Working Paper 97-79
Business Economics Series 11
October 1997

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

PERFECT BAYSEIAN IMPLEMENTATION: ONE ROUND OF SIGNALING IS NOT ENOUGH

Sandro Brusco*

Abstract

We show that there exists a social choice function that cannot be implemented in perfect Bayesian equilibrium if the mechanism has an equilibrium with one round of signaling. The social choice function can however be implemented in perfect Bayesian equilibrium, obviously with an equilibrium reaching the second stage.

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

Perfect Bayesian Implementation: One Round of Signaling is not Enough

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

July 1997

Abstract

We show that there exists a social choice function that cannot be implemented in perfect Bayesian equilibrium if the mechanism has an equilibrium with one round of signalling. The social choice function can however be implemented in perfect Bayesian equilibrium, obviously with an equilibrium reaching the second stage.

1 Introduction

A recent paper by Bergin and Sen (1996) provides sufficient conditions for implementation with extensive form mechanisms when agents have incomplete information. They focus on mechanisms in which at least one of the equilibria yielding the desired social choice function never goes beyond the first stage (they call these 'equilibria with one round of signaling'). In an equilibrium with one round of signaling agents send their messages and then an allocation is selected as a function of the message profile; it is never the case that after a message profile observed with positive probability in equilibrium a new game is reached in which agents are required to issue more messages. To put it simply, when each type of each agent behaves in the way prescribed by the equilibrium the social choice function (scf) is directly implemented. Further stages of the mechanism are only useful to make sure that strategy profiles not yielding the desired scf are not perfect Bayesian equilibria. In other words, stages beyond the first are useful only because they provide profitable deviations (for some type of some agent) against undesired strategy profiles.

The idea of concentrating attention on mechanisms with one round of signaling in the incomplete information case has some intuitive appeal. To start with, restricting to this class of mechanisms simplifies the analysis and helps to find relatively simple sufficient conditions for implementation. Furthermore, the canonical mechanisms used for subgame perfect implementation in complete information environments actually have equilibria with one round of signaling (see Moore and Repullo [1988] and Abreu and Sen [1990]), so that for the complete information case the restriction implies no loss of generality.

Bergin and Sen leave open the question of whether the restriction implies a loss of generality for the incomplete information case, i.e. whether or not it is possible to find a social choice function which is implementable *via* an extensive form mechanism (and adopting some notion of sequential rationality) *but* only with equilibria not having one round of signaling, i.e. equilibria reaching stages beyond the first.

The purpose of this paper is to show that in the incomplete information case *it is not true* that the restriction to equilibria with one round of signaling is without loss of generality. To be more precise we show that when certain restrictions on the way beliefs can be updated are imposed[1], it is possible to find a social choice function which is implementable in perfect Bayesian equilibrium *only* via equilibria reaching the second stage with positive probability. This is done in two steps. First, we show that if there exists an equilibrium with one round of signaling yielding the desired scf, then there is also a non-optimal perfect Bayesian equilibrium. Second, we construct a mechanism implementing the social choice function in perfect Bayesian equilibrium. Obviously, the mechanism cannot have equilibria with one round of signaling.

Surprisingly enough, in the mechanism we propose for implementation the 'truthtelling' equilibrium has no pooling of types at the first stage. Full revelation occurs at the first stage, so that at the second stage there is complete information along the equilibrium path. The reason why additional messages are requested at the second stage is that in this way we can destroy deceptions yielding undesired functions by making sure that no equilibrium exists at that stage when the beliefs are the ones generated by the deception. The same trick does not work when the equilibrium is with one round of signaling because when the second stage is reached only by out-of-equilibrium moves there is much more freedom in assigning beliefs. On the contrary, when the second stage must be reached in equilibrium we are able to pin down the beliefs at that stage under the deception, and design the mechanism in such a way that no equilibrium exists when truthtelling was not adopted at the first stage.

Our result implies that when we look for necessary and sufficient conditions for perfect Bayesian implementation, we cannot ignore the possibility that the equilibrium implementing the scf needs to go beyond the first stage. It would be interesting, for example, to have a complete characterization of scf's implementable in perfect Bayesian equilibrium in economic environments, as we have for Bayesian implementation (see Jackson [1991]). When we do this, however, we have to consider the possibility that the equilibria used go beyond the first stage. Unfortunately this complicates the analysis, especially for necessary conditions. The analysis of mechanisms with one round of signaling can be justified on the ground of simplicity, but we should be aware that we are limiting the set of implementable social choice functions.

[1]Bergin and Sen do not state explicitly the restrictions on the way in which beliefs are formed. They simply take as given that beliefs must obey certain restrictions and, given the restrictions, they provide conditions for implementation in perfect Bayesian equilibrium. The restrictions we adopt in this paper are therefore compatible with their analysis.

The rest of the paper is organized as follows. Section 2 describes the example, detailing preferences for each state of the world, probability distribution, information structure, the scf to be implemented and the notion of equilibrium adopted to predict the outcome of the mechanism. Section 3 shows that, for this example, any mechanism having an equilibrium with one round of signaling yielding the desired scf must have another equilibrium not yielding the scf. In section 4 the mechanism for implementation is built. Concluding remarks are in section 5, and two appendices collect some definitions and proofs.

2 The Example

There are three agents 1, 2 and 3 and four possible allocations, $\{a, b, c, d\}$. Agent 1 can be of two types, t_a^1 and t_b^1. Agent 2 can be of three types, t_a^2, t_b^2 and t_c^2, and agent 3 can be of two types, t_a^3 and t_b^3. The types of agents 1 and 2 are drawn from independent probability distributions which satisfy $\Pr(t_a^1) = \frac{2}{3}$ and $\Pr(t_a^2) = \Pr(t_b^2) = \frac{1}{3}$. The type of agent 3 is correlated with the type of agent two. We have $\Pr\left(t_b^3 \,|t_a^2\right) = \Pr\left(t_b^3 \,|t_b^2\right) = \frac{1}{2}$ and $\Pr\left(t_b^3 \,|t_c^2\right) = 0$.

