JOURNAL OF ECONOMIC THEORY **80**, 185–200 (1998)
ARTICLE NO. ET982392

Unique Implementation of the Full Surplus Extraction Outcome in Auctions with Correlated Types

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

Received June 23, 1997; revised December 20, 1997

In auctions with correlated types it is possible to design mechanisms such that full surplus extraction can be obtained as the outcome of an equilibrium in which agents use (weakly) dominant strategies. However, it is not assured that the outcome is unique. We present an example in which no mechanism can yield the full surplus extraction outcome as the unique Bayesian equilibrium outcome. Next we show that in the standard auction model the multiplicity problem can be fully resolved using sequential mechanisms, i.e., we show that it is possible to obtain the full surplus extraction outcome as the unique perfect Bayesian equilibrium outcome. *Journal of Economic Literature* Classification Numbers: D44; D70.

1. INTRODUCTION

In auctions with correlated types it is possible to design the mechanism in such a way that each participant earns an expected surplus of zero and the object is allocated to the agent with the highest valuation for the object. This is the best possible situation for the seller, and it is referred to as the "full surplus extraction outcome" (FSE outcome). In particular, Crémer and McLean [2] show that when a mild "spanning" condition on the joint probability distribution on types is satisfied then it is possible to design the auction in such a way that the full surplus extraction outcome is obtained as a Bayesian equilibrium in which the agents adopt (weakly) dominant strategies.

The result has been looked upon with some suspicion, as it heavily relies on risk neutrality for the construction of the mechanism. In particular,

* This paper contains results previously included in Brusco [1].

0022-0531/98 $25.00

huge penalties or prizes may have to be awarded to the agents in order to have incentives aligned. Here we do not deal with the plausibility of the mechanism. Rather, we would like to point out another potential weakness. Despite the fact that the FSE equilibrium is in dominant strategies, *it is not necessarily the unique Bayesian equilibrium* of a mechanism yielding the FSE outcome. We will present an example in which every mechanism having the FSE outcome as a Bayesian equilibrium must have another Bayesian equilibrium *which is weakly Pareto superior for the agents*. In other words, the FSE outcome is not uniquely implementable in Bayesian equilibrium in the standard auction model. The fact that the alternative equilibrium is weakly Pareto superior to the one producing the FSE outcome makes the problem even more serious, since a Pareto superior equilibrium is a natural focal point for the agents.

We will show however that the multiplicity problem can be eliminated if we adopt a sequential mechanism and assume that the outcome of the mechanism can be predicted using the notion of perfect Bayesian equilibrium. The basic idea is to ask the agents to announce their type *and* whether they think that some agent is lying. If some agent claims that someone else is lying then a second stage is reached in which a "test" agent is asked to make a choice revealing the true type. The key features of the mechanism are

1. whenever the second stage is reached, there exists a unique equilibrium in strictly dominant strategies in which the truth comes out;

2. given the outcome at the second stage, it is always convenient to denounce an agent who is lying.

We show that it is possible to design the mechanism in such a way that in no equilibrium agents lie.

There are by now a number of applications of sequential mechanisms, see, e.g., Glazer and Ma [4], Jackson and Moulin [6], and Varian [7]. The mechanism presented in this paper is similar to the one in Duggan [3], who deals with the general adverse selection problem in principal/multiagent environments. Duggan also exploits quasi-linearity of preferences to make sure that truth-telling is a dominant strategy at the second stage. Although the mechanism presented in this paper has been developed independently, it can be seen as an application of the mechanism in Duggan. Our setting allows us to simplify the agents' message spaces, giving to them a more natural interpretation.

The rest of the paper is organized as follows. Section 2 presents the general set up. Section 3 presents the example with multiple equilibria, and Section 4 shows that the multiplicity problem can be eliminated with sequential mechanisms.

2. AUCTIONS WITH CORRELATED TYPES

Let us consider the standard auction setting. There is a seller who is trying to sell an object to n risk-neutral bidders. The value of the object to the seller is zero, while the bidders may have positive value. Let S^i be the set of types of bidder i, with S^i finite for each i. An agent of type $s^i \in S^i$ values the object $w(s^i)$. The utility function of agent i is

$$d_i w(s^i) + m_i$$

where $d_i = 1$ ($d_i = 0$) if the object is (not) assigned to bidder i and m_i is the monetary transfer to bidder i (negative in case of payment). The set of states of the world is $S = \times_{i=1}^{n} S^i$, so that a state of the world is a vector $s = (s^1, ..., s^n)$. It is common knowledge that the probability distribution on S is given by π, and $\pi(\cdot \mid s^i)$ denotes the probability distribution on $S^{-i} \times_{j \neq i} S^j$ conditional on agent i being of type s^i. We say that *types are correlated* if there is an agent i, two types $s^i, t^i \in S^i$ and a type profile $s^{-i} \in S^{-i}$ such that $\pi(s^{-i} \mid s^i) \neq \pi(s^{-i} \mid t^i)$. For simplicity, we will assume that $w(s^i) \neq w(s^j)$ for each i, j, i.e., it is never the case that two agents have the same valuation. An *allocation rule* is a collection of $2n$ functions

$$p^1, ..., p^n, h^1, ..., h^n$$

where $p^i \colon S \to [0, 1]$ is the probability that the object is assigned to agent i. Obvious feasibility restrictions on the functions $p^i, ..., p^n$ are that $p^i(t) \geqslant 0$ for each i and t and $\sum_{i=1}^{n} p^i(t) \leqslant 1$ for each t. The function $h^i \colon S \to R$ is the transfer to agent i.

