Soc Choice Welfare (2000) 17: 69–83

Social Choice and Welfare

Collusion, renegotiation and implementation

Sandeep Baliga*, Sandro Brusco

Kellogg Graduate School of Management, 2001 Sheridan Road, Evanston, IL 60208-2009, USA (e-mail: baliga@nwu.edu)
Universidad Carlos III de Madrid, Departamento de Economía de la Empressa, Calle Madrid 126, E-28903 Getafe (Madrid), Spain (e-mail: brusco@emp.uc3m.es)

Received: 27 August 1997 / Accepted: 29 October 1998

Abstract. We study the implementation problem for exchange economies when agents can renegotiate the outcome assigned by the planner and can collude. We focus on the use of sequential mechanisms and present a simple sufficient condition for implementation with renegotiation in strong perfect equilibrium. We present an application to optimal risk sharing, showing that the possibility of collusion and renegotiation does not in general prevent the implementation of efficient allocations.

1 Introduction

Suppose agents share some information, the state, that is unverifiable to an outside party, the planner, and that this information is needed to choose a socially optimal outcome. The planner can design a mechanism, consisting of strategy sets for the agents and an outcome function mapping messages to outcomes, to elicit the state and choose the optimal outcome in equilibrium. We study this *implementation problem* in exchange economies when agents can collude. We allow two types of collusion:

(a) *interim collusion*: strategic coordination by agents within a mechanism;
(b) *ex post collusion (renegotiation)*: renegotiation of the outcome finally identified by a mechanism.

We show that under mild assumptions, interim collusion does not seriously limit the types of social choice functions that can be implemented in exchange economies. It is the force of renegotiation that restricts implementable social

* To whom correspondence should be addressed.

choice functions. We utilize an extensive form mechanism which does not include constructions such as "integer" or "modulo" games, and present an example of optimal risk sharing that shows that the possibility of collusion and renegotiation does not in general prevent the implementation of efficient social choice functions.

The literature on incomplete contracts and the theory of the firm (see Hart [6] for a survey) studies bilateral trade when agents can renegotiate. Canonical results in that literature are not robust to the introduction of third parties. Hart [6] suggests that collusion can, in turn, compromise the use of third parties. We show that this is not true in the model we study and the introduction of third parties can affect the set of social choice functions implementable with renegotiation.

There is one strand of the implementation literature that analyzes the impact of renegotiation and another of collusion. Maskin and Moore [8] introduce the concept of a renegotiation function to capture the exogenous process by which inefficient outcomes chosen by a mechanism are renegotiated. They show that necessary and sufficient conditions for Nash and subgame perfect implementation can simply be translated into their framework by applying the renegotiation function to the standard results. Sjöström [11] shows that in exchange economies the set of Nash implementable social choice rules is equivalent[1] to the set of strong Nash implementable ones. However, he allows the planner to throw away all the goods in the economy following certain messages sent by the agents. This is clearly sensitive to renegotiation. Dutta and Sen [5] provide a necessary and sufficient condition for strong Nash implementation in general environments and extend results in Maskin [7].

The literature on strong Nash implementation only considers normal form mechanisms. In this paper we impose a requisite of subgame perfection on the equilibrium notion, and study implementation with multistage, or sequential, mechanisms. Sequential mechanisms are very useful in economic environments in which agents are able to collude. We will show that, under mild assumptions, sequential mechanisms allow the planner to eliminate 'incentive compatibility for subcoalitions' which has to be taken into account when we use normal form mechanisms.

The rough idea is the following. When a normal form game is used for implementation, we have to make sure that no coalition of agents can profitably deviate from the 'correct' equilibrium. As the game is one shot, this implies that the outcome function must satisfy the property that no subcoalition has a profitable deviation. For example, if we are considering a revelation mechanism and all agents observe the state of the world we have to make sure that no coalition can gain[2] by reporting a different state when the complementary coalition is telling the truth. This problem can be easily dealt

[1] The only restriction is that the social choice function gives a positive quantity of at least one commodity to all agents in all states.

[2] We say that a coalition gains when each member of the coalition is strictly better off.

with in exchange economies when renegotiation is *not* possible. It is sufficient to threaten the agents with destruction of the goods when there is not a unanimous report about the state of the world (this is the construction employed in Sjöström [11]). This makes sure that only unanimous reports are possible, and then the problem of inducing truthtelling can be dealt with exactly in the same way as for Nash implementation.

Things are completely different when renegotiation is allowed. In this case the 'destroy all goods' threat can not be used, since any outcome in which any quantity of any good is destroyed will be renegotiated. The 'incentive compatibility conditions for coalitions' become stringent. We will show however that, under mild conditions, the problem can be eliminated by adopting sequential mechanisms.

