Order Independent Individual Rationality

Laurent Lamy*

20th November 2006

Abstract

We consider the implementation of an economic outcome under complete information when the principal cannot commit to a simultaneous participation game. From a class of sequential participation games, we introduce the concept of implementability under order independent individual rationality. We characterize the set of implementable mechanisms, which is possibly a non-convex set, and we solve the optimal design program: the principal raises a lower revenue but economic efficiency is not damaged.

Keywords: Mechanism Design, Individual Rationality, Imperfect Commitment, Collusion on participation

JEL classification: C72, D62

1 Introduction

The mechanism design paradigm considers that agents are taking their participation decisions simultaneously despite the fact that the principal in many transactions often lacks the ability to commit to close the participation at an exact deadline. In corporate acquisitions and procurement auctions, it is common that the seller violates the announced rules to accept a subsequent better deal. McAdams and Schwarz [14] and Vartiainen [19] consider auction models where the seller is unable to commit not to solicit another round of offers after having publicly disclosed the previous offers. Similarly, in the corruption literature, e.g. Compte et al. [4], the auctioneer may also provide an opportunity for bid readjustments in exchange for a bribe.

We consider the implementation of an economic outcome under complete information when the principal cannot commit to a simultaneous participation game. On the contrary we consider that feasible participation games are such that, sequentially, agents are asked either to accept to participate in the mechanism or to delay their participation decisions. After each acceptation of a given agent, the seller cannot refrain himself from giving a new chance

*Laboratoire d'Economie Industrielle, CREST-INSEE, 28 rue des Saints-Pères 75007 Paris. e-mail: laurent.lamy@ensae.fr

to participate to all the remaining agents which are still eligible for participation. As in the aforementioned positive literature, we consider that participation decisions are publicly observable. A mechanism is implementable if full participation is the only equilibrium of any participation game. In the same spirit as Moldovanu and Winter [16], we require implementation to be independent of the specific structure of the participation game. Following their terminology, the traditional individual rationality constraints are strengthened by requiring *order independent individual rationality*. Thus implementation requires more than the traditional condition that participation is a best-response for agent i given that all the other agents participate. Proposition 5.1 states that order independent individual rationality requires that there is no set of agents $S \subset N$ such that all agents in S prefer the outcome where only agents in $N \setminus S$ participate to the outcome where all agents accept the mechanism. Those new constraints are non-linear and the set of implementable mechanisms is thus in general not convex. Nevertheless, the optimal design program can be simplified as done in Proposition 5.3, our main result: it allows us to separate the choice of the final allocation to the structure of the optimal threats. As under a simultaneous participation game, we obtain that the optimal mechanism is efficient.[1]

Our sequential participation game can be also interpreted as a minimal collusive device for the agents. The main contributions on collusion-proof implementation [12, 13, 3] preclude any collusion on the participation decisions themselves and restrict the collusive activity to the reports. In this literature, the collusion technologies allow agents to fully contract (with monetary transfers) their reports to the principal. Surprisingly, Che and Kim [3] show that optimal noncollusive mechanism can be made collusion-proof in a broad class of circumstances including economic environment with (allocative) externalities. Here our collusive device is much weaker: neither monetary transfers nor binding agreements on the reports are available. Nevertheless, it consists in a form of collusion that includes the participation decisions. We show that in general, except when the framework is negative-externality free, the principal raises a lower revenue at the optimal design under this device. It contrasts with the insights of Pavlov [17] and Che and Kim [2], where the collusion mechanism proposed by a third party takes place before the participation decisions, and where the second best is still implementable with collusion.[2] Those papers consider the auction of a single item in the independent private value framework and thus exclude any kind of external-

[1] Under a simultaneous participation game and complete information, Jehiel et al. [10] shows that there is no loss of generality to consider the stronger dominant strategy implementation concept.

[2] In this line, Dequiedt [6] is an exception: in a binary type environment, he shows that asymmetric information do not prevent bidders to collude efficiently, i.e. to act as a single agent when the third party can manipule the participation decisions.

ity. We thus shed some light on the impact of collusion on participation -any stronger collusive device as the ones in [6, 17, 2] above would only strengthen our results- independently of any informational asymmetry.

The paper is organized as follows. In section 2 we introduce the general allocation problem. Using a simple example, a single item with identity-dependent externalities, section 3 illustrates the main idea of our critic of the traditional mechanism design approach. In section 4 we describe a general class of noncooperative sequential participation games. In section 5 we define our main concept- *order independent individual rationality* -and prove the main results. In section 6 we provide two general examples where our alternative mechanism design approach may be relevant and change some insights. Concluding remarks are gathered in section 7.

2 The Model

Let $N = \{1, 2, \ldots, n\}$ be a set of agents and $A = \{a_1, a_2, \ldots, a_K\}$ be a finite set of possible outcomes. Denote by $\Sigma(N)$ the set of the permutations over the set N. For a given permutation $\sigma : N \to N$, denote by T_i^σ the subset $\{\sigma(1), \sigma(2), \ldots, \sigma(i-1)\}$, i.e. the $i-1$ first smallest agents according to the implicit order defined by σ. We assume that the agents and the principal, characterized by the subscript 0, have quasilinear preferences over outcomes and (divisible) money. Preferences are assumed to be common knowledge. The utility of a player i over outcome $a \in A$ and the money transfer $\mathbf{t}_i$ is:

$$\mathcal{U}_i(a, \mathbf{t}_i) = V_i^a - \mathbf{t}_i.$$

We first describe the class of procedures among which the principal chooses an optimal mechanism. In step 1, the principal designs a mechanism. In a complete information setting, a mechanism, denoted by $(\mathbf{a}, \mathbf{t})$, specifies a final outcome $\mathbf{a}(S)$ and a vector of monetary transfers $\mathbf{t}(S)$ for each possible set of participants $S \subset N$. In step 2, the agents are playing a sequential participation game described in next section. In the previous mechanism design literature, the decisions whether to participate or not in the proposed mechanism are assumed to be taken simultaneously. Here we consider that the principal cannot commit to such a simultaneous participation game: an agent will always have at least one opportunity to participate in the mechanism after each decision to accept the mechanism by an agent. In step 3, the mechanism is implemented according to the participation set $S \subset N$. A mechanism is said to be *feasible* if:

- For each set of participants S, the final outcome belongs to $\mathcal{A}(S)$, the subset of A of accessible or feasible outcome with the consent of agents in S.

