



PARIS SCHOOL OF ECONOMICS
ÉCOLE D'ÉCONOMIE DE PARIS

WORKING PAPER N° 2008 - 55

**Mechanism design with partially-specified
participation games**

Laurent Lamy

JEL Codes: C7, D0, D62

**Keywords: Mechanism design, robust implementation,
surplus extraction, strong Nash equilibrium, Nash
program, partial subgame perfection, imperfect
commitment, collusion on participation**



PARIS-JOURDAN SCIENCES ÉCONOMIQUES
LABORATOIRE D'ÉCONOMIE APPLIQUÉE - INRA



48, Bd JOURDAN – E.N.S. – 75014 PARIS
TÉL. : 33(0) 1 43 13 63 00 – FAX : 33 (0) 1 43 13 63 10
www.pse.ens.fr

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE – ÉCOLE DES HAUTES ÉTUDES EN SCIENCES SOCIALES
ÉCOLE NATIONALE DES PONTS ET CHAUSSÉES – ÉCOLE NORMALE SUPÉRIEURE

Mechanism Design with Partially-Specified Participation Games *

Laurent Lamy[†]

Abstract

This paper considers the implementation of an economic outcome under complete information when the strategic and informational details of the participation game are partially-specified. This means that full participation is required to be a subgame-perfect equilibrium for a large variety of extensive modifications of the simultaneous-move participation game in the same vein as Kalai [Large Robust Games, *Econometrica* 72 (2004) 1631-1665], an implementation concept which is shown to be related to ‘strong Nash’ implementation in the corresponding simultaneous participation game.

We solve the optimal design program: economic efficiency is not damaged but the principal may fail to extract fully agents’ surplus relative to the harsher threats and may have to use divide and conquer strategies that discriminate among symmetric agents. The analysis is extended to implementation under partial subgame-perfection criteria.

Keywords: Mechanism Design, Robust Implementation, Surplus Extraction, strong Nash equilibrium, Nash program, Partial subgame perfection, Imperfect Commitment, Collusion on Participation

JEL classification: C7, D0, D62

*I am grateful above all to my Ph.D. advisor Philippe Jehiel for his continuous support. I would like to thank Yeon-Koo Che, Olivier Compte, Jean Tirole and seminar participants at PSE TOM Seminar, Nasville PET 2007 Conference, the workshop on Implementation of Cooperative Solution Concepts at the 19th Stony Brook festival. Previous versions of this paper circulated under the title "Individual Rationality under Sequential Decentralized Participation Processes" and "Mechanism Design under Strong Nash Equilibrium: Characterization and Non-cooperative Foundations". All errors are mine. This paper is a major revision of Chapter IV of my Ph.D. dissertation.

[†]PSE, 48 Bd Jourdan 75014 Paris. e-mail: lamy@pse.ens.fr

1 Introduction

‘A particular modeling difficulty of noncooperative game theory is the sensitivity of Nash equilibrium to the rules of the game, e.g., the order of players’ moves and the information structure. Since such details are often not available to the modeler or even to the players of the game, equilibrium prediction may be unreliable.’ (Kalai [22], pp 1632). Economic theory usually considers specific games and studies their equilibrium sets. In this way and by means of the revelation principle, the mechanism design paradigm considers direct mechanisms where agents are taking their participation decisions simultaneously. Such an approach implicitly assumes that the principal controls the details of the rules of the underlying proposed game. At first glance, it seems reasonable: all potential participants are invited in separated rooms where they privately report their messages that are then jointly opened by the principal under the scrutiny of a judge. However, participation decisions have a different nature than the report of private signals: it corresponds to the action whether or not to enter the room under our metaphor of the mechanism design paradigm, an action which can e.g. be visible to the other agents before they are taking their own participation decision. A second support for our approach is that even if the potential participants are effectively locked in a room it is often the case that those agents are not corresponding to the real decision-makers from which they take formal instructions. The real participation game would correspond to the one between the decision-makers and which is then out of control from the principal’s perspective. The non-simultaneous nature of the participation decisions is especially relevant if the instructions are resulting from some collegial decisions where the different colleges can keep a close watch on each other. In the same vein as Kalai [22, 23], we consider here that the principal has no idea on the participation game which is played, which will bring us to consider implementation concepts requiring that full participation is an equilibrium not only of the (traditional) simultaneous-move participation game but also for a very large class of extensive games, capturing the lack of knowledge of the principal on the details of how the participation game is played. In other words, full participation is required to be an equilibrium for partially-specified participation games (henceforth PG).¹

¹Relaxing the common knowledge assumptions on the trading game may seem at odd with the original formulation of the ‘Wilson Doctrine’. However in environments where enforcement on the

The reliability of an equilibrium prediction is usually captured by equilibrium refinements while still assuming simultaneous-move. Rationalizability (Tan and Werlang [42]), dominant strategy (Chung and Ely [10]) and ex-post equilibrium (Bergemann and Morris [4]) are examples of detail-free implementation criteria in this research program whose agenda is to relax the common knowledge assumptions on agents' beliefs about another's preferences and information, a program which is also known as the 'Wilson Doctrine' (Wilson [44]). Another kind of equilibrium refinements aims to capture the possibility of cooperation among agents. The most popular ones are coalition-proof equilibrium (Bernheim et al [5]) and STRONG Nash equilibrium (Aumann [3]). This latter concept which is the narrowest requires that no coalition of agents could jointly deviate in a way that benefits all of its members. Coalitions are allowed to use correlated strategies on the contrary to the related concept where coalition are restricted to use pure strategies and which is called here 'strong Nash' equilibrium.² Alternative coalitional concepts are putting some restriction on the set of feasible deviations, e.g. they are themselves also immune to deviations from some of its members, in particular individual deviations. Those refinements are justified as having an 'intuitive' appeal, e.g. coalition-proof equilibria are supposed to reflect that agents can freely discuss while being unable to make binding commitments. Nevertheless, the non-cooperative foundation for such cooperative equilibrium concepts remains unclear.³

In the mechanism design framework with complete information, the present paper makes two contributions.⁴

In a first step, we characterize in Theorem 1 the optimal revenue that can be

details of the participation process seems difficult, we do think that we are remaining in line with the spirit if not the letter of the 'Wilson doctrine'. However, we do not claim that, in general, principals have no control with regards to participation processes. The differences in the online versions of the ascending auction used at Amazon and eBay and analyzed by Ockenfels and Roth [32] is a good example of how small details on the trading rules can have a first-order impact on participation decisions.

²The use of lowercase letters reflects the point that the equilibrium constraints under 'strong Nash' are weaker than under STRONG Nash.

³More recently, Ambrus [1] proposes a new (and somehow weaker) solution concept, coalitional rationalizability, that addresses the issue of coalitional agreements. In Ambrus [2], an epistemic foundation is given for this solution concept and some variations.

⁴We consider that the principal is fully informed. It should not be confused with 'implementation theory' that considers that there is not asymmetric information between agents and that the principal is uninformed. See Dutta and Sen [14] for a characterization of implementable social choice functions under strong Nash equilibrium. In such an environment and with transferable utilities, Pérez-Castrillo and Wettstein [34] exhibit a simple budget-balanced bidding mechanism that implements the efficient allocation under strong Nash equilibrium.

raised under extensively (subgame-perfect) robust implementation criteria and show that it corresponds to the optimal revenue raised under the strong Nash implementation criterium in the simultaneous-move PG. The Coasian logic still applies and we obtain that the optimal mechanism is efficient. Nevertheless, full extraction relative to the harsher threats as in Jehiel et al [20] does not work anymore, in general, in presence of negative externalities. Furthermore, Theorem 1 characterizes the structure of the rents left to the different agents in those extensively robust or strong Nash implementable optimal mechanisms. The proof is in three steps. First we show any extensively robust implementable mechanism is necessary strong Nash implementable. Additional to the usual individual rationality constraints, it requires that there is no set of agents $S \subset N$ such that all agents in S prefer the outcome where only the agents in $N \setminus S$ participate to the outcome where all agents accept the mechanism. Second we characterize in Proposition 4.4 the optimal design under the strong Nash constraints. Those ‘coalitional constraints’ in the mechanism design program are non-linear and the set of implementable mechanisms is thus in general not convex. Nevertheless, the optimal design program can be greatly simplified: we separate in Proposition 4.4 the choice of the final allocation to the structure of the optimal threats. The latter has a *divide and conquer* flavor: it consists in giving the incentive to participate for one agent, say 1, independently of the participation decisions of the other agents. Then given that agent 1 will surely participate, the principal can really threaten another agent, say 2, to use agent 1 in case of non-participation in order to minimize his payoff. Then she threatens the next agent conditionally on the participation of agents 1 and 2 and so on. The optimal design consists thus in finding the optimal ‘order’ to define those threats, which corresponds to a permutation among the agents. Third we show by means of carefully specified out of equilibrium transfers that the strong Nash implementable optimal designs can be made extensively robust implementable.

Despite the huge similarity with Kalai’s framework, we emphasize important differences. First, our analysis deals with robustness concepts that incorporate subgame-perfection. Second, our analysis does not consider the convergence to equilibrium concepts when the number of players grows to infinity but exact equilibria for any number of players. Those two assumptions are crucial in Kalai’s analysis and we can wonder whether much can be said with robustness criteria such that the equi-

libria of most standard familiar games fail to pass the test, not only in the ‘Matching Pennies’, ‘Battle of the sex’ or ‘Chicken’ games but also in the ‘Prisoners’ Dilemma’ that is dominant-strategy solvable. However, in a mechanism design framework, the principal has a lot of flexibility on the final payoffs especially through the monetary transfers such that the existence issue is circumvented.⁵ In a given mechanism, checking that full participation is a subgame perfect equilibrium for all extensive versions à la Kalai seems to be an untractable task in general, though we make an important step by providing general necessary and sufficient conditions in Proposition 4.3 and 4.6 in the first and third step of the proof of Theorem 1. Furthermore, simple examples illustrate that such a robustness criterium may be too strong. It brings us to consider a restricted class of the extensive PGs à la Kalai that will allow us to characterize the set of implementable mechanisms in a tractable way though not modifying the rents in the optimal designs.

In a second step, we consider a slightly different set of extensive games by adding two additional ingredients: first, after an agent is irreversibly committed to participating in the mechanism, all the remaining agents have the opportunity to participate. Second, such an irreversible commitment is always publicly observable. The extension of the participation deadline after a bid submission at Amazon modeled as in Ockenfels and Roth [32] is a good example of such extensive games with *subsequent opportunities and perfect information*. On the contrary, the simultaneous-move PG does not belong to this subclass. We emphasize that this subclass still allows for wide flexibility in the order of players’ moves as well as for informational leakage, commitment and revision possibilities, cheap talk, and more. Theorem 2 establishes that extensive robustness implementation in this subclass is equivalent to strong Nash implementation. Not only do we provide a tractable way to check whether full participation in a given mechanism remains a (and actually the unique) subgame-perfect equilibrium outcome in a large class of extensive versions of the simultaneous-move game, but we also give a non-cooperative foundation of strong Nash implementation in this framework. This equivalence is extended in Theorem 3 where subgame perfection is relaxed to p -subgame perfection with Nash best-responses being required only on histories with a limited depth with regards to the number of players’ that

⁵To obtain that all equilibria of simultaneous-move games become asymptotically extensively robust, Kalai [22] needs an assumption on the final payoffs: players are assumed to be semi-anonymous.

deviate and where strong Nash equilibrium constraints are relaxed to p -strong Nash with ‘coalitional constraints’ being required only for coalition of limited size.

A by-product of our analysis on partially-specified PGs is thus a contribution to the Nash program which aims to bridge the gap between the non-cooperative and cooperative approaches to game theory. The bulk of the works in this area - recently surveyed by Serrano [40]- are devoted to the construction of families of bidding or bargaining games whose equilibria are corresponding to cooperative solution concepts as the Shapley value or the Core.⁶ To the best of our knowledge, the Nash program is silent on the cooperative concepts that are often used in non-cooperative approaches as ‘intuitive’ equilibrium refinements: by relaxing the commitment ability of the principal in the way she controls the PG, we give some theoretical foundations for the use of the strong Nash implementation criteria in the standard mechanism design approach.

Our insights are linked to two main topics in mechanism design: imperfect commitment and the possibility of full surplus extraction. First, relaxing the commitment power of the principal with regards to the rules of the game is precisely the focus of a still growing literature about mechanism design or positive design with imperfect commitment.⁷ In corporate acquisitions and procurement auctions, it is common that the seller violates the announced rules to provide opportunities for bid readjustments (see Compte et al. [11] and McAdams and Schwarz [28]). In electronic auctions and in auction houses, it is common that the seller uses a shill bidder to participate in the mechanism as any other participants (see Lamy [26, 27]). Second, in an incomplete information setup with strictly correlated signals, Crémer and McLean [13] show that the principal can implement the efficient allocation while leaving no informational rents to the agents as in a complete information setup, where full extraction can be reached even in dominant strategies as shown by Jehiel

⁶See Pérez-Castrillo and Wettstein [33] for generalized multi-bidding games and Moldovanu and Winter [31] and Hart and Mas-Collel [17] for alternating-offers bargaining games. It is worthwhile to note that by considering *order independent equilibria*, that is strategy profiles that remain an equilibrium and lead to the same payoff independently of the specification of the order of the moves, Moldovanu and Winter [31] can now be viewed as a pioneering contribution in the research program on game theory with partially-specified games though their analysis still imposes a lot of structure on the family of bargaining games they consider. See also Caruana and Einav [7] for ‘grid invariant equilibria’ and other protocol-free related properties.

⁷Closely related is the literature on commitment failures with regards to future interactions. See Tirole [43] for a survey and Skreta [41] and Zheng [46] for recent contributions in an auction setup, respectively with the impossibility to commit not to re-auction the good or to ban future resale between bidders.

et al. [20]. Heifetz and Neeman [18] show that generic priors on the universal type space do not allow for full surplus extraction in an incomplete information setup. Their insight is that, generically, private information implies informational rents.

With a common concern for robustness, this paper shows that the principal may not be able to fully extract agents' surplus relative to their harsher threats in a complete information setup if the implementation criterium is strengthened. Partial extraction comes from what can be called *coalitional rents*: the 'coalitional participation constraints', either explicit through the strong Nash equilibrium refinements or implicit through the robustness to any extensive PGs are a new channel for imperfect surplus extraction in addition to the well-known 'incentive constraints' that create *informational rents*. Coalitional rents are shown to be driven by negative allocative externalities: the possibility for some agents to hurt their peers by their mere participation.

