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. A mechanism implements a given revenue if it is a subgame-perfect equilibrium outcome 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]. We solve the ‘extensively robust’ optimal design program: revenue maximization is not in conflict with economic efficiency but the principal may fail to extract fully agents’ surplus relative to the harshest threats in presence of externalities. Our implementation criterium gives a foundation for divide and conquer strategies that discriminate among symmetric agents. The solution of the optimal design problem where agents can collude through general commitment devices without monetary transfers is obtained as a by-product.

Keywords: Mechanism Design, Robust Implementation, Full and partial Implementation, Surplus Extraction, Imperfect Commitment, Commitment Devices, Collusion on Participation

JEL classification: C70, C72, 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, Kim Sau Chung, Steve Morris, Alessandro Pavan, Hamid Sabourian, Satoru Takahashi, Olivier Tercieux, Jean Tirole and seminar participants at PSE TOM Seminar, Nashville PET 2007 Conference, the workshop on Implementation of Cooperative Solution Concepts at the 19th Stony Brook festival, Institut Henri Poincaré Paris and at the Recent Developments in Mechanism Design conference (Princeton). This paper is a major revision of Chapter IV of my Ph.D. dissertation which 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.

†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 [23], pp 1632). Economic theory usually considers specific games and studies their equilibrium set. 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 and in a non-cooperative way. 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, for example, be visible to the other agents before they are taking their own participation decisions. Furthermore, 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 emphasis on delegation as a commitment device that may benefit to the players goes back to Schelling [37]. The real participation game would then correspond to the one between the decision-makers which is then out of control from the principal’s perspective. Such a lack of control is especially relevant if the instructions are resulting from some collective decisions where the different groups can keep a close watch on each other.

In this paper, we consider the mechanism design problem under complete information\(^1\): the principal announces a mechanism which specifies economic outcomes and transfers as a function of the full set of messages reported by the agents. Any agent can choose not to participate to the mechanism which guarantees that his transfer with the principal is null and which may also reduce the set of feasible outcomes that

\(^1\)We consider that the principal is fully informed. It should not be confused with “implementation theory” that considers that there is no asymmetric information between agents and that the principal is uninformed as in Maskin [28].
the principal can implement. In the same vein as Jehiel et al. [20] who considers the allocation of a physical good under a non-dumping assumption, a crucial element of our model is that the set of feasible outcomes may depend on the set of participants. The mechanism is then played non-cooperatively by the agents. The current approach in mechanism design considers that the game between the agents is the one where report decisions are taken simultaneously, an approach which implicitly assumes that the principal controls the way agents interact. Under such a paradigm, the principal can achieve full surplus extraction with respect to the harshest threats, i.e. can select the efficient outcome while raising the maximal revenue from each agent according to the threat that she would select the worst outcome for him if he does not participate to the mechanism. Moreover the reliability of the equilibrium prediction can be strengthened insofar as the optimal revenue can be achieved by a mechanism where agents are using dominant strategies as shown by Jehiel et al. [20]. We reexamine this problem when the principal has no idea on the specific participation game which is played among all extensive versions of the simultaneous-move participation game as defined by Kalai [23, 24]. More precisely we adopt a ‘min-max’ approach: the principal achieves a revenue $R$ with a given mechanism if in any participation game there is a subgame-perfect equilibrium which raises at least $R$, a way to capture the lack of knowledge of the principal on the details of how the participation game is played. As it is developed in section 7, there is no fundamental difference in our framework between what can be implemented under a partial implementation criterium as stated above, i.e. the existence of a subgame-perfect equilibrium satisfying a given property, and a full implementation criterium which would require that any subgame-perfect equilibrium satisfies a given property. The main difference is the need to break indifferences and thus the need to use strict incentives to participate under a full implementation criterium. In order to alleviate notation, we thus consider mainly a partial implementation perspective, which may seem awkward with respect to our partially-specified games paradigm. Relaxing the common knowledge assumptions on the trading game may seem at odd with the original formulation of the ‘Wilson Doctrine’ (Wilson [46]) whose agenda is to relax the common knowledge assumptions on players’ beliefs about another’s preferences and information. However in environments where enforcement on the details of the

\[\text{See however remark 7.1 for an argument in favor of partial implementation in our framework.}\]
participation process seems difficult, we do think that we are remaining in line with the spirit if not the letter of the ‘Wilson doctrine’.\(^3\)

The main contribution of the present paper is Theorem 1 where the optimal revenue is characterized under our implementation criterium with partially-specified participation games and which is called *extensively robust implementation*. Full extraction relative to the harshest threats as in Jehiel et al. \cite{20} does not work anymore, in general, in presence of negative externalities. Robustness to all participation games is shown to imply that the principal can not expect to end at an allocation such that a coalition of agents would strictly benefit by jointly not participating to the mechanism with respect to this final allocation. Those explicit coalitional constraints allow us to derive upper bounds on the surplus profile that the principal can extract from the agents. Those bounds are shown to be reached with simple mechanisms without any report to the principal and such that full participation is an equilibrium outcome in any participation game (henceforth PG). The second part of Theorem 1 characterizes the final outcomes and the surplus extraction structure in the optimal mechanisms that belong to this restricted class of simple mechanisms. The Coasian logic still applies and we obtain that optimal mechanisms are efficient. Furthermore, those optimal profiles have 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 threat another agent, say 2, to use agent 1 in case of non-participation in order to minimize his payoff, in some ‘credible’ way. Then she threats the next agent conditionally on the participation of agents 1 and 2 and so on. Seeking optimal mechanism consists thus in finding the optimal ‘order’ to define those threats, which corresponds to a permutation among the agents.

Our insights are linked to three main topics in mechanism design: imperfect commitment, the possibility of full surplus extraction and robustness concerns. 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.\(^4\) In corporate acquisitions and procurement

\(^3\)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 \cite{35} is a good example of how small details on the trading rules can have a significant impact on participation decisions.

\(^4\)Closely related is the literature on commitment failures with regards to future interactions. See
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 [29]). 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 [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. 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. Third, some papers advocate for alternatives to the use of the Bayesian equilibrium concept: dominant strategy (Chung and Ely [10]) or ex-post equilibrium (Bergemann and Morris [4]) emerges on grounds that relax the common knowledge assumption on the distribution of agents’ private signals while still assuming a simultaneous-move PG.

With a common concern for robustness, this paper shows that the principal may not be able to fully extract agents’ surplus relative to their harshest 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’ that emerge implicitly 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. Under a submodularity condition, the relative loss that comes from coalitional rents is shown to be bounded by one half.

The paper is organized as follows. In section 2 we introduce the general allocation problem and our ‘extensively robust’ implementation criterium. Our basic insights and the main ingredients of our proofs are illustrated informally by means of a simple example with two agents in section 3. The optimal design program under extensively robust implementation is solved in section 4, the core of the paper. Section 5 derives conditions for implementation with simple mechanisms, i.e. mecha-
anisms without messages and such that full participation is an equilibrium in any PG. Section 6 is devoted to some links, in terms of the set of implementable revenues and implementable surplus profiles, between extensively robust and strong Nash implementation criteria. The latter plays a key role in our proofs. A by-product of our analysis on partially-specified PGs is thus a partial contribution to the Nash program which aims to bridge the gap between the non-cooperative and cooperative approaches to game theory.\footnote{See Serrano [42] for a survey.} Full implementation criteria are analyzed in section 7. The transposition of our results in term of robustness against general commitment devices is discussed in section 8. Clarifying remarks about our modeling choices are gathered in section 9 while section 10 is devoted to additional comments, including economic motivations and applications for our analysis.

2 The Model

2.1 The Model

Let $N = \{1, 2, \ldots, n\}$ be a set of agents and $A = \{a_1, a_2, \ldots, a_K\}$ be a finite set of possible outcomes. Denote by $\Sigma(N)$ the set of the permutations over the set $N$. For a given permutation $\sigma : N \rightarrow N$, denote by $T_\sigma^i$ the subset $\{\sigma(1), \sigma(2), \ldots, \sigma(i-1)\}$, i.e. the $i-1$ first smallest agents according to the implicit order defined by $\sigma$. In particular, $T_\sigma^{i-1}(i)$ corresponds to the set of agents that are (strictly) smaller than agent $i$ according to the order $\sigma$. Note that $T_\sigma^1 = \emptyset$ and $T_\sigma^n = N$. Denote by $\sharp S$ the cardinality of a set $S \subset N$. We assume that the agents and the principal, characterized by the subscript 0, have quasi-linear 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: $\mathcal{U}_i(a, t_i) = V^a_i - t_i$. As discussed in section 9.3, we emphasize that the quasi-linearity assumption is not critical to our analysis.