- If agent i decides not to participate the principal cannot extract a positive payement from that agent: $\mathbf{t}_i(S) \leq 0$, if $i \in N \setminus S$.
- Transfers are budget-balanced: $\sum_{i=0}^{n} \mathbf{t}_i(S) = 0$, for any $S \subset N$

The second and third restrictions are standard. The first restriction means that some outcome in A may not be feasible if some agents refuse to participate. For example, in the case of the sale of an indivisible good, Jehiel et al. [10] considers that one cannot 'dump' the object on a non-participating agent. We do not impose any specific structure to the feasibility sets $\{\mathcal{A}(S)\}_{S \subset N}$ except that:

Assumption 1 *$\mathcal{A}(S) \subset \mathcal{A}(T)$, whenever $S \subset T$.*

Assumption 1 states that if the consent of the agents in S is enough to implement a given final outcome a, then the extra consent of some agents outside S cannot make this outcome unfeasible. Then, there is no loss of generality to consider that $\mathcal{A}(N) = A$. To simplify the exposition, we assume that, for a given utility level, an agent strictly prefers to participate in the mechanism. With this trick, the set of implementable mechanisms - which is defined in section 5 -is a closed set and has thus an optimal element.

For an agent i and a set of participant $S \subset N \setminus \{i\}$, denote by $a_i^*(S)$ the harsher feasible threat that the principal can inflict on i given that the agents in S have accepted the mechanism: $a_i^*(S) \in Arg\min_{a \in \mathcal{A}(S)} V_i^a$. Denote by $V_i^*(S) = V_i^{a_i^*(S)}$ the corresponding utility level. In mechanism design under simultaneous participation, only the threats $a_i^*(N \setminus \{i\})$ do matter. In the optimal design, if one agent refuses the mechanism, the remaining ones commit to this harsher threat also called 'minmax punishment' as in Jehiel et al. [10] or Dequiedt [6]. On the other hand, in mechanism design under order independent individual rationality, the whole set of the feasible threats $a_i^*(S)$ will play an active role in the design of the optimal mechanism.

Finally, our framework is characterized by the 4-uple: $(N, A, \{V_i^a\}_{i \in N, a \in A}, \{\mathcal{A}(S)\}_{S \subset N})$. Let us define two special subsets among those frameworks: *externality-free* and *negative-externality-free* frameworks.

Definition 1
- *A framework is said to be* externality-free *if for any agent i, the map $a \to V_i^a$ is constant over the set $\mathcal{A}(N \setminus \{i\})$.*
- *A framework is said to be* negative-externality-free *if the optimal threat $V_i^*(S)$ for any agent i is independent of the set of participant $S \subset N \setminus \{i\}$: $V_i^*(S) = V_i^*(\emptyset)$ for any i.*

A framework is said to be externality-free if the agents do not care about the final outcome in the event where they do not participate in the mechanism. For the sale of some goods and under the assumption that a non-participant does not receive any good, it corresponds to the standard case

where agents care only on the set of goods they obtain and in particular are indifferent to the final allocation when they are non-purchaser. Negative-externality-free is less restrictive: it only requires that the principal can credibly threat any agent with the minmax punishment independently to the other participants, i.e. by retaining all goods in the above example.

3 A Simple Example

The following example formalizes the starting examples of Jehiel and Moldovanu [9] and Das Varma [5] where two potential buyers suffering from important reciprocal negative externalities prefer not to participate in the bidding process for a single item and let a third buyer win at a low price.

Let $n = 3$ and $A = \{0, 1, 2, 3\}$ where allocation i corresponds to the allocation of the item to player i. We consider that the seller is able to allocate the item only to participating agents: $\mathcal{A}(S) = \{i | i \in S \cup \{0\}\}$. Let V_i^i be equal respectively to V, v and 0 for $i \in \{1, 2\}$, $i = 3$ and $i = 0$. Let V_i^j be equal to $-\alpha$ if $i, j \in \{1, 2\}$, $i \neq j$ and 0 otherwise. Assume that $V > v > V - \alpha > 0$. Thus the efficient allocation consists in allocating the item to agent 3. Nevertheless, agents 1 and 2 are valuing the item more than agent 3. They are also chosen symmetric only to simplify the exposition. The same kind of results holds in the neighborhood of the parameter values or with huge asymmetries between agents 1 and 2 provided that the reciprocal negative-externalities between them are big enough.

Standard Auctions Consider first a simultaneous participation game as in [9]: the buyers have first the opportunity to decide whether or not they want to participate in the auction. Those decisions are made simultaneously and are publicly revealed before the auction takes place. We consider the first price auction, but the results are similar for any other standard auction as the English button auction considered in [5]. In any equilibrium, the item is sold either to agent 1 or to agent 2. In the unique symmetric equilibrium, agents 1 and 2 both participate with probability 1 and are submitting the bid $V + \alpha$. They are both suffering from a loss of α compared to their profit in the case where they could jointly coordinate themselves not to participate. In our example, non-participation from agent 1 is vain and cannot prevent the purchase by agent 2 in the auction because $V > v$.[3]

Now consider a sequential participation game with agents 1 and 2 such that potential buyers are always proposed to participate after one has decided to participate (the proper formalization is done in section 4). Now it is a subgame perfect equilibrium for agents 1 and 2 not to participate if and only if his 'feared' opponent did so. Under sequential participation, we obtain the paradox that seems to correspond to the stories reported in [9, 5] and

[3] Strategic non-participation as in [9] emerges only if $v > V$ and thus not here.

that cannot emerge in previous models with simultaneous participation: an agent may prefer not to submit a bid though his intrinsic value for the good, i.e. excluding the motivations to outbid resulting from the fear of negative externalities, is greater than the final bid.