The information structure is the following:

- Agent 1 observes her type and whether or not agent 2's type is t_c^2.
- Agent 2 observes her type and furthermore, when her type is t_b^2 she also observes the type of agent 3.
- Agent 3 observes her type and whether or not agent 2 is type t_c^2.

The preferences of the agents depend on the type profile in the following way:

Agent 1

If the type of agent 3 is t_a^3 :

$$t_a^1 : \begin{cases} c \succ a \succ b \succ d & \text{if} \quad t_a^2 \\ c \succ b \succ a \succ d & \text{if} \quad t_b^2 \\ c \succ a \succ b \succ d & \text{if} \quad t_c^2 \end{cases}$$

$$t_b^1 : \begin{cases} c \succ b \succ a \succ d & \text{if} \quad t_a^2 \\ c \succ a \succ b \succ d & \text{if} \quad t_b^2 \\ c \succ b \succ a \succ d & \text{if} \quad t_c^2 \end{cases}$$

If the type of agent 3 is t_b^3 :

$c \succ a \sim b \sim d$

Agent 2

$$t_a^2, t_a^3 : \begin{cases} d \succ c \succ a \succ b & \text{if} \quad t_a^1 \\ d \succ c \succ b \succ a & \text{if} \quad t_b^1 \end{cases} \qquad t_a^2, t_b^3 : \begin{cases} d \succ c \succ b \succ a & \text{if} \quad t_a^1 \\ d \succ c \succ a \succ b & \text{if} \quad t_b^1 \end{cases}$$

$$t_b^2 : \begin{cases} d \succ c \succ b \succ a & \text{if} \quad t_a^1 \\ d \succ c \succ a \succ b & \text{if} \quad t_b^1 \end{cases}$$

$$t_c^2 : \begin{cases} d \succ c \succ b \succ a & \text{if} \quad t_a^1 \\ d \succ c \succ a \succ b & \text{if} \quad t_b^1 \end{cases}$$

Agent 3

If the type is t_a^3 then preferences are as follows:

$$t_a^2 : \begin{cases} d \succ c \succ a \succ b & \text{if } t_a^1 \\ d \succ b \succ a \succ c & \text{if } t_b^1 \end{cases}$$

$$t_b^2 : d \sim c \sim a \sim b$$

$$t_c^2 : \begin{cases} d \succ c \succ b \succ a & \text{if } t_a^1 \\ d \succ a \succ b \succ c & \text{if } t_b^1 \end{cases}$$

If the type is t_b^3 then preferences are:

$d \sim a \sim b \sim c$

For agent 1 we assume that when the type of agent 3 is t_a^3 then the allocation c always gives a utility of 10, the allocation ranked second gives a utility of 1, the allocation ranked third gives a utility of zero and allocation d always gives a utility of -1. If the type of agent 3 is t_b^3 then c is worth 10 and a,b,d are worth zero. For agent 2, we will assume that c gives a utility of 10, d gives a utility of 20, the next-to-last ranked allocation gives a utility of 1, and the allocation ranked last gives a utility of 0. For agent 3, when the type is t_a^3 and the type of agent 2 is t_a^2 or t_c^2 then d is worth 10, the second ranked allocation is worth 2, the third ranked allocation is worth 1 and the last allocation is worth 0. In all other cases each outcome is worth 0.

At last, preferences satisfy the Von Neumann-Morgenstern assumptions. For reader's convenience, the numerical values of the utility functions are summarized in Appendix 1.

The social choice function to be implemented only depends on the types of agents 1 and 2, and it is:

State	Allocation
t_a^1, t_a^2	a
t_a^1, t_b^2	b
t_a^1, t_c^2	a
t_b^1, t_a^2	b
t_b^1, t_b^2	a
t_b^1, t_c^2	b

We will denote by f the scf to be implemented. In our analysis of the implementation problem for this social choice function we restrict our attention to 'stage games', or games with observable actions. Roughly this means that whenever an agent has to move she has perfect information about the moves chosen by other agents at previous stages (see Fudenberg and Tirole (1991), chapter 8, for a formal definition). This simplifies the analysis, since at each information set the relevant uncertainty is only about agents' types, and not about previous actions. Furthermore, we will assume that the equilibrium with one round of signaling is in pure strategies, i.e. at each stage each type takes a given action with probability 1. Bergin and Sen use stage games in their paper, and the equilibrium with one round of signaling they propose is in pure strategies.

We now state informally the restrictions on the way beliefs are formed[2]:

[2]There is no difficulty in formalizing the requirements, but this would need the introduction of lengthy

1. The Bayes' rule is adopted on the equilibrium path.

2. If at stage k agent 1 takes an out of equilibrium action while 2 and 3 takes actions prescribed with positive probability by the equilibrium then the beliefs on agents 2 and 3 at stage $k+1$ are updated according to the Bayes' rule. Analogously, if either 2 or 3 deviate but 1 doesn't then the beliefs on 1 are updated according to the Bayes' rule.

3. If at some stage a type is assigned probability zero then it will have probability zero in every subsequent stage.

The first requirement does not need any comment. The second is a simplified version of the 'no signaling what you don't know condition' (see Fudenberg and Tirole (1991)). The last requirement is more controversial. It has been added not because we think it is a particularly good assumption to impose[3] but simply because it allows us to simplify the analysis. The point of the exercise is to show that for some version of perfect Bayesian equilibrium we can find a social choice function requiring mechanisms with equilibria going beyond the first stage. We do not claim that the version of equilibrium we use is particularly worth of attention, nor we claim that the scf we consider is especially interesting. However, it is important to observe that our example would still be valid under a much weaker requirement than the one described at point 3. The matter is discussed in a remark at the end of section 4.

In the rest of the paper we will look at implementation in perfect Bayesian equilibrium, meaning that at each stage the strategy of each agent has to be optimal given the beliefs and that beliefs are formed obeying the above restrictions.

