) > U^i(a, w_i^*)$.

2. If there exists an action \hat{a}_i such that $U^i((\hat{a}_i, a_{-i}^*), w_i^y) > U^i(a^*, w_i^*)$ then (\hat{a}_i, a_{-i}^*) satisfies condition 1 with respect to (\tilde{a}_i, a_{-i}) .

The condition makes sure that an agent can deviate from an action profile $a \neq a^*$ taking a different action \hat{a}_i and possibly asking for a different compensation scheme (part 1). The same deviation cannot be profitable when the action profile is a^* (part 2).

Last, we have the condition making possible a deviation in the pre-action part.

Condition 3 An action profile a satisfies Condition 3 with respect to (a^*, w^*) if there exists an agent $i(a, a^*)$ (shortly, i), an action profile $\bar{a}(a, a^*)$ (shortly, \bar{a}) and a compensation scheme $w^y(a, a^*)$ (shortly, w^y) such that:

1. $U^i(\bar{a}, w_i^y) > U^i(a, w_i^*)$
2. $U^i(\bar{a}, w_i^y) \leq U^i(a^*, w_i^*)$
3. If an action profile \hat{a} is such that $U^i(\hat{a}, w_i^y) \leq U^i(a, w_i^*)$ then either \hat{a} satisfies Condition 1 with respect to \bar{a} or \hat{a} satisfies Condition 2 with respect to (\bar{a}, w^y) .
4. If there exists an action \hat{a}_i such that $U^i((\hat{a}_i, \bar{a}_{-i}), w_i^y) > U^i(\bar{a}, w_i^y)$ then $(\hat{a}_i, \bar{a}_{-i})$ satisfies condition 1 with respect to \bar{a} .

As we observed above, any mechanism must make sure that the message profile leading to the choice of w^* is not an equilibrium when agents plan to take action profile $a \neq a^*$. This means that any action profile $a \neq a^*$ must satisfy at least one of the conditions stated above. More formally, we have the following result.

Theorem 1 If the pair (a^*, w^*) is implementable, then each action profile $a \neq a^*$ satisfies at least one of conditions 1-3.

Proof. See Appendix. ■

The union of conditions 1-3 is therefore a necessary condition for implementation of a pair (a^*, w^*) . The role played by such condition is therefore analogous to the role played by Maskin monotonicity for Nash implementation or by condition α (see Abreu and Sen (1990)) for subgame perfect implementation. To complete the analogy, we have to show that the condition is also sufficient when there are at least three agents. This is done in the next section.

3 Sufficiency

We now show that when each action profile $a \neq a^*$ satisfies one of conditions 1,2 or 3 and there are at least three agents then we can actually implement the pair (a^*, w^*) .

This is done in the traditional way, building a mechanism having only the right outcome. In the following we will assume for simplicity that there is a sufficiently large amount of money K , and a sufficiently low amount $-\epsilon$, such that for each pair a, \bar{a} we have $U^i(a, K) > U^i(\bar{a}, w_i(v))$

and $U^i(a, -\epsilon) < U^i(\bar{a}, w_i(v))$ for each agent i , for each v and for each compensation scheme w entering in one of conditions 1-3. In other words, it is always possible to ‘reward enough’ and ‘punish enough’ an agent.

Theorem 2 *Suppose that each action profile $a \neq a^*$ satisfies one of conditions 1, 2 or 3. If there are at least 3 agents then the pair (a^*, w^*) can be implemented in subgame perfect equilibrium.*

In the remaining of this section we describe the mechanism for implementation. In the appendix we show that there exists a subgame perfect equilibrium yielding (a^*, w^*) and that no subgame perfect equilibrium has an outcome other than (a^*, w^*) .

Define A^1 as the set of action profiles that satisfy condition 1 with respect to a^* , A^2 the set of actions such that $a \notin A^1$ and Condition 2 is satisfied with respect to (a^*, w^*) , and A^3 the set of actions such that $a \notin A^1 \cup A^2$ and condition 3 is satisfied with respect to (a^*, w^*) . Since (a^*, w^*) is implementable, $A^1 \cup A^2 \cup A^3 = A \setminus a^*$.