The paper is organized as follows. Our basic insights are illustrated by means of a simple example in section 2. In section 3 we introduce the general allocation problem and our implementation criteria. The proper foundation of our restriction to the analysis of direct mechanisms without reports and to the implementation of the full participation outcome is relegated to section 6. The optimal mechanisms under extensively subgame-perfect robust implementation are characterized in section 4 where the optimal mechanisms under strong Nash implementation are also characterized as a key step in the proof. In section 5, PGs with subsequent opportunities and perfect information are considered and extensively robust implementation is shown to be equivalent to strong Nash implementation. Section 7 extends the analysis with partial perfection equilibrium criteria. Additional comments are gathered in section 8, including additional motivations and applications for our analysis.

2 A Simple Example ⁸

Consider the sale of a single object involving identity-dependent externalities among two potential competitors. Bidders 1 and 2 are valuing intrinsically (with regards to the statu quo with no sale) the good V which is assumed to be greater than v the reservation price of the seller. However if he does not obtain the object

⁸A reader eager to get our results may skip this subsection at first reading.

Table 1: ‘Prisoner’s dilemma’

	NP	P
NP	(0,0)	$(-\alpha, V - v)$
P	$(V - v, -\alpha)$	$(-\alpha, -\alpha)$

bidder i ($i = 1, 2$) suffers from a negative allocative externality α when the object is allocated to his opponent j . The allocative externality is supposed to be important enough such that the efficient allocation consists in keeping the object, i.e. $\alpha > V - v$. We consider that the seller is able to allocate the object only to participating agents. In other words, she can not dump the object to a non-participant.⁹

Standard Auctions The buyers have first the opportunity to decide whether or not they want to participate in the auction. Then participation decisions are publicly revealed before a first price auction takes place. The final payoffs as a function of their participation decisions are summarized in Table 1 where NP and P respectively correspond to the nonparticipation and participation decisions.¹⁰

If the participation decisions are modeled as resulting from a simultaneous-move PG as in Jehiel and Moldovanu [19], it is an equilibrium outcome that agents 1 and 2 both participate and are then submitting the bid $V + \alpha$ in the subsequent auction. They are both suffering from a loss of α compared to their profits in the case where they could jointly coordinate themselves not to participate. Now consider a sequential PG between agents 1 and 2 as depicted in the left upper panel of Figure 2. The final payoffs are determined by the set of participants depicted in the brackets at the final nodes of the extensive version. The game corresponds to the sequential game where agent 2 makes his participation decision after being fully informed of the choice of agent 1, but with the slight modification that if agent 2 agrees to participate to the mechanism (action ‘YES’) after agent 1 initially chooses the action ‘NO’ then agent 1 can reconsider his participation decision. The full participation outcome is not a subgame-perfect equilibrium outcome in this extensive version. When agent 2

⁹The reservation price can equivalently be viewed as a reduced form for a third potential bidder with a pure private valuation v . This example formalizes the motivating story in Jehiel and Moldovanu [19] where two potential buyers suffering from important reciprocal negative externalities prefer not to participate in the bidding process and let a third buyer win at a low price.

¹⁰The final payoffs are unchanged for most other standard auction formats as the second price auction or the English button auction.

considers whether to participate, he knows that it is then irreversible and will induce the participation of his opponent in the case he is still not committed to participating. Consequently, when making a participation decision, he compares the outcome where they both participate to the outcome where they both do not participate. The sequential PG offers implicitly a kind of coalitional agreement that makes the nonparticipation decisions the unique equilibrium outcome. We thus obtain a paradox 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.¹¹

The final payoffs of the PG are corresponding to a kind of ‘Prisoner’s dilemma’ where nonparticipation/participation corresponds to the cooperate/deviate actions. It differs however from the standard version where cooperation is a strictly dominated strategy and where the full deviation outcome is the unique rationalizable final outcome. However all the above analysis would remain valid: it illustrates the point that participation being a dominant strategy for any player in the simultaneous-move version does not guarantee that full participation is the equilibrium outcome in all extensive versions.

The Optimal Mechanism Under the simultaneous-move PG and complete information, Jehiel et al. [20] presents an optimal mechanism where participation is a dominant strategy. The optimal mechanism is always efficient and the seller can extract surplus from agents who do not obtain the object by using the optimal threats, i.e. giving the object to the most feared opponent in case of nonparticipation. Here the optimal mechanism raises the revenue 2α : each bidder has to pay α in order to avoid that the seller gives the object to his most feared opponent and the seller keeps the object. However, for some PG that are corresponding to ‘natural’ alteration of the simultaneous-move game, agents 1 and 2 can coordinate their participation decisions by jointly not participating which is Pareto improving. More generally, the seller can never keep the object while extracting a strictly positive surplus (with respect to the harsher threats) from both agents 1 and 2. Otherwise,

¹¹In Jehiel and Moldovanu [19], V is assumed to be smaller than v such that the final payoff outcome when only one of the bidders 1 and 2 participate is $(0, 0)$. On the contrary, nonparticipation from either agent 1 or 2 does not help here and cannot prevent the purchase by his ‘feared opponent’ in the auction. Strategic nonparticipation has thus here a completely different nature.

they could jointly not participate and obtain a null payoff since the seller is assumed to be unable to ‘dump’ the object. In other words, some ‘coalitional participation constraints’ would be violated. 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. We will show that it is actually the optimal extensively robust mechanism and it raises the revenue α .

Additionally to our central insight that the optimal design depends critically on to the way PGs are modeled, the example illustrates several features that are generalized in section 4 when the mechanism is required to be robust for any extensive version of the simultaneous-move PG. First the optimal design implements the efficient outcome. Second those additional constraints may reduce strictly the revenue in presence of negative allocative externalities. 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 cannot be optimal on the contrary to Jehiel et al [21]’s analysis.

3 The Model

3.1 The Model

Let $N = \{1, 2, \dots, n\}$ be a set of agents and $A = \{a_1, a_2, \dots, a_K\}$ be a finite set of possible outcomes. Denote by $\Sigma(N)$ the set of the permutations over the set N . For a given permutation $\sigma : N \rightarrow N$, denote by T_i^σ the subset $\{\sigma(1), \sigma(2), \dots, \sigma(i-1)\}$, i.e. the $i-1$ first smallest agents according to the implicit order defined by σ . Denote by $\#S$ the cardinality of the set $S \subset N$. 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 t_i (to the principal) is: $\mathcal{U}_i(a, t_i) = V_i^a - t_i$.

The principal announces a direct mechanism, denoted by (\mathbf{a}, \mathbf{t}) , that specifies a final outcome $\mathbf{a}(S)$ and a vector of monetary transfers $\mathbf{t}(S)$ for each possible set of

participants $S \subset N$. Monetary transfers are assumed to be deterministic w.l.o.g. since players are assumed to be risk-neutral. A mechanism is said to be *feasible* if:

- For each set of participants S , the final outcome belongs to the set of probability distributions on $\mathcal{A}(S)$, 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 payment from that agent: $\mathbf{t}_i(S) \leq 0$, for all $i \in N \setminus S$.
- 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 outcomes 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. [20] considers that one cannot ‘dump’ the object on a non-participating agent. In the case of exclusionary contracts, Segal and Whinston [38] consider that an incumbent can deter entry only if the number of ‘captured’ agent is above a given threshold. We do not impose any specific structure on 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 $\mathcal{A}(N) = A$. We call an efficient allocation any allocation $a \in A$ that maximizes $\sum_{i=0}^n V_i^a$. For an agent i and a set of participants $S \subset N \setminus \{i\}$, 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: $a_i^*(S) \in \text{Arg min}_{a \in \mathcal{A}(S)} V_i^a$. Denote by $V_i^*(S) = V_i^{a_i^*(S)}$ the corresponding utility level. On the one hand, only the threats $a_i^*(N \setminus \{i\})$ do matter in mechanism design under simultaneous-move participation. In the optimal design, if one agent refuses the mechanism, the remaining ones commit to this harsher threat also called ‘minmax punishment’ as in Caillaud and Jehiel [6] or Jehiel et al. [20]. On the other hand, in extensively robust implementation, the whole set of the feasible threats $a_i^*(S)$ will play an active role in the computation of the optimal mechanism. On the whole, 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 $a \rightarrow V_i^a$ 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 $i \in N$ and $S \subset N \setminus \{i\}$.

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 of the other participants, i.e. by retaining all goods in the above example. Applications where externalities are negative are the general category of interest where the optimal design may be modified by our alternative implementation concepts. Genicot and Ray [15] and Segal [36] discuss extensively related applications that go beyond network externalities in industrial organization.

3.2 Implementation criteria

We first consider implementation concepts under the simultaneous-move PG. In the following, mechanisms are simply labeled as ‘implementable’, ‘strong-Nash implementable’ or ‘STRONG-Nash implementable’ when full participation satisfies the corresponding equilibrium property in the simultaneous-move PG in contrast to the ‘extensively robust implementable’ terminology which will refer to as full participation being an equilibrium outcome in every extensive version, metagame or alteration of the simultaneous-move PG according to Ehud Kalai’s various terminologies of the same idea that the PG is partially-specified.

Definition 2 Full participation is respectively an equilibrium, a strong-Nash equilibrium and a STRONG-Nash equilibrium of the simultaneous-move participation game

if respectively:

$$V_i^{\mathbf{a}(N)} - \mathbf{t}_i(N) - V_i^{\mathbf{a}(N \setminus \{i\})} \geq 0 \text{ for all } i \in N, \quad (1)$$

$$\max_{i \in N \setminus S} \{V_i^{\mathbf{a}(N)} - \mathbf{t}_i(N) - V_i^{\mathbf{a}(S)}\} \geq 0 \text{ for all } S \subset N, \quad (2)$$

$$\text{and } \max_{i \in N \setminus S} \{V_i^{\mathbf{a}(N)} - \mathbf{t}_i(N) - \sum_{S' \in \Delta(S)} [V_i^{\mathbf{a}(S')} - \mathbf{t}_i(S')] \times \mu^S(S')\} \geq 0 \quad (3)$$

for any $S \subset N$ and any probability distribution $\mu^S(\cdot)$ on $\Delta(S) = \{S' | S' \supset S\}$ such that $\mu^S(S')$ denotes the probability that the set of participants is S' .

The constraints in (1) corresponds to the usual Individual Rationality or Participation constraints in mechanism design. The strong-Nash equilibrium concept puts additional restriction by *explicitly* adding some ‘coalitional participation constraints’ (2): for each subset of players $S \subset N$, there is no joint deviations in pure strategies that is profitable for all of its members. In other words there is no set of participants such that all of its members would benefit if they jointly refuse to participate. This concept is indeed slightly weaker than the original concept introduced by Aumann [3] where the immunity to all joint deviations in possible correlated mixed strategies -constraints (3)- is considered and which is labeled as STRONG-Nash equilibrium.

Definition 3 (simultaneous-move implementation) *A mechanism (\mathbf{a}, \mathbf{t}) is implementable, respectively strong Nash [STRONG Nash] implementable if it is feasible and if full participation is an equilibrium, respectively a strong Nash [STRONG Nash] equilibrium, of the simultaneous participation game.*

Any strong Nash implementable mechanism is necessary implementable as ‘coalitional constraints’ are including the usual individual rationality constraints. The converse does not hold in general as it has been illustrated by the PG in section 2 which corresponds to a prisoner’s dilemma. However, in externality-free frameworks it does: the individual rationality constraints (1) imply the strong Nash constraints 2. The payoff tables in Figure 1, with the action NP and P respectively corresponding to nonparticipation and participation, provides additional examples our various simultaneous-move implementation concepts. The ‘Battle of the sex’ and ‘Pure Coordination game’ are STRONG Nash implementable. In the game in the left lower panel, full participation is a strong Nash equilibria. Moreover, participation is a dominant strategy. However, it is not a STRONG Nash equilibrium since

	NP	P
NP	(2,1)	(0,0)
P	(0,0)	(1,2)

‘The Battle of the sex’

	NP	P
NP	(0,0)	(0,6)
P	(6,0)	(2,2)

‘strong-Nash’

	NP	P
NP	(1,1)	(0,0)
P	(0,0)	(2,2)

‘Pure Coordination’

	NP	P
NP	(5,0)	(3,13)
P	(0,0)	(10,1)

‘STRONG-Nash’

Figure 1: Payoff Tables: NP/P for NonParticipation/Participation

the agents would profitably deviate with the correlated strategy profile alternating between (NP, P) and (P, NP) with probability one-half and obtain the expected payoff $(3, 3) > (2, 2)$.

Starting from a mechanism (a, t) , we first describe a large number of variations on how the game may be played. We closely follow Kalai [23] formalization.

Definition 4 *A participation game (PG) of (\mathbf{a}, \mathbf{t}) is any finite extensive game \mathcal{B} (with perfect recall) with the following properties:*

1. *\mathcal{B} includes the (original) players: The players of \mathcal{B} constitute a superset of N .*
2. *Playing \mathcal{B} means playing (\mathbf{a}, \mathbf{t}) : With every final node of \mathcal{B} , (a, t) , there is an associated unique profile of participants S such that $(a, t) = (\mathbf{a}(S), \mathbf{t}(S))$.*
3. *Unaltered payoffs: The payoffs of any player $i \in N$ at every final node (a, t) are the same as their payoffs in the original mechanism, i.e. $\mathcal{U}_i(a, t_i)$.*
4. *Preservation of the original strategies: for any player $i \in N$, every pure strategy has at least one \mathcal{B} -adaptation. That is, a \mathcal{B} -strategy that guarantees ending at a final node corresponding to a final set of participants $S \supset \{i\}$ and a \mathcal{B} -strategy that guarantees ending at a final node with $\{i\} \notin S$ (no matter what strategies are used by the opponents).*

External players are allowed to participate in such variations as in Kalai [23]. In particular the principal herself may be possibly a player of the PG. Our results do not depend on such an assumption, e.g. the optimal design remains unchanged if

the principal is able to commit not to participate in the PG. In the following we will use PG with a unique external player that is labeled as ‘nature’. See Kalai [22] for a panel examples with rounds of revisions, changes in the order of play or nature selecting some moves, illustrating how rich is the set of PGs. Figure 2 depicts other possibilities with two agents. Version 1 has been discussed in section 2 to illustrate how the equilibrium in the simultaneous-move prisoner’s dilemma may not be robust in extensive versions. Version 2 is a mirror version of the extensive version 1 by permuting the role of the action YES and NO which are corresponding to some partial commitment on participating or not to the mechanism.