3 One Round Is Not Enough

In this section we show that whenever a multi-stage mechanism has a perfect Bayesian equilibrium with one round of signaling yielding the desired scf then there must exist another perfect Bayesian equilibrium yielding an undesired scf.

Proposition 1 *If there exists a perfect Bayesian equilibrium with one round of signaling yielding the desired scf, then there exists another perfect Bayesian equilibrium yielding an undesired scf.*

Proof. See Appendix 2. □

The formal proof of proposition 1 is pretty long and involved. We give here a non-rigorous and intuitive account.

---

definitions and new notation.

[3]The condition that beliefs should have a non-expanding support is used, for example, in the definition of perfect sequential equilibrium, see Grossman and Perry (1986). The reader should be warned that this kind of 'support restriction' assumptions may create existence problems and may rule out sensible equilibria. See Madrigal, Tan and Werlang (1987).

The proposed information structure has two common knowledge components: One in which everybody knows that the type of agent 2 is t_c^2 (we call this the t_c^2 component) and the other in which everybody knows that the type of agent 2 is not t_c^2.

In the equilibrium implementing the scf, agents will take different strategies in the two common knowledge components. We show that, whenever there is an equilibrium implementing the scf with one round of signaling (truthtelling equilibrium), there must be another PBE in which all agents follow in both components the same strategy adopted in the t_c^2-component of the truthtelling equilibrium. Thus, in the deception agent 2 always behaves as type t_c^2 did in the truthtelling equilibrium, and agents 1,3 always behave as they do in the truthtelling equilibrium when t_c^2 has been observed.

To make sure that this is an equilibrium we have to select carefully beliefs after each deviation. Beliefs must be such that at each subform the preferences of each agent are such that the original strategy adopted in the truthtelling equilibrium for the t_c^2 component is optimal in both components. This can be done because stages beyond the first are only reached by out-of-equilibrium moves, so that we have considerable freedom in assigning beliefs. This is also the reason why the scf can be implemented if we avoid equilibria with one round of signaling, and instead design the mechanism in such a way that in the truthtelling equilibrium the second stage is reached. In this case we have information revelation at the first stage, and at the second stage beliefs are fully determined. The construction of the alternative equilibrium described above required freedom in assigning beliefs. This freedom is denied when the truthtelling equilibrium reaches the second stage.

4 The Mechanism

We next show that it is possible to build a mechanism implementing the social choice function in perfect Bayesian equilibrium. Clearly, the equilibrium implementing the function cannot be in one round of signaling.

Stage 1. Agents 1 and 2 are asked to announce the types observed, a social choice function and an integer number. Agent 3 is asked to announce the type of agent 2, whether she wants the constant function c be implemented and an integer number. Thus the message spaces are:

- $M^1 = \{t_a^1, t_b^1\} \times \{\{t_a^2, t_b^2\}, t_c^2\} \times N.$
- $M^2 = \{t_a^2, t_b^2, t_c^2\} \times N.$
- $M^3 = \{\{t_a^2, t_b^2\}, t_c^2\} \times \{\emptyset, c\} \times N.$

Notice that the type of agent 3 is irrelevant to the scf to be implemented and it is not reported. Let $M = M^1 \times M^2 \times M^3$, with m as typical element. We will denote a message by agent 1 as $m^1 = (s^1, z, n^1)$, a message by agent 2 as $m^2 = (s^2, n^2)$ and a message by agent 3 as $m^3 = (w, g, n^3)$. Let us define the following sets:

$$A = \left\{m \,\middle|\, s^2 \in w,\ s^2 \in z\right\}$$

$$B = \left\{ m \,\middle|\, m^3 = (\cdot, \emptyset, \cdot) \right\}$$

The set A is the set of messages such that the reports on types agree. The set B is the set of messages such that agent 3 does not ask for c to be implemented.

The message space is partitioned as follows:

- $d_0^a = \{m \,|m \in A \cap B,\ f(s^1, s^2) = a \text{ and } n^i = 0 \text{ for at least two agents}\}$
- $d_0^b = \{m \,|m \in A \cap B,\ f(s^1, s^2) = b \text{ and } n^i = 0 \text{ for at least two agents}\}$
- $d_1 = \{m|m \in A,\ m \notin B, \text{ and } n^1 = n^2 = 0\}$
- $d_2 = \{m|m \notin A \text{ and } w = z \text{ and } n^1 = n^3 = 0\}$
- $d_3 = \{m|m \notin A \text{ and } s^2 \in w \text{ and } n^2 = n^3 = 0\}$
- $d_4 = \{m|m \notin A \text{ and } s^2 \in z \text{ and } n^1 = n^2 = 0\}$
- $d_5 = \left\{m|m \notin d_0^a \cup d_0^b \cup d_1 \cup d_2 \cup d_3 \cup d_4\right\}$

The outcome function is as follows:

- If $m \in d_0^a$ go to Stage 2, game Γ_a.
- If $m \in d_0^b$ go to Stage 2, game Γ_b.
- If $m \in d_1$ then c is implemented when $s^2 \neq t_c^2$. Otherwise, $f(s^1, s^2)$ is implemented.
- If $m \in d_2$ then $f\left(s^1, w\right)$ is implemented if $w = t_c^2$. Otherwise a is implemented.
- If $m \in d_3 \cup d_4$ then $f\left(s^1, s^2\right)$ is implemented.
- If $m \in d_5$ then the agent announcing the highest integer selects the outcome, breaking ties in favor of lower indices.

Stage 2. Depending on the message at the first stage either Γ_a or Γ_b is operated. We first describe Γ_a.

Game Γ_a : Agent 1 announces an element of $\{a, b, Next\} \times N$. Agent 2 announces an element of N. Agent 3 announces an element of N. Let us call q^2 the integer announced by agent 2 and q^3 the integer announced by agent 3. The outcome function is as follows:

- If agent 1 chooses a then implement a.
- If agent 1 chooses b and $q^2 = q^3 = 0$ then implement b.
- If agent 1 chooses $Next$ and $q^2 = q^3 = 0$ then go to the third stage.
- In all other cases the agent announcing the highest integer implements her choice, breaking ties in favor of the agent with the lower index.