The pre-action stage

Each agent announces an action profile $a \in A$ and an integer number. The outcome function is as follows.

- $N - 1$ agents announce $(a^*, 0)$, agent i announces (a, \cdot) . If $a \in A^3$ and $i = i(a, a^*)$ (i.e. i is the agent identified by condition 3) then go to the action stage, defining (\bar{a}, w^y) as the ‘standing message’, where $\bar{a} = \bar{a}(a, a^*)$ is the action profile identified by condition 3 and $w^y = w^y(a, a^*)$ is the compensation scheme identified by condition 3. Otherwise, go to the action stage, defining (a^*, w^*) as the ‘standing message’.
- In all other cases the agent announcing the highest integer receives a large sum K , all other agents receive $-\epsilon$ and the game ends. Ties are broken in favor of the lower index (this tie-breaking rule is used in all the following stages).

The action stage

Let (a, w) be the standing message. Each agent i takes an action and announces an action profile $\tilde{a} \in A$ and an integer number. The outcome function is as follows:

- $N - 1$ agents announce $(a, 0)$, the remaining agent announces (a, \cdot) . Go to the post-action stage with the standing message unchanged.
- $N - 1$ agents announce $(a, 0)$, agent i announces (\hat{a}, \cdot) . Go to the post-action stage defining the standing message as follows:
 - Suppose that \hat{a} satisfies condition 2 with respect to (a, w) and $i = i(a, \hat{a})$, i.e. i is the agent identified by the condition. In this case let $\tilde{a}_i = \tilde{a}_i(a, \hat{a})$ and $w_i^y = w_i^y(a, \hat{a})$ be the action and the individual compensation scheme identified by the condition. The standing message becomes $((\tilde{a}_i, a_{-i}), (w_i^y, w_{-i}))$.

- Otherwise the standing message remains (a, w) .
- In all other cases the agent announcing the highest integer receives a large sum K , all other agents receive $-\epsilon$ and the game ends.

The post-action stage

Each agent announces $\tilde{a} \in A$ and an integer number. The outcome function differs depending on the standing message.

Case 1. The standing message is (a^*, w^*) .

- $N - 1$ agents announce $(a, 0)$, the remaining agent i announces (\tilde{a}, \cdot) such that either \tilde{a} does not satisfy condition 1 with respect to a or $i = j(\tilde{a}, a)$, i.e. i is the agent identified by the condition. In this case:
 - If $a = a^*$ then pay the agents according to w^* .
 - If $a \neq a^*$ then all agents are paid $-\frac{3}{2}\epsilon$ except the agent who announced the highest integer at the action stage, who is paid $-\epsilon$.
- $N - 1$ agents announce $(\hat{a}, 0)$, the remaining agent i announces (\tilde{a}, \cdot) that satisfies condition 1 with respect to \hat{a} and such that $i \neq j(\hat{a}, \tilde{a})$. In this case go to the substage $\Gamma(\hat{a}, \tilde{a})$.
- In all other cases the agent announcing the highest integer receives a large sum K , all other agents receive $-\epsilon$ and the game ends

Case 2. The standing message is $(a, w) \neq (a^*, w^*)$.

- $N - 1$ agents announce $(\hat{a}, 0)$, the remaining agent i announces (\tilde{a}, \cdot) such that either \tilde{a} does not satisfy condition 1 with respect to \hat{a} or $i = j(\tilde{a}, \hat{a})$. In this case:
 - If \hat{a} does not satisfy condition 1 with respect to a then pay the agents according to w (here a and w are given by the standing message).
 - If \hat{a} satisfies condition 1 with respect to a then all agents are paid $-\frac{3}{2}\epsilon$ except the agent who announced the highest integer at the action stage, who is paid $-\epsilon$.
- $N - 1$ agents announce $(\hat{a}, 0)$, the remaining agent i announces (\tilde{a}, \cdot) that satisfies condition 1 with respect to \hat{a} and such that $i \neq j(\hat{a}, \tilde{a})$. In this case go to the substage $\Gamma(\hat{a}, \tilde{a})$.
- In all other cases the agent announcing the highest integer receives a large sum K , all other agents receive $-\epsilon$ and the game ends.