In this paper we will focus on a particular allocation rule, the one that guarantees maximum revenue to the seller. We say that an array $(p^1, ..., p^n, h^1, ..., h^n)$ reaches the full-surplus-extraction (FSE) outcome if the following conditions are satisfied:

1. For each $t \in S$, $\sum_{i=1}^{n} p^i(t) = 1$ and $p^i(t) = 0$ whenever there exists $j \neq i$ such that $w(t^i) < w(t^j)$.

2. For each agent i and type t^i the following equality is satisfied:

$$\sum_{s^{-i} \in S^{-i}} \pi(s^i \mid t^i)[p^i(s^{-i}, t^i)\, w(t^i) + h^i(s^{-i}, t^i)] = 0.$$

The first condition requires that the object be assigned only to the agents who value it most, so that the allocation of the object is efficient. The second condition requires that transfers be arranged in such a way that each type of each agent be left with just her reservation value. The two conditions together imply that the seller is able to extract the full surplus.

Crémer and McLean [2] have shown that, under a mild "spanning" condition on conditional probabilities, it is possible for a principal to design a mechanism yielding the FSE outcome in dominant strategies. The condition requires that, for each agent i, the matrix formed using as rows the $|S^i|$ vectors of conditional probabilities be of rank $|S^i|$ (see Theorem 1 in Crémer and McLean [2]; for reader's convenience the result is reported in the appendix). Despite the fact that the equilibrium is in dominant strategies there is no guarantee that the FSE outcome is the unique Bayesian equilibrium outcome. We present an example in which whenever the FSE outcome is obtained as an equilibrium there must be another Bayesian equilibrium which is weakly Pareto superior from the agents' point of view.

3. THE EXAMPLE

There are two agents. Agent 1 has 3 types, s_0, s_1, and s_2, and agent 2 has 3 types, t_0, t_1, and t_2. We assume

$$w(t_2) > w(t_1) > w(t_0) > w(s_2) > w(s_1) > w(s_0)$$

so that the valuation of agent 2 is always higher than the valuation of agent 1, and the object is always assigned to agent 2. The joint probability distribution is given by:

	t_0	t_1	t_2
s_0	0	0.1	0.1
s_1	0	0.2	0.5
s_2	0.1	0	0

The spanning condition identified in Crémer and McLean [2], Theorem 1, is satisfied. Therefore it is possible to design a mechanism with a dominant strategy equilibrium yielding the FSE outcome. A (normal form) mechanism can be represented as:

- a pair of message spaces (M, K), where M denotes the message space of agent 1, and K the message space for agent 2;
- four functions (p^1, p^2, H_1, H_2), where $p^i\colon M \times K \to [0, 1]$ is the probability of allocating the object to agent i and $H_i\colon M \times K \to R$ is the transfer to agent i. We allow in principle the possibility that for some message pair the object is not allocated, i.e., we allow $p^1(m, k) + p^2(m, k) < 1$.

Suppose now that the mechanism has an equilibrium yielding the FSE outcome (we will refer to this equilibrium as the "correct" one). Let m_2 be the strategy used by type s_2 of agent 1, and k_0 the strategy adopted by type t_0 of agent 2 in the correct equilibrium. We claim that such a mechanism must have at least another equilibrium, in which each type of agent 1 uses m_2 and each type of agent 2 uses k_0. This equilibrium is weakly Pareto superior, from agents' point of view, to the correct one. Agent 1 obtains an expected utility of zero, as in the correct equilibrium. Agent 2 receives the object and pays $w(t_0)$ with probability 1, no matter what her type, so that she obtains a strictly positive surplus when the type is t_1 or t_2.

To show that this is an equilibrium, suppose not. Then there must be a profitable deviation for some type of some agent. Consider first agent 1. Obviously, type s_2 cannot have a profitable deviation, since it is facing the same situation as in the correct equilibrium. Suppose first that the profitable deviation is available for s_1. Then there must be a message $\bar{m}$ such that:

$$p^1(\bar{m}, k_0)\, w(s_1) + H_1(\bar{m}, k_0) > H_1(m_2, k_0).$$

Since $w(s_2) > w(s_1)$ and $p^1(\bar{m}, k_0) \geqslant 0$, this implies:

$$p^1(\bar{m}, k_0)\, w(s_2) + H_1(\bar{m}, k_0) > H_1(m_2, k_0).$$

Notice now that the RHS is the payoff in the correct equilibrium for type s_2. Thus, if there is a strategy $\bar{m}$ that type s_1 can use to break the "bad" equilibrium, the correct equilibrium is also jeopardized, as type s_2 would be given a profitable deviation. It is clear that the same reasoning applies when type s_0 is endowed with the profitable deviation. We therefore conclude that no profitable deviation exists for agent 1.