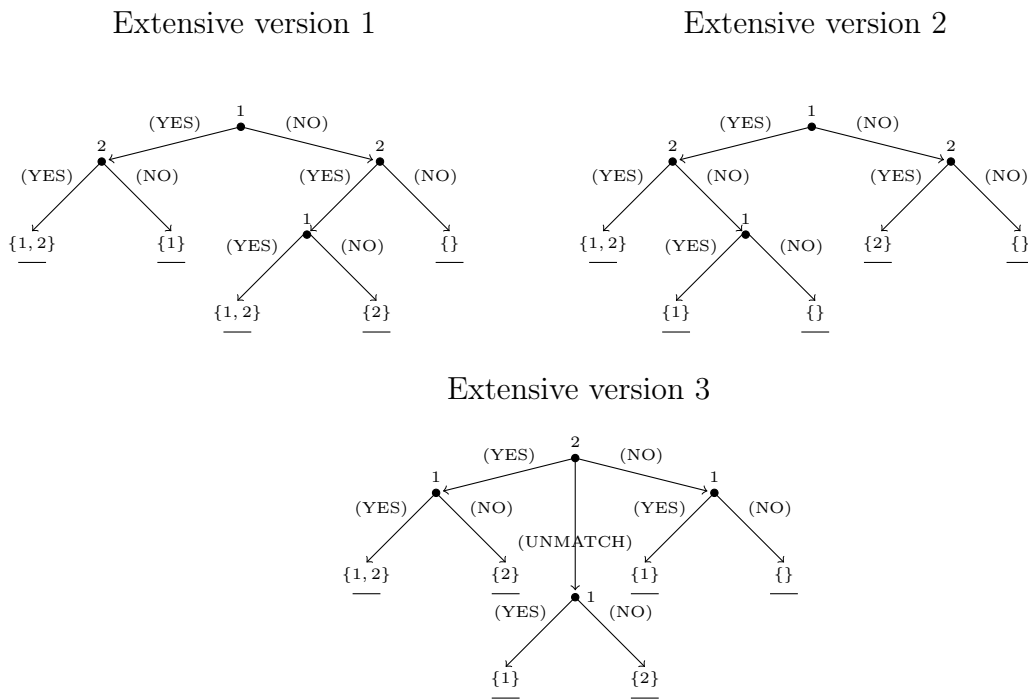


Figure 2: Participation games

The bottom panel of Figure 2 depicts a third extensive version of a PG with two agents. The game is a modification of the sequential game where agent 1 makes his participation decision after being fully informed of the choice of agent 2. At the first stage, agent 2 has an additional action (‘UNMATCH’) which corresponds to commit to chose the opposite action of the one chosen by agent 1. Suppose that the final payoffs are given according to the right bottom table of Figure 1. In the corresponding simultaneous-move game, full participation is a STRONG-Nash equilibrium where agents are using dominant-strategies. Moreover, for the ‘column agent 2’, participation is a super-dominant strategy: the minimum possible payment

when participating is strictly greater than the maximum possible payment under nonparticipation. By the maxmin argument, it is clear that participation has to be his final action. However, it does not imply that the ‘row agent 1’ always play a best response to his opponent super-dominant action as it occurs in the extensive version 3: in any subgame-perfect equilibrium, agent 2 chooses the ‘UNMATCH’ move which is followed by YES and the final payoffs are corresponding to the profile (NP, P) an outcome which is preferred by agent 2 and that he can impose by making the threat not to participate if the other do so.¹² Such a threat is available when we consider the general class of PG à la Kalai. In certain environments we may feel uncomfortable since it require the ability to violate some basic renegotiation-proofness criteria: after observing the irreversible participation decision of agent 1, agent 2 would prefer to participate. It will be a motivation for considering in section 5 the class of games with ‘subsequent opportunities and perfect information’ where such counterintuitive results do not occur. Finally the simultaneous-move PG belongs also to the class of PGs which guarantees that the following definition of an extensively robust implementable mechanism is stronger than the standard notion of an implementable mechanism.

Definition 5 (extensively robust implementation) *A mechanism is extensively subgame-perfect robust implementable if it is feasible and if in every participation game a subgame-perfect equilibrium exists such that full participation is the final outcome.*

Our extensively robust implementation concept differs in two ways from Kalai’s approach. In one way, Kalai’s approach seems more demanding since it does not solely require the existence of an equilibrium leading to the desired strategy profile, as it happens full participation, but that every strategy profile where every agents follow an adaptation of the participation strategy is an equilibrium profile. In another way, Kalai’s approach is less demanding since equilibria in the extensive versions are not required to be subgame-perfect. In our framework where the robustness criterium

¹²In a class of games with endogenous commitment, Caruana and Einav [7] obtain also the possibility for the agent with the super-dominant strategy to discipline his opponent not to use his dominant best response (our payoff table corresponds exactly to their example in section 4.3.2). The commitment of agent 2 to a credible punishment to a deviant agent 1 does not arise through the richness of the extensive versions as here but because actions’ switches are becoming more and more costly over time.

is restricted to a pure strategy equilibrium in complete information, the first difference becomes innocuous since all pure strategy equilibria in the simultaneous-move game survives (but may not be subgame-perfect) in the extensive versions and where agents are following an adaptation of their original strategy in the simultaneous-move game.¹³ The standard implementation concept in the simultaneous-move PG is thus equivalent to the extensively robust without subgame-perfection implementation concept that can be obviously defined in the same way (and exactly as in Kalai's papers). Next proposition recalls the characterization of the optimal mechanisms according to those weaker implementation criteria as a benchmark. Jehiel et al [20] show that such a full surplus extraction result can still be obtained with more stringent implementation criteria, e.g. requiring agents to use dominant strategies.

Proposition 3.1 (Full Surplus Extraction à la Jehiel et al [20]) *A mechanism is implementable if and only if it is extensively robust without subgame-perfection implementable. Any optimal design (\mathbf{a}, \mathbf{t}) is such that:*

- $\mathbf{a}(N)$ is an efficient allocation
- $\mathbf{t}_i(N) = V_i^{\mathbf{a}(N)} - V_i^*(N \setminus \{i\})$ for any $i \in N$.

The optimal revenue is given by: $R_{Full}^* = \max_{\alpha \in A} \{ \sum_{i=0}^n V_i^\alpha \} - \sum_{i=1}^n V_i^*(N \setminus \{i\})$.

3.3 Clarifying remarks

3.3.1 The set of Participation Games

The set of PGs in definition 4 may seem excessively large. We give below three possible additional properties such that our analysis in section 4, 5 and 7 would not be modified under those additional restrictions as the reader can check in our proofs.

Additional Properties:

5. \mathcal{B} is a game of perfect information: all moves (including nature's moves) are publicly observable. 6. Any asymmetry between some agents in a PG should result from asymmetric actions between those agents: at each node the action and information set of players that have used the same strategy in the past should correspond

¹³Kalai considers equilibria of the simultaneous-move game where players' strategies may contain some randomness (resulting from explicit mixed strategies or from private information) and the first difference then matters in his framework.

up to a permutation of their identities. 7. The game does not include any moves from external players (e.g. any nature's moves). Nevertheless, our foundation for our implementation criteria formalized in propositions 6.1 and 6.2 relies on games with nature's moves.

3.3.2 The set of Mechanisms

In the standard mechanism design framework with simultaneous-move participation and with complete information on the payoff structure, the restriction to direct mechanisms that depend only on the set of participants and where full participation is the final equilibrium outcome is w.l.o.g. as it is well-known from the 'Revelation Principle'. The argument is the following. Consider a game \mathcal{B} where a given vector of payoffs is an equilibrium outcome and the nonparticipation strategy has a \mathcal{B} -adaptation. Then build the 'reduced' mechanism where the set of strategies is reduced to be binary: nonparticipation corresponds to the \mathcal{B} -adaptation of nonparticipation while participation corresponds to the equilibrium strategy in \mathcal{B} . The final outcome for a strategy profile s is defined as the final outcome in the original game with the strategies corresponding to s . From assumption (1) and since the original mechanism was feasible, this direct mechanism is feasible. Full participation is also an equilibrium since the new game corresponds to the original one with a truncated set of strategies such that the set of possible deviations is narrower. We emphasize that such a 'Revelation Principle' logic can not be invoked with extensively robust implementation: we do not consider equilibrium outcomes of a given game but of a family of games.

4 Optimal Design with Partially-Specified Participation Games

In a complete information setup, 'coalitional rents' leading to partial surplus extraction are surprising if the principal is able to offer 'multilateral contracts', i.e. contracts that explicitly rely on the set of other agents accepting the contracts, as assumed here. It contrasts with the way 'multilateral contracts' are perceived by the 'bilateral contract' literature, e.g. Genicot and Ray [15] states that multilat-

eral “contracts can effectively create prisoners’ dilemmas among the agents, sliding them into the acceptance of low-payoff outcomes even in the absence of coordination failure”. Our next result challenges this view by considering a new perspective on what could be meant under the terminology ‘coordination failure’. It is true that full surplus extraction can be obtained even if dominant strategy implementation is required as shown by Jehiel et al [20] and thus for various weaker implementation concepts meant to capture the absence of coordination failure as coalition-proofness, rationalizability or coalitional rationalizability. However, if coordination possibilities are reflected by the narrower requirement that the equilibrium should be robust to any extensive specification of the PG, then Theorem 1 shows that the principal may not be able to extract the full surplus.

Theorem 1 *Any extensively (subgame perfect) robust optimal mechanism (\mathbf{a}, \mathbf{t}) is such that:*

- $\mathbf{a}(N)$ corresponds to an efficient allocation
- there exists $\sigma \in \Sigma(N)$ such that $\mathbf{t}_i(N) = V_i^{\mathbf{a}(N)} - V_i^*(\{\sigma(1), \dots, \sigma(\sigma^{-1}(i) - 1)\})$ for any $i \in N$.

The optimal revenue is given by:

$$R_{Partial}^* = \max_{(\alpha, \sigma) \in A \times \sigma(N)} \left\{ \sum_{i=0}^n V_i^\alpha - \sum_{i=1}^n V_i^*(\{\sigma(1), \dots, \sigma(\sigma^{-1}(i) - 1)\}) \right\}. \quad (4)$$

The theorem does not provide explicitly the out of equilibrium outcomes and transfers of an optimal mechanism. The second and third steps of the proof construct an extensively robust mechanism for all solutions (α, σ) of the maximization program (4). This program 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 threatened taken as given the participation decision of the agents that are lower in this order. The optimal choice of α thus coincides with the maximization of the allocative efficiency.

Corollary 4.1 *Extensively (subgame-perfect) robust optimal mechanisms are efficient.*

In general, the possibility to commit to a simultaneous-move PG leads to a strictly greater payoff for the principal since $V_i^*(S)$ is decreasing in S . Under extensively robust implementation the set of optimal implementable threats is reduced to $V_{\sigma(i)}^*(\{\sigma(1), \dots, \sigma(i-1)\})$ 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 $V_i^*(\{\sigma(1), \dots, \sigma(\sigma^{-1}(i) - 1)\})$. We obtain the following corollary:

Corollary 4.2 *In a negative-externality-free framework, the revenue raised under an extensively robust optimal mechanism is equal to the full extraction revenue R_{Full}^* .*

Proof of Theorem 1

First Step

Proposition 4.3 *An extensively robust implementable mechanism is necessary STRONG-Nash implementable*

Consider a mechanism (\mathbf{a}, \mathbf{t}) and suppose that it is extensively robust implementable but not STRONG Nash implementable, i.e. there exists a set $S \subsetneq N$ and a probability distribution $\mu^S(\cdot) \in \Delta(S)$ such that $v_i^{\mathbf{a}(N)} - \mathbf{t}_i(N) < \sum_{S' \in \Delta(S)} [V_i^{\mathbf{a}(S')} - \mathbf{t}_i(S')] \times \mu^S(S')$ for any $i \notin S$ and $\mu^S(N) < 1$. Consider a PG where all the agents in S are deciding first and irreversibly whether to participate or not in the mechanism in a first step while the agents in $N \setminus S$ do not play at this stage. In a second step the remaining agents in $N \setminus S$ are playing a PG given the decisions of the agents in S while the agents in S do not play. The PG can thus be written as the sequence of two games with perfect information and with a distinct set of players: G_1 with the set S and G_2 with the set $N \setminus S$ for respectively the first and second steps. Consider the following modified PG: insert between G_1 and G_2 a sequential delegate/veto game between the agents in $N \setminus S$ in the case where all the agents in S have chosen to participate in G_1 . At each node of this intermediary sequential game, the agents in $N \setminus S$ have to chose between agreeing to delegate their participation decision such that the mutually profitable proposal $\mu^S(\cdot)$ is implemented and vetoing the proposal. If all the agents in S agree to delegate, then the final participation set is $S' \supset S$ with probability $\mu^S(S')$. On the contrary, if one agent vetoes the proposal, then the

agents in $N \setminus S$ are playing the mechanism G_2 . Denote by \mathcal{B}_1 [\mathcal{B}_2] the original [modified] whole PG. First note that the modified PG is actually a PG: a \mathcal{B}_2 -adaptation of the participation strategy for an agent $i \in S$ (respectively $i \in N \setminus S$) is the strategy consisting in playing a \mathcal{B}_1 -adaptation of the participation strategy (respectively vetoing the proposal and then playing in the continuation game after a veto a \mathcal{B}_1 -adaptation of the participation strategy). Consider an equilibrium of the game \mathcal{B}_2 such that full participation is the final outcome. It implies that all the agents in S are accepted the mechanism and that at the intermediary stages the proposal is vetoed with probability 1. Consider the last node of the sequential veto game in the case where all the previous agents in $N \setminus S$ have accepted the proposal then the last agent should accept the proposal. By backward induction the best response of each agent is to accept the proposal, which raises a contradiction with full participation being an equilibrium outcome in \mathcal{B}_2 .

Second Step

Proposition 4.4 *Any strong Nash optimal mechanism (\mathbf{a}, \mathbf{t}) is such that:*

- $\mathbf{a}(N)$ corresponds to an efficient allocation
- there exists $\sigma \in \Sigma(N)$ such that $\mathbf{t}_i(N) = V_i^{\mathbf{a}(N)} - V_i^*(T_{\sigma^{-1}(i)}^\sigma)$ for any $i \in N$.

The optimal revenue is given by: $\max_{(\alpha, \sigma) \in A \times \Sigma(N)} \left\{ \sum_{i=0}^n V_i^\alpha - \sum_{i=1}^n V_i^*(T_{\sigma^{-1}(i)}^\sigma) \right\}$.

The strong Nash optimal design program is:

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

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.

We simplify this program by showing that we can restrict ourselves w.l.o.g. 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 $S \subsetneq N$ 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: if the set of participants is S , the principal will inflict the minmax punishment to the agent $j(S, \sigma)$.

Definition 6 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 $S \subsetneq N$
- $\mathbf{t}_i(N) = V_i^\alpha - V_i^*(T_{\sigma^{-1}(i)}^\sigma)$, for any $i \in N$
- $\mathbf{t}_i(S) = 0$, if $S \subsetneq N$, for any $i \in N$.

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 optimal design with simultaneous participation.

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

Lemma 4.1 Any (α, σ) - optimal threat mechanism is strong Nash implementable.