Game Γ_b is the same as Γ_a, except that the roles of a and b are exchanged. For completeness, here is the description of Γ_b.

Game Γ_b: Agent 1 announces an element of $\{a, b, Next\} \times N$. Agent 2 announces an element of N. Agent 3 announces an element of N. The outcome function is:

- If agent 1 chooses b then implement b.
- If agent 1 chooses a and $q^2 = q^3 = 0$ then implement a.
- If agent 1 chooses $Next$ and $q^2 = q^3 = 0$ then go to the third stage.
- In all other cases the integer game is operated.

Stage 3. Agent 2 announces an element of $\{a, b\} \times N$ and agent 3 announces an element of $\{a, b\} \times N$. The choice of the agent announcing the highest integer is implemented, breaking ties in favor of agent 2.

We first prove that this mechanism has a perfect Bayesian equilibrium yielding the desired scf.

Proposition 2 *There exists a perfect Bayesian equilibrium yielding the desired social choice function.*

Proof. We describe the equilibrium.

- *First Stage.* Each agent tells the truth about the types she observes and announces $n^i = 0$. Agent 3 announces $g = \emptyset$.
- *Second Stage.* Beliefs are as follows:
  - Beliefs on agent 1 put probability 1 on the type announced at Stage 1.
  - For agent 2, if t_c^2 has been observed then there is no need to define beliefs. Suppose agents 1 and 3 observe $\{t_a^2, t_b^2\}$. Then beliefs put probability 1 on the type announced by 2, unless t_c^2 is announced. In this case the belief is equal to the prior.
  - Beliefs are identical to the prior for agent 3.

Strategies are the following:

- Γ_a: Agent 1 picks $(a, 0)$ if she told the truth at Stage 1 and $(b, 0)$ otherwise. Agent 2 announces $q^2 = 0$ and agent 3 announces $q^3 = 0$.
- Γ_b: Agent 1 picks $(b, 0)$ if she told the truth at Stage 1 and $(a, 0)$ otherwise. Agent 2 announces $q^2 = 0$ and agent 3 announces $q^3 = 0$.

- *Third Stage.* We first describe beliefs. Stage 3 can only be reached if agents 2 and 3 play according to the equilibrium strategy at stage 2, and the strategy does not depend on types. Thus the beliefs on 2 and 3 must be identical to the one held at Stage 2. Agent 1 is the deviator, and we assume beliefs are unchanged. Strategies at Stage 3 are the following:
  - Suppose t_c^2 has been observed. If t_a^1 has been announced at Stage 1 then agents 2 and 3 both announce $(b, 0)$. If t_b^1 has been announced at Stage 1 then agents 2 and 3 both announce $(a, 0)$.
  - Suppose agent 3 has observed $\{t_a^2, t_b^2\}$. Then either the belief on agent 2 is $\Pr(t_a^2) = 1$ or $\Pr(t_b^2) = 1$. If t_a^1 has been announced at Stage 1 then:
    * If $\Pr(t_a^2) = 1$ is the belief on agent 2 then both types of agent 3 announce $(a, 0)$. Type t_a^2 of agent 2 announces $(a, 0)$ and type t_b^2 announces $(b, 0)$.
    * If $\Pr(t_b^2) = 1$ is the belief on agent 2 then both types of agent 3 announce $(b, 0)$. Type t_a^2 of agent 2 announces $(a, 0)$ and type t_b^2 announces $(b, 0)$.

    If t_b^1 has been announced at Stage 1 then:
    * If $\Pr(t_a^2) = 1$ is the belief on agent 2 then both types of agent 3 announce $(b, 0)$. Type t_a^2 of agent 2 announces $(b, 0)$ and type t_b^2 announces $(a, 0)$.
    * If $\Pr(t_b^2) = 1$ is the belief on agent 2 then both types of agent 3 announce $(a, 0)$. Agent 2 always announces $(a, 0)$.

It is obvious that beliefs are consistent, so we only need to consider optimality of the strategies.

At stage 3, each type of agent 3 is always obtaining the preferred outcome in the set $\{a, b\}$, given her beliefs. The strategy is therefore optimal. Type t_a^2 is indifferent between the constant (with respect to the type of agent 3) function a and the constant function b. Constant functions are the only feasible outcomes given the strategy of agent 3, so any strategy is optimal. Type t_b^2 always obtains the preferred outcome in the set $\{a, b\}$, given her beliefs on agent 1.

At stage 2, deviations by agent 2 or agent 3 can change the outcome only if agent 1 has not told the truth at Stage 1, a zero-probability event. As for agent 1, the strategy described is clearly optimal.

At stage 1, agent 1 can change the outcome only if she lies about her type, which by incentive compatibility is not profitable. Agent 2 can only place the message in d_2. This does not change the outcome if her true type is t_c^2, and yields the constant function a otherwise. For type t_b^2 this is a strictly worse outcome, and type t_a^2 is indifferent between such an outcome and the scf. We conclude that at Stage 1 no profitable deviations exists for agent 2. At last, consider agent 3. Type t_b^3 is completely indifferent among all outcomes. Type t_a^3 cannot change the outcome when t_c^2 is the true type of agent 2. If $\{t_a^2, t_b^2\}$ is observed then agent 3 can obtain that the constant function c be implemented. This gives the same expected utility as the equilibrium strategy, and it is therefore not a profitable deviation. $\square$

We now show that there are no equilibria with a different outcome.

Proposition 3 *There is no equilibrium in which the message is not in set $d_0^a \cup d_0^b$ at the first stage, and there is no equilibrium in which agent 1 chooses Next or b with positive probability at game Γ_a, and Next or a at game Γ_b.*

Proof. If the choice by 1 is different from a with positive probability at Γ_a then agent 2 can obtain d with positive probability by announcing a sufficiently high number. The best response for 1 is to announce an higher number whenever a is not selected in order to obtain c. Thus the integer game is triggered, and no equilibrium exists. The reasoning is analogous for Γ_b.