Consider now agent 2. Again, no profitable deviation can exist for type t_0. Suppose first that type t_1 of agent 2 has a profitable deviation $\bar{k}$. Then it must be the case that:

$$p^2(m_2, \bar{k})\, w(t_1) + H_2(m_2, \bar{k}) > w(t_1) + H_2(m_2, k_0).$$

Again, $1 \geqslant p^2(m_2, \bar{k})$ and $w(t_1) > w(t_0)$ imply that $\bar{k}$ would also be a profitable deviation for type t_0 against the "good" equilibrium. The same reasoning applies to type t_2. We conclude that no profitable deviation exists, so that the proposed strategy profile is indeed an equilibrium. Therefore, it is impossible to obtain the FSE outcome as the unique outcome supported by a Bayesian equilibrium. In other words, the FSE outcome is not implementable in Bayesian equilibrium.

What is happening is that the Bayesian monotonicity condition, which is necessary for Bayesian implementation (see Jackson [5]), is violated for the particular allocation we want to implement. The Bayesian monotonicity condition requires that for every deception[1] there exist a deviation which is profitable when the deception is used but unprofitable for every type of every player in the truth-telling equilibrium.

We have shown that, under the deception we consider, a deviation can be profitable only if it is also a profitable deviation for some type in the truth-telling equilibrium. This implies that the Bayesian monotonicity condition is violated.

Remark. The example heavily exploits the special information structure, and in particular the existence of separate common knowledge components. It turns out that in this standard auction setup it is relatively easy to construct examples in which every normal form mechanism yielding the FSE outcome must have multiple equilibria whenever there are separate common knowledge components. This includes the important special case of complete information.[2] It is much harder (we suspect impossible) to produce such examples when there are no separate common knowledge components. To produce examples in which all normal form mechanisms have multiple equilibria it is crucial to find a deception $\alpha: S \rightarrow S$ such that the function given by the FSE outcome does not satisfy the Bayesian monotonicity condition with respect to α. When there are separate common knowledge components such deception can frequently be constructed making the agents behave, in a certain common knowledge component, as if they were in a different common knowledge component (this is what happens in our example, with the agents pretending to be in the bottom-left component when they are in the top-right component). These deceptions are difficult to destroy, since profitable deviations against the deception usually turn out to be profitable deviations against the truth-telling equilibrium as well (again, this is exactly what happens in our example). When there is a single common knowledge component there is no obvious candidate deception for violating the Bayesian monotonicity condition.

4. IMPLEMENTATION WITH SEQUENTIAL MECHANISMS

The previous section has shown that even in the case in which the conditions contained in Theorem 1 of Crémer and McLean [2] are satisfied, the multiplicity problem cannot be eliminated.

[1] Section 4 contains a formal definition of deception.

[2] The example presented in this section can be modified to generate a "complete information" example. It is sufficient to eliminate s_0 and t_2 and assign positive probability only to the pairs (s_1, t_1) and (s_2, t_0). Many other examples can be produced in this way, including examples in which the object is allocated to different bidders in different common knowledge components.

We now show that whenever the conditions contained in Theorem 2 of Crémer and McLean [2] are satisfied,[3] the FSE outcome can always be implemented as the only outcome supported by a perfect Bayesian equilibrium. Theorem 2 of Crémer and McLean [2] provides conditions ensuring that there exists a mechanism having the FSE outcome as a Bayesian equilibrium, although not necessarily unique. Obviously the conditions are weaker than the ones contained in Theorem 1, ensuring the existence of a dominant strategy mechanism. We establish in this section that whenever it is possible to build a normal form mechanism having the FSE outcome as a Bayesian equilibrium then it is possible to build a sequential mechanism having the FSE outcome as the *unique* outcome supported by a perfect Bayesian equilibrium. The mechanism used is a simple two-stage mechanism, involving no "integer" or "modulo" games (but see the remark after the theorem on the use of "open set" tricks in the proof).

In order to describe the mechanism, we first need a definition. A *deception for agent i* is a mapping $\alpha^i\colon S^i \to \Delta S^i$. It can be interpreted as a reporting strategy adopted by agent i. We will denote by $\alpha^i(t^i \mid s^i)$ the probability of announcing t^i given that the type is s^i. Let A^i be the space of all possible deceptions for agent i and define $A^{-i} = \times_{j \neq i} A^j$ and $A = \times_{i=1}^n A^i$. The notation $\bar{\alpha}^i$ will be used to denote the identity deception or truth-telling, i.e., $\bar{\alpha}^i(t^i \mid s^i) = 1$ if $s^i = t^i$ and $\bar{\alpha}^i(t^i \mid s^i) = 0$ otherwise. We will also use the notation $\alpha(t \mid s)$ to denote the probability that agents jointly issue a report t when the true type profile is s, and the notation $\alpha^{-i}(t^{-i} \mid s^{-i})$ with obvious meaning.

We can proceed to show that it is possible to find a sequential mechanism having the FSE outcome as the unique perfect Bayesian equilibrium outcome. Notice again that we are making the simplifying assumption that $w(s^i) \neq w(t^j)$ for each pair $s^i \in S^i$ and $t^j \in S^j$.