Proof 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 $S \subsetneq N$ and the agent $j(S, \sigma)$ who does not belong to S . We have:

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

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 $T_{\sigma^{-1}(j(S, \sigma))}^\sigma = \{\sigma(1), \dots, \sigma(j(S, \sigma) - 1)\} \subset S$ (the inclusion comes from the definition of $j(S, \sigma)$). Thus we have proved that the strong Nash inequalities hold. **CQFD**

Then we show in Proposition 4.5 that, for any strong Nash 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 while extracting more surplus from all the agents. As a corollary, there

is no loss of generality to look at the rent extraction profile for an (α, σ) - optimal threat mechanism when we are characterizing the rent extraction profile of optimal mechanism.

Proposition 4.5 *For any strong Nash implementable mechanism (\mathbf{a}, \mathbf{t}) , there exists a strong Nash 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 and more surplus for all the agents.*

Proof 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

- $\sigma(1) = \text{Arg max}_{i \in N} \{V_i^{\mathbf{a}(N)} - \mathbf{t}_i(N) - V_i^{\mathbf{a}(\emptyset)}\}$ (initial step)
- $\sigma(i) = \text{Arg max}_{i \in N \setminus \{\sigma(1), \dots, \sigma(i-1)\}} \{V_i^{\mathbf{a}(N)} - \mathbf{t}_i(N) - V_i^{\mathbf{a}(\{\sigma(1), \dots, \sigma(i-1)\})}\}$ (inductive step).

The map σ is by definition a permutation. From lemma 4.1, 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}) , while extracting 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:

$$\mathbf{t}_i^{(\alpha, \sigma)}(N) = V_i^{\mathbf{a}(N)} - V_i^*(T_{\sigma^{-1}(i)}^\sigma) \geq V_i^{\mathbf{a}(N)} - V_i^{\mathbf{a}(T_{\sigma^{-1}(i)}^\sigma)} \geq \mathbf{t}_i(N). \quad (5)$$

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 $V_i^*(\cdot)$ and since $a(T_{\sigma^{-1}(i)}^\sigma) \in \mathcal{A}(T_{\sigma^{-1}(i)}^\sigma)$. (\mathbf{a}, \mathbf{t}) being strong Nash implementable implies that the strong Nash inequalities (2) are satisfied, in particular for the set $T_{\sigma^{-1}(i)}^\sigma$, i.e.

$$\max_{j \in N \setminus \{\sigma(1), \dots, \sigma(\sigma^{-1}(i)-1)\}} \{V_j^{\mathbf{a}(N)} - \mathbf{t}_j(N) - V_j^{\mathbf{a}(\{\sigma(1), \dots, \sigma(\sigma^{-1}(i)-1)\})}\} \geq 0.$$

The construction of $\sigma(i)$ guarantees that the expression in the ‘max’ is positive for $j = \sigma(i)$, i.e. $V_{\sigma(i)}^{\mathbf{a}(N)} - V_{\sigma(i)}^{\mathbf{a}(T_{\sigma^{-1}(i)}^\sigma)} \geq \mathbf{t}_{\sigma(i)}(N)$. Finally, we have proved the last inequality in equation (5). 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 also higher in the (α, σ) -optimal threat mechanism we have constructed than in (\mathbf{a}, \mathbf{t}) . **CQFD**

Third Step

Next proposition states that the existence of an order such that, for all agents, the full participation outcome is the best outcome among all possible outcomes conditional on the consent of the smaller agents according to this order is a sufficient condition for being extensively robust implementable.

Proposition 4.6 *For a given mechanism (\mathbf{a}, \mathbf{t}) , suppose that there exists a permutation $\sigma \in \Sigma(N)$ such that $U_i(N) \geq U_i(S)$ for all set $S \supset T_{\sigma^{-1}(i)}^\sigma$ and for any agent $i \in N$, where $U_i(S) = V_i^{\mathbf{a}(S)} - \mathbf{t}_i(S)$, then the mechanism (\mathbf{a}, \mathbf{t}) is extensively robust implementable.*

Proof Consider a finite extensive game \mathcal{B} . We build a strategy profile $\gamma = (\gamma_1, \dots, \gamma_n)$ for the agents in N in the following way. For all $i \in N$ denote by H_i^σ the set of histories from agent i 's point of view occurring with possible probability when the agents in T_i^σ are all assumed to play a \mathcal{B} -adaptation of the participation strategy. Consider the following game \mathcal{B}' among the agents in N and an external player that is indifferent to the final outcome: the game corresponds exactly to \mathcal{B} except that at the nodes where a given player i makes a choices and that belong to H_i^σ then it is the external player that makes player i 's choice and that for any $i \in N$.

In other words, it is the game as all agents $\sigma(i)$, $i \in N$, were 'forced' by an external player to play an adaptation of the participation strategy on the histories H_i^σ . Note that this 'artificial' game \mathcal{B}' is not a participation game but is a finite extensive game with perfect recall. From Selten [39] such a game has thus at least one subgame-perfect equilibrium. Moreover, since the external player is indifferent to the final outcomes, there exists a subgame-perfect equilibrium for any strategy profile of the external player, in particular the one where he plays a \mathcal{B} -adaptation of the participation strategy of the original player. Consider such an equilibrium profile γ' of \mathcal{B}' . This profile specifies strategies for all agent i for histories that do not belong to H_i^σ . Finally define γ_i as corresponding to a \mathcal{B} -adaptation of the participation strategy on any history $h \in H_i^\sigma$ and to γ'_i otherwise. We claim that the strategy profile γ is a subgame-perfect equilibrium. Out of the equilibrium path best response to some beliefs is guaranteed since γ' is supposed to be an equilibrium. On the equilibrium path, consider a deviation by agent $i \in N$: the best outcome he can

reach is $\max_{S \supset T_i^\sigma} U_{\sigma(i)}(S)$ (given the equilibrium belief that all the agents in T_i^σ will finally participate) which is smaller than his equilibrium outcome $U_{\sigma(i)}(N)$. **CQFD**

We conclude by joining the three steps. Consider an optimal strong Nash mechanism. From proposition 4.5 the same rent extraction profile can be raised with an (α, σ) - optimal threat mechanism. Denote by α^*, σ^* the corresponding optimal allocation and ‘order’. Define the mechanism $(\mathbf{a}^*, \mathbf{t}^*)$ as the (α^*, σ^*) - optimal threat mechanism except for the out of equilibrium transfers $\mathbf{t}_i(S)$ for $S \subsetneq N$ which are set such that $\mathbf{t}_i(S) = V_i^{\mathbf{a}^*(S)} - V_i^*(T_{\sigma^*^{-1}(i)}^\sigma)$ which guarantees that the sufficient condition from proposition 4.6 is satisfied (with an equality). Thus the extensively robust implementable mechanism $(\mathbf{a}^*, \mathbf{t}^*)$ implements the rent structure from this optimal strong Nash implementable mechanism. We obtain Theorem 1 after noting that any extensively robust implementable mechanism is STRONG Nash and thus strong Nash implementable. **CQFD**

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}) .¹⁴ 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, it is w.l.o.g. to consider mechanisms that are symmetric if the 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 set strong Nash implementable mechanisms is not convex in general, except in externality-free frameworks. Moreover, the set of optimal strong Nash implementable design are not necessary convex, except in negative externality-free frameworks. As an example of such a non-convexity, we can come back to the simple example of section 2. There are two optimal way to extract agents’ surplus each corresponding to the two permutations among two agents. However, in any (strict) mixture of those two optimal

¹⁴The 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. $(1 - \lambda)$).

mechanisms, each agents would make a loss with regards to the allocation if they both do not participate.

4.1 Partial versus Full implementation

In this paper, we mainly consider ‘partial implementation’ concepts, i.e. full participation being *an* equilibrium. Let us discuss ‘full implementation’ concepts, i.e. full participation being *the* unique equilibrium outcome. The formal way to adapt the simultaneous-move implementation criteria in definition 3 is left to the reader. The point that the partial and full simultaneous-move standard implementation criterion differ is well known. For the more stringent strong Nash and STRONG Nash implementation criteria, the difference is illustrated by the possible multiplicity of those equilibrium concepts as illustrated by the ‘Battle of the sex’ game (Fig. 1.). For definition 8, the natural concept for extensively robust full implementation is the following: a mechanism is ‘extensively robust full subgame-perfect implementable’ requires that full participation is *the* unique subgame-perfect equilibrium outcome for any PG. Under extensively robust implementation criteria, partial and full implementation are not equivalent concepts as can be checked with the ‘Pure Coordination’ game in Figure 1. To simplify, we assume that, for a given utility level, an agent strictly prefers to participate in the mechanism and to use an adaptation of the participation strategy. With this trick, the sets of full implementable mechanisms are closed sets and have thus an optimal element. In a nutshell, we assume:

Assumption 2 *If an agent’s expected utility is the same under a given strategy and an adaptation of the participation strategy, we assume that his preferences break strictly in favor of the adaptation of the participation strategy.*

Under extensively robust implementation criteria, the optimal designs that we derived in the third step of the proof of Theorem 1 are also full implementable under assumption 2. The agent $\sigma(1)$ is indifferent to all final outcomes and will then surely use an adaptation of the participation strategy in equilibrium from assumption (2). Then given that the agent $\sigma(2)$ is sure that agent $\sigma(1)$ uses such a strategy in equilibrium, he will also chose an adaptation of the participation strategy and so on. Under our various simultaneous-move implementation criteria, it can also be easily checked that full implementation requirement does not lower the principal’s revenue.

Table 2: Surplus Extraction according to the class of games for partial and full implementation concepts.

Class of Participation Games:		Surplus Extraction (structure of the design)
Simultaneous Nash, Dominant Strategy Coalition-Proof, Rationalizable Coalitionally Rationalizable	Partially-Specified a Nash Equilibrium	R_{Full}^* : Full (Prisoner Dilemma)
strong Nash STRONG Nash	a subgame-perfect Nash Equilibrium, the unique subgame-perfect Nash Equilibrium	$R_{Partial}^*$: Partial (Divide & Conquer)

Our surplus extraction results derived in this section are summarized in Table 2.

5 A non-cooperative Foundation of strong Nash Implementation

In this section we consider a slightly different set of PGs. First we enlarge the set of PGs by considering that some agents may be non-strategic and are then playing an adaptation of their participation strategy. Thus property 4. of definition 4 is satisfied only for a subset of N , called the active or strategic agents whereas the agents in the complementary set are ‘forced’ to participate. Second we narrow the set of PGs by considering that any irreversible decision to participate from one agent is followed by a participation subgame where all the remaining active agents are perfectly informed of such a decision and where their strategies in the original game are also preserved meaning in particular that they still have an opportunity to participate. This subclass of PG is labeled as participation games with subsequent opportunities and perfect information (henceforth PG with SO&PI), a terminology chosen with regard to this second alteration which has a non-commitment flavor which makes sense in light of the implicit impossibility for the seller to commit to multi-stage games as argued in section 6.¹⁵

Definition 7 *A participation game with subsequent opportunities and perfect information (PG with SO&PI) of the mechanism (\mathbf{a}, \mathbf{t}) is any finite extensive game \mathcal{B} (with perfect recall) satisfying the properties 1., 2. and 3. of the definition of a participation game of (\mathbf{a}, \mathbf{t}) and with property 4. being replaced by:*

4'. Some nature’s moves determine a partition (S_1, S_2) of N where S_1 is a set of non-strategic agents that are ‘forced’ to participate at the end and S_2 is the

¹⁵The first alteration comes from technical motivations as it will be clarified below.

set of strategic agents who are satisfying the ‘preservation of the original strategies’ property 4. from definition 4.

Moreover the additional ‘subsequent opportunities and perfect information’ property is required: after some players are definitely committed to participating in the mechanism independently of the opponents’ actions, then the subgame between the remaining players is such that it is common knowledge that those players participates in the mechanism and that the remaining players belonging to S_2 still have a \mathcal{B} -adaptation of the original strategies (no matter what strategies are used by the opponents).

Whereas the extensive version 1 of Figure 2 provides an example of a PG with SO&PI, the extensive versions 2 and 3 fail to satisfy the SO&PI property. Version 2 corresponds to a simple switch of the participation and nonparticipation decisions from version 1: after agent 1 playing ‘NO’ and agent 2 playing ‘YES’, no opportunity is left to agent 1 to revise his participation decision. Due to the symmetry between versions 1 and 2, it appears that the SO&PI property creates an asymmetry between the action consisting in agreeing and refusing the mechanism.

Definition 8 *A mechanism is extensively [full] subgame-perfect robust implementable with subsequent opportunities and perfect information if it is feasible and if in every participation game with subsequent opportunities and perfect information a [the unique] subgame-perfect equilibrium exists such that full participation is the final outcome.*

Note that the simultaneous-move game is not a PG with SO&PI. However, the property that “any extensively robust implementable mechanism with subsequent opportunities and perfect information is implementable” is still satisfied due to property 4’ which opens the possibility that a given player’s opponents are all non-strategic guarantees that participation is a best reply when all the other players participate.

The following proposition provides a characterization of extensively robust implementation in the class of PGs with SO&PI which is a great simplification from a practitioner’s point of view: the implicit (uncountable) set of constraints from extensively robust implementation criteria can be replaced by the finite set of constraints corresponding to the ‘strong Nash’ equilibrium constraints.

Theorem 2 *The strong Nash implementation criterium is equivalent to the following two implementation criteria in the class of participation games with subsequent opportunities and perfect information:*

- *extensively robust implementable*
- *extensively robust full subgame-perfect implementable*

The equivalence between the full implementation property with the strong Nash equilibrium property is surprising with regards to the possible multiplicity of equilibria satisfying the strong Nash property. Consider for example the ‘Battle of the Sex’ (Fig 1.), the game is completely symmetric with respect to the labeling of the actions (up to a permutation of the label of the agents), the strong Nash property of the full participation equilibrium is thus also satisfied by the full non-participation equilibrium. However, if you consider a PG with SO&PI and in particular those where all agents are active, all equilibrium outcomes involve full participation leaving thus no room for the non-participation equilibrium of the simultaneous-move PG. This result comes from the irreversible nature of participation decisions in the PGs with SO&PI: if the ‘row player’ chooses to participate, he knows that the ‘column player’ will have the opportunity to ‘renegotiate’ his previous participation decisions: he will be able to chose to participate and will actually do it since it is a best-reply to his irreversible commitment to participate.

Since the ‘strong Nash implementable’ optimal design program has been solved in Proposition 4.4, we immediately obtain the solutions for the ‘extensively robust implementable’ and ‘extensively robust full subgame-perfect implementable’ programs which thus correspond to the ones from the larger class of PGs in Theorem 1. In an externality-free framework, the individual rationality constraints 1 are sufficient for the strong Nash constraints 2 and thus being implementable is equivalent to being strong Nash implementable. We obtain thus the following corollary:

Corollary 5.1 *Any extensively robust implementable mechanism with subsequent opportunities and perfect information is implementable. In an externality-free framework, the converse holds: a mechanism that is implementable is extensively robust implementable with subsequent opportunities and perfect information.*

Proof of Theorem 2

‘Sufficiency’ part The sufficiency part is proved by induction on the cardinality of the set of agents that are not (irreversibly) committed to participating to the mechanism. The induction hypothesis H_k for $k \in [0, n]$ is:

H_k : Consider any set $S \subset N$ such that $\#S = k$, if the strong Nash constraints $\max_{i \in S'} \{V_i^{\mathbf{a}(N)} - \mathbf{t}_i(N) - V_i^{\mathbf{a}(N \setminus S')}\} \geq 0$ are satisfied for any $S' \subset S$, then full participation is the only subgame-perfect equilibrium outcome for any PG with SO&PI given the consent of the agent in $N \setminus S$.