The principal announces a direct mechanism, denoted by $(a, t)$, that specifies a final outcome $a(m) \in A$ and a vector of monetary transfers $t(m) \in \mathbb{R}^n$ for each possible set of messages $m = (m_1, \ldots, m^n) \in \mathcal{M}^n$, where $\mathcal{M} = \mathcal{M}_P \cup \{m_{NP}\}$ with $\mathcal{M}_P$ denoting a finite message space that participating agents can report to the principal and $m_{NP}$ corresponding to the nonparticipation action. For $S \subset N$ denote by $m^S$
(resp. \( m_{NP}^S \)) the vector of final messages reported by the agents in \( S \) (the vector of messages where all the agents in \( S \) do not participate). Let \( \mathcal{S}(m) = \{ i \mid m^i \in \mathcal{M}_P \} \) be the set of participants that corresponds to the vector of messages \( m \). We call \textit{simple mechanisms} the direct mechanisms where the final outcome and the monetary transfers depend solely on \( \mathcal{S}(m) \) the set of participants. In a simple mechanism \((a, t)\), let \( a(S) \) and \( t(S) \) denote -with a mild abuse of notation- the final outcome and the vector of monetary transfers if the set of final participants is \( S \). 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 \textit{feasible} if for any \( m \in \mathcal{M}^n \):

- For each set of participants \( \mathcal{S}(m) \), the final outcome belongs to the set of probability distributions on \( \mathcal{A}(\mathcal{S}(m)) \), the subset of \( \mathcal{A} \) of accessible or feasible outcomes with the consent of agents in \( \mathcal{S}(m) \).
- If agent \( i \) decides not to participate, the principal cannot extract a positive payment from that agent: \( t_i(m) \leq 0 \) if \( m^i = m_{NP} \).
- Transfers are budget-balanced: \( \sum_{i=0}^n t_i(m) = 0 \).

The second and third restrictions are standard. The first restriction means that some outcomes in \( \mathcal{A} \) 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 [40] 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:

\textbf{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. Up to this stage our model did not exclude that participation decisions could be an integral part of the final outcome. However, assumption (1) implicitly excludes environments where participation decisions are costly, as in Bernstein and Winter [5], where the optimal contracting schemes may involve partial participation. Then, there is no loss of generality to consider that \( \mathcal{A}(N) = \mathcal{A} \). We call an efficient allocation any allocation \( a \in \mathcal{A} \) that maximizes \( \sum_{i=0}^n V^a_i \). For an
agent \(i\) and a set of participants \(S \subset N \setminus \{i\}\), denote by \(a_i^*(S)\) the harshest feasible threat that the principal can inflict on \(i\) given that the agents in \(S\) have accepted the mechanism: \(a_i^*(S) \in \text{Arg}\min_{a \in A(S)} V^a_i\). 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 to participate in the mechanism, the principal is committed to use the remaining agents to impose this harshest 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 quadruplet: \((N, A, \{V^a_i\}_{i \in N, a \in A}, \{A(S)\}_{S \subset N})\).

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 \(a \rightarrow V^a_i\) is constant over the set \(A(N \setminus \{i\})\).

- A framework is said to be negative-externality-free if, for any agent \(i\), the utility level \(V_i^*(S)\) corresponding to the harshest threat given the set of participant \(S \subset N \setminus \{i\}\) is independent of \(S\): \(V_i^*(S) = V_i^*(\emptyset)\) 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 about 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.\(^6\)

\(^6\)Genicot and Ray [16] discuss extensively related applications that go beyond industrial organization applications as the traditional allocation of a patent or a licence considered by Jehiel and Moldovanu [19].
The crucial assumption of the model is the restriction to direct, i.e. one-shot, mechanism: we argue that it makes sense in the perspective that the principal has a very limited commitment power, i.e. she cannot commit to any multi-stage game, contrary e.g. to Moore and Repullo [34] for the full implementation literature. In particular she is unable to commit to a continuation game that depends on some initial participation decisions. In a nutshell, she cannot commit not to change the rule of the game after observing some participation report, but can rather only commit to direct mechanisms.\(^7\) In this perspective, the final outcomes in \(A\) should not be viewed as an irreversible economic outcome but should rather 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 [17]. 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 by inviting sequentially each agent individually and committing to impose the tougher threat in case of nonparticipation through the continuation games in the later stages designed to provide the right incentives to participate to the remaining agents.\(^8\)

### 2.2 Simultaneous-move implementation concepts

We first introduce simultaneous-move implementation concepts that do not correspond to what this paper is interested in per se but that correspond to useful benchmarks. Mechanisms are simply labeled as ‘implementable’ or ‘strong-Nash implementable’ when full participation satisfies the corresponding equilibrium property in the simultaneous-move PG.

**Definition 2** Full participation is an equilibrium, respectively a strong-Nash equilibrium, of the simultaneous-move participation game in a simple mechanism \((\mathbf{a}, \mathbf{t})\) if respectively:

\[
IR(i) : V_{i}^{a(N)} - t_{i}(N) - V_{i}^{a(N\backslash\{i\})} \geq 0, \text{ for all } i \in N, 
\]

\[
CP(S) : \max_{i \in S} \left\{ V_{i}^{a(N)} - t_{i}(N) - V_{i}^{a(N\backslash S)} \right\} \geq 0 \text{ for all } S \subset N. 
\]

\(^7\)This perspective is especially relevant in environments where the designer would compete with multiple principals.

\(^8\)However, such continuation games may violate some basics renegotiation-proofness constraints.
The constraints (1) correspond 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 \( S \subseteq N \), 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 strictly 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 is considered and which will be labeled later as STRONG-Nash equilibrium and discussed in section 5.

**Definition 3 (simultaneous-move implementation)** A simple mechanism \((a, t)\) is implementable [resp. strong Nash implementable] if it is feasible and if full participation is an equilibrium [resp. a strong Nash equilibrium] of the simultaneous participation game. If so, we say that the mechanism \((a, t)\) implements [resp. strong Nash implements] the revenue \(V^a_0(N) + \sum_{i=1}^n t_i(N)\) and the surplus profile \((V^a_1(N) - t_1(N), \ldots, V^a_n(N) - t_n(N))\).

Any strong Nash implementable mechanism is necessary implementable as the ‘coalitional constraints’ are including the usual individual rationality constraints which correspond to \(CP(\{i\}), i \in N\). The converse does not hold in general as it is illustrated by prisoner’s dilemmas. However, in externality-free frameworks it does: the individual rationality constraints (1) imply any strong Nash constraints 2. The resolution of the strong Nash optimal design program will play a central role in our analysis.

### 2.3 Extensively robust implementation

The ‘extensively robust implementable’ terminology refers to an outcome that is 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. 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 [24]'s formalization.
Definition 4. A participation game (PG) of \((a, t)\) is any finite extensive game \(\mathcal{B}\) (with perfect recall) with the following properties:

1. \(\mathcal{B}\) includes the (original) players: The set of players of \(\mathcal{B}\) constitutes a superset of \(N\).

2. Playing \(\mathcal{B}\) means playing \((a, t)\): With every pair \((a, t)\) associated with a final node of \(\mathcal{B}\), there is an associated unique message profile \(m\) such that \((a, t) = (a(m), t(m))\).

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. \(U_i(a, t_i)\).

4. Preservation of the original strategies: for any player \(i \in N\), every pure strategy \(m \in \mathcal{M}_P \cup \{m_{NP}\}\) 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 reports \(m\) such that \(m^i = m\).

External players are allowed to participate in such variations as in Kalai [24]. 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 mainly consider PGs with a unique external player that is labeled as ‘nature’. See Kalai [23] for a panel examples with rounds of revisions, changes in the order of play or nature selecting some moves, illustrating how rich the set of PGs is. Figures 1, 3 and 5 hereafter depict three examples of PGs for the two agents case: Figures 1 and 5 are PGs that are limited to simple mechanisms, while reports are allowed in the PG depicted in Figure 3. The final payoffs are thus determined by the set of participants (Figure 1 and 5) or the final set of messages (Figure 3) that are given in the brackets at the final nodes of the game tree. The black points are corresponding to the decision nodes where players (possibly an external player labeled N for ‘nature’) are choosing an action depicted by an arrow. The extensive versions in Figure 1 and 3 are discussed more carefully in section 3. Those examples are games with perfect information. We emphasize that the set of PGs is not a subset of extensive games with perfect information: definition 4 allows for general extensive games with imperfect information.
Definition 5 (extensively robust implementation) A revenue $R$ is extensively robust implementable if there is a feasible mechanism such that for every participation game there exists a subgame-perfect equilibrium that guarantees the expected revenue $R$ for almost all nature’s move.\(^9\)\(^10\)

Our simultaneous-move implementation concept has required that full participation is an equilibrium in a simple mechanism. To ‘implement’ a given revenue $R$ we could have seek more generally for feasible direct mechanisms (with reports) such that there exists an equilibrium that guarantees the expected revenue $R$ in the simultaneous-move participation game. However, 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 the simultaneous-move PG $B$ and an equilibrium outcome in a given mechanism $(a, t)$. Then build the ‘reduced’ mechanism where the set of strategies is reduced to be binary: nonparticipation corresponds to the $B$-adaptation of nonparticipation in the original game while participation corresponds to the equilibrium strategy in $B$. Full participation leads thus to the equilibrium outcome in the original game. From assumption (1) and since the original mechanism was feasible, this 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 since we do not consider equilibrium outcomes of a given game but of a family of games and furthermore, while we assume complete information on the final payoff structure, we do not assume complete information on the game that is played between agents. That is the reason why we had to return to the more basic definition of an implementable

\(^9\)We emphasize that this criterium 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. 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. In 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 [31], while still assuming that it was common knowledge that the game was a PG.