The Optimal Mechanism Under a simultaneous participation game, Jehiel et al. [10] presents an optimal mechanism where participation is a strictly dominant strategy. The optimal mechanism is efficient and the seller can extract surplus from agents who do not obtain the object by using the optimal threats $a_i^*(N\setminus\{i\})$ for each agent i. Here the efficient allocation is to allocate the object to agent 3 and the optimal mechanism raises the revenue $v+2\cdot\alpha$: each non-purchaser has to pay α in order to avoid that the seller gives the object to his most feared opponent. But what can be implemented if agents 1 and 2 can coordinate their participation decision thanks to a sequential participation process? Then the seller can not allocate the object to agent 3 and extract a strictly positive surplus from both agents 1 and 2. In particular, she cannot threat simultaneously agents 1 and 2 with their respective tougher threat. Otherwise, they could jointly not participate and obtain a null payoff since the seller is assumed to be unable to 'dump' the object. To maximize her revenue, the seller should use a *divide and conquer* strategy: it consists in giving the incentive to participate for one agent, say 1, independently of the participation decision of agent 2. Then given that agent 1 participates, she could really threat agent 2 to allocate the object to agent 1 in case of non-participation. Indeed we will show that it is the optimal mechanism and it raises the revenue $v+\alpha$. It illustrates several features that are generalized in section 5: first, the optimal selling procedure is still efficient under the *order independent individual rationality* constraints; second, those constraints reduce the revenue. Finally, we find surprisingly that although agents 1 and 2 are symmetric, they should not be treated in a symmetric way in an optimal mechanism. That is the reason why standard auctions that are intrinsically symmetric were leading to joint non-participation.

4 A Participation Game

We describe a simple sequential participation procedure based on a given mechanism $(\mathbf{a},\mathbf{t})$. Suppose that the agents in S have already accepted the mechanism, then the remaining agents are playing a participation game where each agent has a least once the possibility to accept the mechanism or to delay his decision. If all those agents do not accept the mechanism, then the participation game stops and the outcome $(\mathbf{a}(S),\mathbf{t}(S))$ is implemented. Otherwise, if at least one agent i accepts the mechanism, then his acceptance is followed by a participation game given the consent of $S\cup\{i\}$. Observe that, in this informal described situation, the order according to

which agents are approached to accept or not the mechanism has not been specified. Indeed, each order generates a different extensive form game. We wish to compare final outcomes in these different games, and therefore we proceed to a formal description of the participation games.

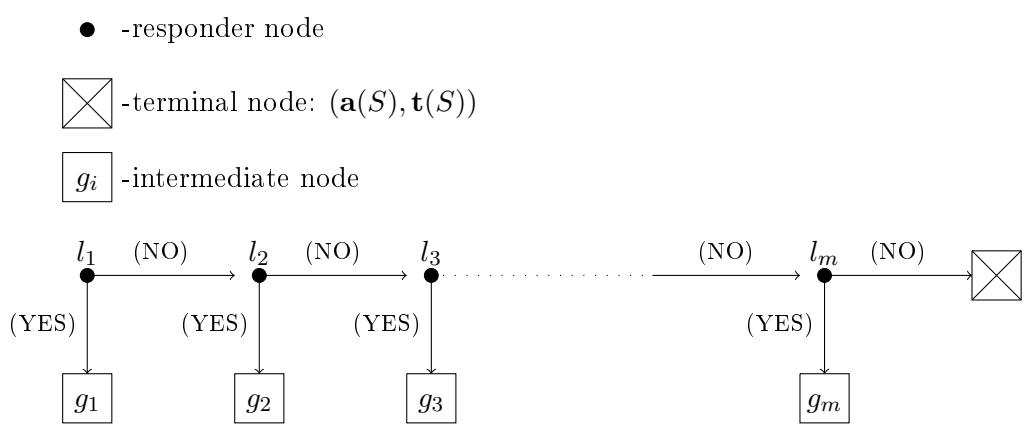

Figure 1

For a given mechanism $(\mathbf{a}, \mathbf{t})$, we define recursively the set of participation games as a function of the cardinality of the set of the agents that have already accepted the mechanism. We denote by $\mathcal{G}(\mathbf{a}, \mathbf{t}, S)$ the set of participation games if the agents in S have accepted the mechanism. If $S = N$, this set corresponds to the (unique) degenerate game where agents make no choice and the final outcome $(\mathbf{a}(N), \mathbf{t}(N))$ is implemented. If $S \subsetneq N$, we consider a participation game $g = ((\mathbf{a}, \mathbf{t}), S, \{l_i\}_{i=1,\dots,m}, \{g_i\}_{i=1,\dots,m})$ where $(\mathbf{a}, \mathbf{t})$ is a feasible mechanism, S is the set of the agents that have previously accepted the mechanism, $\{l_i\}_{1,\dots,m}$ is an ordered list of agents such that for any $j \in N \setminus S$, there exists $i^* \in \{i = 1, \dots, m\}$ such that $j = l_{i^*}$ and g_i is a participation game in $\mathcal{G}(\mathbf{a}, \mathbf{t}, S \cup \{l_i\})$ which is properly defined by the induction hypothesis. See Figure 1.