The initial hypothesis H_0 is a tautology. Now consider a set $S \subset N$ such that $\#S = k+1$ and assume that the strong Nash constraints $\max_{i \in S'} \{V_i^{\mathbf{a}(N)} - \mathbf{t}_i(N) - V_i^{\mathbf{a}(N \setminus S')}\} \geq 0$ are satisfied for any $S' \subset S$. Assume however that there exists a PG with SO&PI given the consent of the agent in $N \setminus S$ and a subgame perfect equilibrium outcome that differs from full participation. From the induction hypothesis H_k and since the game is supposed to belong to the class of PGs with SO&PI, the only candidates to be a final equilibrium outcomes are either $N \setminus S$ or N .¹⁶ From the strong Nash constraint relative to the coalition S , no agent in S is strictly better under the final outcome with the set of participants $N \setminus S$. Full participation is then the unique equilibrium outcome (since we have assumed that indifference breaks in favor of participation). We conclude after noting that the hypothesis H_n corresponds to strong Nash implementation implying extensively robust full subgame-perfect implementation.

‘Only If’ Part Consider a mechanism (\mathbf{a}, \mathbf{t}) and suppose that it is not strong Nash implementable. Consider then a ‘minimal’ coalition S such that the strong Nash constraint with respect to the set of participant S is violated. That is:

- $V_i^{\mathbf{a}(S)} > V_i^{\mathbf{a}(N)} - \mathbf{t}_i(N)$, for all $i \in S$
- $\max_{i \in N \setminus S'} \{V_i^{\mathbf{a}(N)} - \mathbf{t}_i(N) - V_i^{\mathbf{a}(S')}\} \geq 0$ for any $S' \subsetneq S$.

We then carefully build a PG such that full participation is never a subgame-perfect equilibrium outcome which proves that (\mathbf{a}, \mathbf{t}) is not extensively robust implementable. The PG \mathcal{B} we build has the following structure: first all agents in $N \setminus S$ are forced irreversibly (by nature’s move) to participate to the mechanism and their participation decisions are then common knowledge; second a carefully designed PG with SO&PI is played with the remaining agent in S given the consent of the agents in

¹⁶If an additional agent participate, then the other agent can not commit to any behavior, the key element of our refinement to PGs with SO&PI.

$N \setminus S$. Such a game is defined by induction on the cardinality of the set of the agents that have already accepted the mechanism.

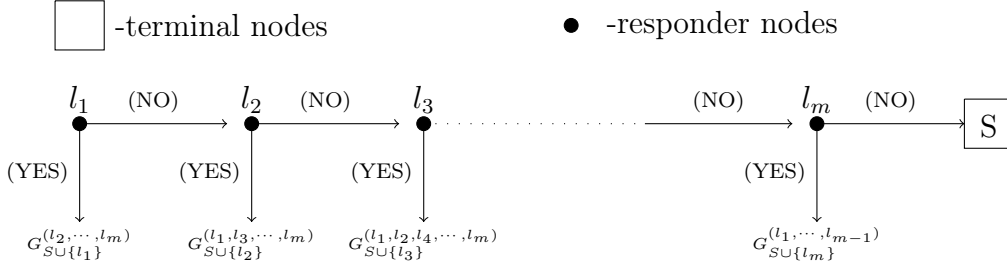


Figure 3: Tree of $G_S^{(l_1, \dots, l_m)}$

For a given mechanism (\mathbf{a}, \mathbf{t}) , denote by $G_S^{(l_1, \dots, l_m)}$ the PG between the agents in the ordered list (l_1, \dots, l_m) and given the consent of the agents in $S = N \setminus \{l_1, \dots, l_m\}$ which is properly defined by the following induction hypothesis. See Figure 3 for the tree depicting the game $G_S^{(l_1, \dots, l_m)}$. Version 1 of Figure 1 corresponds to the special case $G_\emptyset^{(1,2)}$. In the following $m = \#N \setminus S$.

There are three kinds of positions in $G_S^{(l_1, \dots, l_m)}$:

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 potential participant with the initiative.
2. Intermediate nodes of the form $G_{S \cup \{l_i\}}^{(l_1, \dots, l_{i-1}, l_{i+1}, \dots, l_m)}$, which corresponds to a participation subgame given the additional consent of agent l_i to the set S .
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 $G_{S \cup \{l_i\}}^{(l_1, \dots, l_{i-1}, l_{i+1}, \dots, l_m)}$, agents have no choice and the game moves to the responder node $(l_i, S \cup \{l_i\})$ for $i \geq 2$ ($(l_2, S \cup \{l_1\})$ for $i = 1$) if $m > 1$ or moves to the terminal node $(\mathbf{a}, \mathbf{t}, N)$ if all agents give their consent, i.e. if $m = 1$. 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 $i < m$, where l_{i+1} is the smaller index in $N \setminus S$ that is bigger than l_i . It 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. 3.

2. $(\mathbf{a}, \mathbf{t}, S)$ if $i = m$ which means that agent l_m refuses participation and the game ends at this terminal node. It corresponds to the array (NO) at the extreme right of Fig. 3.
3. $G_{S \cup \{l_i\}}^{(l_1, \dots, l_{i-1}, l_{i+1}, \dots, l_m)}$ which means that agent l_i accepts the mechanism and the game moves to the intermediate node $G_{S \cup \{l_i\}}^{(l_1, \dots, l_{i-1}, l_{i+1}, \dots, l_m)}$. It corresponds to the arrays (YES) in Fig. 3.

We also assume that the game is of perfect information meaning that all moves are publicly observed.

By applying the H_{m-1} induction hypothesis to the games $G_{S \cup \{l_i\}}^{(l_1, \dots, l_{i-1}, l_{i+1}, \dots, l_m)}$ ($i \in N \setminus S$) which are PGs with SO&PI, where S is the ‘minimal’ coalition such that the all the strong Nash constraints with respect to strict subset of S are satisfied, we obtain that choosing to participate in the responder nodes (l_i, S) will necessary lead to full participation at equilibrium for any $i \in N \setminus S$. By backward induction from l_m to l_1 we obtain then that all agents in $\{l_1, \dots, l_m\}$ prefer not to accept the mechanism at the nodes (l_i, S) . Then full participation is not an equilibrium outcome in the game $G_S^{(l_1, \dots, l_m)}$. **CQFD**

The proof of Theorem 2 should be slightly adapted if Property 6. is added to the definition of PGs (see subsection 3.3.1). In the ‘Only If’ part of Theorem 2 the games $G_S^{(l_1, \dots, l_m)}$ are not symmetric. They could be replaced by games where the players (l_1, \dots, l_m) are simultaneously asked m -times whether they accept to pre-commit to participate and where only m positive pre-commitment are validating the participation strategy. The sequential nature of the pre-commitments allows us to get rid of the bad coordination equilibria resulting from simultaneous play.

6 Foundation of our implementation criteria

First we argue that the restriction to direct, i.e. one-shot, mechanism makes sense in the perspective that the principal has a very limited commitment power: she cannot commit to any multi-stage game on the contrary to the full implementation literature, e.g. she is unable to commit to a continuation game that depends on some initial participation decisions. In other words, she cannot commit not to change the rule of the game after observing some participation report, but can rather only

commit to direct mechanisms. The final outcomes in A should be interpreted as a reduced form approach that captures possible commitment from the principal to certain given actions (possibly mixed) as in Gomes and Jehiel [16]. If she can commit to any multi-stage game (while leaving room for ‘coalitional constraints’ at each stage), the principal would be able to reach full surplus extraction à la Jehiel et al [20] by inviting sequentially each agent individually and committing to impose the tougher threat in case of nonparticipation through the continuation game in the later stages.

Second, since agents may be privately informed on the PG they are playing in the general class of PGs we consider, we can wonder whether it could be profitable for the principal to propose ‘general direct mechanisms’ where the participants are reporting some types. Let \mathfrak{M}_P denote the finite message space of participating agents. Then a ‘general direct mechanism’ denoted by (\mathbf{a}, \mathbf{t}) is a mapping that specifies a final outcome $\mathbf{a}(m)$ and a vector of monetary transfers $\mathbf{t}(m)$ for each possible set of reports $m \in (\mathfrak{M}_P \cup \{m_{NP}\})^n$, where m_{NP} corresponds to the nonparticipation action. Let $\mathfrak{M} = \mathfrak{M}_P \cup \{m_{NP}\}$. The definition of a PG \mathcal{B} for a general direct mechanism (\mathbf{a}, \mathbf{t}) can be naturally adapted from definitions 4 by requiring in particular that any report $m \in \mathfrak{M}$ has at least one \mathcal{B} -adaptation. The corresponding extensively robust implementation concept extends in an obvious way.

Third, to implement some revenue, it is not clear that there is no loss of generality to consider mechanisms where full participation is always an equilibrium outcome. The equilibrium outcome may depend both on the history of play of the agents that may use mixed strategies and on the PG in hand while still raising at least the targeted revenue.

Though the ‘Revelation Principle’ paradigm does not apply in our framework, we give an answer to the two last objections. Next proposition show that the optimal revenue $R_{Partial}^*$ that can be raised under our restricted implementation criterium and that is characterized in Theorem 1 remains optimal in the wider class of mechanisms with possibly some reports and a set of participants that may depend on the PG in hand. We give thus a foundation for the implementation criterium considered in the analysis developed in section 4. The proofs in this section are relegated in the appendix.

Proposition 6.1 *There is no feasible general direct mechanism and no revenue $R >$*

$R_{Partial}^*$ such that, for any participation game and any positive measure on nature's moves, there exists an equilibrium leading to an expected revenue that is bigger than R .¹⁷

The following proposition is the analog of proposition 6.1 for PG with SO&PI. For simplification purposes we do not consider general direct mechanisms but only consider implementation beyond the full participation equilibrium.¹⁸

Proposition 6.2 *There is no feasible direct mechanism and no revenue $R > R_{Partial}^*$ such that, for any participation game in the class of participation games with subsequent opportunities and perfect information and any positive measure on nature's moves, there exists an equilibrium leading to an expected revenue that is bigger than R .*¹⁹

We emphasize that this foundation is not a worst case scenario with regards to both the participation games and the players moves. The nature's move is nevertheless included in the worst case scenario, more precisely on a motive measure. Nature's moves could have been incorporated in the definition of a PG game which would have allowed for non-rational expectations or heterogenous beliefs that we wanted to avoid to impose directly for clarification purposes.²⁰

7 Partial subgame-perfection

In proposition 3.1 we argued that the standard implementation criterium is equivalent to the extensively robust criterium *without subgame perfection*. It can be easily shown that this equivalence still holds when we consider the class of PGs with SO&PI. On the other hand, theorem 2 establishes that the strong Nash implementation criterium is equivalent to the extensively (subgame-perfect) robust criterium in

¹⁷PGs are assumed to be finite which guarantees that nature's moves are isomorphic to an Euclidian space. The measure in the proposition refers then to the Lebesgue measure.

¹⁸The way to adapt the definition 7 of a PG with SO&PI for general direct mechanisms is less straightforward and more subject to discussions. If we consider that nonstrategic agents may be forced to play any action $m \in \mathfrak{M}_P$ then the proof of proposition 6.2 can be adapted to general direct mechanisms.

¹⁹See footnote 17

²⁰In the previous versions of the paper, such a foundation was obtained with a sledgehammer argument: by relaxing directly the common knowledge assumptions on the game that is played, more precisely by enlarging the set of possible beliefs à la Mertens-Zamir [30], while still assuming that it was common knowledge that the game was a PG with SO&PI.

the class of PGs with SO&PI. In a nutshell, the two extreme requirements in term of explicit ‘coalitional constraints’ -robustness to individual deviations or to deviations from any coalitions- are corresponding exactly to the two extreme rationality requirements in term of subgame-perfection when robustness to all extensive versions of the simultaneous-move PG is considered. Nash equilibrium imposes best responses only on the equilibrium path while subgame perfection requires best responses for any possible histories. We consider below partial subgame-perfection where best responses are imposed only on histories that are ‘closed to the equilibrium path’. The closeness to the equilibrium path for a given agent is reflected by the number of opponent players that have surely deviated from their equilibrium strategy given his current information. Though similar, our ‘ p -subgame-perfect’ equilibrium concept differs from Kalai and Neme [24]’s concept where it is the total number of deviations that is considered.²¹

Definition 9 *For a given PG, a strategy $f = (f_1, \dots, f_m)$ is a p -subgame-perfect equilibrium (p a nonnegative integer) if for all $i = 1, \dots, m$, f_i is a p -perfect response to f , i.e. for any history h_i that belongs to the set of histories where player i is sure that less than p of his opponents players have deviated then $f_i|_{h_i}$ is a best-response.*

Basic property: In any game with m active players, being a $m - 1$ -subgame perfect equilibrium is equivalent with being a subgame perfect equilibrium.

We naturally adapt our extensively robust implementation concepts to partial subgame perfection concept by replacing ‘subgame-perfect’ by ‘ p -subgame-perfect’ in the corresponding definitions. The following theorem establishes an equivalence between being extensively p -subgame-perfect robust implementable and the robustness of the simultaneous-move PG to coalition of size $p + 1$ which is labeled as $p + 1$ -strong Nash implementable.

Theorem 3 *A mechanism (\mathbf{a}, \mathbf{t}) is extensively p -subgame-perfect robust implementable in the class of participation games with subsequent opportunities and perfect information if and only if (\mathbf{a}, \mathbf{t}) is $(p + 1)$ -strong Nash implementable, i.e.*

$$\max_{i \in N \setminus S} \{V_i^{\mathbf{a}(N)} - \mathbf{t}_i(N) - V_i^{\mathbf{a}(S)}\} \geq 0 \text{ for all } S \subset N \text{ with } \#(N \setminus S) \leq p + 1. \quad (6)$$

²¹Example 1 in the supplementary material clarifies why we do not consider their concept.

When $p < n - 1$, such an equivalence between p -strong Nash implementation and extensively $p - 1$ -subgame perfect robust implementation holds naturally only for the partial implementation concept and not the full implementation concept on the contrary to Theorem 2.²²

Proof of Theorem 3

‘Sufficiency’ Part Suppose that (\mathbf{a}, \mathbf{t}) is not p -strong Nash and consider then a ‘minimal’ coalition S such that the strong Nash constraint is violated in the same way as in the proof of the ‘Only If’ part of Theorem 2. Consider then exactly the same PG with SO&PI where only the agents in S are active players. In that game we have shown previously that full participation is never a subgame perfect equilibrium outcome and thus never a $p - 1$ -subgame perfect equilibrium outcome from the basic property.