Consider now the first stage. If the outcome is in $d_1 \cup d_3 \cup d_4 \cup d_5$ with positive probability then 2 can announce n^2 sufficiently high and obtain d with positive probability. But in this case agent 1 will announce an higher integer to get c. If the outcome is in d_2 then 1 can announce $n^1 > 0$ and obtain c, causing 2 to announce an higher integer to obtain d. No equilibrium exists. □

Proposition 4 *There is no equilibrium in which the correct outcome is not implemented when agent 2 is of type t_c^2.*

Proof. From proposition 3 we know that in equilibrium the second stage is reached with probability 1 and agent 1 chooses the same allocation prescribed by the type profile announced at the first stage. If t_c^2 is the type of agent 2 then agent 1 knows exactly the payoff-relevant type profile. Thus, she will pick a when her type is t_a^1 and are b when her type is t_b^1 and the correct outcome is chosen in any equilibrium. □

Proposition 5 *There is no equilibrium in which the type of agent 2 is in the set $\{t_a^2, t_b^2\}$ and the types of agent 1 do not fully separate at Stage 1.*

Proof. If the set $\{t_a^2, t_b^2\}$ has been observed, agent 3 only cares about what happens when t_a^2 is the true type. By proposition 3, equilibrium outcomes can only be a or b. Let now p_a be the probability that the outcome is a when the type of agent 1 t_a^1 and the type of agent 2 is t_a^2, and p_b the probability that the outcome is b when the type of agent 1 is t_b^1 and the type of agent 2 is t_a^2. Then the expected utility in equilibrium for type t_a^3 of agent 3 when agent 2 is of type t_a^2 is:

$$U = \frac{2}{3}p_a + \frac{1}{3}\left(2p_b + (1 - p_b)\right) = \frac{1 + p_b + 2p_a}{3}$$

Now observe that by asking that the constant function c be implemented agent 3 can obtain an expected utility of $\frac{4}{3}$. Therefore, for this to be an equilibrium it must be the case that:

$$1 + p_b + 2p_a \geq 4$$

This is only possible if $p_a = p_b = 1$, i.e. type t_a^1 always reaches Γ_a when t_a^2 is the true type and type t_b^1 always reaches Γ_b when t_a^2 is the true type. Therefore, full separation occurs. □

Proposition 6 *There is no equilibrium in which agent 2 is in the set $\{t_a^2, t_b^2\}$ and the two types of agent 2 do not fully separate at Stage 1.*

Proof. We now know that 1 is separating. We first prove that it cannot be the case that the two types of agent 2 send the same message with probability 1, by showing that if this is the case then agent 1 will choose *Next* at stage 2. This can never happen in equilibrium.

To see this, assume that agent 1 sends the message signaling that her type is t_a^1, and assume that Γ_a is reached in this case (the argument is analogous if Γ_b is reached). Then, for this to be an equilibrium, it must be a best response for agent 1 to choose a. What happens if 1 chooses *Next*? The belief on agent 1 at stage 3 must be $\Pr(t_a^1) = 1$ by the assumption of non-expanding support. Furthermore, in any equilibrium all types of agents 2 and 3 must announce 0. Since stage 3 is reached only if 2 and 3 take the equilibrium action, beliefs on 2 and 3 must be the same as at stage 2. Since 2 is pooling we must have $\Pr(t_a^2) = \frac{1}{2}$. Agent 3 must pool in every equilibrium, so we also have $\Pr(t_a^3) = \frac{1}{2}$.

The equilibrium outcome at stage 3 must be that a is selected when the type of agent 2 is t_a^2 and b is selected when the type of agent 2 is t_b^2. This follows from the fact that it cannot be the case that a function constant with respect to the types of agent 2 is the equilibrium outcome, since agent 3 would prefer a and type t_b^2 would want b. Thus the outcome must be a function that selects a when the type of agent 2 is t_a^2 and b when the type is t_b^2. It is immediate to see that this is better for agent 1 than choosing a at Γ_a.

We have therefore proved that there is no equilibrium in which t_a^2 and t_b^2 completely pool. Notice that this implies that when the type of agent 2 is in $\{t_a^1, t_a^2\}$ it can never be the case that agent 2 claims to be t_c^2. If this happened it would be necessary for agents 1 and 3 to report t_c^2 as well, in order to make sure that the outcome is in $d_0^a \cup d_0^b$. But this would imply complete pooling of t_a^2 and t_b^2.

We now show that no partial pooling occur. First observe that when the type of agent 2 is t_a^2 then it must be the case that a is obtained whenever the type of agent 1 is t_a^1 and b is obtained whenever the type of agent 1 is t_b^1. Otherwise, it would be a profitable deviation for agent 3 to ask that the constant function c be implemented (remember that the set $\{t_a^2, t_b^2\}$ is reported with probability 1). On the other hand, since agent 1 is separating, it must be the case that type t_b^2 can obtain that b be implemented when the type of agent 1 is t_a^1 and a be implemented when the type of agent 1 is t_b^1. This implies that the two types of agent 2 fully separate. □

Proposition 4 implies that the correct outcome is obtained whenever the type of agent 2 is t_c^2, and propositions 5 and 6 imply that the correct outcome is obtained when the type of agent 2 is in the set $\{t_a^2, t_b^2\}$. This concludes the proof. □

Remark. The only point of the proof in which we use the 'no expanding support' restriction on beliefs is in proposition 6. The assumption ensures that, if the belief at Stage 2 is $\Pr(t_a^1) = 1$ and agent 1 takes action *Next*, then at stage 3 beliefs will still be $\Pr(t_a^1) = 1$. However, a much weaker assumption would suffice.

In general, the following requirement would be sufficient: Suppose that a given stage k beliefs put probability 1 on a type s^i of player i, and suppose that stage $k+1$ is reached because of a deviation by agent i. Then the other players should revise their beliefs on agent i only if it is not possible to find a perfect Bayesian equilibrium at Stage $k+1$ such that type s^i of agent i is strictly better off. In other words, when probability 1 is put on s^i then beliefs change only if it is impossible to find a justification for the deviation by type s^i. This is clearly a much weaker restriction than 'no expanding support'.