The substage $\Gamma(\hat{a}, \tilde{a})$. Let j be the agent and (w_j^x, w_j^y) the two compensation schemes identified by condition 1. Each agent announces either Y or N and an integer number. The outcome function is as follows.

- If $N - 1$ agents announce $(Y, 0)$ and agent j announces (Y, \cdot) then each agent is paid -2ϵ and j is paid w_j^x .
- If $N - 1$ agents announce $(Y, 0)$ and agent j announces (N, \cdot) then each agent is paid -2ϵ and j is paid w_j^y .
- If at least $N - 1$ agents announce $(N, 0)$ then agent i (the agent who ‘disagreed’ at the post-action stage) is paid a large sum K and all other agents are paid -2ϵ .
- In all other cases the agent announcing the highest integer receives a large sum K , all other agents receive -2ϵ and the game ends. Ties are broken in favor of the lower index.

The basic idea of the mechanism is to make sure that in all equilibria the agents repeatedly confirm that the action profile a^* is taken. The structure of the mechanism makes then sure that no action profile $a \neq a^*$ can be taken when such a sequence of messages is issued. This is done by exploiting conditions 1-3 to make sure that whenever $a \neq a^*$ is taken in equilibrium some agent has a profitable deviation at some point.

4 Conclusion

We have provided necessary and sufficient conditions for the implementation of an action profile a^* and a vector of compensation schemes w^* , thus obtaining a full characterization of the implementable pairs (a^*, w^*) . This constitutes the natural counterpart of the necessary and sufficient condition for implementation in the traditional (‘hidden information’) implementation framework. The ‘hidden action’ case presents some interesting differences with respect to the traditional case. We have shown that mechanisms implementing action profiles have a natural sequential structure centered on the ‘action taking’ stage, and that necessary and sufficient conditions can be obtained looking at some ‘reversal of preferences’ in different parts of the mechanism. Previous work has been more ‘ad hoc’ in the sense that special pairs (a^*, w^*) were considered and a mechanism for implementation was found. In general, these mechanisms rely on some reversal of preferences at what we have called ‘post-action stage’. We have shown that in fact implementation may be possible under more general conditions, involving reversal of preferences at the ‘action stage’ or at the ‘pre-action stage’.

Appendix

Proof of theorem 1.

Suppose that there exists an extensive form mechanism implementing (a^*, w^*) in subgame perfect equilibrium. A formal description of an extensive form mechanism is beyond the scope of this paper, and we simply refer to Selten (1975). Here we will introduce only the notation which is helpful in understanding the proof.

Let us define g as the outcome function of the extensive form game. Therefore, g maps each action profile and each sequence of messages observed into the same action profile and a vector of compensation schemes. We will denote by M the set of all possible sequences of messages that can be observed, M_1 the set of possible sequences of messages that can be observed before the action stage, M_a the set of possible messages that can be observed at the action stage and M_2 the set of possible sequences of messages that can be observed at the post-action stage. Thus $g(a, m) = (a, w)$, with $m \in M$. For a given action profile a and messages $m_1 \in M_1$, $m_a \in M_a$ we define $M_2(m_1, m_a)$ as the set of message profiles observable at the post-action stage and $g_2(m_2 | (m_1, m_a))$ as the outcome function for this subgame. Notice that the outcome function is independent of a since this is not observable by the planner. However, a is observed by the players, so that they can condition their strategies on a .

A strategy profile σ describes the behavior taken by the agents at each information set. A given σ yields a sequence of messages $m \in M$ and an action profile. In turn, the sequence of messages yields through g a compensation scheme.

Since the mechanism implements (a^*, w^*) , there exists a strategy profile σ^* yielding this outcome. Let us call m_1^* the vector of messages sent at the pre-action stage when agents follow σ^* . Since the strategy profile yields (a^*, w^*) we have that σ^* yields the pair (a^*, m_a^*) at the action stage (that is, following m_1^* the action profile a^* is taken and a message m_a^* is issued) and a message profile m_2^* at the post action stage when (m_1^*, a^*, m_a^*) have been observed. Then $g_2(m_2^* | (m_1^*, m_a^*)) = w^*$, that is when messages m_1^* , m_a^* and m_2^* are issued the selected compensation scheme is w^* .