‘Only If’ Part Consider a p -strong Nash implementable mechanism. Denote by $H_i(S)$ the set of histories for player i such that exactly the players in the set S (including himself) are not definitely committed to participating to the mechanism. For any PG with SO&PI \mathcal{B} , for any player i and immediately after the release of some opponents’ participation such that $h \in H_i(S)$, consider $f_i^P(S)$ (respectively $f_i^{NP}(S)$) a \mathcal{B} -adaptation of the participation strategy (the non-participation strategy) which are existing from our definition of a PG with SO&PI. Consider the following strategy, denoted by f_i , for player i , in any history $h \in H_i(S)$ with $\#S \geq p + 1$ follow the strategy $f_i^P(S)$. In an history $h \in H_i(S)$ with $\#S \leq p$, follow the strategy $f_i^P(k)$ if $V_i^{\mathbf{a}(N)} - \mathbf{t}_i(N) - V_i^{\mathbf{a}(S)} \geq 0$ and $f_i^{NP}(S)$ otherwise. We claim that $f = (f_1, \dots, f_n)$ is a $p - 1$ -subgame-perfect equilibrium. Note that it leads surely to the full participation outcome. PGs are not in general games of ‘perfect information’ and the ‘one-shot deviation principle’ does not apply. However, we are considering a strategy profile where each player strategy depends only on a public information, as it happens the set of agents that has agreed to participate whose public information nature comes from the SO&PI property, and where it is thus sufficient for checking the global optimality of a player strategy to check its robustness according to deviations on local histories where the set of participants has not changed. Consider a deviation for player i in an history $h \in H_i(S)$ with $\#S \geq p + 1$, it modifies the final outcome only if at least p other players have deviated from their equilibrium strategies and it

²²By means of the proof of Theorem 2, we can precisely build a game with a subgame perfect equilibrium such that full participation is never an equilibrium outcome.

is thus a $p - 1$ perfect (weakly) best response. Consider a deviation for player i in an history $h \in H_i(S)$ with $\#S \leq p$.

Consider first the case where $V_i^{\mathbf{a}(N)} - \mathbf{t}_i(N) - V_i^{\mathbf{a}(S)} \geq 0$. Since the final outcome is either full participation (if at least one the players in S agrees to participate) or full non-participation of the player in S , then the adaptation of the participation strategy at this stage is a dominant response for i (possibly a weakly best response) and thus a $p - 1$ best response independently of any possible deviations in the initial stage of the game. Consider then the case where $V_i^{\mathbf{a}(N)} - \mathbf{t}_i(N) - V_i^{\mathbf{a}(S)} < 0$, the adaptation of the non-participation is a dominant response. Finally we have checked that players' strategies are $p - 1$ best responses at the equilibrium profile. **CQFD**

We emphasize that we provide a foundation for the strong Nash constraints and not the STRONG Nash constraints. The former are corresponding to the notion of 'viable coalitions' discussed by Schelling [35]. While having some intuitive appeal their theoretical status was unclear up to this point and present the weakness that if the idea is that some exogenous coalition could be at work then there is no reason to exclude more sophisticated devices in the coalition that would allow randomization and correlated strategies. Being immune to such critics the STRONG Nash constraints would have seemed more satisfactory at first glance.

We do not have a 'nice' and easily interpretable characterization of p -strong Nash optimal designs for $1 < p < n$ as in propositions 3.1 and 4.4 for the cases $p = 1$ and $p = n$ and thus not of extensively $p - 1$ -subgame perfect robust optimal designs. The p -strong Nash optimal design program is discussed in the supplementary material: a mild simplification in the vain of Proposition 4.4 is given and the difficulties for any further characterization are illustrated by an example.

8 Additional Comments

As illustrated by the example in section 2, an important class of applications where our 'coalitional constraints' are binding are auctions with negative externalities as in Jehiel et al [19, 20, 21]. This final section briefly discusses other general applications where our implementation criteria may provide new benchmarks.

8.1 Collusion in mechanism design

Our analysis can be useful in two aspects for the literature on collusion in mechanism design.

First robustness to partially specified PG can be also interpreted as a robustness criterium against collusive devices for the agents when neither binding agreements nor transfers are available, making our worst case scenario implementation criteria more relevant if we have in mind that a third party can design the PG. Main contributions on collusion-proof implementation as Laffont and Martimort [25] and Che and Kim [9] 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 [9] show that optimal non-collusive mechanism can be made collusion-proof in a broad class of circumstances including economic environment with allocative externalities. On the one hand our collusive device is much weaker: neither monetary transfers nor binding agreements on the reports are available. On the other hand, it includes participation decisions. We show that in general extensively robust optimal design raises a strictly lower revenue than standard optimal design.²³

Second, in the recent mechanism design literature on collusion as in Che and Kim [9], one agent (or 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 [6] 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. The reluctance to adopt fully the standard mechanism design approach to model collusion may come from the seemingly excessive commitment power that it implies and which is slightly softened under our stronger implementation criteria.

Let us discuss those differences in a simple example under complete information:

²³This result contrasts with the insights of Che and Kim [8] 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. We emphasize that [8] considers a negative-externality free framework: the sale of a single item in the independent private value framework.

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 collusion mechanism, which specifies the quantities produced by each participant and balanced monetary transfers among participants for each possible set of participant, is proposed by one firm, say 1. Under complete information, all the different models lead to the collusive outcome in the optimal mechanism. Nevertheless, the distribution of the profits from collusion that can be implemented are different according to the model for collusion. Under veto power, an assumption that is often made, each firm is guaranteed to obtain her non-cooperative profit $1/16$. The proposer is able to capture all the rents from collusion $\Pi_{col} - \Pi_{nc} = 1/16$. At the other extreme, if a non-participant can be punished by the minmax punishment, then nonparticipant can be threatened by the 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: firm 1 manages to extract all the surplus from trade ($1/4$) from both firms by threatening each to flood the market with the help of the other one. With our model, the maximal surplus that firm 1 can extract is intermediate: she can extract the 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.

8.2 Environments with imperfect commitment on future interactions

Our analysis brings a new benchmark in environments where the underlying allocation problem is negative-externality-free while the current environment is not negative-externality-free: seemingly pure private value environments may entail negative externalities insofar as the principal lacks the ability to commit not to propose a new mechanism if the first one fails to work. E.g. for the allocation of a pure private good, McAfee and Vincent [29] and Skreta [41] assume that the seller cannot commit never to attempt to resell the good if she fails to sell it. More generally,

externalities are the norm in environment with imperfect commitment with respect to future interaction, i.e. when long-term contracts are not available.

Gomes and Jehiel [16] 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 negative result compared with the Coasian intuition depends critically on the model for collusion. If collusion is modeled by means of the extensively robust implementation criterium, then the transposition of corollary 4.1 in their framework restores efficiency: all Markov Perfect Equilibria of the economy with general spot contract that are extensively robust are efficient, entailing an immediate move to the efficient state, where it remains forever. However, 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.

8.3 Binary games with externalities

Finally, our noncooperative foundation for the ‘strong Nash’ equilibrium concept is not limited to mechanism design but can be exported to any games with binary actions if one presents a kind of irreversibility and when there is perfect information on such participation choices. Such games occur often in economics (adoption of a technology, entry in a market) and also in politics (ratification of a treaty, recognition of a country). The irreversibility is captured by an irreversible fixed cost that is related to one action and/or to asymmetric switching costs: defectors from a contract or a political agreement are severely punished by explicit (termination penalties) or implicit (reputation effects) costs. For binary games without this irreversibility and

where the modeler would not be comfortable with the restriction to PG with SO&PI (see Schelling [35] for more examples in this area), propositions 4.3 and 4.6 can be exported to check whether an equilibrium is robust.

8.4 Related insights with bilateral contracts with externalities

We emphasize that our efficiency insight contrasts strongly with the recent literature on optimal bilateral contracts with externalities where the design of the final outcome and the optimal threats cannot be separated and may thus lead to a trade-off between efficiency and revenue extraction in complete information. On the contrary, our insight in favor of discrimination -even in symmetric framework- is already present in Genicot and Ray [15], Segal and Whinston [38], Segal [37] and Winter [45]. Nevertheless, in those papers, discrimination comes from both the bilateral nature of the contractual relationship and from severe coordination failures (unique implementation).

Genicot and Ray [15] and Segal and Whinston [38] consider also frameworks where contracting decisions may be sequential. However, the principal is assumed to have full knowledge on the timing of the game contrary to our approach. In Segal and Whinston [38], the principal can commit not to reapproach some agent after some refusal and sequentiality then enlarges the set of payoffs that the principal can implement since it breaks agents' coordination. In Genicot and Ray [15], the principal lacks the commitment to never approach an agent to whom an offer has already been made and sequentiality then weakens the principal's bargaining power.

8.5 Incomplete Information

We have restricted our analysis to a complete information setup with respect to agents' preferences. It is left for further research how to extend the notion of extensively robust implementation in incomplete information setups in order to analyze the interactions with the incentive compatibility constraints. The main issue is whether the 'coalitional constraints' are beneficial or not to the welfare. As for the concept of ratifiability introduced by Cramton and Palfrey [12], incomplete information requires a careful treatment of how agents revise their beliefs relative to

the participation decisions of their opponents.

References

- [1] A. Ambrus. Coalitional rationalizability. *Quarterly Journal of Economics*, 121(2):903–930, 2006.
- [2] A. Ambrus. Theories of coalitional rationality. *mimeo Harvard*, 2006.
- [3] R. Aumann. *Acceptable Points in General Cooperative n-Person Games*, pages 322–354. Collected Papers. MIT Press, 2000.
- [4] D. Bergemann and S. Morris. Robust mechanism design. *Econometrica*, 73(6):1771–1813, 2005.
- [5] B. D. Bernheim, B. Peleg, and M. D. Whinston. Coalition-proof nash equilibria 1. concepts. *Journal of Economic Theory*, 42(1):1–12, 1987.
- [6] B. Caillaud and P. Jehiel. Collusion in auctions with externalities. *RAND J. Econ.*, 29(4):680–702, 1998.
- [7] G. Caruana and L. Einav. A theory of endogenous commitment. *Review of Economic Studies*, 75(1):99–116, 2008.
- [8] Y.-K. Che and J. Kim. Optimal collusion-proof auctions. *Unpublished manuscript, Columbia University*, 2006.
- [9] Y.-K. Che and J. Kim. Robustly collusion-proof implementation. *Econometrica*, 74:1063–1107, 2006.
- [10] K.-S. Chung and J. Ely. Foundations of dominant strategy mechanisms. *Rev. Econ. Stud.*, 74:447–476, 2007.
- [11] O. Compte, A. Lambert-Mogiliansky, and T. Verdier. Corruption and competition in procurement auctions. *RAND J. Econ.*, 36(1):1–15, 2005.
- [12] P. Cramton and T. Palfrey. Ratifiable mechanisms: Learning from disagreement. *Games Econ. Behav.*, 10:255–283, 1995.
- [13] J. Crémer and R. McLean. Full extraction of the surplus in bayesian and dominant strategy auctions. *Econometrica*, 56(6):1247–1257, 1988.
- [14] B. Dutta and A. Sen. Implementation under strong equilibrium : A complete characterization. *Journal of Mathematical Economics*, 20(1):49–67, 1991.
- [15] G. Genicot and D. Ray. Contracts and externalities: How things fall apart. *Journal of Economic Theory*, 127(1):71–100, 2006.
- [16] A. Gomes and P. Jehiel. Dynamic processes of social and economic interactions on the persistence of inefficiencies. *J. Polit. Economy*, 113:626–667, 2005.

- [17] S. Hart and A. Mas-Colell. Bargaining and value. *Econometrica*, 64:357–380, 1996.
- [18] A. Heifetz and Z. Neeman. On the generic (im)possibility of full surplus extraction in mechanism design. *Econometrica*, 74(1):213–233, 2006.
- [19] P. Jehiel and B. Moldovanu. Strategic nonparticipation. *RAND J. Econ.*, 27(1):84–98, 1996.
- [20] P. Jehiel, B. Moldovanu, and E. Stacchetti. How (not) to sell nuclear weapons. *Amer. Econ. Rev.*, 86(4):814–829, 1996.
- [21] P. Jehiel, B. Moldovanu, and E. Stacchetti. Multidimensional mechanism design for auctions with externalities. *J. Econ. Theory*, 85:258–293, 1999.
- [22] E. Kalai. Large robust games. *Econometrica*, 72:1631–1665, 2004.
- [23] E. Kalai. Structural robustness of large games. *mimeo Northwestern University*, 2006.
- [24] E. Kalai and A. Neme. The strength of a little perfection. *Int. J. Game Theory*, 20(4):335–355, 1992.
- [25] J.-J. Laffont and D. Martimort. Mechanism design with collusion and correlation. *Econometrica*, 68:309–342, 2000.
- [26] L. Lamy. Competition between auction houses: a shill bidding perspective. *mimeo*, 2008.
- [27] L. Lamy. The shill bidding effect versus the linkage principle. *Journal of Economic Theory*, page forthcoming, 2008.
- [28] D. McAdams and M. Schwarz. Credible sales mechanisms and intermediaries. *Amer. Econ. Rev.*, 97(1):260–276, 2007.
- [29] P. McAfee and D. Vincent. Sequentially optimal auctions. *Games Econ. Behav.*, 18:246–276, 1997.
- [30] J. Mertens and S. Zamir. Formulation of bayesian analysis for games with incomplete information. *International Journal of Game Theory*, 14:1–29, 1985.
- [31] B. Moldovanu and E. Winter. Order independent equilibria. *Games Econ. Behav.*, 9:21–34, 1995.
- [32] A. Ockenfels and A. Roth. Last and multiple bidding in second price internet auctions: Theory and evidence concerning different rules for ending an auction. *Games and Economic Behavior*, 55:297–320, 2006.
- [33] D. Perez-Castrillo and D. Wettstein. Bidding for the surplus: A non-cooperative approach to the shapley value. *Journal of Economic Theory*, 100:274–294, 2001.