There are three kinds of positions in $g \in \mathcal{G}(\mathbf{a}, \mathbf{t}, S)$:

1. Responder nodes of the form (l_i, S), where $S \subset N$ is the set of the agents that have previously accepted the mechanism and $l_i \in N \setminus S$ is the identity of the agent with the initiative.

   If $S \subsetneq N$, $\{l_i\}_{i=1,\dots,m}$ and $\{g_i\}_{i=1,\dots,m}$ are not empty and the responder node (l_1, S) is called the *initiator* of the participation game.

2. Intermediate nodes of the form g_i, where g_i is a participation game in $\mathcal{G}(\mathbf{a}, \mathbf{t}, S \cup \{l_i\})$.

3. Terminal nodes of the form $(\mathbf{a}, \mathbf{t}, S)$ where S is the set of the agents that have previously accepted the mechanism $(\mathbf{a}, \mathbf{t})$.

At an intermediate node g_i, agents have no choice and the game moves to the initiator of the game g_i or moves to the terminal node $(\mathbf{a}, \mathbf{t}, N)$ if all agents give their consent. At a terminal node $(\mathbf{a}, \mathbf{t}, S)$, the game ends and the outcome $(\mathbf{a}(S), \mathbf{t}(S))$ is implemented.

At any responder position (l_i, S) there is the choice:

1. (l_{i+1}, S) if $i < m$ which means that agent l_i delays participation and l_{i+1} becomes the new responder. It corresponds to the two first arrays (NO) at the left of Fig. 1.

2. $(\mathbf{a}, \mathbf{t}, S)$ if $i = m$ which means that agent l_i refuses participation and the game ends at this terminal node. It corresponds to the array (NO) at the extreme right of Fig. 1.

3. g_i which means that agent l_i accepts the mechanism and the game moves to the intermediate node g_i. It corresponds to the arrays (YES) in Fig. 1.

The closure of the participation game after a finite number of delays may seem incoherent with our paradigm that agents do not decide whether they accept or reject the mechanism but rather that they have to decide either to accept or to delay their acceptance decision. Indeed it is left to reader to check that our following analysis is unchanged with the related infinite participation games, i.e. if $\{l_i\}_{i=1,\ldots,m}$ with $m = \infty$ is an infinite ordered list of agents, where an infinite delay for the agents in $N \setminus S$, given that agents in S gave their consent, results in the implementation of the outcome $(\mathbf{a}(S), \mathbf{t}(S))$. Contrary to Moldovanu and Winter [16]'s analysis which is confined to pure stationary equilibria, such an analysis would rely on subgame perfect equilibria in its whole degree of generality. Participation games with a finite number of nodes have been chosen to ease the backward induction argument and the presentation.

5 Optimal Design under Order Independent Individual Rationality

We now define a rationality constraint that removes, in our framework, the dependence of the participation decision on the exact structure of the participation game.

Definition 2

- *A mechanism* $(\mathbf{a}, \mathbf{t})$ *is* order independent individually rational *if* $(\mathbf{a}(N), \mathbf{t}(N))$ *is the final outcome in any subgame perfect equilibrium of any participation game* $g \in \mathcal{G}(\mathbf{a}, \mathbf{t}, S)$.

- *A mechanism is* implementable *if it is feasible and order independent individually rational.*

- *A mechanism* $(\mathbf{a}, \mathbf{t})$ *is* implementable under simultaneous participation *if it is feasible and if* $(\mathbf{a}(N), \mathbf{t}(N))$ *is the final outcome in any equilibrium under simultaneous participation.*

Note the difference of our order independent concept with the 'order independent equilibrium' concept of Moldovanu and Winter [16]. In a nutshell, [16] considers *strategy profiles* that are *an* equilibrium independently of the specific structure of their coalition formation game. Here we consider *final outcomes*, more precisely the outcome $(\mathbf{a}(N), \mathbf{t}(N))$ derived from full participation, that are the equilibrium outcome of *any* equilibrium independently of the specific structure of the participation game. We are indeed more demanding by requiring that this final outcome $(\mathbf{a}(N), \mathbf{t}(N))$ is the only equilibrium outcome for any participation game after that some agents give their consent.

Proposition 5.1 *A mechanism* $(\mathbf{a}, \mathbf{t})$ *is implementable if and only if it is feasible and for any* $S \subset N$

$$\max_{i \in N \setminus S} \{V_i^{\mathbf{a}(N)} - \mathbf{t}_i(N) - V_i^{\mathbf{a}(S)}\} \geq 0 \tag{1}$$

Proof 1 *We first prove the 'Only if' part. Suppose that* $(\mathbf{a}, \mathbf{t})$ *is implementable and that there exists a subset* S *such that* $V_i^{\mathbf{a}(N)} - \mathbf{t}_i(N) - V_i^{\mathbf{a}(S)} < 0$ *for any agent* $i \in N \setminus S$*. Then consider a subgame* $g \in \mathcal{G}(\mathbf{a}, \mathbf{t}, S)$*. At the node* (l_m, S)*, the responder's best response is to refuse the mechanism (if he accepts, the only equilibrium outcome is full participation since we have assumed that* $(\mathbf{a}, \mathbf{t})$ *is implementable). By backward induction, each responder's best response at the nodes* (l_i, S) *is to delay and move to* (l_{i+1}, S)*. Consequently, any subgame perfect equilibrium of the game* g *leads to the non-participation of the agents in* $N \setminus S$ *which raises a contradiction.*