\(^10\)PGs 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.
revenue for the more stringent extensively robust implementation criterion.

Since the simultaneous-move PG belongs to the class of PGs, the set of implementable revenues under extensively robust implementation is smaller than under the standard implementation requirement. Though any kind of ‘Revelation Principle’ is not available under extensively robust implementation, we show in the analysis below that when we are seeking for an optimal mechanism we can restrict ourselves w.l.o.g. to ‘extensively robust simply implementable’ mechanisms.

**Definition 6 (extensively robust simple implementation)** A mechanism is extensively robust simply implementable if it is simple, feasible and if in every participation game a subgame-perfect equilibrium exists such that full participation is the final outcome. If so, we say that the revenue \( V_a(N) + \sum_{i=1}^{n} t_i(N) \) and the surplus profile \( (V_1^{a(N)} - t_1(N), \ldots, V_n^{a(N)} - t_n(N)) \) are extensively robust simply implementable.

**Remark 2.1** We could use an alternative concept as perfect equilibrium (Selten [41]). It may seem that we are then too permissive by allowing for subgame-perfect equilibria which are not perfect. Indeed, the optimal extensively robust implementable design we build in the second step of the proof of Theorem 1 fails to be perfect in some PGs due to indifferences: in equilibrium, agents are indifferent between participating and not participating but if the other agents tremble then the indifference breaks down in favor of nonparticipation. In the same way as going from partial implementation to full implementation can be obtained by ‘breaking the ties’ with infinitesimal additional transfers and giving thus strict incentives to participate as mentioned earlier and developed subsequently in section 7, ‘breaking the ties’ would allow us to implement under the stronger requirement of the existence of a perfect equilibrium any strictly smaller revenue than a given implementable one or any strictly greater agents’ surplus profile than a given simply implementable one.\(^{11}\)

### 2.4 The full extraction benchmark

Our extensively robust implementation concepts differ in an important way from Kalai’s approach since we are requiring that equilibria leading to our desired outcome to be subgame-perfect. Without subgame-perfection, all pure strategy equilibria in

\(^{11}\)See also footnote 25.
the simultaneous-move game survive in the extensive versions: agents are following an adaptation of their original strategy in the simultaneous-move game.\textsuperscript{12} The standard implementation concept in the simultaneous-move PG is thus equivalent to the extensively robust implementation concept without subgame-perfection that can be obviously defined in the same way. Next proposition recalls the characterization of the optimal mechanisms according to those weaker implementation criteria as a benchmark: it reaches full surplus extraction with respect to the harshest threats.

Proposition 2.1 (Full Surplus Extraction à la Jehiel et al. [20]) A simple mechanism is implementable if and only if it is extensively robust simply implementable without subgame-perfection.\textsuperscript{13} Any optimal simple mechanism \((a, t)\) is such that:

- \(a(N)\) is an efficient allocation,
- \(t_i(N) = V^{a(N)}_i - V^{*}_i(N \setminus \{i\})\) for any \(i \in N\).

The optimal revenue is given by:

\[
R^*_{\text{Full}} = \max_{a \in A} \left\{ \sum_{i=0}^{n} V^{a_i} - \sum_{i=1}^{n} V^{*}_i(N \setminus \{i\}) \right\}.
\]

Jehiel et al. [20] show that such a full surplus extraction result can still be obtained with a more stringent implementation criterium where agents are required to use dominant strategies and thus for various weaker implementation concepts meant to capture the absence of coordination failure as coalition-proofness, rationalizability or if the final outcome is required to be the one in any correlated equilibrium.

3 A Simple Example\textsuperscript{14}

Consider the sale of a single object involving externalities among two potential competitors. Let \(A = \{0, 1, 2\}\) where outcome \(i\) corresponds to the assignment to

\textsuperscript{12}Kalai mainly considers equilibria of the simultaneous-move game where players’ strategies contain some randomness (resulting from explicit mixed strategies or from private information) such that it is not straightforward that any equilibrium of the original game remains an equilibrium (possibly not subgame-perfect) in the extensive versions.

\textsuperscript{13}By “extensively robust [simply] implementable without subgame-perfection”, we mean the definition of “extensively robust [simply] implementable” where we replace ‘a subgame-perfect equilibrium’ by ‘a Nash equilibrium’.

\textsuperscript{14}A reader eager to get our general results may skip this section.
Table 1: Prisoner’s dilemma

<table>
<thead>
<tr>
<th></th>
<th>NP</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP</td>
<td>(0,0)</td>
<td>(−α₁,V)</td>
</tr>
<tr>
<td>P</td>
<td>(V,−α₂)</td>
<td>(−α₁ + \frac{\epsilon}{2},−α₂ + \frac{\epsilon}{2})</td>
</tr>
</tbody>
</table>

agent \(i\). Bidders 1 and 2 are valuing intrinsically (with regards to the status 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 bidder \(i\) \((i = 1, 2)\) suffers from a negative allocative externality \(\alpha_i > 0\) when the object is allocated to his opponent.

According to our notations we have: \(V^i = V\) for \(i = 1, 2\), \(V^0 = v\) and \(V^j = -\alpha_i\) for \(i, j \in \{1, 2\}, i \neq j\). Consider w.l.o.g. that \(\alpha_2 \geq \alpha_1\). Allocative externalities are supposed to be important enough such that the efficient economic outcome consists in keeping the object: \(\alpha_1 > V - v\). We consider also that the seller is able to allocate the object only to participating agents, i.e. \(\mathcal{A}(S) = S\).

**Jehiel et al. [20]’s mechanism**  We consider first the simple mechanism derived by Jehiel et al. [20] where participation is a strictly dominant strategy for both agents in the simultaneous-move PG and where the full participation outcome is efficient and leads to full surplus extraction with respect to the harshest threats up to \(\epsilon > 0\). Each participant \(i\) has to pay \(\alpha_i - \epsilon/2\) in order to avoid that the seller gives the object to his opponent under full participation, the seller keeps the object and raises the revenue \(v + \alpha_1 + \alpha_2 - \epsilon\). Agents’ final payoffs as a function of the participation decisions are given in Table 1, where NP and P respectively correspond to the nonparticipation and participation decisions. The payoff matrix has the structure of a ‘Prisoner’s dilemma’ where participation corresponds to defection.

If participation decisions are modeled as resulting from a simultaneous-move PG as in the standard mechanism design paradigm, the unique equilibrium outcome is that both agents participate and are suffering respectively from the losses \(\alpha_1 - \epsilon/2\) and \(\alpha_2 - \epsilon/2\) compared to their profits in the case where they could jointly coordinate themselves not to participate and where the no sale assignment then prevails. Such an insight does not hold anymore if we consider a sequential PG, as the one depicted in Figure 1, that can be viewed as a ‘simple’ alteration of the simultaneous-move game. The game corresponds to the sequential game where agent 2 makes his par-
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 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 after agent 1 initially chooses the action ‘NO’, agent 2 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. Under our extensively robust implementation criterium, we obtain that the seller does not implement any revenue strictly greater than \( v \) under the optimal mechanism proposed by Jehiel et al. [20].