The rest of the paper is organized as follows. Section 2 presents the general framework in which we discuss the problem of implementation under collusion and renegotiation. In Sect. 3 we deal with the case of two agents. While it is known that the two agent case poses special problems for implementation, it also allows us to ignore the collusion by subcoalitions of agents. The main result in this section is that the set of social choice functions is only restricted by renegotiation, not by collusion. Given renegotiation, a social choice function is implementable in strong perfect equilibrium if and only if it is implementable in subgame perfect equilibrium. Section 4 deals with the general case of three or more agents. We provide sufficient conditions for a social choice function to be implementable despite the possibility of collusion and renegotiation. The main condition is one of 'preference reversal', familiar from the subgame perfect implementation literature. The mechanism we propose has a finite number of stages and a finite strategy space at each stage. Our results are illustrated by an application in Sect. 5. Concluding remarks are in Sect. 6.

2 The general framework

There is a set of agents $\mathcal{N} = \{1, 2, \ldots, N\}$ with $N \geq 2$. The finite set of possible states of the world is Θ. There is a total endowment $\omega \in \Re_+^m$ of m perfectly divisible private goods. An allocation $x \in \Re^{mN}$ is a collection $x = (x^1, \ldots, x^N)$, where $x^i \in \Re^m$. A feasible allocation $x = (x^1, \ldots, x^N)$ must be such that $\sum_{i=1}^N x^i \leq \omega$. Let $X(\omega)$ be the set of feasible allocations given endowment ω where, by convention, we also include in this set all lotteries over feasible allocations. We denote by x_j^i the quantity of good j allocated to agent i under x. Each agent i has von Neumann – Morgenstern preferences defined over $\Re^m$ (her own consumption of the private goods). These preferences are state-dependent and are represented by a utility function $U^i(x^i, \theta)$. We assume that for each θ and i the utility function U^i is continuous and strictly increasing in each good.

The state of the world is observable to all the agents but unobservable to an outside party. A social choice function (scf) f is a function $f : \Theta \to X(\omega)$. We will denote by $f^i(\theta)$ the vector of goods assigned by the scf f to agent i

in state θ, and we say that a scf f is θ-efficient if it is Pareto-efficient in all states.

As the state is observable but unverifiable, a mechanism must be designed to elicit the state from the agents. We are concerned with collusion in such a mechanism.

2.1 The notion of collusion

Collusion can take effect at several points in the mechanism. *Interim collusion* takes place within a mechanism: it is the coordination of agents' strategies in order to obtain a certain outcome of the mechanism. *Ex post collusion* (or *renegotiation*) takes place after the mechanism has identified an outcome: it is the choice of an outcome other than the one identified by the mechanism through voluntary trade of the assigned allocations.

We will allow for interim and ex post collusion of a particular type: we assume that agents can collude inside the mechanism, meaning that a group of agents can agree to modify jointly their messages, but they cannot specify how they will ex post collude until the mechanism has identified an outcome. In particular, this structure does not allow a group of agents to agree on a trade ex post conditional on certain messages to be sent[3] (i.e. strategies of the type 'I will give you a certain amount after the outcome has been identified if you issue a given message'). The basic idea is that agents cannot enforce ex ante agreements to exchange goods in a given way conditional on the behavior taken in the mechanism.

A possible justification for this assumption is that the planner is the one who controls the legal system, and she may decide that contracts interfering with truthtelling are void and unenforceable. Agreements conditioning the transfer of the good on a particular outcome of the mechanism are simply not allowed. Since agents cannot rely on explicit contracts, they can only exchange implicit, self-enforcing promises to carry out certain trades after the mechanism has identified an outcome. But such promises are credible only if they would be accepted by the agents anyway once the uncertainty is resolved. No ex ante commitment is possible, and the situation is equivalent to the one we consider, i.e. the trading of the allocations is considered only after the outcome has been identified. This philosophy would also suggest considering a self-enforcing form of interim collusion such as perfect coalition-proof Nash equilibrium. However, since we are proving possibility results, not impossibility ones, we adopt strong perfect equilibrium as our solution concept.[4]

---

[3] Sjöström [11] is the first to make the point that a form of individual rationality must be imposed on ex post trade if agents cannot write comprehensive side-contracts. For an examination of the case where agents can write comprehensive contracts that include both interim and ex post collusion see Baliga [3].

[4] In fact, our results also go through for the notion of perfect coalition-proof Nash equilibrium.

After the mechanism has identified the outcome, however, the agents are free to renegotiate it. In particular, given any allocation x decided by the mechanism the agents can trade whatever (feasible) amount they desire. This assumption includes the fact that no good can be wasted, as agents would agree to redistribute among themselves any good designed for destruction.