The sufficiency part is proved by induction on the cardinality of the set of the agents that have already accepted the mechanism. The initial step where this set has the cardinality n *is immediate. Now consider that all agents in* $S \subsetneq N$ *have accepted the mechanism and suppose that* $\max_{i \in N \setminus S} \{V_i^{\mathbf{a}(N)} - \mathbf{t}_i(N) - V_i^{\mathbf{a}(S)}\} \geq 0$*. By the induction hypothesis, we obtain that every agents accept the mechanism in any subgame perfect equilibrium of any subgame* $\{g_i\}_{i=1,\ldots,m}$ *of the participation game* $g \in \mathcal{G}(\mathbf{a}, \mathbf{t}, S)$*. It remains to show that, for any game* $g \in \mathcal{G}(\mathbf{a}, \mathbf{t}, S)$*, it cannot belong to any equilibrium path that all agents refuse the mechanism at the responder nodes* $\{l_i\}_{i=1,\ldots,m}$*. In such a case, the agent* i *such that* $V_i^{\mathbf{a}(N)} - \mathbf{t}_i(N) - V_i^{\mathbf{a}(S)} \geq 0$ *has a profitable deviation: he accepts (with probability one) the mechanism when he is the responder, i.e. for a responder node such that* $l_k = i$*, which exists from the structure of the participation game.*

The inequality (1) with $S = N \setminus \{i\}$ corresponds to the standard individual rationality constraint of agent i in the standard mechanism design approach under a simultaneous participation game. Thus the lack of commitment in the participation game results -as expected- in a limitation of the set of implementable mechanisms. On the other hand, in an externaltity-free framework, the standard individual rationality constraints $V_i^{\mathbf{a}(N)} + \mathbf{t}_i(N) \geq V_i^{\mathbf{a}(N\setminus\{i\})}$ imply that $V_i^{\mathbf{a}(N)} + \mathbf{t}_i(N) - V_i^{\mathbf{a}(S)} \geq 0$ for any S and any $i \in N \setminus S$, the inequalities (1) are thus satisfied. Those points are summed up in the following corollary.

Corollary 5.2 *Any implementable mechanism is implementable under simultaneous participation. In an externaltity-free framework, the converse holds: a mechanism that is implementable under simultaneous participation is implementable.*

In the previous literature on mechanism design (with possibly incomplete information), the set of constraints that makes a mechanism implementable, i.e. feasibility, incentive compatibility and individual rationality constraints, results from inequalities that are linear according to the mechanisms $(\mathbf{a}, \mathbf{t})$.[4] Thus the set of the mechanisms that are implementable is a convex set. Moreover, the payoff of the principal depends linearly on the mechanism. From an optimal design perspective, there is thus no loss of generality to consider mechanisms that are symmetric if agents are symmetric. Suppose that a given asymmetric mechanism m is optimal. Then consider the permutations m_σ of this mechanism where $\sigma \in \Sigma(N)$. By symmetry, those mechanisms *implement* the same revenue for the principal. Finally, the mechanism $\frac{1}{n!}\sum_{\sigma\in\Sigma(N)} m_\sigma$ implements the same revenue in a symmetric way. On the contrary, the order independent individual rationality constraint results from inequalities involving the maximum of some linear maps and is thus not linear. Let us reconsider our simple example to illustrate the possible non-convexity of the set of implementable mechanisms.

Example 5.1 ***A simple example (suite)*** *Let* $\mathbf{a}(S) = 1$, $\mathbf{t}_1(S) = V$ *and* $\mathbf{t}_i(S) = 0$ *if* $i \neq 1$ *in the event where* $1 \in S$ *and* $2 \notin S$. *Let* $\mathbf{a}(S) = 0$, $\mathbf{t}_i(S) = 0$ *for any* $i \in N$ *in the event where* $1 \notin S$. *Let* $\mathbf{a}(S) = 3$, $\mathbf{t}_1(S) = \alpha$, $\mathbf{t}_2(S) = 0$ *and let* $\mathbf{t}_3(S) = v$, *if* $S = \{1, 2, 3\}$ *and* $\mathbf{a}(S) = 0$, $\mathbf{t}_1(S) = \alpha$, $\mathbf{t}_i(S) = 0$ *for any* $i \in N$ *in the event where* $S = \{1, 2\}$. *It is easily checked that this mechanism is feasible. Agents 1 and 3 obtain the same utility level independently of the final set of participants. Thus the inequalities (1) are satisfied if either 1 or 3 belongs to* $N \setminus S$. *Thus, it remains to check that*

[4] The implicit space structure according to which linearity applies is the following. For two mechanisms, $(\mathbf{a}, \mathbf{t})$ and $(\mathbf{a}', \mathbf{t}')$ and a real number $\lambda \in [0, 1]$, the mechanism $\lambda \cdot (\mathbf{a}, \mathbf{t}) + (1 - \lambda) \cdot (\mathbf{a}', \mathbf{t}')$ is the mechanism that implements the mechanism $(\mathbf{a}, \mathbf{t})$ (respectively $(\mathbf{a}', \mathbf{t}')$) with probability λ (resp. $(1 - \lambda)$).

the inequality (1) is satisfied if $S = \{1, 3\}$. *Finally, the mechanism* $(\mathbf{a}, \mathbf{t})$ *is implementable. The mechanism* $(\mathbf{a}', \mathbf{t}')$ *where the roles of 1 and 2 have been switched is implementable by symmetry. Now consider the mechanism where, at a terminal node, each mechanisms* $(\mathbf{a}, \mathbf{t})$ *and* $(\mathbf{a}', \mathbf{t}')$ *are implemented with probability one half. This mechanism is of course feasible. Nevertheless, it is not order independent individually rational. The constraint (1) with* $S = \{3\}$ *is violated. If agents 1 and 2 do not jointly participate, they obtain a null payoff. On the contrary, under full participation, their expected payoff is* $-\frac{\alpha}{2}$. *Indeed the (efficient) mechanism* $(\mathbf{a}, \mathbf{t})$ *is the optimal design as an application of proposition 5.3.*

There is no loss of generality to invite all agents to the mechanism since the set of feasible allocations does not shrink when some participants are added (Assumption 1). The optimal design program is thus:

$$\max_{(\mathbf{a},\mathbf{t})} V_0^{\mathbf{a}(N)} + \sum_{i=1}^{n} \mathbf{t}_i(N)$$

subject to

$$\forall S \subset N, \max_{i \in N \setminus S} \{V_i^{\mathbf{a}(N)} - \mathbf{t}_i(N) - V_i^{\mathbf{a}(S)}\} \geq 0,$$

where $(\mathbf{a}, \mathbf{t})$ is a feasible mechanism.