Extensively robust optimal mechanisms

A. Simple Mechanisms  Figure 1’s PG illustrates more generally that there is no simple mechanism such that full participation is an equilibrium in any PG and such that the final surplus of both agents are strictly negative. Otherwise, they could

---

This paper adopts a normative perspective, however the robustness according to all PGs can be illuminating from a positive perspective. In standard auctions, we obtain with Figure 1’s PG a paradox that cannot emerge in previous models with simultaneous-move PGs: an agent may prefer not to submit a bid though his intrinsic value for the good is greater than the reservation price \( v \). \( v \) can equivalently be viewed as a reduced form for a third potential bidder with a pure private valuation \( v \) such that 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. Strategic nonparticipation has thus here a completely different nature than in Jehiel and Moldovanu [19], where \( V \) was assumed to be smaller than \( v \) such that the final assignment, when only one of the bidders participates in the auction, was that the seller keeps the object.
jointly not participate and obtain a null payoff, which is strictly Pareto improving, since the seller is assumed to be unable to ‘dump’ the object to a non-participant. Surplus profiles that are sustainable in this way have to satisfy: the individual rationality constraints (IR), agent $i$’s surplus ($i = 1, 2$) should be greater than $-\alpha_i$, and also the coalitional participation constraint (CP) requiring that the surplus of at least one agent is positive. Consequently, the seller can never keep the object while extracting a strictly positive surplus (with respect to the harshest threats) from both agents 1 and 2 since some ‘coalitional participation constraints’ would be violated. The surplus profiles of the agents satisfying those constraints are depicted by the shaded area of Figure 2: this set is called the strong-Nash implementable surplus profiles. Among them we can distinguish two minimal surplus profiles labeled as $A$ and $B$. Conversely, those surplus profiles can be implemented as the equilibrium outcome of any PG by means of a divide and conquer strategy. For the profile $A$, which is then the optimal extraction profile, it consists in giving the incentive to participate for agent 1 independently of the participation decision of agent 2. Then given that agent 1 surely participates, she could really threat agent 2 to allocate the object to agent 1 in case of non-participation. The order of the threats is then $(1, 2)$. The seller’s revenue is $v + \alpha_2$. For the profile $B$, it corresponds to the analog of the above mechanism after swapping the order of the threats which becomes then $(2, 1)$. Table 2 provides explicitly the payoff matrix of simple mechanisms that implement the profiles $A$ and $B$. Consider the left panel that implements $A$. In any PG, a strategy profile such that agent 1 uses a strategy that guarantees him to participate independently of the strategy of her opponent while agent 2 uses also such a strategy provided that she faces an history that is compatible with agent 1 playing his equilibrium strategy. The existence of such a strategy profile comes directly from the main ingredient of the formal definition of PGs where each agent is assumed to be able to mimic any pure-strategy of the simultaneous-move game under any circumstances. Agent 1 has no incentive to deviate since his equilibrium payoff is the maximum he could expect among all final allocations. Given that agent 1 participate for sure, agent 2 has also no incentive to deviate since participation is a best-response to agent 1’s participation.
B. General Mechanisms  

The above discussion refers to simple mechanisms with full participation. Indeed, by means of more complex mechanisms with reports to the seller -we could conjecture that the seller may benefit from the agents reporting information either about the PG in hand or about the actions played by his opponents- and without imposing that full participation is the unique equilibrium outcome, the seller has no way to implement a better revenue than \( v + \alpha_2 \) as shown by the extensive version depicted in Figure 3. The first stage (‘Report Stage’) corresponds to the sequential game where each agent \( i \) may either report a message \( m^i \) that belongs to \( \mathcal{M}_P \) (the set of possible messages) or choose not to participate, which corresponds to the message \( m_{NP} \). This stage is a simple alteration of the simultaneous-move game: in particular, each agent can mimic any pure strategy of the original simultaneous-move game. Then the remaining stage of the game is a sequence of ‘veto games’ among a subset of the participants where the proposal is an irreversible joint deviation not to participate while the status quo is the final outcome according to the current chosen messages. The proposal prevails only if the agents that are asked to veto or accept it choose the later action (‘ACCEPT’). The availability of the ‘VETO’ action at each node of this stage guarantees that the ‘mimicking’ property still holds in the whole extensive game. At this stage, nature (player N) is moving to choose the veto game. Consider the node after agent 1 playing \( m^1 \in \mathcal{M}_P \) and agent 2 playing \( m^2 \in \mathcal{M}_P \): nature then chooses with probability \( 1 - \epsilon \) that the game ends immediately (extreme
left arrow) and with probability $\epsilon/3$ among the three set of participants $\{1\}$ (left arrow), $\{2\}$ (middle arrow) and $\{1, 2\}$ (right arrow). The agents in the selected set are playing the veto game: after any veto action, the game definitely ends at the current set of reports without any further veto proposal. On the contrary, if all selected agents have accepted the proposal, then the remaining non-selected agents whose current reports belong to $\mathcal{M}_p$ may be involved in a subsequent veto game. E.g. after the reports $(m_1, m_2) \in \mathcal{M}_p$, and after nature’s left move, then agent 1 is playing a veto game: if he vetoes then the whole game ends at $(m_1, m_2)$; if he accepts the proposal, then the current set of reports becomes $(m_{NP}, m_2)$ and the continuation game is a veto game with agent 2 which is the same game as the one that he would have played if the set of reports after the first stage of the game where directly $(m_{NP}, m_2)$ and which is depicted as $VG_\epsilon((m_{NP}, m_2))$ in the upper panel of Figure 3. Importantly, if $\epsilon$ is small enough then agent 1 considers only his final payoffs in the final outcomes $\{m_1, m_2\}$ and $\{m_{NP}, m_2\}$ (provided they are not equal) to make his optimal decision in the above veto stage. For any revenue $r > v + \alpha_2$ and any mechanism, we show that there exists a sufficiently small $\epsilon > 0$ such that in any equilibrium of Extensive version 2, the final expected revenue is strictly below $r$ on a positive measure with respect to nature’s move. Let $\nu = \frac{r - [v + \alpha_2]}{2} > 0$ and pick $\epsilon \in (0, 1/2)$ such that $\nu \cdot (1 - \epsilon) + \epsilon \cdot \alpha_2 > 0$. Let us consider the four kinds of final outcomes.

- If the final outcome is $\{m_{NP}, m_{NP}\}$, then the final revenue is $v$ which is smaller than $v + \alpha_2$.

- If the final outcome is $\{m_1, m_{NP}\}$, then two cases can occur: either the final revenue is smaller than $v + \alpha_2$ or it is greater which implies that agent 1’s surplus is strictly negative (because he either obtains the object for the price $v + \alpha_2 > V$ or does not obtain it while paying $\alpha_2 > 0$ to the seller). In this latter case, agent 1 would have chosen the ACCEPT action if he would have had previously the opportunity to reconsider his participation decision while the current reports were $\{m_1, m_{NP}\}$.

- If the final outcome is $\{m_{NP}, m_2\}$, the point is similar: either the final revenue is smaller than $v + \alpha_2$ or agent 2 would have chosen the ACCEPT action if he would have had previously the opportunity to reconsider his participation
Table 2: Mechanisms implementing the ‘Divide & Conquer’ surplus profiles

<table>
<thead>
<tr>
<th>1 \ 2</th>
<th>profile A</th>
<th>profile B</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP</td>
<td>(0,0)</td>
<td>(0,0)</td>
</tr>
<tr>
<td>P</td>
<td>(-α₁,0)</td>
<td>(-α₁,0)</td>
</tr>
<tr>
<td></td>
<td>(0,-α₂)</td>
<td>(0,-α₂)</td>
</tr>
</tbody>
</table>

decision while the current reports were \(\{m_{NP}, m^2\}\).

- Finally, if the final outcome is \(\{m^1, m^2\}\), then two cases can occur: either the final revenue is strictly smaller than \(r\) or the final surplus profile fails to be strong-Nash implementable. In this latter case, we have more precisely: either one of the individuality constraints fails from at least \(\nu\) which guarantees that the corresponding agent would have chosen the ACCEPT action if he would have had previously the opportunity to reconsider his participation decision while the current reports were \(\{m^1, m^2\}\) (with probability \((1-\epsilon)\), the veto game stops after the ACCEPT action such that he raises his payoff by at least \(\nu\), with probability \(\epsilon\), the veto game continues and his payoff can not subsequently shrink from more than \(\alpha_2\)); or the coalitional constraint fails to hold which guarantees that both agents would have chosen the ACCEPT action if they both would have had previously the opportunity to reconsider their participation decisions while the current reports were \(\{m^1, m^2\}\).

On the whole we obtain for any final outcome: either it raises a final revenue which is smaller than \(v + \alpha_2\) or it would not have been reached in equilibrium if a well-chosen set of agents would have been selected in the previous nature’s move.

Since \(\epsilon > 0\) and since nature’s moves are drawn independently at each node of the tree, the measure on nature’s moves such that the well-chosen set of agents is selected at each node is strictly positive. We conclude that in this PG the seller’s expected revenue can not be larger than \(v + \alpha_2\) on a positive measure on nature’s moves.

Additionally to our central insight that the set of implementable revenues depends critically on to the way PGs are modeled in presence of negative allocative externalities, 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, optimal mechanisms are efficient. Second, we can re-
strict ourselves to simple mechanisms with full participation w.l.o.g. when seeking for an optimal design as it will be proved by a generalization of extensive version 2 for any number of agents. Third, optimal simple mechanisms have to be asymmetric even in symmetric environments as in the limiting case $\alpha_2 = \alpha_1$.