To see that the requirement would be sufficient, let's go back to the proof of proposition 6. Suppose that types t_a^2, t_b^2 pool at Stage 1 (e.g. they both adopt the strategy intended for type t_c^2, as it happens in the deception equilibrium of the previous section) and that we end up in Γ_a with $\Pr\left(t_a^1\right) = 1$. For this to be an equilibrium, agent 1 must choose that a be implemented. A deviation to *Next* would be profitable for agent 1 if the belief at Stage 3 following the deviation were $\Pr\left(t_a^1\right) > .5$. Thus, the behavior of agent 1 can be 'justified', and agents 2,3 do not need to revise their beliefs. In fact, we can allow 'minor' revisions leading to $\Pr\left(t_a^1\right) > .5$.

5 Conclusion

It is instructive to see why the deception described in section 3 is not a perfect Bayesian equilibrium of the mechanism just described. The deception was, roughly: agent 1 tells the truth about her type and claims that 2 is t_c^2; agents 2 and 3 confirm the claim that t_c^2 is the true type of agent 2. If any agent deviates, beliefs are chosen so that preferences about constant functions are such that behaving as if t_c^2 had actually been observed is optimal at further stages.

In our mechanism, we show that the deception can be broken in proposition 6, first part. The crucial step is that in our mechanism at the second stage the beliefs under the deception are exactly defined, and there is no freedom in selecting them. Under the beliefs that should be held in the deception no equilibrium exists, since agent 1 has a profitable deviation from the only action that can possibly be part of an equilibrium.

On the other hand, if the equilibrium had only one round of signaling the second stage would only be reached out of equilibrium, and beliefs could be assigned to make sure that an equilibrium exists (and the deviation is not profitable).

Appendix 1

Summary of agents' preferences.

Agent 1 If the type of agent 3 is t_a^3:

	a	b	c	d
t_a^1, t_a^2	1	0	10	-1
t_a^1, t_b^2	0	1	10	-1
t_a^1, t_c^2	1	0	10	-1
t_b^1, t_a^2	0	1	10	-1
t_b^1, t_b^2	1	0	10	-1
t_b^1, t_c^2	0	1	10	-1

If the type of agent 3 is t_b^3:

	a	b	c	d
	0	0	10	0

Agent 2 If the type is t_a^2:

	a	b	c	d
t_a^1, t_a^3	1	0	10	20
t_a^1, t_b^3	0	1	10	20
t_b^1, t_a^3	0	1	10	20
t_b^1, t_b^3	1	0	10	20

If the type of agent 2 is either t_b^2 or t_c^2 then preferences only depend on the type of agent 1. We have:

	a	b	c	d
t_a^1	0	1	10	20
t_b^1	1	0	10	20

Agent 3 If the type is t_a^3 and the type of agent 2 is not t_b^2:

	a	b	c	d
t_a^1, t_a^2	1	0	2	10
t_a^1, t_c^2	0	1	2	10
t_b^1, t_a^2	1	2	0	10
t_b^1, t_c^2	2	1	0	10

If either the type is t_b^3 or the type of agent 2 is t_b^2 then:

	a	b	c	d
	0	0	0	10

Appendix 2

Proof of proposition 1.

Let G be a multistage mechanism, and assume that it has an equilibrium with round of signaling (we will call this 'the original truthtelling equilibrium'). For agent 1, let m^1_{hz} be the action taken at the first stage in the truthtelling equilibrium when she is of type t^1_h and it has been observed that the type of agent 2 is in the set z, with $h = a, b$ and $z \in \{\{a, b\}, \{c\}\}$. For agent 2, let m^2_{jk} be the message sent at the first stage when agent 2's type is t^2_j, with $j = a, b, c$ and the type of agent 3 is t^3_k, with $k \in \{a, b\}$ (notice that the type of agent 3 is reported only if $j = b$). At last, for agent 3 let m^3_{xy} be the message sent when agent 3's type is t^3_x and it has been observed that the type of agent 2 is in the set $y \in \{\{a, b\}, \{c\}\}$.

Before discussing the deception equilibrium, let us observe that it can never be the case that in the subform following a deviation by agent 1 the outcome c is obtained, since in this case agent 1 would have a profitable deviation (notice that agent 1 strictly prefers any social choice function in which c is obtained at some state of the world to the social choice function to be implemented). Analogously, agent 2 can never obtain c or d through a deviation.

We now describe the deception equilibrium, specifying the strategy profiles and beliefs for each history.

First stage

Agent 1: Type t^1_a of agent 1 sends message m^1_{ac} with probability 1. Type t^1_b sends message m^1_{bc} with probability 1. In other words, agent 1 always claims that the type of agent 2 is t^2_c and tells the truth about her type.

Agent 2 All types send message m^2_c with probability 1. In other words, agent 2 always claims that her type is t^2_c.

Agent 3 Message m^3_{ac} is sent with probability 1, i.e. agent 3 always claims that her type is t^3_a and that the type of agent 2 is t^2_c.

Further stages.

Notice that, since the truthtelling equilibrium has one round of signaling, whenever $\hat{m} = (m^1_{hz}, m^2_{kj}, m^3_{jz})$ with $k \in z$ (i.e. the reports about agents' types agree) then the game ends and the allocation $f(t^1_h, t^2_k)$ is implemented. We now deal with the other cases, describing in each case beliefs and strategies and proving that beliefs are consistent and the strategy profile constitutes a Bayesian equilibrium at each stage given the beliefs.

First observe that whenever t^2_c is the true type of agent 2 then this is common knowledge, and it is also common knowledge that the type of agent 3 is t^3_a. Furthermore, notice that in this case the message profiles which are issued are exactly the ones observed in the truthtelling equilibrium (remember that 1 is telling the truth about her type). We will assume that when t^2_c has been observed everything, i.e. beliefs on agent 1 and strategies at further stages, is as in the original truthtelling equilibrium. Since strategies at the first stage are as in the original equilibrium, it is obvious that beliefs at further stages are acceptable and strategies at each stage are best responses.