Without loss of generality, assume that if at some subgame a unilateral deviation is observed then the worst possible subgame perfect equilibrium for the deviator is selected.

We now show that if there is an action profile $a \neq a^*$ that satisfies none of conditions 1,2 and 3 then there exists a subgame perfect equilibrium yielding (a, w^*) , a contradiction.

Consider such an action profile a and the following strategy profile:

- At the pre-action stage each agent follows strategy σ_1^{*i} .
- If m_1^* is observed after the pre-action stage then each agent takes action a_i (rather than a_i^*) and issues message m_a^{*i} (that is, the same message as in the truthtelling equilibrium). In all other cases behave as in the original equilibrium.
- At the post-action stage, use strategy $\sigma_2^i(m_1^*, a^*, m_a^*)$ if (m_1^*, m_a^*, a) has been observed. In all other cases, behave as in the original equilibrium. In other words, whenever all agents take the actions prescribed in action profile a and issue the message (m_1^*, m_a^{*i}) then behave as if the observed action profile were a^* . Otherwise, behave as in the original equilibrium.

This strategy profile results in the action profile a and a sequence of messages (m_1^*, m_a^*, m_2^*) . Since this is the message observed in the ‘right’ equilibrium, the compensation scheme w^* is selected. Therefore, the overall outcome is $(a, w^*) \neq (a^*, w^*)$, and the strategy profile cannot be a subgame perfect equilibrium. We now prove that if a does not satisfy 1,2 or 3 then this is in fact a subgame perfect equilibrium, a contradiction.

Suppose not. Then, at some subgame, a profitable deviation exists. Given the definition of the strategy profile we observe that:

- no deviation is possible in subgames following $m_1 \neq m_1^*$;
- no deviation is possible in subgames following (m_1^*, m_a) if there is an agent i such that $m_a^i \neq m_a^{i*}$;
- no deviation is possible after observing an action profile $\hat{a} \neq a$.

The reason is that in all those subgames exactly the original equilibrium strategy is followed. Deviations can therefore occur:

1. At the post-action stage in some subgame following (m_1^*, m_a^*, a) ;
2. At the action stage following a message m_1^* .
3. At the pre-action stage.

Suppose first that the deviation occurs at a subgame following the sequence of messages and actions (m_1^*, m_a^*, a) , and let us call j the agent with a profitable deviation. Suppose that by following strategy $\sigma_2^j(m_1^*, m_a^*, a^*)$ the outcome is w^x (notice that the subgame need not be reached in equilibrium), and that a deviation by agent i yields a compensation scheme w^y . Since the deviation is profitable we have $U^j(a, w_j^y) > U^j(a, w_j^x)$. On the other hand, the same deviation is not profitable in the original equilibrium, so that $U^j(a^*, w_j^x) \geq U^j(a^*, w_j^y)$. The implication is that condition 1 is satisfied. Since we assumed the contrary, no profitable deviation exists in subgames following (m_1^*, m_a^*, a) . We are therefore assured that the strategy profile is a subgame perfect equilibrium for each subgame at the post-action stage. More in general, observe that whenever action profile a does not satisfy condition 1 with respect to another action profile a' then the set of subgame perfect equilibrium following (m_1, a, m_a) is identical to the set of equilibria following (m_1, a', m_a) . We will use repeatedly this observation later.

Suppose next that a profitable deviation is available at the action stage. We show that in this case condition 2 must be satisfied, generating a contradiction. The existence of a profitable deviation implies that there exists an agent i , an action $\tilde{a}_i \neq a_i$ and a message \tilde{m}_a^i such that the agent is better off taking the action \tilde{a}_i , issuing the message \tilde{m}_a^i and then obtaining the resulting compensation scheme. Following a deviation \tilde{a}_i , the observed action profile is (\tilde{a}_i, a_{-i}) . Let $\tilde{m}_2 = \sigma_2(m_1^*, (\tilde{m}_a^i, m_a^{*-i}), (\tilde{a}_i, a_{-i}))$ and $w^y = g(m_1^*, (\tilde{m}_a^i, m_a^{*-i}), \tilde{m}_2)$. Then the profitability of the deviation implies:

$$U^i((\tilde{a}_i, a_{-i}), w_j^y) > U^i(a, w_i^*)$$

that is, part 1 of condition 2 is satisfied. To see that part 2 is also satisfied, suppose that there exists an action \hat{a}_i such that (\hat{a}_i, a_{-i}^*) does not satisfy condition 1 with respect to (\tilde{a}_i, a_{-i}) . The implication is that the set of subgame perfect equilibria following a given set of messages (m_1, m_a) must be the same. Now recall that w_i^y has been taken to be the worst subgame perfect equilibrium outcome for agent i . If $U^i((\hat{a}_i, a_{-i}^*), w_i^y) > U^i(a^*, w_i^*)$ then agent i has a profitable deviation in the right equilibrium, namely take action \hat{a}_i and issue message \bar{m}_a^i at the message stage. We conclude that for every action \hat{a}_i , either $U^i((\hat{a}_i, a_{-i}^*), w_i^y) \leq U^i(a^*, w_i^*)$ or (\hat{a}_i, a_{-i}^*) satisfies condition 1 with respect to (\tilde{a}_i, a_{-i}) .

At last, a profitable deviation may be available only at the pre-action stage. Again, we show that in this case condition 3 has to be satisfied, a contradiction.

Let \bar{m}_1 be the message resulting at the pre-action stage as a consequence of a profitable deviation by agent i (other agents may send messages other than in the right equilibrium because there may be substages at the pre-action stage, and agents may react to an out of equilibrium announcement). As a consequence of message \bar{m}_1 the action profile \bar{a} and the message profile \bar{m}_a results at the action stage, and the message $\bar{m}_2 = \sigma_2(\bar{m}_1, \bar{m}_a, \bar{a})$ is issued at the post-action stage. These are the same action profiles and messages sent in the right equilibrium. Define $w^y = g(\bar{m}_1, \bar{m}_a, \bar{m}_2)$ and observe that, since this is a profitable deviation, we have $U^i(\bar{a}, w_i^y) > U^i(a, w_i^*)$. Since the deviation is not profitable in the truthtelling equilibrium we have $U^i(\bar{a}, w_i^y) \leq U^i(a^*, w_i^*)$. This implies that parts 1 and 2 of condition 3 are satisfied. To show that part 3 is satisfied as well, consider an action profile \hat{a} such that $U^i(\hat{a}, w_i^y) \leq U^i(a, w_i^*)$. Remember that (\bar{a}, w_i^y) was the worst subgame perfect equilibrium outcome for agent i (the agent who first deviated at the pre-action stage). Since $U^i(\hat{a}, w_i^y) \leq U^i(a, w_i^*) < U^i(\bar{a}, w_i^y)$, the following cannot be a subgame perfect equilibrium for the game starting at the action stage after message \bar{m}_1 :

- At the action stage each agent takes action \hat{a}_i and issues message \bar{m}_a^i .
- After the action stage, if (\hat{a}, \bar{m}) is observed then follow strategy $\sigma_2(\bar{m}_1, \bar{a}, \bar{m}_a)$. Otherwise follow the original strategy of the equilibrium yielding (\bar{a}, w^y) .

If \hat{a} does not satisfy either condition 1 with respect to \bar{a} or condition 2 with respect to (\bar{a}, w^y) then this is in fact a subgame perfect equilibrium, a contradiction. The reasoning is analogous to the one used before. If \hat{a} does not satisfies condition 1 with respect to \bar{a} then no profitable deviation can exist at the post-action stage, as it would also be a profitable deviation in the original equilibrium. Thus, a profitable deviation must exist at the action stage. However, if \hat{a} does not satisfy condition 2 with respect to (\bar{a}, w^y) then any profitable deviation against \hat{a} is also a profitable deviation against \bar{a} .