Nevertheless, in this form, the program is hardly tractable and it is unclear whether the optimal design is efficient. We simplify the program by showing that there is no loss of generality to restrict the maximisation to a subclass of implementable mechanisms which are fully characterized by a couple $(\alpha, \sigma) \in A \times \Sigma(N)$. Let us introduce a last useful notation: for a given set $S \subset N$ and a permutation $\sigma \in \Sigma(N)$, denote by $j(S, \sigma)$ the smallest agent according to the order σ that is not belonging to S. Formally, $j(S, \sigma) = \max \{j \in N | T_j^\sigma \subset S\}$. This agent plays a key role in the subclass that we define below and such that if the set of participants is S, the principal will inflict the minmax punishment to the agent $j(S, \sigma)$.

Definition 3 *For* $(\alpha, \sigma) \in A \times \sigma(N)$, *we define the* (α, σ)*- optimal threat mechanism as the mechanism* $(\mathbf{a}, \mathbf{t})$ *defined in the following way:*

- $\mathbf{a}(N) = \alpha$
- $\mathbf{a}(S) = a^*_{j(S,\sigma)}(S)$, *if* $S \subsetneq N$
- $\mathbf{t}_i(N) = V_i^\alpha - V_i^*(T^\sigma_{\sigma^{-1}(i)})$
- $\mathbf{t}_i(S) = 0$, *if* $S \subsetneq N$

Those mechanisms can be interpreted in the following way: take one agent, $\sigma(1)$, and give him the incentive to participate independently to the participation decision of the other agents by using the optimal threat among $\mathcal{A}(\emptyset)$; then take another agent, $\sigma(2)$, and give him the incentive to participate taken as given that $\sigma(1)$ surely participates and independently to the participation decisions of the other agents in $N \setminus \{\sigma(1)\}$ by using the optimal threat among $\mathcal{A}(\{\sigma(1)\})$; and so on. In particular, for the last agent, $\sigma(N)$, in this new order σ, the principal uses the optimal threat in $\mathcal{A}(N \setminus \{\sigma(N)\})$ as in the standard literature with simultaneous participation.

We first show that this restricted class of mechanisms is a subset of the implementable mechanisms.

Lemma 5.1 *Any (α, σ)- optimal threat mechanism is implementable.*

Proof 2 *It is immediately feasible by definition of $a^*_{j(S,\sigma)}(S)$ which is the minmax punishment for agent $j(S,\sigma)$ given the participation set S. Consider $S \subset N$ and the agent $j(S,\sigma)$ who does not belong to S. We have:*

$$V^{\mathbf{a}(N)}_{j(S,\sigma)} - \mathbf{t}_{j(S,\sigma)}(N) - V^{\mathbf{a}(S)}_{j(S,\sigma)} = V^*_{j(S,\sigma)}(T^{\sigma}_{\sigma^{-1}(j(S,\sigma))}) - V^*_{j(S,\sigma)}(S) \geq 0$$

The equality comes from the definition of $\mathbf{t}_{j(S,\sigma)}(N)$ and because $\mathbf{a}(S) = a^*_{j(S,\sigma)}(S)$. The inequality is satisfied because $T^{\sigma}_{\sigma^{-1}(j(S,\sigma))} = \{\sigma(1), \ldots, \sigma(j(S,\sigma) - 1)\} \subset S$ (the inclusion comes from the definition of $j(S,\sigma)$). Thus we have proved that the inequality (1) holds for any $S \subset N$.

Then we show that there is no loss of generality to look for an (α, σ)-optimal threat mechanism to solve the optimal design program.

Proposition 5.3 *For any implementable mechanism $(\mathbf{a}, \mathbf{t})$, there exists an implementable mechanism that belongs to the class of (α, σ)- optimal threat mechanisms and that raises at least the same utility level for the principal. The optimal design program becomes:*

$$\max_{(\alpha,\sigma) \in A \times \sigma(N)} \left\{ \sum_{i=0}^{n} V^{\alpha}_i - \sum_{i=1}^{n} V^*_i(\{\sigma(1), \ldots, \sigma(\sigma^{-1}(i) - 1)\}) \right\} \qquad (2)$$

Proof 3 *For a given mechanism $(\mathbf{a}, \mathbf{t})$, we define a corresponding (α, σ)- optimal threat mechanism in the following way: $\alpha = \mathbf{a}(N)$, σ is defined by induction such that $\sigma(1) = Arg \max_{i \in N} \{V^{\mathbf{a}(N)}_i - \mathbf{t}_i(N) - V^{\mathbf{a}(\emptyset)}_i\}$ (initial step) and $\sigma(i) = Arg \max_{i \in N \setminus \{\sigma(1), \ldots, \sigma(i-1)\}} \{V^{\mathbf{a}(N)}_i - \mathbf{t}_i(N) - V^{\mathbf{a}(\{\sigma(1), \ldots, \sigma(i-1)\})}_i\}$ (inductive step). The map σ is by definition a permutation. From lemma 5.1, the (α, σ)- optimal threat mechanism is implementable. It remains to*

show that it raises a greater utility for the principal than the original mechanism $(\mathbf{a}, \mathbf{t})$*. More precisely, the principal implements the same economic outcome and extracts more surplus from each agent. Let* $\mathbf{t}_i^{(\alpha,\sigma)}(N)$ *be the transfer for agent* i *in the* (α, σ)*- optimal threat mechanism at equilibrium. We have:*

$$\mathbf{t}_i^{(\alpha,\sigma)}(N) = V_i^{\mathbf{a}(N)} - V_i^*(T^\sigma_{\sigma^{-1}(i)}) \geq V_i^{\mathbf{a}(N)} - V_i^{\mathbf{a}(T^\sigma_{\sigma^{-1}(i)})} \geq \mathbf{t}_i(N)$$