Let us first describe formally the mechanism used by Crémer and McLean for (weak) Bayesian implementation of the full surplus extraction outcome. Let $p^i\colon S \to [0, 1]$ be the probability of assigning the object to i given announcement s. We want the efficient allocation to be implemented, so $p^i(s) = 1$ if $w(s^i) > w(s^j)$ for each $j \neq i$ and $p^i(s) = 0$ otherwise. The payment for agent i is a function $H_i\colon S \to R$ satisfying the properties:

$$\sum_{s^{-i} \in S^{-i}} \pi(s^{-i} \mid s^i)[p^i(s^i, s^{-i})\, w(s^i) - H_i(s^i, s^{-i})] = 0$$

for each i and $s^i \in S^i$, and:

$$\sum_{s^{-i} \in S^{-i}} \pi(s^{-i} \mid t^i)[p^i(s^i, s^{-i})\, w(t^i) - H_i(s^i, s^{-i})] \leqslant 0$$

[3] For reader's convenience, the theorem is reported in the appendix.

whenever $t^i \neq s^i$. In the appendix we report the results by Crémer and McLean ensuring that appropriate functions H_i satisfying the two properties exist. Now consider the following two-stage mechanism.

First Stage. Each agent announces her type and the deception adopted, i.e., $M_1^i = S^i \times A$ is the message space of agent i at stage 1. Let $m_1^i = (\hat{s}^i, \hat{\alpha})$ be the announcement of agent i, and let $\hat{s} = (\hat{s}^1, ..., \hat{s}^n)$. For a given reporting strategy α and a given announcement $\hat{s}$ let us define:

$$\Pr^{\alpha}(t^i \mid \hat{s}) = \frac{\sum_{s^{-i} \in S^{-i}} \alpha(\hat{s} \mid s^{-i}, t^i)\, \pi(s^{-i}, t^i)}{\sum_{s \in S} \alpha(\hat{s} \mid s)\, \pi(s)}$$

whenever $\sum_{s \in S} \alpha(\hat{s} \mid s)\, \pi(s) > 0$. Thus, $\Pr^{\alpha}(t^i \mid \hat{s})$ is the probability that agent i is of type t^i when the deception α is used by the agents and the report $\hat{s}$ has been issued. For a given vector of announcements $m_1 = (m_1^1, ..., m_1^n)$, define agent i^* as the agent with the lowest index reporting a deception $\hat{\alpha}$ such that $\hat{\alpha}^{-i^*} \neq \bar{\alpha}^{-i^*}$, $\sum_{s \in S} \hat{\alpha}(\hat{s} \mid s)\, \pi(s) > 0$ and $\Pr^{\hat{\alpha}}(\hat{s}^{i^*+1} \mid \hat{s}) \neq 1$. Thus, i^* is the agent with the lowest index satisfying the following properties:

1. i^* claims that the other agents are using an untruthful deception $\hat{\alpha}^{-i^*} \neq \bar{\alpha}^{-i^*}$.

2. The announcement $\hat{s}$ occurs with positive probability when the deception $\hat{\alpha}$ is used.

3. With positive probability, agent i^*+1 has been lying, i.e., she issued report $\hat{s}^{i^*+1}$ and her true type was $s^{i^*+1} \neq \hat{s}^{i^*+1}$.

Of course, for some announcement i^* will not be well defined. For the cases in which it is defined, let $j^* = i^* + 1$ (or $j^* = 1$ if $i^* = n$). The outcome function is as follows:

- If i^* is not well defined then each agent i receives the object with probability $p^i(\hat{s})$ and receives a transfer $H_i(\hat{s})$, where p^i and H_i are the functions adopted in the Crémer–McLean mechanism. Notice that this case occurs when each agent announces $\bar{\alpha}$, i.e., each agent announces that all bidders are telling the truth.

- If i^* is well defined then each agent $i \neq j^*$ receives a transfer $H_i(\hat{s})$, and the game moves to the second stage. Agent j^* pays a large sum $\bar{K}$. The amount of the sum is chosen so that each agent j^* is strictly worse off when the second stage is reached.

Second Stage. For the given deception $\hat{\alpha}$ announced by agent i^* and for the given announcement $\hat{s}$, define $\bar{t}^{j^*}$ as the type with the highest valuation such that $\hat{\alpha}^{j^*}(\hat{s}^{j^*} \mid \bar{t}^{j^*}) > 0$ and $\bar{t}^{j^*} \neq \hat{s}^{j^*}$, i.e., the highest type who, according to the deviation denounced by i^*, is announcing $\hat{s}^{j^*}$ with positive probability (notice that, since the announcement of $j*$ is not fully revealing, the type $\bar{t}^{j*}$ is well defined). There are two possibilities.