Part 4 of condition 3 is satisfied because \bar{a} is a subgame perfect equilibrium. Observe that if $(\hat{a}_i, \bar{a}_{-i})$ does not satisfy condition 1 with respect to \bar{a} then, following an identical set of messages, the same set of subgame perfect equilibria follows. Also recall that w_i^y is the outcome of the worst possible subgame perfect equilibrium for agent i following \bar{a} . This also implies that it is the worst possible subgame perfect equilibrium outcome when the action profile is $(\hat{a}_i, \bar{a}_{-i})$, since the set of possible outcomes is the same. If $U^i((\hat{a}_i, \bar{a}_{-i}), w_i^y) > U^i(\bar{a}, w_i^y)$ than agent i has a profitable deviation against the equilibrium yielding (\bar{a}, w^y) , a contradiction. ■

Proof of theorem 2

The proof is divided into two lemmas. Lemma 1 shows that there is a subgame perfect equilibrium with outcome (a^*, w^*) . Lemma 2 shows that there is no subgame perfect equilibrium with outcome other than (a^*, w^*) .

Lemma 1 *There exists a subgame perfect equilibrium with outcome (a^*, w^*) .*

Proof. We show that the following strategy profile is a subgame perfect equilibrium yielding (a^*, w^*) .

PRE-ACTION STAGE: Each agent announces $(a^*, 0)$.

ACTION STAGE: When the standing message is (\hat{a}, \hat{w}) each agent i takes action \hat{a}_i and announces $(\hat{a}, 0)$.

POST-ACTION STAGE: If a is the true action profile observed then announce $(a, 0)$. If a game $\Gamma(\hat{a}, \tilde{a})$ is reached then announce $(Y, 0)$ if agent j weakly prefers w_j^x to w_j^y . Otherwise announce $(N, 0)$.

We now check, working backward, that this is a subgame perfect equilibrium.

At $\Gamma(\hat{a}, \tilde{a})$ the only agent who can change the outcome is j . It is clear that the prescribed strategy is optimal.

At the post-action stage an agent i can deviate announcing a false action profile \hat{a} that satisfies condition 1 with respect to the action profile and such that $j(a, \hat{a}) \neq i$. Since the other agents are announcing the true action profile, the equilibrium at $\Gamma(a, \hat{a})$ is that everybody announces $(Y, 0)$, making the deviant agent worse off.

Consider now the action stage. Suppose first that the standing message is (a^*, w^*) . Taking an action $a_i \neq a_i^*$ and announcing (a^*, \cdot) is clearly not profitable since, given the equilibrium at the post-action stage, it yields at most $-\epsilon$. Taking action a_i and issuing message (\hat{a}, \cdot) , yielding as standing message (\hat{a}, w^y) is not profitable either. This yields at most $-\epsilon$ if (a_i, a_{-i}^*) satisfies condition 1 with respect to \hat{a} , and a utility inferior to $U^i(a^*, w^*)$ otherwise (this is a consequence of part 2 of condition 2).

If the standing message is (\bar{a}, w^y) then the same reasoning applies. If $a_i \neq \bar{a}_i$ is taken and (\bar{a}, \cdot) is announced then the agent obtains at most $-\epsilon$ if (a_i, \bar{a}_{-i}) satisfies condition 1 with respect to \bar{a} . If the condition is not satisfied, then by part 4 of condition 3 the agent is not better off.

At last, no deviation at the pre-action stage is profitable, by part 2 of condition 3. ■

Lemma 2 *There is no subgame perfect equilibrium with an outcome other than (a, w^*) .*

Proof. All equilibria must have the following structure:

- All agents announce $(a^*, 0)$ at the pre-action stage.
- If (\hat{a}, \hat{w}) is the standing message at the action stage then all agents announce $(\hat{a}, 0)$.

- When the action profile is a has been taken then all agents announce unanimously an action profile \hat{a} at the post action stage such that \hat{a} does not satisfy condition 1 with respect to a .
- Whenever a given message is issued with positive probability at the action stage, it is confirmed at the post-action stage. If not, the integer game would be triggered at the action stage.
- If a substage $\Gamma(\hat{a}, \tilde{a})$ is reached and agent $j(\hat{a}, \tilde{a})$ strictly prefers $w_j^y(\hat{a}, \tilde{a})$ to $w_j^x(\hat{a}, \tilde{a})$ then the only equilibrium is that all agents announce $(N, 0)$.

The implication is that in all equilibria the action stage is reached with probability 1 with standing message (a^*, w^*) , the announcement is confirmed at the action stage and then again at the post action stage. We have only to show that, when this sequence of messages occurs, no action profile $a \neq a^*$ can be part of an equilibrium.