2.2 Definitions and assumptions

We now define our notion of collusion more formally. We begin by introducing renegotiation. We will follow Maskin and Moore [8] in assuming that the outcome of the renegotiation process can be represented by a function $h : X(\omega) \times \Theta \to X\ (\omega)$. Thus, $h\ (x, \theta)$ denotes the allocation resulting after renegotiation[5] when the initial allocation is x and the state of the world is θ. When an allocation x is deterministic, we will write $h(x, \theta) = (h^1(x, \theta), \ldots, h^N(x, \theta))$, where $h^i(x, \theta) \in \Re^m$ denotes the consumption vector allocated to agent i after renegotiation has taken place. The following assumptions on h and agents' preferences will be maintained throughout the paper:

Assumption 1. *If an allocation $b = (b^1, \ldots, b^N)$ is Pareto efficient in state θ and there is an allocation $a = (a^1, \ldots, a^N)$ such that $U^i(a^i, \theta) \geq U^i(b^i, \theta)$ for all $i \in N$, then $b = a$.*

Assumption 2. *When allocation x is deterministic, $U^i(h^i(x, \theta), \theta) \geq U^i(x^i, \theta)$ for each x, θ and i.*

Assumption 3. *When allocation x is deterministic, $h(x, \theta)$ is Pareto efficient for each x and θ.*

These are similar to the assumptions made in Maskin and Moore [8].

Let us now come to collusion inside the mechanism. As mentioned in the introduction, we will focus on the use of extensive form games in the implementation problem. Hence, our notion of interim collusion is strong perfect equilibrium. Given a game of complete information, a strategy profile is a *strong Nash equilibrium* if no coalition of agents can deviate and make all its members strictly better off. Given an extensive form game Γ with complete information, we say that a strategy profile is a *strong perfect equilibrium* if for each proper subgame the strategy profile is a strong Nash equilibrium.[6] That is, strong perfect equilibrium incorporates a form of coalitional sequential rationality. Also, notice that the notion of equilibrium requires that a strategy profile at each stage be robust even against an alternative strategy profile by a group of agents which is not self-enforcing in following subgames. It is therefore demanding.

---

[5] An alternative to this approach is to specify the renegotiation process explicitly as an extensive form game. See Aghion, Dewatripont and Rey [2] and Rubinstein and Wolinsky [10] for a discussion of the relation between the two approaches.

[6] We refer the reader interested in a formal definition to Brusco [4].

We say a scf f is *strong perfect implementable with renegotiation h* (in short, *SPE-h* implementable) if there exists an extensive form game such that for each θ and for each strong perfect equilibrium in state θ the outcome is an allocation x such that $f(\theta) = h(x, \theta)$. We will say that a scf f is *subgame perfect implementable with renegotiation h* (in short, *spe-h* implementable) if there exists an extensive form game such that for each θ and for each subgame perfect equilibrium in state θ the outcome is an allocation x such that $f(\theta) = h(x, \theta)$.

We now establish some simple results which will be used in the rest of the paper.

Lemma 1. *Suppose that x is a Pareto efficient allocation under* θ. *Then* $h(x, \theta) = x$.

Proof. Immediate consequence of Assumptions 1 and 2. □

Lemma 2. *Suppose that x is such that* $x^i = \omega$ *for some i. Then* $h(x, \theta) = x$ *for each* θ.

Proof. Strict monotonicity of preferences in each state of the world implies that x^i is Pareto efficient for each θ. The result then follows from Lemma 1. □

We will use the notation $x(i)$ to denote the allocation where the whole endowment is given to agent i, i.e. $x(i)$ is the allocation such that $x^i(i) = \omega$ and $x^j(i) = 0$ if $j \neq i$.

3 The two agent case

In this section, we consider the case $N = 2$. We argue that, in our setting, a social choice function is implementable in strong perfect equilibrium if and only if it is implementable in subgame perfect equilibrium. The reason for the equivalence is that in the case $N = 2$ the only difference between the two solution concepts is that the grand coalition of the two players is allowed to deviate. However, renegotiation ensures that any subgame perfect equilibrium is Pareto efficient, which in turn implies that the grand coalition cannot deviate and improve the utility of both members. Hence, interim collusion has no bite in the two agent case. This simple observation is formally proven in the next theorem.

Theorem 1. *If* $N = 2$ *then a scf f is SPE-h implementable iff it is spe-h implementable.*

Proof. If $N = 2$ the notions of spe and SPE coincide for extensive form games with the property that for each strategy profile the outcome is Pareto-efficient. In this case it is impossible for the grand coalition to improve the utility of both players through a coalitional deviation. Thus, the extra requirement imposed by the SPE notion has no bite. We now only need to observe that

renegotiation ensures that the outcome is Pareto efficient for each strategy profile. □

We now report the necessary and sufficient conditions for subgame perfect implementation with renegotiation offered by Maskin and Moore [8] and refer the interested reader to their paper for proofs:

Theorem 2. *Assume that* $N = 2$.