From this point on we will concentrate the analysis on the case in which it is common knowledge that the type of agent 2 is in the set $\{t_a^2, t_b^2\}$. We specify beliefs and strategies for each deviation.

Agent 1 deviates In this case the message profile observed at stage 1 is $(\widetilde{m}^1, m_c^2, m_{ac}^3)$, with $\widetilde{m}^1 \notin \{m_{ac}^1, m_{bc}^1\}$. When this happens agents 2 and 3 believe $\Pr\left(t_a^1 | \widetilde{m}^1\right) = \frac{1}{2}$. Beliefs on agents 2 and 3 are equal to the prior, conditional on $\{t_a^2, t_b^2\}$ having been observed, i.e. $\Pr(t_a^2) = \Pr(t_b^2) = \frac{1}{2}$ and analogously for agent 3. The strategy profile is the following:

AGENT 1: at each stage following a given history H, let $m_a^1(H)$ be the message selected by the type t_a^1 of agent 1 in the original truthtelling equilibrium *when type t_c^2 of agent 2 has been observed*, and let $m_b^1(H)$ be the corresponding action for type t_b^1. Furthermore, let $\mu_a^1(H)$ be the probability assigned to type t_a^1 in the original equilibrium when t_c^2 has been observed and after history H. Then, in the deception equilibrium both types of agent 1 choose $m_a^1(H)$ if $\mu_a^1(H) \geq .5$ and $m_b^1(H)$ otherwise. In other words, agent 1 always chooses the action prescribed to the type with the highest probability in the truthtelling equilibrium when the same history H and t_c^2 have been observed.

AGENT 2: after each history H all types of agent 2 choose the same action selected by type t_c^2 in the truthtelling equilibrium.

AGENT 3: after each history H all types of agent 3 choose the same action selected by t_a^3 when type t_c^2 has been observed in the truthtelling equilibrium.

We assume that beliefs remain the same after each history. This is necessary along the equilibrium path, since each type of each agent is taking the same action, and it is clearly acceptable off the equilibrium path.

We have to show that there is no subgame in which some type of some agent has a profitable deviation in any subgame. Consider first agent 1 and observe that at each subgame she can only obtain constant functions (all types of agents 2 and 3 adopting the same strategies). Among constant functions, all types of agent 1 are indifferent between a and b, strictly prefer c and rank d last. Thus, if a profitable deviation exists it must be the case that after some history a stage is reached where either the outcome is d and agent 1 can obtain a different outcome or the outcome is a or b and agent 1 can obtain c. Now observe that agent 1 is facing at each stage the same action profile as in the original equilibrium for the case in which t_c^2 is common knowledge. It is immediate to see that the existence of such a deviation would imply that a profitable deviation is available to agent 1 in the original truthtelling equilibrium, a contradiction.

Consider now agent 2. Only constant functions can be obtained and, given the beliefs on agent 1, preferences among constant functions for all types are: $d \succ c \succ a \sim b$. A profitable deviation can only exist if, at a given stage, the outcome is a or b and it is possible to obtain c or d, or the outcome is c and it is possible to obtain d.

We can show that this would also be a profitable deviation in the original truthtelling equilibrium. The reason why we cannot use the straightforward argument used for agent 1 is that agent 2 is not facing the same action profile as in the original equilibrium. Consider a subgame reached after history H, and suppose $\mu_a^1(H) > .5$ (recall that this is the probability attached to type t_a^1 after history H in the original equilibrium). In the original equilibrium it is possible that, after history H, types t_a^1 and t_b^1 take different actions, say $m_a^1(H)$ and $m_b^1(H)$. In the equilibrium we propose beliefs are different (each type of agent 1 has probability $\frac{1}{2}$) and the action profile chosen by both types is $m_a^1(H)$ with probability 1. However, we can still prove that the original action taken by agent 2 in the case in which t_c^2 is common knowledge is a best response. Notice that in the original equilibrium, it cannot be the case that at a given stage the outcome is c for a type of agent 1 and not for the other type as well, since otherwise that type would deviate. We can assume that whenever in the original equilibrium a strategy profile at a given subgame yields c then both types of agent 1 take the same action. This is without loss of generality, since even if the original equilibrium prescribes different actions we can construct another equilibrium in which both types take the same action. Similarly, if the outcome is d for one type of agent 1 it must be d for the other type as well, since it must mean that 1 cannot reach any other outcome.

The implication is that whenever the outcome is c or d agent 2 is facing the same strategy profile as in the original equilibrium when t_c^2 is the true type, so that no profitable deviation exists. We can therefore restrict attention to subgames where the outcome is a or b under the deception equilibrium.

After an history H for which the outcome is expected to be a or b in the deception equilibrium, it must be the case that the outcome was in the set $\{a, b\}$ for both types of agent 1 in the original truthtelling equilibrium when t_c^2 was observed. Suppose now that a profitable deviation exists in the deception equilibrium, moving the outcome from a or b to c or d. Then the same deviation in the original truthtelling equilibrium would move the outcome from a or b to c or d with probability greater than .5, and at worst it moves the outcome to the less preferred between a and b with probability less than .5. This implies that type t_c^2 would deviate in the original equilibrium, a contradiction.

At last, let us consider agent 3. Given the beliefs on agent 1, all types of agent 3 are indifferent among the constant functions a, b, c. If a profitable deviation exists then at some stage it is possible to reach the outcome d, but this would also be a profitable deviation in the original truthtelling equilibrium.

Agent 2 deviates In this case the message profile observed at stage 1 is $(m_{hc}^1, \widetilde{m}^2, m_{ac}^3)$, with $\widetilde{m}^2 \neq m_c^2$. Agent 1 believes $\Pr(t_b^2, t_b^3 | \widetilde{m}^2) = 1$. Agent 3 believes $\Pr(t_b^2 | \widetilde{m}^2) = 1$. Beliefs on agent 1 are the same as in the truthtelling equilibrium.