1. $w(\bar{t}^{j*}) > w(\hat{s}^{j*})$. In this case agent $j*$ selects between having the object or a transfer K^{α}, with $w(s_{+}^{j*}) > K^{\alpha} > w(\hat{s}^{j*})$, where s_{+}^{j*} is the type of agent $j*$ with the next higher valuation with respect to $\hat{s}^{j*}$. If $j*$ chooses to receive the object then it is evident that agent $i*$ was right in denouncing the deception. Thus agent $i*$ receives a positive transfer $\bar{T}^{y} = [1-(1-\bar{\gamma}(\hat{\alpha}, \hat{s}))^2]$, where $\bar{\gamma}(\hat{\alpha}, \hat{s})$ is defined as:

$$\bar{\gamma}(\hat{\alpha}, \hat{s}) = \sum_{\{t^{j*} \in S^{j*} \mid w(t^{j*}) > w(\hat{s}^{j*})\}} \Pr^{\hat{\alpha}}(t^{j*} \mid \hat{s}),$$

i.e., the conditional probability that the valuation of agent i is greater than $w(\hat{s}^{j*})$ given that the announcement has been $\hat{s}$ and that the deception $\hat{\alpha}$ has been used. In case $j*$ chooses K^{α}, agent $i*$ was wrong in denouncing the deception and receives a transfer $\bar{T}^{n} = -\bar{\gamma}(\hat{\alpha}, \hat{s})^2$.

The object is allocated as follows. With probability $1-\varepsilon_{\hat{\alpha}, \hat{s}}$ the object is allocated according to the first period announcement, i.e., $p^{i}(\hat{s})$. In this case no additional transfers are made (in particular, the decision of agent $j*$ has no effect other than on the transfer to agent $i*$). With probability $\varepsilon_{\hat{\alpha}, \hat{s}}$ agent $j*$ is given what she has chosen, i.e., the object or the transfer. In case $j*$ chooses the object, she pays an amount θ such that $w(\hat{s}_{+}^{j*}) - \theta > K^{\alpha}$. If $j*$ has chosen the transfer, then the object is destroyed. The probability $\varepsilon_{\hat{\alpha}, \hat{s}} > 0$ is chosen to satisfy the inequality:

$$\varepsilon_{\hat{\alpha}, \hat{s}} w(\bar{s}^{i*}) < \bar{\gamma}(\hat{\alpha}, \hat{s})[1-(1-\bar{\gamma}(\hat{\alpha}, \hat{s}))^2] - (1-\bar{\gamma}(\hat{\alpha}, \hat{s}))\, \bar{\gamma}(\hat{\alpha}, \hat{s})^2$$

where $\bar{s}^{i*}$ is the type of agent $i*$ with the highest possible valuation. (Notice that the RHS is the expected value of the transfer for agent $i*$ when this agent is revealing her type.)

2. $w(\bar{t}^{j*}) < w(\hat{s}^{j*})$. In this case agent $j*$ selects between having the object or a transfer K^{α}, with $w(\hat{s}^{j*}) > K^{\alpha} > w(\bar{t}^{j*})$. If $j*$ chooses to receive the transfer then it is evident that agent $i*$ was right in denouncing the deception. Thus agent $i*$ receives a positive transfer

$$\underline{T}^{y} = 1-(1-\underline{\gamma}(\hat{\alpha}, \hat{s}))^2$$

where $\underline{\gamma}(\hat{\alpha}, \hat{s})$ is defined as

$$\underline{\gamma}(\hat{\alpha}, \hat{s}) = \sum_{\{t^{j*} \in S^{j*} \mid w(t^{j*}) < w(\hat{s}^{j*})\}} \Pr^{\hat{\alpha}}(t^{j*} \mid \hat{s}).$$

If j^* chooses to receive the object, agent i^* receives a transfer $\underline{T}^n$ equal to:

$$\underline{T}^n = -\underline{\gamma}(\hat{\alpha}, \hat{s})^2$$

Rules for allocation of the object are the same as in the previous case, except that agent j^* does not pay for the object: With probability $1-\varepsilon_{\hat{\alpha},\hat{s}}$ the decision of agent j^* is irrelevant and with probability $\varepsilon_{\hat{\alpha},\hat{s}}$ agent j^* is given what she has chosen, i.e., the object or the transfer. If j^* has chosen the transfer, then the object is destroyed.

The mechanism may look quite complicated, but the underlying idea is rather simple. Basically, each agent is asked to report her type *and* to announce whether she believes that everybody else is reporting honestly. If everybody claims that agents are reporting honestly then the Crémer–McLean allocation is implemented. In case one agent reports that a deception is being used, then the claim is "tested" by going to the second stage and offering to some other agent the choice between having the object and a transfer. The choice of the agent reveals whether a lie was told at the first stage. If it turns out that a lie was told then the agent who reported the use of a deception is given a prize, otherwise she has to pay a fine. Payoffs are designed in such a way that whenever a deception is used, it is optimal for some agent to denounce the deception. Furthermore, whenever a deception is denounced, it is optimal for the other agents to tell the truth. These two facts together make sure that no deception can be used in equilibrium. An interesting feature of the mechanism is that what happens in the second stage does not depend on the particular beliefs assigned. At the second stage only agent j^* moves, selecting the transfer or the object depending only on her type. In any perfect Bayesian equilibrium, agent j^* uses a *strictly* dominant strategy at stage 2.[4] At last, we point out that there are no restriction on the number of players, i.e., the mechanism also works for the case $n=2$.

We will now show that the mechanism yields the FSE outcome as the unique perfect Bayesian equilibrium outcome. We will refer to the condition stated by Theorem 2 in Crémer and McLean [2], allowing full extraction in Bayesian equilibrium, as condition CM2 (see the appendix for a statement of the condition).