1) If a scf f is implementable in Nash or subgame perfect equilibrium with renegotiation h, then the following holds: For all ordered pairs of preference profiles (θ, ϕ) *there exists a possibly random outcome* $a(\theta, \phi)$ *(in short, a) such that*:

$$\begin{aligned} &U^2(f^2(\theta), \theta) \geq U^2(h^2(a, \theta), \theta)(*) \quad \textit{and} \\ &U^1(f^1(\phi), \phi) \geq U^1(h^1(a, \phi), \phi)(**) \end{aligned} \tag{1}$$

If the Pareto frontier is linear in all states of the world, then a scf f is implementable in Nash or subgame perfect equilibrium with renegotiation h if for all ordered pairs of preference profiles (θ, ϕ) *there exists a possibly random outcome* $a(\theta, \phi)$ *satisfying* $(*)$ *and* $(**)$.
2) A scf f is spe-h implementable if there exists a random function $a : \Theta \to X\ (\omega)$ *and an outcome* $e \in X\ (\omega)$ *such that for all* $\theta \in \Theta$, *(I)* $h(a(\theta), \theta)$ *is Pareto optimal and* $h(a(\theta), \theta) = f(\theta)$; *(II) for all* ϕ *such that* $h(a(\theta), \phi) \neq f(\phi)$ *there exists an agent* $i = i(\theta, \phi)$ *and a pair of outcomes* $b(\theta, \phi)$ *and* $c(\theta, \phi)$ *such that*

$$U^i(h(b(\theta, \phi)), \theta) \geq U^i(h(c(\theta, \phi)), \theta) \quad \textit{and}$$

$$U^i(h(b(\theta, \phi)), \phi) < U^i(h(c(\theta, \phi)), \phi)$$

and (III)

$$U^i(x, \theta) > U^i(e, \theta) \quad i = 1, 2$$

for all $x \in B$ *where B is the union of all* $a(\pi)$, $\pi \in \Theta$ *together with all* $b(\pi, \pi')$ *and* $c(\pi, \pi'), \pi, \pi' \in \Theta$ *where all these are defined.*

We now turn to the more interesting case $N > 2$.

4 Three or more agents

The main difference between the 2 agent case and the general case is that when $N \geq 3$ the Pareto efficiency of the renegotiation process does not imply that a proper subset of agents cannot profitably deviate from a subgame perfect equilibrium strategy profile. Therefore, the equivalence between spe and SPE is not immediate. In this section, we provide a sufficient condition for SPE-*h*

implementation which is a mild strengthening of a necessary and sufficient condition for spe-h implementation. Therefore, interim collusion does not significantly reduce the set of implementable social choice functions in exchange economies.

Assumption 4. *For each pair of states (θ, ψ) it is possible to find at least two triplets $(i, a_{(i)}, b_{(i)})$ and $(j, a_{(j)}, b_{(j)})$, where $k \in \{i,j\}$ denotes agents and $(a_{(k)}, b_{(k)})$ are allocations such that*:

$$U^k(h^k(a_{(k)}, \theta), \theta) > U^k(h^k(b_{(k)}, \theta), \theta)$$

$$U^k(h^k(b_{(k)}, \psi), \psi) > U^k(h^k(a_{(k)}, \psi), \psi)$$

Furthermore, for each pair (θ, ψ) it is possible to select agents (i, j) so that $N \notin \{i, j\}$.

An agent j satisfying the assumption for the pair (θ, ψ) will be called "a test agent for the pair (θ, ψ)". Some condition on "preference reversal" is necessary for implementation with extensive form games (see Moore and Repullo [9] and Abreu and Sen [1]). We require that in fact there are at least *two* test agents for each pair (θ, ψ), and that one specific agent N be excluded from the set of possible test agents. Agent N can be interpreted as a "dummy" agent added to the environment to prevent collusion. Notice that Assumption 4 does not make reference at all to the scf f to be implemented.

We make two more assumptions[7] on the social choice function f.

Assumption 5. $f^i(\theta) > 0$ *for each θ and each $i \neq N$.*

Assumption 6. *Suppose that for a given θ there exists a ψ such that for each $i < N$ the following inequality holds*:

$$U^i(h^i(f(\psi), \theta), \theta) > U^i(h^i(f(\theta), \theta), \theta) \tag{2}$$

Then it is the case that for agent N the following inequality holds:

$$U^N(h^N(f(\theta), \psi), \psi) \leq U^N(h^N(f(\psi), \psi), \psi) \tag{3}$$

Assumption 5 says that in each state of the world the scf allocates a strictly positive quantity of some good to each agent except possibly agent N. This is a mild assumption, and if a scf f does not satisfy it, it is still possible to find a scf f' arbitrarily close to f (where 'close' means $\|f(\theta) - f'(\theta)\| < \varepsilon$ for each θ, with $\|\cdot\|$ being the Euclidean distance in R^{mN}) and satisfying the assumption. Assumption 5, together with the finiteness of the state space Θ, allows us to find a $\bar{p} > 0$ such that given any agent $i < N$ and pair of states (θ, ψ) the fol-

[7] In interpreting Assumption 5, recall that if $x = (x_1, \ldots, x_m)$ and $y = (y_1, \ldots, y_m)$ are vectors in R^m then $x > y$ means $x_i \geq y_i$ for each i and $x_i > y_i$ for some i.

lowing inequalities are satisfied:

$$\begin{aligned} &\bar{p}U^i(x^i(i),\theta) + (1-\bar{p})U^i(0,\theta) < U^i(f^i(\psi),\theta) \\ &(1-\bar{p})U^i(x^i(i),\theta) + \bar{p}U^i(0,\theta) > U^i(f^i(\psi),\theta), \end{aligned} \tag{4}$$

where recall that $x(i)$ denotes the allocation where the whole endowment is given to agent i.