AGENT 1. After each history H all types of agent 1 choose the same action as in the truthtelling equilibrium when type t_c^2 has been observed.

AGENT 2. After each history H all types of agent 2 choose the same action selected by type t_c^2 in the truthtelling equilibrium.

AGENT 3. After each history H all types of agent 3 choose the same action selected when type t_c^2 has been observed in the truthtelling equilibrium.

After each history H beliefs on agent 1 are the same as in the original equilibrium when t_c^2 is common knowledge. Beliefs on agents 2 and 3 remain the same after each history. Beliefs on agent 1 are acceptable because they were in the original equilibrium. Beliefs on agents 2 and

3 must be unchanged along the equilibrium path and are acceptable off the equilibrium path.

Optimality for agent 1 follows from the fact that she can only obtain constant functions, and she is indifferent among the constant functions a, b or d. Profitable deviations must reach c, but in this case the original equilibrium would be broken.

Consider now agent 2. Only functions which are constant in the type of agent 3 can be obtained. This implies that for any given type of agent 1, agent 2 is indifferent between a and b. A profitable deviation requires that at a given subform the outcome be moved from the set $\{a, b\}$ to $\{c, d\}$ or from c to d for at least one type of player 1. Since beliefs on 1 and strategies are the same as in the original equilibrium when t_c^1 has been observed, this would also be a profitable deviation in the original equilibrium.

Type t_b^2 has the same preferences as t_c^2, has the same beliefs as in the original equilibrium and is facing the same action profile. Thus, the strategy is optimal.

As for agent 3, given the beliefs on agent 2 she is indifferent among outcomes a, b, c, d.

Agent 3 deviates In this case the message profile observed at stage 1 is $(m_{hc}^1, m_c^2, \tilde{m}^3)$, with $\tilde{m}^3 \neq m_{ac}^3$. Agent 1 and type t_a^2 of agent 2 believe $\Pr\left(t_a^3 \,|\tilde{m}^3\right) = 0$ (remember that type t_b^2 observes the type of agent 3). Agents 2 and 3 believe $\Pr(t_h^1) = 1$. Agents 1 and 3 believe $\Pr(t_a^2) = \Pr(t_b^2) = \frac{1}{2}$ (beliefs over 1 and 2 are obtained using the Bayes' rule).

Strategies are as follows:

AGENT 1. After each history H all types of agent 1 choose the same action selected in the truthtelling equilibrium when type t_c^2 has been observed.

AGENT 2. After each history H all types of agent 2 choose the same action selected by type t_c^2 in the truthtelling equilibrium.

AGENT 3. After each history H type t_b^3 of agent 3 choose the same action selected by type t_a^3 when type t_c^2 has been observed in the truthtelling equilibrium. Type t_a^3 selects at each stage a best response to the strategy profile of agents 1 and 2.

Beliefs on agent 1 are as in the truthtelling equilibrium when t_c^2 has been observed after each history. This is acceptable since the same strategies are being used. For agents 2 and 3 beliefs remain the same after each history. This is necessary along the equilibrium path and acceptable outside.

Given the beliefs, a profitable deviation exists for agent 1 only when it is possible to obtain c. Since she is facing the same strategy profile as in the truthtelling equilibrium when t_c^2 has been observed, this would be a profitable deviation in that case as well.

The behavior of agent 2 is optimal because she believes $\Pr(t_b^3) = 1$, so that her preferences when the type is t_a^2 are the same preferences as if t_c^2 were her type (this is also true independently of beliefs on agent 3 for type t_b^2). Thus, she is facing the same action profile as in the original truthtelling equilibrium for the case in which t_c^2 is common knowledge and she has the same preferences, implying that the original actions are still a best response.

At last, consider agent 3. Optimality is obvious for type t_a^3, and type t_b^3 has is indifferent among all outcomes.

The cases in which two or more agents deviate can be treated in a similar way.

The only thing which is left to show is that there is no profitable deviation at the first stage.

No profitable deviation exists for agent 1 because after a deviation she can at most obtain a constant function a or b with respect to the type of agent 2, and she is indifferent between the two constant functions (this is because the strategies of agents 2 and 3 are constant with respect to types, and c cannot be an outcome because it would destroy the original equilibrium). Agent 2 cannot change the outcome through a deviation, i.e. the outcome must be a whenever agent 1 announces a type t_a^1 and b whenever t_b^2 is announced. If not, type t_c^2 would have a profitable deviation in the original equilibrium. At last, consider a deviation by agent 3. A deviation can never lead to d. She gets the same social choice function that she obtains in the truthtelling equilibrium when t_c^2 is observed. It is immediate to see that if the deviation is profitable under the deception then it would also be profitable in the truthtelling equilibrium when t_c^2 is observed.

References

[1] Dilip Abreu and Arunava Sen (1990) 'Subgame Perfect Implementation: A Necessary and Almost Sufficient Condition', *Journal of Economic Theory*, **50**:285-299.

[2] James Bergin and Arunava Sen (1996) 'Extensive Form Implementation in Incomplete Information Environments', mimeo, Queen's University.

[3] Drew Fudenberg and Jean Tirole (1991) '*Game Theory*', MIT Press, Cambridge,.

[4] Sanford Grossman and Motty Perry (1986) 'Perfect Sequential Equilibrium', *Journal of Economic Theory*, **39**: 97-119.

[5] Mathhew O. Jackson (1991) 'Bayesian Implementation', *Econometrica*, **59**: 461-477.

[6] John Moore and Rafael Repullo (1988) 'Subgame Perfect Implementation', *Econometrica*, **56**: 1191-1220.

[7] Vicente Madrigal, Tommy C.C. Tan and Sérgio Ribeiro da Costa Werlang (1987) 'Support Restrictions and Sequential Equilibria', *Journal of Economic Theory*, **43**: 329-334.