THEOREM 1. *If condition CM2 is satisfied then it is possible to obtain the FSE outcome as the unique outcome supported by a perfect Bayesian equilibrium.*

Proof. First observe that there is an equilibrium in which each agent reports truthfully the type and says that everybody else is reporting truthfully, and whenever the second stage is reached agent j^* selects the

[4] The mechanism proposed by Duggan [3] shares the same feature.

preferred outcome. Incentive compatibility implies that no agent can profitably deviate by reporting untruthfully her type while still maintaining that the other agents are reporting truthfully. Let us consider deviations in which an agent of type s^i claims that a deception α is used and (possibly) changes the report about her type. In this case at the second stage agent j^* makes a choice confirming her initial report, so that the deviating agent obtains a negative transfer, and the deviation is unprofitable.

We next show that there is no equilibrium in which a deception is used. We first present three lemmas which will constitute the main argument of the proof.

LEMMA 1. *In every perfect Bayesian equilibrium, if a type t^2 of agent 2 is using a deception then each type s^1 of agent 1 such that $\pi(t^2 \mid s^1) > 0$ denounces a deception.*

Proof. Suppose that in the equilibrium the agents are reporting their types according to the deception α, and suppose further that agent 2 is not reporting the truth, i.e., $\alpha^2 \neq \bar{\alpha}^2$. In particular, let t^2 be a type who is not reporting the truth with probability 1. Let s^1 be a type of agent 1 such that $\pi(t^2 \mid s^1) > 0$, and let $\hat{s}^1$ be a type of agent 1 such that $\alpha^1(\hat{s}^1 \mid s^1) > 0$, i.e., $\hat{s}^1$ is reported with positive probability by type s^1. We will show that it cannot be optimal for type s^1 of agent 1 not to denounce a deception, i.e., it cannot be optimal to announce $(\hat{s}^1, \bar{\alpha})$.

Let $\hat{\alpha}^1$ be a deception for agent 1 such that $\hat{\alpha}^1(\hat{s}^1 \mid s^1) = 1$ and $\hat{\alpha}^1(\hat{s}^1 \mid t^1) = 0$ if $t^1 \neq s^1$ (this is a deception such that $\hat{s}^1$ is only announced by type s^1), and define $\hat{\alpha} = (\hat{\alpha}^1, \alpha^{-1})$.

Suppose now that agent 1 announces $(\hat{s}^1, \hat{\alpha})$ instead of $(\hat{s}^1, \bar{\alpha})$. We will show that for each message having positive probability agent 1 is not worse off, and for some message having strictly positive probability agent 1 is strictly better off. There are two different cases:

- The message $\hat{s}^{-1}$ announced by the other agents is such that $\Pr^{\hat{\alpha}}(\hat{s}^2 \mid \hat{s}) = 1$. In this case the outcome is unchanged.
- The message $\hat{s}^{-i}$ announced by the other agents is such that $\Pr^{\hat{\alpha}}(\hat{s}^2 \mid \hat{s}) < 1$. This happens with strictly positive probability, and in this case agent 1 becomes agent i^* and the second stage is reached.

We now show that in this second case agent 1 is strictly better off. It suffices to compare the payoff obtained when all other agents announce $\bar{\alpha}$; the case in which some other agent is also denouncing a deception is treated along similar lines and omitted here.

Suppose first that $w(\bar{t}^2) > w(\hat{s}^2)$. In any perfect Bayesian equilibrium, agent 2 will choose the object if $w(s^2) > w(\hat{s}^2)$, and the transfer otherwise.

Notice that, according to the deception denounced by agent 1, the announcement $\hat{s}^1$ fully reveals type s^1, so that the conditional probability $\bar{\gamma}(\hat{\alpha}, \hat{s})$ used in the mechanism is computed using the same information held by agent i. In other words, let $\Pr^{\alpha}(w(s^2) > w(\hat{s}^2) \mid \hat{s}^{-1}, s^1)$ be the probability that the valuation of agent 2 is greater than $w(\hat{s}^2)$ given that the type of agent 1 is s^1, agents other than 1 report their types according to deception α^{-1} and the announcement by other agents is $\hat{s}^{-1}$. Then we have that:

$$\Pr^{\alpha}(w(s^2) > w(\hat{s}^2) \mid \hat{s}^{-1}, s^1) = \bar{\gamma}(\hat{\alpha}, \hat{s}). \tag{1}$$

Agent 1 still receives the transfer $H(\hat{s})$, and additionally at the second stage she obtains an expected transfer Tr equal to:

$$E(Tr \mid \hat{s}) = \Pr^{\alpha}(w(s^2) > w(\hat{s}^2) \mid \hat{s}^{-1}, s^1)[1 - (1 - \bar{\gamma}(\hat{\alpha}, \hat{s}))^2] \\ - (1 - \Pr^{\alpha}(w(s^2) > w(\hat{s}^2) \mid \hat{s}^{-1}, s^1))\, \bar{\gamma}(\hat{\alpha}, \hat{s})^2.$$