The first equality results from the definition of $\mathbf{t}_i^{(\alpha,\sigma)}(N)$ *and that* $\alpha = \mathbf{a}(N)$*. The first inequality comes from the definition of the map* $V_i^*(.)$ *and since* $a(T^\sigma_{\sigma^{-1}(i)}) \in \mathcal{A}(T^\sigma_{\sigma^{-1}(i)})$*. The last inequality results from our subtle construction of* σ *and the inequality (1) for the set* $T^\sigma_{\sigma^{-1}(i)}$*. This latter inequality states that* $\max_{j \in N \setminus \{\sigma(1),\ldots,\sigma(\sigma^{-1}(i)-1)\}} \{V_j^{\mathbf{a}(N)} - t_j(N) - V_j^{\mathbf{a}(\{\sigma(1),\ldots,\sigma(\sigma^{-1}(i)-1)\})}\} \geq 0$ *if* $(\mathbf{a}, \mathbf{t})$ *is implementable. The construction of* $\sigma(i)$ *guarantees that the expression in the 'max' is positive for* $j = \sigma(i)$*, i.e.* $V_{\sigma(i)}^{\mathbf{a}(N)} - V_{\sigma(i)}^{\mathbf{a}(T^\sigma_{\sigma^{-1}(i)})} \geq \mathbf{t}_{\sigma(i)}(N)$*. To sum up, we have proved that* $\alpha = \mathbf{a}(N)$ *and* $\mathbf{t}_i^{(\alpha,\sigma)}(N) \geq \mathbf{t}_i(N)$ *for all agents. The utility level of the principal is thus higher in the* (α, σ)*-optimal threat mechanism we have constructed than in* $(\mathbf{a}, \mathbf{t})$*.*

The optimal program (2) allows us to separate the choice of the final outcome α to the choice of the optimal threat structure, which is indeed reduced to the choice of a permutation that specifies the order according to which agents will be threat taken as given the participation decision of the agents that are lower in this order. The optimal choice of α thus coincides with the maximisation of the allocative efficiency.

Corollary 5.4 *Optimal order independent individually rational feasible mechanisms are efficient.*

The expression (2) of the utility level of the principal should be compared with the standard expression under simultaneous participation:

$$\max_{\alpha \in A} \left\{ \sum_{i=0}^{n} V_i^\alpha \right\} - \sum_{i=1}^{n} V_i^*(N \setminus \{i\}) \tag{3}$$

In general, the possibility to commit to a simultaneous participation game leads to a greater payoff for the principal since $V_i^*(S)$ is decreasing in S. Under order independent individual rationality and in an (α, σ)-optimal threat mechanism, the set of implementable threats is reduced to $V^*_{\sigma(i)}(\{\sigma(1), \ldots, \sigma(i-1)\})$ for the agent $\sigma(i)$. Nevertheless, in a negative-externality-free framework, the optimal threat $V_i^*(N \setminus \{i\})$ against agent i requires economic an outcome a that is always feasible independently to the set

of participant, i.e. $a \in \mathcal{A}(\emptyset)$, and is thus always equal to $V_i^*(\{\sigma(1), \ldots, \sigma(\sigma^{-1}(i)-1)\})$. We obtain the following corollary:

Corollary 5.5 *In a negative-externality-free framework, the optimal revenue under simultaneous participation can be implemented under order independent individually rationality.*

6 Examples

As illustrated by our starting example, an important class of applications where our new rationality constraints are binding is auctions with negative externalities as in [9, 10, 11, 5, 7]. The scope of application may seem relatively restricted since the optimal design is unchanged in a negative-externality-free framework. Our two following general examples show how order independent individual rationality may be fruitful first to model general collusion mechanisms and second contracting in dynamic environments when long-term contracts are not available.

6.1 Example 1: General collusion mechanisms

To avoid confusion with the collusive interpretation of the order independent individual rationality concept, we emphasize that, in this example, we consider the relevance of applying it in the design of the collusive mechanism itself. In most of the aforementioned literature on collusion, a third party proposes a mechanism that can be vetoed by each agent. When an agent breaks the collusion process, the game is played in a non-cooperative way under passive-beliefs. Thus contrary to the mainstream mechanism design literature, the principal is significantly limited in the way she can punish non-participants. In an auction framework, Caillaud and Jehiel [1] relax slightly this veto power assumption by also considering the case where a defection leads to a collusive report from the agents that are remaining in the collusion process. Dequiedt [6] considers that the remaining agent can commit to the harsher punishment if the other agent refuses the collusion mechanism. The reluctance to adopt the standard mechanism design approach to model collusion may come from the seemingly excessive commitment power that it requires and which is slightly softened under our approach.