Simple ‘Veto Games’:

\[ V_G(m^1, m_{NP}) \]

\[ V_G(m_{NP}, m^2) \]

\[ 1 - \epsilon \]

\[ \epsilon \]

\[ (VETO) \]

\[ (ACCEPT) \]

\[ (m^1, m_{NP}) \]

\[ (m_{NP}, m_{NP}) \]

\[ (m^1, m_{NP}) \]

\[ (m_{NP}, m_{NP}) \]

Figure 3: Extensive version 2
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 [16] states that multilateral “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’: 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 Optimal extensively robust implementable mechanisms are leading to efficient outcomes and the optimal revenue $R^*_{\text{Partial}}$ is given by:

$$R^*_{\text{Partial}} = \max_{(\alpha, \sigma) \in A \times \Sigma(N)} \left\{ \sum_{i=0}^{n} V^\alpha_i - \sum_{i=1}^{n} V^*_i(\{\sigma(1), \ldots, \sigma^{-1}(i) - 1\}) \right\}. \quad (4)$$

The optimal revenue can be raised as the full participation outcome of an extensively robust simply implementable mechanism. Such mechanisms $(\mathbf{a}, \mathbf{t})$ satisfy:

(i) $\mathbf{a}(N)$ corresponds to an efficient allocation,

(ii) there exists $\sigma \in \Sigma(N)$ such that $t_i(N) = V^\mathbf{a}(N)_i - V^*_i(\{\sigma(1), \ldots, \sigma^{-1}(i) - 1\})$ for any $i \in N$,

(iii) $\sigma$ maximizes the principal’s revenue:

$$\sigma \in \text{Arg} \max_{\sigma \in \Sigma(N)} \left\{ \sum_{i=0}^{n} V^\mathbf{a}(N)_i - \sum_{i=1}^{n} V^*_i(\{\sigma(1), \ldots, \sigma^{-1}(i) - 1\}) \right\}.$$
Conversely, from any simple mechanism such that \( a(N) \) and \( t(N) \) are satisfying the properties (i) and (ii), an extensively robust simply implementable optimal mechanism can be built by a proper specification of the ‘out of equilibrium’ final outcomes and transfers \((a(S), t(S))\) for \( S \subset N \).

The theorem does not provide explicitly the out of equilibrium outcomes and transfers of an optimal mechanism. The second step of the proof constructs explicitly an extensively robust simply implementable optimal mechanism. More generally, for any \((\alpha, \sigma) \in \mathcal{A} \times \Sigma(N)\), we are constructing extensively robust simply implementable mechanism \((a, t)\) such that \( a(N) = \alpha \) and \( t_i(N) = V_i^{\mathcal{A}(N)} - V_i^*(\{\sigma(1), \ldots, \sigma(\sigma^{-1}(i) - 1)\}) \) which implements thus the revenue \( \sum_{i=0}^{n} V_i^\alpha - \sum_{i=1}^{n} V_i^*(\{\sigma(1), \ldots, \sigma(\sigma^{-1}(i) - 1)\}) \).

In this limited class of mechanisms we build, the choices of the principal are reduced to the choice of the final outcome \( \alpha \) and 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 impact on the implemented revenue of those two instruments that characterizes the optimal design program is additive, in particular the optimal choice of \( \alpha \) coincides with the maximization of the allocative efficiency.

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 \( V_{\sigma(i)}^*(\{\sigma(1), \ldots, \sigma(i-1)\}) \) for the agent \( \sigma(i) \). Nevertheless, in a negative-externality-free framework, the optimal threat \( V_i^*(N \setminus \{i\}) \) against agent \( i \) requires an economic outcome \( a \) that is always feasible independently to the set of participants, i.e. \( a \in \mathcal{A}(\emptyset) \), and is thus always equal to \( V_i^*(\{\sigma(1), \ldots, \sigma(\sigma^{-1}(i) - 1)\}) \). We obtain the following corollary:

**Corollary 4.1** In a negative-externality-free framework, the revenue raised under an extensively robust optimal mechanism is equal to the full extraction revenue \( R_{\text{Full}}^* \).

We can wonder about how to measure and possibly bound the gap between the optimal revenue \( R_{\text{Partial}}^* \) under our robustness criterium and the usual optimal revenue \( R_{\text{Full}}^* \). To this extend, a natural reference for the revenue is the revenue with respect to the (always credible) threats \( V_i^*(\emptyset) \) denoted by \( R_{\text{Ref}}^* = \max_{\alpha \in \mathcal{A}} \{\sum_{i=0}^{n} V_i^\alpha\} - \sum_{i=1}^{n} V_i^*(\emptyset) \leq R_{\text{Partial}}^* \). The difference \( R_{\text{Full}}^* - R_{\text{Ref}}^* \) can be viewed as the surplus extracted by an
optimal design of the threats. Then let \( \eta = \frac{R^*_{\text{Partial}} - R^*_{\text{Ref}}}{R^*_{\text{Full}} - R^*_{\text{Ref}}} \in [0, 1] \) denote the share of the optimal surplus that can be extracted in a robust way.\(^{16}\) Equivalently, \( 1 - \eta \) can be viewed as the relative loss that comes from the coalitional rents. From (3) and (4), we have

\[
\eta = \frac{\max_{\sigma \in \Sigma(N)} \sum_{i=1}^{n} \left( V^*_i(\emptyset) - V^*_i(T^\sigma_{\sigma^{-1}(i)}) \right)}{\sum_{i=1}^{n} (V^*_i(\emptyset) - V^*_i(N \setminus \{i\}))}. \tag{5}
\]

In general, it is straightforward to check from (5) that \( \eta \geq 1/n \) but that any tighter bound would fail.\(^{17}\) A tighter bound can be established with a mild additional structure. In next proposition, we assume that the functions \( S \to V^*_i(S) \) are submodular, i.e. \( V^*_i(S \cup \{j\}) - V^*_i(S) \leq V^*_i(S' \cup \{j\}) - V^*_i(S') \) if \( S' \subset S \subset N \setminus \{i\} \) and \( j \notin S \). It means that agents are substitutes in the perspective of the possible threats they are allowing from their participation decisions.

**Proposition 4.2** If the function \( S \to V^*_i(S) \) is submodular for any \( i \in N \), then \( \eta \geq 0.5 \).

**Proof** From a standard convexity inequality that is derived from the submodularity assumption, we obtain for any permutation \( \sigma \in \Sigma(N) \) that \( V^*_i(\emptyset) - V^*_i(T^\sigma_{\sigma^{-1}(i)}) \geq \frac{T^\sigma_{\sigma^{-1}(i)}(V^*_i(\emptyset) - V^*_i(N \setminus \{i\})).\(^{18}\) Summing those inequalities from \( i = 1 \) to \( n \) and taking the maximum on both side, we obtain:

\[
R^*_{\text{Partial}} - R^*_{\text{Ref}} = \max_{\sigma \in \Sigma(N)} \sum_{i=1}^{n} (V^*_i(\emptyset) - V^*_i(T^\sigma_{\sigma^{-1}(i)})) \geq \max_{\sigma \in \Sigma(N)} \sum_{i=1}^{n} \left( \frac{T^\sigma_{\sigma^{-1}(i)}(V^*_i(\emptyset) - V^*_i(N \setminus \{i\}))}{N-1} \right) \tag{6}
\]

Let \( \hat{\sigma} \in \Sigma(N) \) be a permutation such that \( a_i := V^*_{\hat{\sigma}(i)}(\emptyset) - V^*_{\hat{\sigma}(i)}(N \setminus \{\hat{\sigma}(i)\}) \) is nondecreasing in \( i \). The right term in (6) equals \( \sum_{i=1}^{n} \frac{a_i}{n-1} \cdot a_i \), which is larger than \( \frac{1}{2} \sum_{i=1}^{n} \frac{a_i}{n} = \frac{1}{2} (R^*_{\text{Full}} - R^*_{\text{Ref}}) \) since \( a_i \) is nondecreasing in \( i \). QED

The rest of this section is devoted to the Proof of Theorem 1 in three steps.

**Preliminary Step: The strong Nash optimal design program**

\(^{16}\)If \( R^*_{\text{Full}} = R^*_{\text{Ref}} \), then let \( \eta = 1 \). In such a case (which would occur in a negative-externality-free framework), we have \( R^*_{\text{Full}} = R^*_{\text{Partial}} \).

\(^{17}\)Examples where \( \eta = 1/n \) can be build in the following way. Consider symmetric agents such that \( V^*_i(S) = V^*_j(S(j \cap i)) \) for any \( i, j \) and \( S \subset N \setminus \{i\} \) and where \( S(j \cap i):= S \) if \( j \notin S \) and \( S(j \cap i):= S \cup \{i\} \setminus \{j\} \) otherwise. Consider then that \( V^*_i(S) = V^*_i(\emptyset) \) for any \( S \subset N \setminus \{i\} \) and \( V^*_i(N \setminus \{i\}) < V^*_i(\emptyset) \). Note that the submodularity assumption considered subsequently fails if \( n > 2 \) in such examples.