The LHS in the first inequality is the expected utility of a lottery giving $x^i(i) = \omega$ with probability $\bar{p}$ and 0 with probability $(1-\bar{p})$. Assumption 5 ensures that it is possible to find $\bar{p} > 0$ small enough such that, no matter what allocation x is chosen, the expected utility of this lottery is lower, in any state of the world, than the utility of any other allocation which is prescribed by the scf. The LHS in the second inequality is the expected utility of a lottery giving $x^i(i)$ with probability $(1-\bar{p})$ and 0 with probability $\bar{p}$. Assumption 5 ensures that it is possible to find $\bar{p} > 0$ small enough such that the expected utility of this lottery is greater than the utility of any socially optimal outcome. These facts will be used in the construction of the mechanism.

Assumption 6 states that if it is the case that all agents except N are better off by obtaining allocation $f(\psi)$ rather than $f(\theta)$ when the state of the world is θ, then agent N is worse off: Notice that Pareto efficiency of the renegotiation function implies that when (2) is satisfied then $U^N(h^N(f(\psi),\theta),\theta) < U^N(h^N(f(\theta),\theta),\theta)$, so that agent N, if given the opportunity, will ask for $f(\theta)$ rather than $f(\psi)$. Inequality (3) implies that this opportunity will not be used when the state of the world is actually ψ. This allows us to build the mechanism in such a way that agent N can reveal a deception in which the first $N-1$ agents claim that the state of the world is ψ when it is actually θ. At the same time agent N has no interest in obtaining $f(\theta)$ when the state of the world is ψ. A simple case in which Assumption 6 is satisfied is when $N > 3$ and $f^N(\theta) = 0$ for each θ, i.e. agent N is a 'dummy' player added by the planner who does not really participate in the trading process. In this case, Pareto efficiency implies that inequality (2) cannot be satisfied for all $i < N-1$, so that Assumption 6 is automatically satisfied. We will expand on this observation in Remark 2 after the theorem.

Theorem 3. *If a scf f is θ-efficient, and satisfies assumptions 4, 5 and 6, then f is SPE-h implementable.*

Proof. We define the mechanism for implementation. Notice that since f is θ-efficient we have $f(\theta) = h(f(\theta),\theta)$ for all θ.

The mechanism we adopt for implementation is the following.

Stage 0 This stage is made up of N substages.

Substage 0.1 Agent 1 announces $\theta_1 \in \Theta$. Go to next substage.

Substage 0.2 Agent 2 announces $\theta_2 \in \Theta$. If $\theta_2 = \theta_1$ go to the next substage. Otherwise, go to Stage 1 and play the game $\Gamma(\theta_1,\theta_2,2)$.

. . . . . .

Substage 0.i Agent i announces $\theta_i \in \Theta$. If $\theta_i = \theta_1$ go to the next substage. Otherwise, go to Stage 1 and play the game $\Gamma(\theta_1, \theta_i, i)$.

. . . . . .

Substage 0.N-1 Agent $N-1$ announces $\theta_{N-1} \in \Theta$. If $\theta_{N-1} = \theta_1$ go to the next substage. Otherwise, go to Stage 1 and play the game $\Gamma(\theta_1, \theta_{N-1}, N-1)$.

Substage 0.N Agent N announces $\theta_N \in \Theta$. There are 3 cases:

1. If $\theta_N = \theta_1$ the game stops and $f(\theta_1)$ is implemented.
2. If $\theta_N \neq \theta_1$ and $U^N(h^N(f(\theta_N), \theta_1), \theta_1) \geq U^N(h^N(f(\theta_1), \theta_1), \theta_1)$ then $f(\theta_1)$ is implemented.
3. If $\theta_N \neq \theta_1$ and $U^N(h^N(f(\theta_N), \theta_1), \theta_1) < U^N(h^N(f(\theta_1), \theta_1), \theta_1)$ then $f(\theta_N)$ is implemented.

Stage 1 The game $\Gamma(\theta_1, \theta_i, i)$ is as follows: Let j be a test agent for the pair (θ_1, θ_i) such that $j \neq i$ and let a and b be two allocations such that

$$U^j(h^j(a, \theta_1), \theta_1) > U^j(h^j(b, \theta_1), \theta_1) \quad U^j(h^j(b, \theta_i), \theta_i) > U^j(h^j(a, \theta_i), \theta_i)$$

We distinguish two cases.