- [34] D. Perez-Castrillo and D. Wettstein. Choosing wisely: A multibidding approach. *American Economic Review*, 92(5):1577–1587, 2002.
- [35] T. Schelling. Hockey helmets, concealed weapons, and daylight saving: A study of binary choices with externalities. *The Journal of Conflict Resolution*, 17(3):381–428, 1973.
- [36] I. Segal. Contracting with externalities. *Quarterly Journal of Economics*, 114(2):337–388, 1999.
- [37] I. Segal. Coordination and discrimination in contracting with externalities: divide and conquer? *Journal of Economic Theory*, 113(2):147–181, 2003.
- [38] I. R. Segal and M. D. Whinston. Naked exclusion: Comment. *American Economic Review*, 90(1):296–309, 2000.
- [39] R. Selten. Reexamination of the perfectness concept for equilibrium points in extensive games. *Int. J. Game Theory*, 4(1):25–55, 1975.
- [40] R. Serrano. Fifty years of the nash program, 1953-2003. *Investigaciones Economicas*, 29(2):219–258, 2005.
- [41] V. Skreta. Sequentially optimal mechanisms. *Rev. Econ. Stud.*, 73(4):1085–1111, 2006.
- [42] T. Tan and S. R. Werlang. The bayesian foundations of solution concepts of games. *Journal of Economic Theory*, 45(2):370–391, 1988.
- [43] J. Tirole. Incomplete contracts: Where do we stand? *Econometrica*, 67(4):741–782, 1999.
- [44] R. Wilson. *Game-Theoretic Analysis of Trading Processes*, pages 33–70. Advances in Economic Theory: Firth World Congress. T. Bewley. Cambridge University Press, 1987.
- [45] E. Winter. Incentives and discrimination. *American Economic Review*, 94(3):764–773, 2004.
- [46] C. Z. Zheng. Optimal auction with resale. *Econometrica*, 70(6):2197–2224, 2002.

Appendix

A Proof of Proposition [6.1]

We first introduce some additional notation. For $m = (m_1, \dots, m_n) \in \mathfrak{M}^n$ and $S \subset N$ denote by m^S the vector of final messages reported by the agents in S . For $S \subset N$, denote m_{NP}^S the vector of messages where all the agents in S do not participate. Let $\mathfrak{S}(m) = \{i | m_i \in \mathfrak{M}_P\}$ be the set of participants that corresponds to the vector of messages m . Let $K = \max_{i \in N} \max_{a, a' \in A} V_i^a - V_i^{a'}$ be an upper bound on the maximal threats that can be imposed to a nonparticipant by subsequent changes in the other agents reports, e.g. subsequent nonparticipating decisions. For $S \subset N$ let $\mathfrak{P}(S)$ be the set of the subsets of S .

Consider a general direct mechanism denoted by (\mathbf{a}, \mathbf{t}) that specifies a final outcome $\mathbf{a}(m)$ and a vector of monetary transfers $\mathbf{t}(m)$ for each possible set of reports $m \in \mathfrak{M}^n$. Suppose that there exists a final outcome with a final set of messages m such that the principal raises a strictly higher revenue than $R_{\text{partial}}^* + n \cdot \epsilon$ for a given set of reports, the following preliminary lemma establishes that there is a subset of the participants such that all of its members would benefit at least ϵ from a joint deviation.

Lemma A.1 *Consider a final set of messages m such that the revenue of the principal is strictly bigger than $R_{\text{partial}}^* + n \cdot \epsilon$. Then there exists a subset $S \subset \mathfrak{S}(m)$ such that $V_i^{\mathbf{a}(m^{N \setminus S}, m_{NP}^S)} - (V_i^{\mathbf{a}(m)} - \mathbf{t}_i(m)) > \epsilon$, for any $i \in S$.*

In the following we call such a set $S \subset \mathfrak{S}(m)$ a set of ϵ -deviators.

Proof Suppose on the contrary that for all $S \subset \mathfrak{S}(m)$

$$\max_{i \in S} V_i^{\mathbf{a}(m)} - \mathbf{t}_i(m) - V_i^{\mathbf{a}(m^{N \setminus S}, m_{NP}^S)} \geq -\epsilon. \quad (7)$$

Then build a direct mechanism $(\mathbf{a}^*, \mathbf{t}^*)$ with no reports in the following way:

- $\mathbf{a}^*(S) = \mathbf{a}(m^{(N \setminus \mathfrak{S}(m)) \cup S}, m_{NP}^{\mathfrak{S}(m) \setminus S})$, for all $S \subset N$
- $\mathbf{t}_i^*(S) = 0$, for all $i \notin (S \cap \mathfrak{S}(m))$ and $S \subset N$
- $\mathbf{t}_i^*(S) = \mathbf{t}(m) - \epsilon$, for all $i \in S \cap \mathfrak{S}(m)$ and $S \subset N$.

We check that the following inequalities are satisfied for any $S \subset N$:

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

If $S \cap \mathfrak{S}(m) \neq \emptyset$, it is immediately satisfied since any $i \in S \cap \mathfrak{S}(m) \neq \emptyset$ is indifferent to the final outcome. For $S \subset \mathfrak{S}(m)$, it comes from the inequalities (7). Finally we have proved that the mechanism $(\mathbf{a}^*, \mathbf{t}^*)$ is strong Nash implementable. However it raises a revenue $R - n \cdot \epsilon$ that is thus strictly bigger than $R_{Partial}^*$ which raises a contradiction with proposition 4.4. **CQFD**

We will now raise a contradiction when a revenue R strictly higher than $R_{Partial}^*$ is raised for any PGs with an explicit construction of a PG.

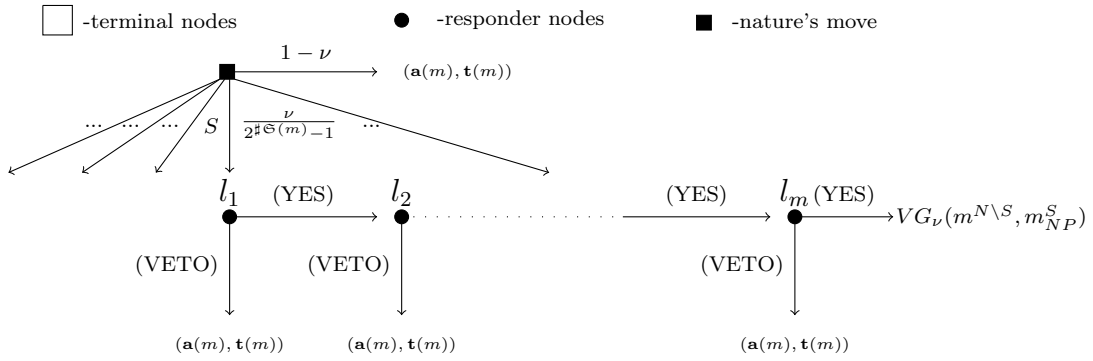


Figure 4: Tree of $VG_\nu(m)$

For any general direct mechanism (\mathbf{a}, \mathbf{t}) , we now construct by induction a finite extensive ‘veto game’ parameterized by $\nu \in (0, 1)$, for any current set of messages m , and denoted by $VG_\nu(m)$. The induction is on the number of current participants, i.e. the cardinal of the set $\mathfrak{S}(m)$. See Figure 4. When this set is empty, the allocation $(\mathbf{a}(m), \mathbf{t}(m))$ is implemented without any moves. Now consider that $\mathfrak{S}(m)$ contains at least one element. The game starts with a nature’s move each corresponding to an element of $\mathfrak{P}(\mathfrak{S}(m))$. The probability to move to the node corresponding to \emptyset is equal to $1 - \nu$ while the other moves are equally probable with probability $\frac{\nu}{2^{|\mathfrak{S}(m)} - 1}$. At the node \emptyset the game ends and the statu quo allocation $(\mathbf{a}(m), \mathbf{t}(m))$ is implemented. At a node $S = \{s_1, \dots, s_k\} \in \mathfrak{P}(\mathfrak{S}(m)) \setminus \{\emptyset\}$, the k agents in S are sequentially asked whether they accept a proposal consisting in a joint deviation where they do not participate or veto the proposal.²⁴ If there is at least one veto, the game ends at the statu quo allocation. If all the agents in S accept the proposal then the continuation

²⁴The order in the sequence does not matter.

game is the ‘veto game’ $VG_\nu(m^{N \setminus S}, m_{NP}^S)$ which is properly defined by the induction hypothesis.

Let $\epsilon \in (0, \frac{R - R_{Partial}^*}{n})$ and $\nu \in (0, 1/2)$ such that $\epsilon \cdot (1 - \nu) - \nu K > 0$. Consider a PG \mathcal{B} and then the modified game such that at each final node with a final set of messages m of the original game \mathcal{B} , the final set of messages m is publicly disclosed and the game $VG_\nu(m)$ is appended as a continuation game. The modified game is also a PG.

In the veto games $VG_\nu(m)$, consider a node where the current set of messages is m such that $\sum_{i=0}^n V_i^{a(m)} - t_i(m) \geq R_{Partial}^* + n \cdot \epsilon$ and the sequential veto game between the set of participants S such that S is a set of ϵ -deviators : any subgame perfect equilibrium strategy profile is such that the agents in S are accepting the joint deviation where the agents in S jointly do not to participate at each node where all the previous responders are accepted the proposal. Consider the history where all the agents in S except the last one have accepted the proposal to deviate. If he accepts the proposal, then with probability $1 - \nu$ he will surely win at least ϵ whereas with probability ν subsequent defections may change the final allocation. In the worse case, he will loose K . Thus the last agent in the sequence in the veto game should find it strictly profitable to accept the proposal. By backward induction, it is true for all the agents in S .

With probability at least $(\frac{\nu}{2^n - 1})^n > 0$, nature’s moves are such that the game never ends at a final message m such that the principal’s revenue is strictly bigger than $R_{Partial}^* + n \cdot \epsilon$: in at most n veto proposal stages, nature selects the node \emptyset if the revenue under the current set of messages is lower than $R_{Partial}^* + n \cdot \epsilon$ while it selects a set of ϵ -deviators otherwise (a set whose existence is guaranteed by lemma A.1). Note that in the case where all agents definitely do not participate, then the revenue is necessarily lower than $R_{Partial}^*$ since no transfer can help the principal. Finally, on a positive measure on nature’s move, the final outcome in any equilibrium of this modified game always raises an expected revenue that is strictly lower than R which raises a contradiction.

Remark Note that the proof extends immediately to general direct mechanisms where nonparticipants can send messages that may impact the final outcome while still imposing that the final allocation does belong to $\mathcal{A}(\mathfrak{S}(m))$. Note also that we do not use the point that the message space is finite. The generalization to

arbitrary message spaces would only require to adapt the current definition of PGs that assumes that the game is finite.

B Proof of Proposition [6.2]

Consider a direct mechanism (\mathbf{a}, \mathbf{t}) such that the expected revenue R is strictly bigger than $R_{Partial}^*$ for an equilibrium of any PG with SO&PI and on any positive measure with respect to nature's moves. Let $H = \max_{i \in N, S, S' \subset N} (V_i^{\mathbf{a}(S)} - \mathbf{t}_i(S)) - (V_i^{\mathbf{a}(S')} - \mathbf{t}_i(S'))$ be an upper bound on the maximal gain that an agent can expect by a deviation.

First note that from property 4' of the definition of PG with SO&PI, this class includes PGs where the agents are forced to participate which implies that:

$$\sum_{i=0}^n V_i^{\mathbf{a}(N)} - \mathbf{t}_i(N) > R_{Partial}^*.$$

From lemma A.1 there exists a set $S \subset N$ of ϵ -deviators ($\epsilon > 0$). Now consider S_1 a set of ϵ_1 -deviators ($\epsilon_1 > 0$) such that this set that is minimal, i.e. such that there is no strict non empty subset of S_1 that is a set of ϵ' -deviators ($\epsilon' > 0$). Then consider the PG such that all agents in $N \setminus S_1$ are forced to play which is publicly disclosed and the remaining agents are then playing a PG with SO&PI. In the same way as in Theorem 2 the only equilibrium outcome is nonparticipation since the participation of a single agent will trigger the full participation outcome. We thus obtain that:

$$\sum_{i=0}^n V_i^{\mathbf{a}(N \setminus S_1)} - \mathbf{t}_i(N \setminus S_1) \geq R_{Partial}^*.$$

Repeating the same argument conditional on the fact that the agents in S_1 will surely not participate there is then a minimal set of deviators S_2 among $N \setminus S_1$ with the associated $\epsilon_2 > 0$. By induction, we construct a partition (S_1, S_2, \dots, S_k) of N such that given that the agents in $S_1 \cup \dots \cup S_{i-1}$ are forced to participate while the agents in $S_{i+1} \cup \dots \cup S_k$ are forced not to participate, then in any PG with SO&PI among the agents S_i the unique equilibrium outcome is nonparticipation.

Let $\epsilon = \min_{i=1, \dots, k} \epsilon_i$. Let $\alpha \in (0, 1)$ such that $\epsilon \cdot (1 - \alpha) - \alpha H > 0$. The nature's moves will have the following underlying structure: with probability $\alpha^i \cdot (1 - \alpha)$ the set of nonstrategic agents will be $F_i = N \setminus \{S_1 \cup S_i\}$ for $1 \leq i < k - 1$ and with the remaining probability α^k all agents will be strategic $F_k = \emptyset$. We use the convention

that $F_0 = N$.

We construct by induction a family of participation games $G_{F_j, S}^{\{l_1, \dots, l_m\}}(\alpha)$, parameterized by α , where F_j ($j = 1, \dots, k$) is the set of nonstrategic agents as selected by the nature's moves and $S \subset S_j$, is a remaining set of agents that gave their consent. Finally $\{l_1, \dots, l_m\}$ the set of remaining potential participants. See Figure 5 for the tree depicting the game $G_{F_j, S}^{\{l_1, \dots, l_m\}}(\alpha)$. The notation reflects the proximity to the family of games $G_S^{\{l_1, \dots, l_m\}}$ introduced in section 4. The induction is on m the cardinal of the set of remaining potential participants. When there is no remaining potential participants, i.e. $m = 0$, the game $G_N^\emptyset(\alpha)$ coincides with the game G_N^\emptyset , i.e. ends at the full participation outcome.