\(^{18}\)Note that \( S \to V^*_i(S) \) submodular implies that the function \( H : [1, n] \cap \mathbb{Z} \to \mathbb{R} \) defined by \( H(j) = V^*_i(T^\sigma_j) \) for \( j < \sigma^{-1}(i) \) and \( H(j) = V^*_i(T^\sigma_{j+1} \setminus \{i\}) \) for \( j \geq \sigma^{-1}(i) \) is concave.
Proposition 4.3  Any strong Nash optimal mechanism \((a, t)\) is such that:

- \(a(N)\) corresponds to an efficient allocation
- there exists \(\sigma \in \Sigma(N)\) such that \(t_i(N) = V^{a(N)}_i - V^*_i(T^\sigma_{\sigma^{-1}(i)})\) for any \(i \in N\).

The optimal revenue is given by:

\[
R^*_\text{strong} = \max_{(\alpha, \sigma) \in A \times \Sigma(N)} \left\{ \sum_{i=0}^{n} V^{\alpha}_i - \sum_{i=1}^{n} V^*_i(T^\sigma_{\sigma^{-1}(i)}) \right\}.
\]

The strong Nash optimal design program is:

\[
R^*_\text{strong} = \max_{(a, t)} V^{a(N)}_0 + \sum_{i=1}^{n} t_i(N)
\]
subject to \(\max_{S \subset N} \{ V^{a(N)}_i - t_i(N) - V^{a(N)\setminus S}_i \} \geq 0, \forall S \subset N\), where \((a, t)\) 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 \(S \subset N\) and a permutation \(\sigma \in \Sigma(N)\), denote by \(j(S, \sigma)\) the smallest agent according to the order \(\sigma\) that is not belonging to \(S\). Formally, \(j(S, \sigma) = \sigma(\max\{ j \in N | T^\sigma_j \subset S \})\). In particular, the definition of \(j(S, \sigma)\) guarantees that \(T^\sigma_{\sigma^{-1}(j(S, \sigma))} \subset S\): agent \(j(S, \sigma)\) can be effectively punished by all the agents than are smaller than him according to the order \(\sigma\) if the set of participants is \(S\). 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 7  For \((\alpha, \sigma) \in A \times \Sigma(N)\), we define the \((\alpha, \sigma)\)-optimal threat mechanism as the mechanism \((a, t)\) defined in the following way:

- \(a(N) = \alpha\)
- \(a(S) = a^*_j(S, \sigma)(S), \text{ if } S \subsetneq N\)
- \(t_i(N) = V^{\alpha}_i - V^*_i(T^\sigma_{\sigma^{-1}(i)})\), for any \(i \in N\)
- \(t_i(S) = 0\), if \(S \subsetneq N\), for any \(i \in N\).

Recall that \(T^\sigma_{\sigma^{-1}(i)}\) corresponds to the set of agents that are smaller than agent \(i\) according to the order \(\sigma\). 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 decisions of the other agents by using the optimal threat among \( A(\emptyset) \);
then take another agent, \( \sigma(2) \), and give him the incentive to participate taken as given that \( \sigma(1) \) surely participates and independently of the participation decisions of the other agents in \( N \setminus \{ \sigma(1) \} \) by using the optimal threat among \( A(\{ \sigma(1) \}) \);
and so on. In particular, for the last agent, \( \sigma(N) \), in this new order \( \sigma \), the principal uses the optimal threat in \( A(N \setminus \{ \sigma(N) \}) \) 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 \((\alpha, \sigma)\)-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 \( S \subseteq N \) and the agent \( j(S,\sigma) \) who does not belong to \( S \). We have:

\[
V^{a(S)}_{j(S,\sigma)}(N) - t_{j(S,\sigma)}(N) - V^{a(\emptyset)}_{j(S,\sigma)} = V^{a(S)}_{j(S,\sigma)}(T_{\sigma^{-1}(j(S,\sigma))}) - V^{a(S)}_{j(S,\sigma)}(S) \geq 0.
\]

The equality comes from the definition of \( t_{j(S,\sigma)}(N) \) and because \( a(S) = a^{*}_{j(S,\sigma)}(S) \). The inequality is satisfied because \( T_{\sigma^{-1}(j(S,\sigma))} \subset S \). Thus we have proved that the strong Nash inequalities hold. **QED**

Then we show in Proposition 4.4 that, for any strong Nash implementable mechanism \((a, t)\), there exists an implementable mechanism that belongs to the class of \((\alpha, \sigma)\)-optimal threat mechanisms and that raises at least the same revenue for the principal. As a corollary, there is no loss of generality to look at the rent extraction profile for an \((\alpha, \sigma)\)-optimal threat mechanism when we are characterizing the rent extraction profile of optimal mechanisms.

**Proposition 4.4** For any strong Nash implementable mechanism \((a, t)\), there exists a strong Nash implementable mechanism that belongs to the class of \((\alpha, \sigma)\)-optimal threat mechanisms and that raises at least the same revenue for the principal.

**Proof** For a given mechanism \((a, t)\), we define a corresponding \((\alpha, \sigma)\)-optimal threat mechanism in the following way: \( \alpha = a(N) \), \( \sigma \) is defined by induction such that

- \( \sigma(1) \in Arg \max_{j \in N} \{ V^{a(N)}_{j} - t_{j}(N) - V^{a(\emptyset)}_{j} \} \) (initial step)
• $\sigma(i) \in \text{Arg max}_{j \in N \backslash \{\sigma(1), \ldots, \sigma(i-1)\}} \{V_{j}^{a(N)} - t_{j}(N) - V_{j}^{a(\sigma(1), \ldots, \sigma(i-1))}\}$ (inductive step).

The map $\sigma$ is by definition a permutation. From lemma 4.1, the $(\alpha, \sigma)$-optimal threat mechanism is implementable. We show then that the new mechanism extracts more surplus from each agent. Let $t_{i}^{(\alpha, \sigma)}(N)$ be the transfer for agent $i$ in the $(\alpha, \sigma)$-optimal threat mechanism with full participation. We have:

$$t_{i}^{(\alpha, \sigma)}(N) = V_{i}^{a(N)} - V_{i}^{a(T_{\sigma-1}(i))} \geq V_{i}^{a(N)} - V_{i}^{a(T_{\sigma-1}(i))} \geq t_{i}(N). \quad (7)$$

The first equality results from the definition of $t_{i}^{(\alpha, \sigma)}(N)$ and that $\alpha = a(N)$. The first inequality comes from the definition of the map $V_{i}^{*}(.)$ and since $a(T_{\sigma-1}(i)) \in A(T_{\sigma-1}(i))$. $(a, t)$ being strong Nash implementable implies that the strong Nash inequalities (2) are satisfied and more specifically for the set of deviator $N \backslash T_{\sigma-1}(i)$, the set of agents that are strictly smaller than $i$ according to the order $\sigma$, i.e.

$$\max_{j \in N \backslash \{\sigma(1), \ldots, \sigma(\sigma^{-1}(i))-1\}} \{V_{j}^{a(N)} - t_{j}(N) - V_{j}^{a(\sigma(1), \ldots, \sigma(\sigma^{-1}(i))-1))} \geq 0.$$

Since $i \notin T_{\sigma-1}(i)$, $i$ belongs himself to the set of deviators $N \backslash T_{\sigma-1}(i)$. Moreover, among this set, agent $i$ is the one that has the smallest utility shift from the joint deviation as it is guaranteed by the way the order $\sigma$ is built. Consequently this shift is negative for agent $i$, i.e. $V_{\sigma(i)}^{a(N)} - V_{\sigma(i)}^{a(T_{\sigma-1}(i))} \geq t_{\sigma(i)}(N)$. Finally, we have proved the last inequality in equation (7). To sum up, we have proved that $\alpha = a(N)$ and $t_{i}^{(\alpha, \sigma)}(N) \geq t_{i}(N)$ for all agents. The revenue of the principal is thus higher in the $(\alpha, \sigma)$-optimal threat mechanism we have constructed than in $(a, t)$. QED

**Remark 4.1** The proof of Proposition 4.4 establishes a slightly stronger result: for any strong Nash implementable mechanism $(a, t)$, there exists a strong Nash implementable mechanism that belongs to the class of $(\alpha, \sigma)$-optimal threat mechanisms and that extracts more surplus from the agents.

This paper is mainly focused on the implementation of some revenue. If we adopt the broader perspective to implement some surplus profiles, then we obtain that $(\alpha, \sigma)$-optimal threat mechanism are corresponding to the minimal surplus profiles for the agents that are strong Nash implementable, i.e. such that there is no surplus profile leaving less surplus to all agents and strictly less surplus for at least
one agent. Furthermore, it is easily shown that the set of strong Nash implementable surplus profiles corresponds exactly to the surplus profiles that are bigger than those latter minimal profiles. See Figure 2 for an illustration and section 6 for formal results.