Using 1, it is immediate to see that $E(Tr \mid \hat{s}) > 0$. On the other hand, agent 1 can lose the object with at most probability $\varepsilon_{\alpha, \hat{s}}$. Since, by construction, $\varepsilon_{\alpha, \hat{s}} w(\bar{s}^1) < E(Tr \mid \hat{s})$ (again making use of 1), agent 1 is strictly better off denouncing the deviation. An analogous reasoning holds for the case $w(\bar{t}^2) < w(\hat{s}^2)$. We conclude that it cannot be optimal for agent 1 to announce $(\hat{s}^1, \bar{\alpha})$ whenever agent 2 is not expected to tell the truth with probability 1. ▮

LEMMA 2. *There is no perfect Bayesian equilibrium in which agent 2 does not tell the truth.*

Proof. We have seen that agent 1 will denounce a deception by agent 2 whenever it is used. Suppose agent 2 does not tell the truth. For a given announcement $\hat{s}^2$ which does not reveal the type of agent 2, consider agent $\bar{t}^2$ (the highest type announcing $\hat{s}^2$), and assume first that $w(\bar{t}^2) > w(\hat{s}^2)$. We will show that it cannot be optimal for this type to announce $\hat{s}^2$. By announcing $\hat{s}^2$, agent 2 ends up choosing the object and paying θ. By announcing $\bar{t}^2$, three things may happen.

- If $\bar{t}^2$ is revealing then the second stage is not reached and agent 2 is strictly better off.
- If $\bar{t}^2$ is also announced by a type $\bar{\bar{t}}^2$ such that $w(\bar{\bar{t}}^2) > w(\bar{t}^2)$ then the second stage is reached and agent 2 selects a transfer $K_{\alpha} > w(\bar{t}^2)$, so that she is strictly better off.
- In all other cases, the second stage is reached and agent 2 selects the object without paying any price.

We conclude that agent $\bar{t}^2$ is strictly better off telling the truth, so that it can never happen that $\hat{s}^2$ is announced by a type with a higher valuation. This in particular implies that the announcement s_0^2 (the type of agent 2 with the lowest valuation) either is fully revealing or it never occurs. In both cases, all types of agent 2 are able to avoid the payment of the fine $\bar{K}$ by announcing s_0^2. We conclude that no type of agent 2 pays the fine. But this in turn implies that all announcements are fully revealing, contradicting the fact that agent 2 is not telling the truth. ▮

LEMMA 3. *There is no perfect Bayesian equilibrium in which agent* 2 *tells the truth and agent* 1 *denounces a deception leading to the second stage with positive probability.*

Proof. Suppose that agent 1 is denouncing a deception and let $\hat{s}$ be an announcement occurring with positive probability and leading to the second stage. Assume that the deception denounced by agent 1 implies that $\hat{s}^2$ can be issued by higher types of agent 2 (the case in which $\hat{s}^2$ is only issued by lower types is analogous). Since agent 2 is telling the truth, agent 1 receives a transfer equal to

$$-\bar{\gamma}(\hat{\alpha}, \hat{s})^2,$$

and possibly loses the object with probability $\varepsilon_{\hat{\alpha}, \hat{s}}$. In order to maximize this quantity agent 1 should select $\bar{\gamma}(\hat{\alpha}, \hat{s}) = 0$ (this would also guarantee $\varepsilon_{\hat{\alpha}, \hat{s}} = 0$). But this contradicts the assumption that the second stage is reached. Notice that what is exploited here is an "open set" problem: An agent incorrectly announcing a deception would like to be as close as possible to the truth, without reaching it. ▮

The three lemmas imply that every perfect Bayesian equilibrium has the following characteristics:

1. agent 2 tells the truth;
2. agent 1 does not report a deception.

Once these facts have been established, it is immediate to observe that we can apply the same argument to agents 2 and 3, and then inductively to all agents up to the pair $(n-1, n)$. This allows us to conclude that in every perfect Bayesian equilibrium:

1. agents 2, 3,..., n tell the truth;
2. agents 1, 2,..., $n-1$ do not report a deception.

The last step is to recognize that Lemmas 1, 2, and 3 can now be applied to agents 1 and n: agent 1 must tell the truth in every perfect Bayesian equilibrium, since otherwise agent n would denounce her, and agent n cannot be reporting a deception when agent 1 tells the truth. We conclude that truth-telling is the only equilibrium. ∎

Remark. The mechanism does not include integer or modulo games, but there is one point in the proof in which an "open set" trick is used. The problem is in eliminating the equilibria in which every agent is correctly reporting the type, but at the same time every agent claims that an untruthful deception is used (Lemma 3). To understand the point better, consider the case $n=2$, and suppose that agent 2 is announcing correctly her type but is also (incorrectly) claiming that an untruthful deception is used by agent 1. If agent 1 does not announce that agent 2 is using an untruthful deception then agent 2 ends up being i^* and agent 1 is j^*, so that she has to pay the penalty $\bar{K}$. In order to avoid the penalty, agent 1 should claim that an untruthful deception is used by agent 2. This is turn makes it optimal the strategy of agent 2. Thus, there seems to be an equilibrium in which both agents are reporting correctly their types and at the same time incorrectly denouncing untruthful reporting by the other agent. The reason why this is not an equilibrium is that agent 1 has no best response to agent 2. Given the payoff structure, agent 1 would like to go as close as possible to the real reporting strategy (i.e., truth-telling) of agent 2, but at the same time she has to claim that 2 is not reporting truthfully. There is therefore an "open set" problem, and no best response exists for agent 1. This makes sure that the proposed strategy profile is not an equilibrium.