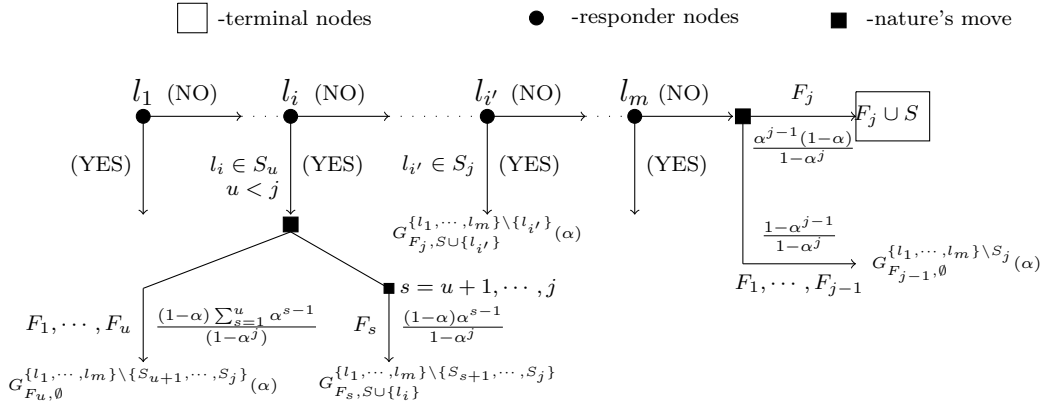


Figure 5: Tree of $G_{F_j, S}^{\{l_1, \dots, l_m\}}(\alpha)$, $j = 1, \dots, k$ and $S \subset S_j$

There are four kinds of positions in $G_{F_j, S}^{\{l_1, \dots, l_m\}}(\alpha)$:

1. Responder nodes of the form (l_i, F_j, S) , where $F_j \subset N$ is the set of the nonstrategic agents that have previously selected by nature's move, $S \subset S_j$ a set of agent (whose strategic or nonstrategic nature is still not determined by the nature) that have moved to accept the mechanism and $l_i \in N \setminus \{F_j \cup S\}$ is the identity of the potential participant with the initiative.
2. Intermediate nodes of the form $G_{F_s, S}^{\{l_1, \dots, l_{m'}\}}(\alpha)$ or $G_{F_s, S}^{\{l_1, \dots, l_{m'}\}}$, where $m' < m$ and $s \leq j$ which corresponds to a participation subgame given the additional consent of some agents and possibly more nonstrategic agents.
3. Nature's moves where a partition in the sets F_s , $s = 1, \dots, j$ is selected.
4. 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 $G_{F_s, S}^{\{l_1, \dots, l_{m'}\}}(\alpha)$ or $G_{F_s, S}^{\{l_1, \dots, l_{m'}\}}$, agents have no choice and the game moves to the responder node (l_1, F_s, S) for $m' > 0$ or moves to the terminal node $(\mathbf{a}, \mathbf{t}, N)$ if all agents give their consent, i.e. if $m = 0$. At a terminal node $(\mathbf{a}, \mathbf{t}, S)$, the game ends and the outcome $(\mathbf{a}(S), \mathbf{t}(S))$ is implemented. At a nature's move, the probabilities with respect to the different draws are depicted in Figure 5 and are coherent with the underlying nature's move structure described above. At any responder position (l_i, F_j, S) there is the choice:

1. (l_{i+1}, S) if $i < m$, where l_{i+1} is the smaller index in $N \setminus S$ that is bigger than l_i . It means that agent l_i delays participation and l_{i+1} becomes the new responder. It corresponds to the three first arrays (NO) at the left of Fig. 5.
2. The nature's move with the partition $F_j, \{F_s\}_{s=1, \dots, j-1}$ and the probabilities as depicted in Fig. 5 if $i = m$ which means that agent l_m refuses participation and. It corresponds to the array (NO) at the extreme right of Fig. 5.
3. $G_{F_j, S \cup \{l_{i'}\}}^{\{l_1, \dots, l_m\} \setminus \{l_{i'}\}}(\alpha)$ which means that agent $l_i \in S_j$ accepts the mechanism and the game moves to the intermediate node $G_{F_j, S \cup \{l_{i'}\}}^{\{l_1, \dots, l_m\} \setminus \{l_{i'}\}}(\alpha)$. It corresponds to the third array (YES) in Fig. 5.
4. The nature's move with the partition $F_j, \dots, F_{u+1}, \{F_s\}_{s=u, \dots, j-1}$ and the probabilities as depicted in Fig. 5 which means that agent $l_i \in S_u$ ($u < j$) accepts the mechanism. It corresponds to the second array (YES) in Fig. 5.

We also assume that the game is of perfect information meaning that all moves are publicly observed. Finally our construction guarantees that $G_{F_j, S}^{\{l_1, \dots, l_m\}}(\alpha)$ are PG with SO&PI. Note that the nature's moves are independent drawn according to the set of nonstrategic agents are independently from their actions: only the timing of the draws dependent of those actions.

We show by induction that the unique equilibrium strategy profile in $G_{F_j \cup S}^{\{l_1, \dots, l_m\}}(\alpha)$ is such that:

1. If $S = \emptyset$, the strategy of the agents $i \in N \setminus S$ is 'NO'.
2. If $S \neq \emptyset$, the strategy of agent $i \in S_u$ is 'NO' if $u < j$ and 'YES' for at least one agent in $\{l_1, \dots, l_m\} \cap S_j$.

Such a strategy profile leads to the following final set of participants: if $S = \emptyset$, F_u with probability $\frac{\alpha^{u-1}(1-\alpha)}{1-\alpha^j}$ for $u = 1, \dots, j$; if $S \neq \emptyset$, F_{j-1} with probability $\frac{\alpha^{j-1}(1-\alpha)}{1-\alpha^j} + \frac{\alpha^{j-2}(1-\alpha)}{1-\alpha^j}$ and F_u with probability $\frac{\alpha^{u-1}(1-\alpha)}{1-\alpha^j}$ for $u = 1, \dots, j-2$.

The induction is on m . If $m = 0$, the induction hypothesis is satisfied. Consider first the case where $S = \emptyset$. Consider l_i such that $l_i \in S_u$, $u < j$. If l_i deviates from his equilibrium profile by choosing ‘No’ it has an impact on the final outcome only in the events where nature choose F_s for $s = u, \dots, j$. In the other events, the final outcome would be F_s for $s = 1, \dots, u-1$ independently of his action. Moreover, for $s = u$ which occurs with probability $\frac{\alpha^{u-1}(1-\alpha)}{1-\alpha^j}$, the final outcome will be then F_{u-1} instead of F_u which costs at least ϵ for l_i as our construction for S_j guarantees. Moreover, for $s = u+1, \dots, j$ which occurs with probability $\frac{\sum_{s=u+1}^j \alpha^{s-1}(1-\alpha)}{1-\alpha^j}$ the gain for l_i is bounded by H . We conclude that the deviation is not profitable since $\sum_{s=u+1}^j \alpha^{s-1}H + \alpha^{u-1}\epsilon < 0$ as it can be easily checked from our choice for α . For the case where $l_i \in S_j$ the argument is the same: the difference is only that the potential gains with the term in H is not absent.

Consider now the case $S \neq \emptyset$. For l_i such that $l_i \in S_u$ with $u < j$ the argument is exactly the same as above. For $l_i \in S_j$, we now work conditional of the cases where the final set of nonstrategic agents is F_k with $k = j-1, j$. Suppose now that there are some histories where all the l_i , $i = 1, \dots, m$ such that $l_i \in S_j$ do not participate and where the final set of participants is then $F_j \cup S$, while the remaining histories are necessarily ending with the set of participants F_{j-1} from our induction hypothesis. and consider the agent l_i in S_u that prefers strictly the F_{j-1} outcome to the $F_j \cup S$. Such an agent exists from the way we construct the set S_u . This agents should strictly benefit from choosing the ‘YES’ action which raises a contraction and thus guarantees that one agent in l_i , $i = 1, \dots, m$ such that $l_i \in S_j$ will participate.

Finally, with probability α^{k-1} the set of nonstrategic agent is \emptyset (and thus on a positive measure on nature’s moves), the final set of participants in any equilibrium of the game $G_{F_k, \emptyset}^{\{1, \dots, n\}}(\alpha)$ is then \emptyset and thus the final revenue is necessarily lower than $R_{Partial}^*$ which raises a contradiction.

Supplementary Material to “Mechanism Design with Partially-Specified Participation Games”

Laurent Lamy*

1 Example 1

We provide a simple example that gives the intuition why Theorem 3 would not hold if we use the notion of p -subgame perfection according to Kalai and Neme [1]’s terminology when Property 5. (see subsection 7) is added to the definition of PGs. We consider that all the agents are symmetric: their final payoffs depend on the total number of participants and on their participation decisions according to Table A. E.g. an agent that is the only participant obtains 4 while an agent that participates with a single other participant obtains 0. The mechanism is 2-strong Nash implementable but not 3-strong Nash implementable: the full participation outcome is Pareto-dominated by the full non-participation outcome. On the contrary, given the consent of at least one agent, the full participation outcome is Pareto-dominant among all possible outcome among the remaining potential participants.

Consider the following strategy for a given agent: at each node chose an adaptation of the participation strategy except in the (out of equilibrium) case where it is common knowledge that two agents have initially deviated from their equilibrium strategies and are not committed to be participating while the third agent has accepted the mechanism. In such a case, the two remaining agents are playing an adaptation of the non-participation strategy. The strategy profile where all agents are using the strategy profile defined above is a 2-subgame perfect equilibrium for any PG with property 5. Note that in such a strategy profile agents are following strategies that depend only on public information and thus we can invoke the weak

*PSE, 48 Bd Jourdan 75014 Paris. e-mail: lamy@pse.ens.fr

Number of Participants	0	1	2	3
Participant	-	4	0	2
Non-Participant	3	1	1	-

Table A

version of the ‘one-shot deviation principle’ argued in the proof of Theorem 3 to check the equilibrium property. The only kind of histories where the equilibrium strategies are not best-response are the ones where two agents initially deviate while the third one accept the mechanism and one of the two remaining participant accepts the mechanism which corresponds to a third deviation.

2 The p -strong Nash optimal design program

Denote by \mathfrak{B} the set of reflexive binary relations over N . For a given element $B \in \mathfrak{B}$, let $S_i^B = \{j | iBj, j \neq i\}$. S_i^B should be viewed as the set of agents threatening agent i according to the relation B . A specific class of relations are total order which have been previously characterized by a permutation σ . The binary relation corresponding to the order σ is such that: jBi if and only if $\sigma^{-1}(j) \geq \sigma^{-1}(i)$. Next proposition gives a mild characterization of p -strong Nash optimal mechanisms: the final allocation is efficient while the surplus extracted from each agent equals his payoff under the final allocation minus his payoff under some harsher feasible threat according to some set of opponents.

Proposition *Any p -strong Nash optimal mechanism (\mathbf{a}, \mathbf{t}) is such that:*

- $\mathbf{a}(N)$ corresponds to an efficient allocation
- there exists a binary relation $B \in \mathfrak{B}$ such that for any $i \in N$, $\mathbf{t}_i(N) = V_i^{\mathbf{a}(N)} - V_i^*(S_i^B)$.

The optimal revenue is given by: $\max_{(\alpha, \mathfrak{B}) \in A \times \mathfrak{B}} \{ \sum_{i=0}^n V_i^\alpha - \sum_{i=1}^n V_i^(S_i^B) \}$*

As a corollary, we obtain that both corollaries 4.1 and 4.2 still hold under partial subgame-perfect implementation criteria.

Proof The p -strong Nash optimal design program is:

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

subject to $\max_{i \in N \setminus S} \{V_i^{\mathbf{a}(N)} - \mathbf{t}_i(N) - V_i^{\mathbf{a}(S)}\} \geq 0$ for all $S \subset N$ with $\#(N \setminus S) \leq p + 1$, where (\mathbf{a}, \mathbf{t}) is a feasible mechanism.

Consider (\mathbf{a}, \mathbf{t}) a p -strong Nash optimal mechanism. Suppose first that $\mathbf{a}(N)$ is not an efficient allocation. Let $(\mathbf{a}', \mathbf{t}')$ be the mechanism defined in the following way:

- $\mathbf{a}'(N)$ is an efficient allocation
- $\mathbf{t}'_i(N) = \mathbf{t}_i(N) + V_i^{\mathbf{a}'(N)} - V_i^{\mathbf{a}(N)}$, for any $i \in N$
- $\mathbf{a}'(N) = \mathbf{a}(N)$ and $\mathbf{t}'(N) = \mathbf{t}(N)$ for any $S \subsetneq N$.

$(\mathbf{a}', \mathbf{t}')$ is feasible since (\mathbf{a}, \mathbf{t}) is feasible. $(\mathbf{a}', \mathbf{t}')$ satisfies the p -strong Nash constraint since (\mathbf{a}, \mathbf{t}) does and since $V_i^{\mathbf{a}'(N)} - \mathbf{t}'_i(N) - V_i^{\mathbf{a}'(S)} = V_i^{\mathbf{a}(N)} - \mathbf{t}_i(N) - V_i^{\mathbf{a}(S)}$. The revenue raised under $(\mathbf{a}', \mathbf{t}')$ is strictly bigger than the revenue raised under (\mathbf{a}, \mathbf{t}) since the welfare is strictly bigger while agents' surplus are unchanged. We have thus raised a contradiction with regards to the optimality of (\mathbf{a}, \mathbf{t}) in the class of p -strong Nash implementable mechanisms.

Suppose that there exists $i \in N$ such that $\mathbf{t}_i(N) \neq V_i^{\mathbf{a}(N)} - V_i^*(S)$ for any $S \subset N \setminus \{i\}$. Then let $(\mathbf{a}'', \mathbf{t}'')$ be the mechanism exactly identical to (\mathbf{a}, \mathbf{t}) except that $\mathbf{t}''_i(N) = \mathbf{t}_i(N) + \epsilon$. For $\epsilon > 0$ sufficiently small then $(\mathbf{a}'', \mathbf{t}'')$ is p -strong Nash implementable mechanism and raises a revenue that is strictly bigger than (\mathbf{a}, \mathbf{t}) which raises a contradiction. **CQFD**

The characterization in the above proposition is quite coarse and illustrates how the characterization in proposition 4.4 reduces the complexity of the maximization program: the maximization moves from the set of reflexive binary relations over N to the set of total orders. In the special case where all agents are symmetric, any order leads to the same revenue making the maximization in proposition 4.4 to be vacuous. On the contrary, the maximization over reflexive binary relations is not immediate even in this simple case.

One could conjecture that the set of reflexive binary relations to characterize an optimal mechanism could be significantly shrunk by combining the structure of the threats of propositions 3.1 and 4.4, e.g. shrunk to the set of total preorders. The following example illustrates that it can be even worse: the transitivity of the binary relation $B \in \mathfrak{B}$ in the proposition is not guaranteed.

Consider $N = 3$ and a mechanism (\mathbf{a}, \mathbf{t}) such that $U_i(S) = 0$ if $i \in S$, $U_i(N \setminus \{i\}) = -m$ and $U_i(\{j\}) = -m - \epsilon$ for any $i \in N$, where $U_i(S) = V_i^{\mathbf{a}(S)} - \mathbf{t}_i(S)$. Suppose that

$m > \epsilon > 0$. It can be checked that optimal 2-strong Nash implementable mechanisms are such that the reflexive binary relation B in the above proposition is a cycle and thus violates the transitivity property: $S_i^B = \{j\}$, $S_j^B = \{k\}$ and $S_k^B = \{i\}$ for i, j, k distinct elements in N . The corresponding optimal revenue equals $3m$ which lies strictly between $3(m + \epsilon)$ the revenue under 1-strong Nash implementable mechanism and $2m + \epsilon$ the revenue under 3-strong Nash implementable mechanisms.

References

- [1] E. Kalai and A. Neme. The strength of a little perfection. *Int. J. Game Theory*, 20(4):335–355, 1992.