First Step: An upper bound on the optimal revenue

Let \( K = \max_{i \in N} \max_{a, a' \in A} [V^a_i - V^{a'}_i] \) 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 \( \Psi(S) \) denote the set of the subsets of \( S \). Consider a general feasible mechanism denoted by \((a, t)\) that specifies a final outcome \( a(m) \) and a vector of monetary transfers \( t(m) \) for each possible set of reports \( m \in M^n \). Suppose that there exists a final outcome with a final set of messages \( m \) such that the principal raises a strictly higher revenue than \( R_{\text{Partial}} + n \cdot \nu \) for the set of reports \( m \) and for a given \( \nu > 0 \), the following preliminary lemma establishes that there is a subset of the participants such that all of its members would benefit at least \( \nu \) from a joint deviation while leaving unchanged the reports of the other agents.

Lemma 4.2 Consider a final set of messages \( m \) such that the revenue of the principal \( R = V^a_m(m) + \sum_{i=1}^n t_i(m) \) is strictly bigger than \( R_{\text{Partial}} + n \cdot \nu \). Then there exists a subset \( S \subset \Theta(m) \) such that \( V^a_{m(m^S \setminus e, m^S_N, m^S)} - (V^a_{m(m)} - t_i(m)) > \nu \), for any \( i \in S \).

In the following we call such a set \( S \subset \Theta(m) \) a set of \( \nu \)-deviators.

Proof Suppose on the contrary that for all \( S \subset \Theta(m) \):

\[
\max_{i \in S} V^a_{m(m)} - t_i(m) - V^a_{m(m^S \setminus e, m^S_N, m^S)} \geq -\nu. \tag{8}
\]

Then build a simple mechanism \((a^*, t^*)\) in the following way: for any set of participants \( S \subset N \),

- \( a^*(S) = a(m^{(N \setminus \Theta(m)) \cup S}, \Theta(m) \setminus S) \),
- \( t^*_i(S) = 0 \), for all \( i \notin (S \cap \Theta(m)) \),
- \( t^*_i(S) = t_i(m) - \nu \), for all \( i \in S \cap \Theta(m) \).

It is the simple mechanism derived from \((a, t)\) as if we could impose non-participation to all agents in \( N \setminus \Theta(m) \), leave only the choice between reporting the message \( m^i \)
and $m^i_{NP}$ to any agent $i \in \mathcal{S}(m)$ and reduce by $\nu$ the transfers of the participants to the principal. We check that the coalitional participation constraints $CP(S)$ are satisfied for any $S \subset N$:

$$\max_{i \in S} V_i a^*(N) - t_i^*(N) - V_i a^*(N \setminus S) \geq 0. \quad (9)$$

Consider first $S \subset N \setminus \mathcal{S}(m)$, then any subsequent deviation among the participants would not have any impact on the final outcome and the final transfers and inequality (9) is thus satisfied as an equality. Consider now $S$ such that $S \cap \mathcal{S}(m) \neq \emptyset$ and let $S' = S \cap \mathcal{S}(m)$. Applying inequality (8) to $S'$ and replacing the expression $a(m)$ and $t(m)$ as functions of $(a^*, t^*)$ implies that $\max_{i \in S'} V_i a^*(N) - t_i^*(N) - V_i a^*(N \setminus S') \geq 0$. Since $a^*(N \setminus S) = a^*(N \setminus S')$, we obtain that inequality (9) is satisfied. Finally we have proved that the mechanism $(a^*, t^*)$ is strong Nash implementable. However it raises a revenue $R - (\sharp \mathcal{S}(m)) \cdot \nu$ that is thus strictly bigger than $R^*_{Partial} = R^*_{strong}$ which raises a contradiction with Proposition 4.3. QED

We now establish that $R^*_{Partial}$ is an upper bound on the set of extensively robust implementable revenues. Suppose on the contrary that a general direct mechanism $(a, t)$ implements the revenue $R > R^*_{Partial}$.

![Figure 4: Tree of $VG_\epsilon(m)$](image)

We first construct by induction a finite extensive ‘veto game’ parameterized by $\epsilon \in (0, 1)$, for any current set of messages $m$, and denoted by $VG_\epsilon(m)$. The induction is on the number of current participants, i.e. the cardinal of the set $\mathcal{S}(m)$. See Figure 4. When this set is empty, the allocation $(a(m), t(m)) = (a(m_{NP}), t(m_{NP}))$ is implemented without any moves. Now consider that $\mathcal{S}(m)$ contains at least one element. The game starts with a nature’s move each corresponding to an element of $\mathcal{P}(\mathcal{S}(m))$. The probability to move to the node corresponding to $\emptyset$ is equal to $1 - \epsilon$.
while the other moves are equally probable with probability $\frac{\epsilon}{2\#S(m)}$. At the node $\emptyset$ the game ends and the status quo allocation $(a(m), t(m))$ is implemented. At a node $S = \{l_1, \ldots, l_k\} \in \mathcal{P}(S(m)) \setminus \{\emptyset\}$, 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.\footnote{The order in the sequence does not matter. Take for example $l_1 < \cdots < l_k$.} If there is at least one veto, the game ends at the status quo allocation. If all the agents in $S$ accept the proposal then the continuation game is the ‘veto game’ $V G_{\epsilon}(m^{N\setminus S}, m^{S}_{NP})$ which is properly defined by the induction hypothesis.

Let $\nu \in (0, \frac{R - R_{Partial}}{n})$ and $\epsilon \in (0, 1/2)$ such that $\nu \cdot (1 - \epsilon) - \epsilon K > 0$. Consider a PG $B$ and then the modified game such that at each final node with a final set of messages $m$ of the original game $B$, the final set of messages $m$ is publicly disclosed and the game $V G_{\epsilon}(m)$ is appended as a continuation game. The modified game is also a PG since the veto action is always available. Figure 3 depicts the example with two agents where the original game $B$ is the sequential PG that first ask agent 1 to make a report and then the remaining agent 2.

In the veto games $V G_{\epsilon}(m)$, consider a node where the current set of messages is $m$ such that $V^*_{0} + \sum_{i=1}^{n} t_i(m) \geq R_{Partial} + n \cdot \nu$ and the continuation game after nature has chosen the set of participants $S$ and such that $S$ is a set of $\nu$-deviators which exists from lemma 4.2. In this subgame, any subgame perfect equilibrium strategy profile is such that the agents in $S$ are accepting the joint deviation not to participate at each node where all the previous responders have 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 $1 - \epsilon$ he will surely win at least $\nu$ whereas with probability $\epsilon$ 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 $S$.

For a given set of final messages $m$, there are two possibilities: either it raises a final revenue which is smaller than $R_{Partial} + n \cdot \nu$ or there is a set of $\nu$-deviators, a set whose existence is guaranteed by lemma 4.2. With a strictly positive probability\footnote{This probability can be easily bounded by $(\frac{\epsilon^2}{2^n-1})^n > 0$}, nature’s moves are such that: the set $\emptyset$ is selected if the revenue under the current set of messages is lower than $R_{Partial} + n \cdot \nu$ while it selects a set of $\nu$-deviators otherwise.
Note that in the case where all agents definitely do not participate, then the revenue is necessarily lower than \( R_{\text{Partial}}^* \) and thus strictly lower than \( R \) since no transfer can help the principal. From the previous paragraph, a set of \( \nu \)-deviators would accept to jointly non-participate if they receive this opportunity in the participation game. On the whole, under such nature’s moves, then the game never ends at a final message \( m \) such that the principal’s revenue is strictly bigger than \( R_{\text{Partial}}^* + n \cdot \nu \). 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.\(^{21}\)

**Remark** The PGs we build may seem very complex. First note that it works for any general direct mechanism. For a specific mechanism, much simpler PGs can be built as Extensive version 1 which allows cooperation in prisoner’s dilemma. Second, if we are considering implementation with simple mechanisms and such that full participation is the only equilibrium not only in the set of PGs from definition 4 but also in PGs where some players are forced to participate, then the set of PGs we build to check that \( R_{\text{Partial}}^* \) is an upper bound on the optimal revenue is much simpler. In particular it does not involve any nature’s move but only the participation games where agents are asked sequentially whether they are willing to participate and in which any (irreversible) participation decision will offer another opportunity to participate to the remaining agents that are not already committed to participate in the mechanism (see the working paper version [26] for more details). In this restricted class of PGs we allow any order in the players’ moves which has thus a similar flavor as the robustness concern of Moldovanu and Winter [32] to the order of players’ moves in sequential bargaining games.

**Second Step: optimal extensively robust implementable mechanisms**

We first show that the previous upper bound is reached with some extensively robust simply implementable mechanisms and more generally we characterize the optimal mechanisms in this class. At this stage we use two results about extensively robust simply implementable mechanisms which are derived independently in section 21.