Suppose that there is an equilibrium in which at the action stage the agents take an action profile $a \neq a^*$. It can never be the case that action a satisfies condition 1 with respect to a^* . The reason is that at the post action stage one agent could announce the true action profile, thus reaching $\Gamma(\hat{a}, \tilde{a})$ and obtaining K .

If a satisfies condition 2 then agent i identified by the condition has a profitable deviation, namely take action \tilde{a}_i and announce (a, \cdot) . This implies that the standing message becomes $((\tilde{a}_i, a_{-i}), (w_i^y, w_{-i}^*))$. After such a message profile, the only equilibria at the post-action stage are that all agents announce an action profile $(\hat{a}, 0)$ which does not satisfy condition 1 with respect to (\tilde{a}_i, a_{-i}) , so that w_i^y is implemented and agent i is better off.

At last, suppose a satisfies condition 3. In this case the agent i identified by Condition 3 can profitably deviate at the pre-action stage announcing $(a, 0)$. Following such an announcement, in all subgame perfect equilibria following it must be the case that all agents announce unanimously $(\bar{a}, 0)$ at the action stage and confirm the message at the post-action stage, so that w^y is actually implemented. It remain to show that the action profile \hat{a} taken in equilibrium is such that $U^i(\hat{a}, w^y) > U^i(a, w^*)$. Suppose not, so that $U^i(\hat{a}, w^y) \leq U^i(a, w^*)$. Then, part 3 of condition 3 tells us that \hat{a} either satisfies condition 1 with respect to \bar{a} or it satisfies condition 2 with respect to (\bar{a}, w^y) . In the first case, an unanimous announcement of $(\bar{a}, 0)$ cannot be an equilibrium at the post-action stage. In the second case, it is immediate to see that there is an agent with a profitable deviation at the action stage. ■

References

- [1] Abreu, D. and A. Sen (1990) 'Subgame Perfect Implementation: A Necessary and Almost Sufficient Condition', *Journal of Economic Theory*, **50**: 285-299.
- [2] Abreu, D. and H. Matsushima (1992) 'Virtual Implementation in Iteratively Undominated Strategies: Complete Information', *Econometrica*, **60**: 993-1008.

- [3] Arya, A. and J. Glover (1997) 'A Simple Forecasting Mechanism for Moral Hazard Settings', *Journal of Economic Theory*, **66**: 507-521.
- [4] Arya, A., J. Glover and J. Hughes (1997) 'Implementing Coordinated Team Play', *Journal of Economic Theory*, **74**: 218-232.
- [5] Brusco S. (1977) 'Implementing Action Profiles with Sequential Mechanism', mimeo, Universidad Carlos III de Madrid.
- [6] Demski, J. and D. Sappington (1984) 'Optimal Incentive Contracts with Multiple Agents', *Journal of Economic Theory*, **33**: 152-171.
- [7] Dutta, B. and A. Sen (1991) 'A Necessary and Sufficient Condition for Two Person Nash Implementation', *Journal of Economic Theory*, **50**: 285-299.
- [8] Ma, C. (1988) 'Unique Implementation of Incentive Contracts with Many Agents', *Review of Economic Studies*, **55**: 555-572.
- [9] Ma, C. , J. Moore and S. Turnbull (1988) 'Stopping Agents from Cheating', *Journal of Economic Theory*, **46**: 355-372.
- [10] Mookherjee, D. (1984) 'Optimal Incentive Schemes with Many Agents', *Review of Economic Studies*, **51**: 433-446.
- [11] Moore, J. and R. Repullo (1988) 'Subgame Perfect Implementation', *Econometrica*, **56**: 1191-1220.
- [12] Moore, J. and R. Repullo (1990) 'Nash Implementation: A Full Characterization', *Econometrica*, **58**: 1083-1099.
- [13] Selten, R. (1975) 'Re-examination of the Perfectness Concept for Equilibrium Points in Extensive Games', *International Journal of Game Theory*, **9**: 1-12.
- [14] Sjöström, T. (1996) 'Implementation and Information in Teams', *Economic Design*, **1**: 327-341.