1. If $j = 1$ then 1 chooses between a and b. If a is chosen then $x(N)$ is implemented with probability $(1 - \bar{p})$ and a is implemented with probability $\bar{p}$. If b is chosen then $x(i)$ is implemented with probability $(1 - \bar{p})$ and b is implemented with probability $\bar{p}$.
2. If $j \neq 1$ the game is as follows: Agent 1 selects between $x(i)$ and $x(N)$, agent j selects between a and b. The outcome function is as follows:
   - With probability $(1 - \bar{p})$ the allocation selected by agent 1 is implemented; with probability $\frac{\bar{p}}{2}$ the allocation chosen by the test agent is implemented; for the remaining $\frac{\bar{p}}{2}$ probability, the allocation is $x(1)$ if 1 selected $x(i)$ and j selected b or if 1 selected $x(N)$ and j selected a, and $x(N)$ otherwise.

We show that this mechanism implements f. We first show that there exists a strong perfect equilibrium supporting the desired outcome. The strategy profile we propose is the following:

1. At Substage 0.i, with $i < N$, each agent tells the truth no matter what happened at the previous substages. At substage 0.N agent N tells the truth if all other agents have told the truth, and chooses a best response to the announcement of the remaining $N-1$ agents otherwise.
2. If a game $\Gamma(\theta_1, \theta_i, i)$ is played when the true state of the world is ψ, agent 1 always selects $x(i)$ if $U^j(h^j(b, \psi), \psi) > U^j(h^j(a, \psi), \psi)$ and $x(N)$ otherwise. The test agent j chooses her preferred allocation.

We now show that the proposed strategy profile is a strong Nash equilibrium at each subgame. Consider the game $\Gamma(\theta_1, \theta_i, i)$ first. Let us assume that the test agent j is not agent 1 (the other case can be treated similarly). Clearly,

no profitable deviation is available for the test agent j: she receives zero with probability $\left(1-\frac{\bar{p}}{2}\right)$ no matter what her choice, and the optimal strategy is to select the preferred allocation to be implemented with probability $\frac{\bar{p}}{2}$. As for gent 1, the best she can do given the expected choice of j is to 'match' the choice so that $x(1)$ is implemented with probability $\frac{\bar{p}}{2}$. Finally, observe that the coalition of agents 1 and j cannot strictly improve the utility of both agents since j is already obtaining the best possible outcome of the subgame.

Now consider Stage 0. Suppose the true state is θ. First observe that Pareto optimality implies that no deviation by the coalition of all agents is possible. First consider a deviation by the coalition of the first $N-1$ agents. If they agree on an untruthful report ψ then either at least one of them will not be better off or, by Assumption 6, agent N is able to increase her utility by announcing the truth and getting $f(\theta)$ implemented. This implies that it is not possible to obtain outcome $f(\psi)$ when the true state of the world is $\theta \neq \psi$. This in turn implies that any deviating coalition must reach Stage 1. Next observe that agent 1 cannot be part of any deviating coalition, since any outcome at Stage 1 is inferior to the equilibrium one. Thus, given that in the proposed strategy profile agent j chooses her most preferred outcome in Stage 1, agent 1 will always tell the truth at Stage 0 and select $x(N)$ whenever her report is challenged. This implies that any deviation from truthtelling at Stage 0 by any coalition excluding 1 invariably leads to $x(N)$ with probability $(1-\bar{p})$, and it cannot be profitable. Suppose now that, in Stage 0, all previous agents to agent $i>1$ have lied and have announced the state ψ when the true state is θ. Then, if agent i can persuade all following agents $j>i$ to also announce ψ, she can get a payoff of $U^i(f^i(\psi),\theta)$; in Stage 1, if she announces θ, given the strategy of the test agent for (ψ,θ) and agent 1, she can receive a payoff of at least $(1-\bar{p})U^i(x^i(i),\theta)$ which she prefers by the relation in (4). We conclude that no deviation from the proposed strategy profile is profitable, and it is a SPE.

We next show that this is the only SPE. It is enough to show that this is the only subgame perfect equilibrium (spe), since the set of strong perfect equilibria is contained in the set of subgame perfect equilibria. First observe that in any subgame perfect equilibrium whenever the game $\Gamma(\theta_1,\theta_i,i)$ is reached, the only spe for the subgame is that agent j chooses the preferred allocation and agent 1 'matches' the choice of j (i.e. she chooses $x(N)$ when j selects b and $x(i)$ when j selects a).