APPENDIX

This appendix briefly reviews the results contained in Crémer and McLean [2] used in the text. It is only intended to help the reader to go through the paper without having to reach the original article.

The first theorem provides conditions for the existence of a mechanism having an equilibrium in dominant strategies yielding the FSE outcome.

THEOREM 1 (CM1). *An information structure (S, π) guarantees full extraction of the surplus by a dominant strategy auction if and only if for each agent i there do not exist $\{\rho^i(s^i)\}_{s^i \in S^i}$, not all equal to zero, such that*:

$$\sum_{s^i \in S^1} \rho^i(s^i)\, \pi(s^{-i} \mid s^i) = 0 \qquad \textit{for all} \quad s^{-i} \in S^{-i}.$$

The mechanism used for implementation is a second price auction augmented by a lottery depending only on the announcements by other agents. The condition in CM1 is a spanning condition that enables the seller to build, for each agent i, a lottery $H_i\,(s^{-i})$ having for each type s^i an expected value equal to the expected surplus in a second price auction.

The second theorem provides conditions for the existence of a mechanism having a Bayesian equilibrium yielding the FSE outcome.

THEOREM 2 (CM2). *An information structure (S, π) guarantees full extraction of the surplus by a Bayesian auction if and only if for each agent i there does not exist $s^i \in S^i$ and a family $\{\rho^i(t^i)\}_{t^i \in S^i \setminus s^i}$ such that*:

(a) $\rho_i(t^i) \geqslant 0$ *for all* $t^i \in S^i \setminus s^i$, *and*

(b) $\pi(s^{-i} \mid s^i) = \sum_{t^i \neq s^i} \rho^i(t^i)\, \pi(s^{-i} \mid t^i)$ *for all* $s^{-i} \in S^{-i}$.

The condition is similar to the one for dominant strategy implementation, but we now allow for the lottery to depend on the announcement of agent i. It assures that for each agent i and type s^i it is possible to build a lottery, possibly depending on the announcement of agent i, such that truthtelling is optimal (i.e., by announcing a different type the agent is never better off) and the expected payoff is zero.

When the condition is satisfied, the Farkas' Lemma ensures that for each i we can find a function $h_i(s)$ such that

$$\sum_{s^{-i} \in S^{-i}} \pi(s^{-i} \mid s^i)\, h_i(s^i, s^{-i}) = 0, \tag{2}$$

$$\sum_{s^{-i} \in S^{-i}} \pi(s^{-i} \mid t^i)\, h_i(s^i, s^{-i}) < 0. \tag{3}$$

Define now

$$H_i(s) = \sum_{t^{-i} \in S^{-i}} \pi(t^{-i} \mid s^i)\, p^i(s^i, t^{-i})\, w^i(s^i) - \theta^i(s^i)\, h_i(s)$$

where p^i are the function obtained when the efficient allocation is implemented and each $\theta^i\,(s^i)$ is a real member. Then 2 implies that, when the other agents are truthtelling, announcing the truth yields:

$$\sum_{t^{-i} \in S^{-i}} \pi(t^{-i} \mid s^i)[\,p^i(s^i, t^{-i})\, w^i(s^i) - H_i(s^i, s^{-i})] = 0.$$

Notice that this is true independently of the choice of $\theta^i\,(s^i)$. This implies that we can choose the $\theta^i\,(s^i)$ large enough so that

$$\sum_{t^{-i} \in S^{-i}} \pi(t^{-i} \mid t^i)[\,p^i(s^i, t^{-i})\, w^i(t^i) - H_i(s^i, s^{-i})] \leqslant 0$$

for each $t^i \neq s^i$.

ACKNOWLEDGMENTS

I thank participants to the conference on "New Directions in Market and Games" held in Toronto in October 1995 and to the XXI Simposio de Análisis Económico held in Barcelona in December 1996 for helpful comments. I remain responsible for errors.

REFERENCES

1. S. Brusco, "Perfect Bayesian Implementation in Economic Environments," Working Paper 322.95, IAE–UAB working paper series, 1995.
2. J. Crémer and R. McLean, Full extraction of the surplus in Bayesian and dominant strategy auctions, *Econometrica* **56** (1988), 1247–1257.
3. J. Duggan, "An Extensive Form Solution to the Adverse Selection Problem in Principal/Multi-Agent Environments," Mimeo, Queen's University, 1996.
4. J. Glazer and C. A. Ma, Efficient allocation of a "Prize"-King Solomon's dilemma, *Games Econ. Behav.* **1** (1989), 222–233.
5. M. O. Jackson, Bayesian implementation, *Econometrica* **59** (1991), 461–477.
6. M. O. Jackson and H. Moulin, Implementing a public project and distributing its cost, *J. Econ. Theory* **57** (1992), 125–140.
7. H. R. Varian, A solution to the problem of externalities when agents are well-informed, *Amer. Econ. Rev.* **84** (1994), 1278–1293.