Let us discuss those differences in a simple example under complete information: a symmetric triopoly under Cournot competition. Each firm has a constant null marginal cost and a maximum capacity $q_{max} = 0.5$. Inverse demand is given by $P = 1 - Q$, where Q denotes the total quantity supplied. Without collusion, the quantity supplied by each firm in equilibrium is equal to $1/4$ and the corresponding total profit of the triopoly is $\Pi_{nc} = 3/16$. The collusive outcome corresponds to the total production $Q = 1/2$ and the joint profit $\Pi_{col} = 1/4$. Suppose that a general collusion mechanism (which

specifies the quantities produced by each participant and balanced monetary transfers among participants) is proposed by one firm, say 1. Under complete information, all the different models, leads to the collusive outcome in the optimal mechanism. Nevertheless, the repartition of the profits from collusion are very different according to the model for collusion. Under veto power, each firm is guaranteed to obtain her non-cooperative profit 1/16. The proposer manages to capture all the profits from collusion $\Pi_{col} - \Pi_{nc}$. At the other extreme as in [6], a non-participant can be punished by the minmax punishment which leads here to a null payoff: the two remaining participants commit to produce $q = 0.5$ which leads to a null price. Nevertheless, this mechanism may seem poorly convincing since firm 1 manages to extract all surplus from both firms by threatening each to flood the market with the help of the other one. With our model, in the optimal mechanism, firm 1 can extract full surplus only to one firm and has to leave the surplus 1/36 to the other one, the profit corresponding to the Cournot outcome after the commitment to produce $q = 0.5$ by firm 1. Thus she should use a divide and conquer strategy.

6.2 Example 2: Dynamic processes of social and economic interactions

Gomes and Jehiel [8] consider a model of dynamic interactions in complete information where, at each period, an agent is selected to make an offer to a subset of the other agents to move the state of the economy. They do not only assume that long-term contracts are not available but also restrict the analysis to simple-offer contracts where each approached agent can veto the proposed move. Indeed, as they emphasize, this restriction is with no loss of generality if a third party can coordinate the approached agents by means of a 'strong' collusion contract with transfers. With general contracts -i.e. without any form of collusion- the economy moves immediately to the efficient state. On the contrary, with simple-offer contracts, efficiency is no longer guaranteed. This last (negative) result depends critically on the model for collusion. If collusion is modeled by order independent individual rationality, then the transposition of corollary 5.4 in their framework restaures efficiency: all Markov Perfect Equilibria of the economy with general spot contract that are order independent individually rational are efficient, entailing an immediate move to the efficient state, where it remains forever. Note however that, under our milder collusion device, the expected payoff of the selected proposer is lower than with general contracts. At the other extreme, under a mildly stronger form of collusion where the third party can also contract with non-approached agents and where collusion is not observable by the proposer, the economy also moves immediately to the efficient state.

7 Concluding Remarks

We relax the commitment ability of the principal in some minimal way and give some theoretical fondations for such a refinement of the standard mechanism design approach. The scope of application may seem relatively restricted since the optimal design is unchanged in a negative-externality-free framework. Nevertheless, jointed with other commitment failures as the unability to commit not to propose a new mechanism if the first one fails to work, e.g. the unability to commit never to attempt to resell the good if she fails to sell it as in McAfee and Vincent [15] and Skreta [18], order independent individual rationality may have some bite even in pure private value framework that are externality-free. For example, in a procurement auction, the designer may be unable to set a high reserve price since this would trigger a joint boycott of the main market participants that will force the designer to propose a new mechanism.

Finally, we have restricted attention to a complete information setup. It is left for further research how to extend the notion of implementation under order independent rationality constraints in incomplete information, analyse the interactions with the incentive compatibility constraints and ask whether this constraint is beneficial or not to the welfare.

References

[1] B. Caillaud and P. Jehiel. Collusion in auctions with externalities. *RAND J. Econ.*, 29(4):680–702, 1998.

[2] Y.-K. Che and J. Kim. Optimal collusion-proof auctions. *Unpublished manuscript, Columbia University*, 2006.

[3] Y.-K. Che and J. Kim. Robustly collusion-proof implementation. *Econometrica*, 74:1063–1107, 2006.

[4] O. Compte, A. Lambert-Mogiliansky, and T. Verdier. Corruption and competition in procurement auctions. *RAND J. Econ.*, 36(1):1–15, 2005.

[5] G. Das Varma. Standard auctions with identity-dependent externalities. *RAND J. Econ.*, 33(4):689–708, 2002.

[6] V. Dequiedt. Efficient collusion in optimal auctions. *J. Econ. Theory*, page doi: 10.1016/j.jet.2006.08.003, 2006.

[7] N. Figueroa and V. Skreta. The role of outside options in auction design. Unpublished Manuscript, March 2006.

[8] A. Gomes and P. Jehiel. Dynamic processes of social and economic interactions on the persistence of inefficiencies. *J. Polit. Economy*, 113:626–667, 2005.

[9] P. Jehiel and B. Moldovanu. Strategic nonparticipation. *RAND J. Econ.*, 27(1):84–98, 1996.

[10] P. Jehiel, B. Moldovanu, and E. Stacchetti. How (not) to sell nuclear weapons. *Amer. Econ. Rev.*, 86(4):814–829, 1996.

[11] P. Jehiel, B. Moldovanu, and E. Stacchetti. Multidimensional mechanism design for auctions with externalities. *J. Econ. Theory*, 85:258–293, 1999.

[12] J.-J. Laffont and D. Martimort. Collusion under asymmetric information. *Econometrica*, 65:875–911, 1997.

[13] J.-J. Laffont and D. Martimort. Mechanism design with collusion and correlation. *Econometrica*, 68:309–342, 2000.

[14] D. McAdams and M. Schwarz. Credible sales mechanisms and intermediaries. *Amer. Econ. Rev.*, forthcoming.

[15] P. McAfee and D. Vincent. Sequentially optimal auctions. *Games Econ. Behav.*, 18:246–276, 1997.

[16] B. Moldovanu and E. Winter. Order independent equilibria. *Games Econ. Behav.*, 9:21–34, 1995.

[17] G. Pavlov. Colluding on participation decisions. Unpublished Manuscript, Boston University, 2004.

[18] V. Skreta. Sequentially optimal mechanisms. *Rev. Econ. Stud.*, 73(4):1085–1111, 2006.

[19] H. Vartiainen. Auction design without commitment. Unpublished Manuscript, August 2002.