Consider now substage $0.N$. If all previous agents have told the truth then agent N can obtain a different allocation only by making himself worse off. We conclude that if the first $N-1$ agents have told the truth then the correct allocation will be implemented. Next observe that at Substage $0.i$, agent i must obtain outcome $x(i)$ with probability $(1-\bar{p})$ whenever the agents with a lower index have claimed a false state of the world. This can be achieved by announcing the true θ and reaching $\Gamma(\theta_1,\theta,i)$ (any other announcement such that the test agent strictly prefers b would also work) and this makes her strictly better off by (4). On the other hand, if all agents before agent i have told the truth then agent i will tell the truth. If she lies, Stage 1 is reached $x(N)$

is implemented with at least probability $(1-\bar{p})$, while by telling the truth, $f(\theta)$ is implemented (this is obvious if $i = N-1$ and follows by backward induction if $i < N-1$). This argument holds for agent 1 as well. We conclude that truthtelling is the unique spe, and, given our earlier argument, the unique SPE. □

The idea of the mechanism is simple. Each agent i can reveal whether the previous agents were telling a lie. Any such revelation is checked using the test agent. A choice of b indicates that i was right in claiming the previous agents were lying, so she is awarded $x(i)$, while a choice of a indicates that the claim was not correct, and agent i is punished by choosing allocation $x(N)$ with high probability which leaves her with an endowment of zero. The reason agent 1 is brought into the picture at stage 1 is to avoid collusion between two test agents. If the allocation selected at stage 1 depended only on the choice of the test agent then the following mixed strategy deviation from the truthtelling equilibrium would in principle be possible: Agent i and j agree that, with probability $\frac{1}{2}$, agent i claims a false state that makes j the test agent; in turn, j validates the lie (i.e. she makes a choice implying that agent i denouncement was correct). With the remaining $\frac{1}{2}$ probability the roles are reversed. This deviation yields 'almost' a lottery between $x(i)$ and $x(j)$ with equal probability[8], which may improve the welfare of the two agents. To break this kind of collusive agreement we introduce agent 1. Essentially, agent 1 'bets' on the choice of the test agent, begin rewarded with $x(1)$ with probability $\frac{\bar{p}}{2}$ when the choices coincide. This gives agent 1 the incentive to tell the truth whenever she expects the test agent to do the same, which must be the case in equilibrium. Since the allocation to be implemented depends "mostly" on the announcement of 1, this eliminates the possibility of bilateral collusion between two test agents. Notice that the mechanism is designed in such a way that agent 1 can never benefit from a deviation at Stage 0, since she is always worse off when Stage 1 is reached, which in turn implies that 1 will never join a coalitional deviation from the truthtelling equilibrium.

Remark 1. As previously discussed, we have ruled out deviations where a change of strategy is linked to an ex post transfer of goods. Otherwise, we should worry about all possible promises of reallocating the goods after the deviation. For example, agent 1 and 2 could collude to get $x(2)$ implemented with probability next to 1, supporting this collusive agreement with a promise to transfer to 1 a sufficiently large amount of goods in the renegotiation process. Also, there can be collusion between some agent i and agent N, where agent i sends the game to Stage 1 for a promise on the part of agent N that she will transfer most of the goods to her in the renegotiation game. If these kinds of comprehensive side contracts are possible, implementation is much more difficult.

[8] The 'almost' comes from the fact that the allocations are implemented with probability $(1-\bar{p})$ rather than with probability 1.

Remark 2. If we assume that $N > 2$ and $f^N(\theta) = 0$ for each θ (agent N is a 'dummy' agent) then the mechanism can be simplified by eliminating Substage $0.N$. In this case agent N is never asked to report, and in fact we don't even need to assume that she knows the state of the world. The role of agent N would be that of a pure 'money dump', an agent to whom the goods are given when the other agents do not agree on the state of the world. In fact, the presence of agent N would in this case substitute for the possibility of destroying the goods. Notice also that adding a third party to a two agent model means that a social choice function satisfying Assumptions 4 and 5 but not the conditions identified in part (1) of Theorem 2 is SPE-h implementable (recall that Assumption 6 is automatically satisfied when a dummy agent is added to the environment). Therefore, in our model, addition of third parties expands the set of implementable allocations under mild assumptions even though agents can collude.

Remark 3. It is obvious from the proof that the desired social choice function is also the unique outcome if we use the notion of subgame perfect equilibrium, rather than strong perfect equilibrium, to predict the outcome of the game. Therefore, our mechanism double implements f in subgame perfect equilibrium and strong perfect equilibrium. Moreover, double implementation in subgame perfect equilibrium plus the fact that a strong perfect equilibrium is also a perfect coalition-proof Nash equilibrium implies that we also implement the desired social choice function in perfect coalition-proof Nash equilibrium.

5 An application

We can now present a simple application to the case of risk sharing. Consider a group of $N - 1 > 2$ agents who have a fixed amount of resources $\omega \gg 0$ to share. The utility of agent i depends on her consumption $x^i \in \Re^m$ and on the realization of a random variable $\theta \in \Theta$, with Θ finite and $p(\theta) > 0$ for each θ. Furthermore, we assume that the marginal utility of the first unit of each good is infinity in each state of the world. More formally:

$$\lim_{x_j^i \to 0} \frac{\partial U^i(x^i, \theta)}{\partial x_j^i} = +\infty \quad \text{for each } \theta \in \Theta \text{ and } i \in N - 1. \tag{5}$$

Agents are risk-averse and they would like to write an ex ante contract guaranteeing optimal risk sharing. Let a scf f be a function which shares risk optimally ex ante and observe that (5) implies that Assumption 5 is satisfied. Furthermore, we add an extra agent who does not necessarily observe the state of the world and receives zero units of each good at every state, so that Assumption 6 is also satisfied. Finally, we have to make assumptions on the function h. We assume that the agents Nash bargain whenever they are faced with the possible implementation of an inefficient outcome. Therefore, $h(a, \theta)$ is the Nash bargaining solution for state θ when the *status quo* is allocation a.

In order to apply our Theorem 3 we have to make sure that Assumption 4 is satisfied. It turns out that the following assumption is sufficient.

Assumption 7. *Given any pair of states (θ, ψ) it is possible to find at least two triplets (i, a^i, b^i) and (j, a^j, b^j), where $k \in \{i, j\}$ denotes an agent and (a^k, b^k) are elements of $\Re^m$ (that is, allocations relative to agent k) such that:*

$$U^k(a^k, \theta) > U^k(b^k, \theta) \quad U^k(b^k, \psi) > U^k(a^k, \psi)$$

Furthermore, for each pair (θ, ψ) it is possible to select agents (i, j) so that $N \notin \{i, j\}$.

This is clearly a weak assumption. It is not related either to the function to be implemented or to the renegotiation function characterizing the problem, and it will always be satisfied when the U^i are continuous and for every pair of states (θ, ψ) there are at least two agents whose utility functions change across the two states.

To see that Assumption 7 implies Assumption 4 when h is given by the Nash bargaining solution, we prove a more general result.

Lemma 3. *Assume that the renegotiation function h satisfies the following property:*

$$U^i(a^i, \theta) > U^i(b^i, \theta) \Rightarrow U^i(h^i((a^i, \mathbf{x}^{-i}), \theta), \theta) > U^i(h^i((b^i, \mathbf{x}^{-i}), \theta), \theta)$$

for each i and θ.

Then Assumption 7 implies Assumption 4.

Proof. Let (i, a^i, b^i) and (j, a^j, b^j) be the two triplets identified by Assumption 7. Define $x_{(k)} = (x^k, \mathbf{0}^{-k})$ for $k = i, j$ and $x^k = a^k, b^k$, where, $\mathbf{0}^{-k}$ indicates that all agents other than k are assigned the vector 0. It is then immediate to check that the triplets $(i, a_{(i)}, b_{(i)})$ and $(j, a_{(j)}, b_{(j)})$ satisfy Assumption 4. □

The property stated in Lemma 3 is quite natural. It requires that when the initial allocations of agents other than i are left unchanged and agent i is given an initial allocation that she prefers, then agent i will be in a better position in the renegotiation process, and she will end up with an higher utility.

We now simply observe that the Nash bargaining solution satisfies the property stated in Lemma 3. We can therefore conclude that when Assumption 7 is satisfied optimal risk sharing can be achieved when the agents renegotiate the outcome according to the Nash bargaining solution.

6 Conclusion

We have analyzed the problem of implementation in an exchange economy under the assumption that agents can collude when they play the mechanism and can renegotiate the outcome, i.e. they can reopen trade after the goods have been assigned by the planner. It turns out that under assumptions slightly stronger than the ones adopted for subgame perfect implementation, it

is possible to implement a social choice function. Thus, collusion and renegotiation do not pose a serious problem to implementation. The mechanism we propose for implementation is a sequential one with a finite number of stages and a finite strategy space at each stage. Furthermore, the mechanism double implements the social choice function in strong perfect equilibrium and subgame perfect equilibrium.

References

[1] Abreu D, Sen A (1990) Subgame Perfect Implementation: A Necessary and Almost Sufficient Condition. J Econ Theory 50: 285–299
[2] Aghion P, Dewatripont M, Rey P (1994) Renegotiation Design with Unverifiable Information. Econometrica 62: 257–282
[3] Baliga S (1996) Universally Collusion-Proof Contracts. Mimeo, Northwestern University
[4] Brusco S (1997) Implementing Action Profiles when Agents Collude. J Econ Theory 73: 395–424
[5] Dutta B, Sen A (1991) Implementation under Strong Equilibrium. J Math Econ 20: 49–67
[6] Hart O (1995) Firms, Contracts and Market Structure. Oxford University Press, Oxford
[7] Maskin E (1985) The Theory of Implementation in Nash Equilibrium: A Survey. In: Hurwicz L, Schmeidler D, Sonnenschein H (eds) Social Goals and Social Organization. Cambridge University Press, Cambridge
[8] Maskin E, Moore J (1998) Implementation and Renegotiation. Forthcoming. Rev Econ Studies (forthcoming)
[9] Moore J, Repullo R (1988) Subgame Perfect Implementation. Econometrica 56: 1191–1220
[10] Rubinstein A, Wolinsky A (1992) Renegotiation-Proof Contracts and Time Preferences. Ame Econ Rev 82: 600–614
[11] Sjöström T (1994) Credibility and Renegotiation of Outcome Functions in Implementation. Jpn Econ Rev 47: 157–169