\(^{21}\)Remark that we do not really use in the proof the point that the message space is finite. E.g. the use of kinds of integer games à la Maskin [28] would not allow higher implementable revenues. The generalization to arbitrary message spaces would only require to adapt the current definition of PGs that assumes that the game is finite.
that is exclusively devoted to simple mechanisms. Proposition 5.1 establishes that extensively robust simply implementable mechanisms are necessary strong-Nash implementable. The necessary conditions on optimal extensively robust simply implementable mechanisms (in Theorem 1) are thus a corollary of the conditions derived in Proposition 4.3 on optimal strong-Nash implementable mechanisms. Consider then an optimal strong Nash mechanism characterized by an allocation \( \alpha^* \) and an order \( \sigma^* \). From Proposition 4.4 the same revenue can be raised with an \((\alpha^*, \sigma^*)\)-optimal threat mechanism. Define the mechanism \((a^*, t^*)\) as the \((\alpha^*, \sigma^*)\)-optimal threat mechanism except for the out of equilibrium transfers \( t_i(S) \) for \( S \supset T_{\sigma - 1(i)}^\sigma \) with \( i \in S \) which are set such that \( t_i(S) = V_i^{a(S)} - V_i^{a(S)}(T_{\sigma - 1(i)}^\sigma) \). In such a simple mechanism, we have \( \mathcal{U}_i(a(S), t_i(S)) = \mathcal{U}_i(a(N), t_i(N)) \) for any \( S \supset T_{\sigma - 1(i)}^\sigma \) which guarantees that the sufficient condition that guarantees that simple mechanisms are extensively robust simply implementable (which is derived in Proposition 5.2) is satisfied in a way such that the required inequalities stand as equalities. Finally the extensively robust implementable mechanism \((a^*, t^*)\) implements the same revenue as our original optimal strong Nash implementable mechanism.

Second, we end the proof of Theorem 1 by proving that any extensively robust implementable mechanism is efficient. The argument follows exactly the same logic as the first step of our whole proof. If the final outcome is inefficient while raising the revenue \( R_{\text{Partial}}^* \) then there exists a set of \( \nu \)-deviators for \( \nu \) sufficiently small (the analog of lemma 4.2 for inefficient outcomes). Such an outcome will never occur on a positive measure of nature’s move in PGs where the veto games \( VG_\epsilon(m) \) have been appended and with \( \epsilon \) sufficiently small.

5 Simple implementation

This section is devoted to simple mechanisms: we derive a sufficient condition for being extensively robust implementable and a necessary condition for being extensively robust simply implementable. Those results are much stronger than what was needed for the proof of Theorem 1 where the focus was limited to optimal mechanisms. Indeed, those conditions can be useful more generally for games when the modeler is not comfortable with the simultaneous-move assumption and wants to check that an equilibrium is ‘extensively robust’, i.e. remains a subgame-perfect
equilibrium outcome in any alteration of the simultaneous-move game. The following propositions with simple mechanisms can be exported immediately for binary games with externalities that occur often in economics (adoption of a technology, entry in a market) or in politics (ratification of a treaty, recognition of a country).

**Proposition 5.1 (A necessary condition)** Any extensively robust simply implementable mechanism is necessary STRONG-Nash implementable: full participation is a STRONG-Nash equilibrium of the simultaneous-move participation game, i.e.

$$\max_{i \in S} \{V^a_i(N) - t_i(N) - \sum_{S' \in \Delta(S)} (V^a_i(N \setminus S') - t_i(N \setminus S')) \times \mu^S(S')\} \geq 0$$

for any $S \subset N$ and any probability distribution $\mu^S(.)$ on $\Delta(S) = \{S'|S' \subset S\}$ such that $\mu^S(S')$ denotes the probability that the set of participants is $S'$. As a corollary, any extensively robust simply implementable mechanism is necessary strong-Nash implementable.

**Proof** Consider a mechanism $(a, t)$ and suppose that it is extensively robust simply implementable but not STRONG Nash implementable, i.e. there exists a set $S \neq \emptyset$ and a probability distribution $\mu^S(.) \in \Delta(S)$ such that $V^a_i(N) - t_i(N) < \sum_{S' \in \Delta(S)} (V^a_i(N \setminus S') - t_i(N \setminus S')) \times \mu^S(S')$ for any $i \in S$ and $\mu^S(\emptyset) < 1$. Consider a PG defined in the following way: in the first stage the agents are playing a PG $B_1$ (e.g. the simultaneous-move PG), in a second stage a sequential veto game between the agents in $S$ is appended to the first stage if the final outcome correspond to full participation of the game $B_1$. At each node of this sequential veto game, the agents in $S$ have to choose between agreeing to delegate their participation decision such that the mutually profitable proposal $\mu^S(.)$ is implemented and vetoing the proposal. If all the agents in $S$ agree to delegate, then the final set of non-participants is $S' \subset S$ with probability $\mu^S(S')$. On the contrary, if one agent vetoes the proposal, then the final outcome corresponds to full participation. Denote by $B_2$ the modified game. First note that $B_2$ is actually a PG: a $B_2$-adaptation of the participation strategy

---

---

22 Additionally to the subgame-perfection criterium, we emphasize an additional important difference with Kalai [23, 24] which makes his robustness approach much more demanding if it were combined with subgame-perfection. Kalai does not solely require the existence of an equilibrium leading to a desired outcome profile, e.g. full participation, but that every strategy profile where every agents follow an adaptation of the final outcome is an equilibrium profile.

23 See Schelling [38] for more examples in this area. Interestingly, he cares also about the stability of prisoner’s dilemma to coalitional deviations.
The STRONG-Nash property of the full participation outcome in a simple mechanism is not a sufficient condition for extensively robust implementation as it is illustrated by the ‘Battle of the sex’ game (left panel of Table 3) with the basic sequential participation game where agent 1 chooses first to participate or not which gives him the commitment power to select his preferred equilibrium. More surprising is the failure of the extensively robust implementation property for the game in the middle panel of Table 3: in the 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 in any PG. However, it does not imply that the ‘row agent 1’ always play a best response to his opponent super-dominant action in any PG. Figure 5 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

<table>
<thead>
<tr>
<th></th>
<th>1\2</th>
<th>NP</th>
<th>P</th>
<th></th>
<th>1\2</th>
<th>NP</th>
<th>P</th>
<th></th>
<th>1\2</th>
<th>NP</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>NP</td>
<td>(2,1)</td>
<td>(0,0)</td>
<td>P</td>
<td></td>
<td>NP</td>
<td>(5,0)</td>
<td>(3,13)</td>
<td>P</td>
<td></td>
<td>NP</td>
<td>(1,1)</td>
</tr>
<tr>
<td>P</td>
<td>(0,0)</td>
<td>(1,2)</td>
<td>P</td>
<td></td>
<td>P</td>
<td>(0,0)</td>
<td>(10,1)</td>
<td>P</td>
<td></td>
<td>P</td>
<td>(0,0)</td>
</tr>
</tbody>
</table>

‘The Battle of the sex’  ‘STRONG-Nash’  ‘Pure Coordination’

Table 3: Payoff Tables, NP/P for NonParticipation/Participation
has an additional action (‘UNMATCH’) which corresponds to commit to choose the opposite action of the one chosen by agent 1. If the final payoffs are given according to the middle panel of Table 3, then, there is a unique 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. Such a threat is available when we consider the general class of PG à la Kalai.\(^{24}\)

![Figure 5: Extensive version 3](image)

In a given mechanism, checking that full participation is a subgame perfect equilibrium for all extensive versions à la Kalai seems to be a very difficult task in general. Next proposition makes nevertheless an important step by showing 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. Pure coordination games as the one in the right panel of Table 3 satisfy this sufficient condition.

**Proposition 5.2 (A sufficient condition)** For a given simple mechanism \((a, t)\), suppose that there exists a permutation \(\sigma \in \Sigma(N)\) such that \(U_i(a(N), t_i(N)) \geq U_i(a(S), t_i(S))\) for all set \(S \supset T_{\sigma^{-1}(i)}\) and for any agent \(i \in N\), then the mechanism \((a, t)\) is extensively robust simply implementable.

**Proof** Consider a finite extensive game \(B\). We build a strategy profile \(\gamma = (\gamma_1, \cdots, \gamma_n)\) for the agents in \(N\) in the following way. For all \(i \in N\) denote by \(H_i^\sigma\)

\(^{24}\)In 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 their 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.
the set of histories from agent $i$’s point of view occurring with positive probability when the agents in $T_\sigma_i$ are all assumed to play a $B$-adaptation of the participation strategy. Consider the following game $B'$ among the agents in $N$ and an external player that is indifferent to the final outcome: the game corresponds exactly to $B$ except that at the nodes where a given player $i$ makes a choice and that belong to $H_\sigma^i$ then it is the external player that makes player $i$’s choice and this applies for any $i \in N$. Note that this ‘artificial’ game $B'$ is not a participation game but is a finite extensive game with perfect recall. 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 $B$-adaptation of the participation strategy of the original players. In other words, it is the game as if all agents $\sigma(i)$, $i \in N$, were ‘forced’ by an external player to play an adaptation of the participation strategy on the histories $H_\sigma^i$. From Selten [41] such a game has thus at least one perfect equilibrium and a fortiori one subgame-perfect equilibrium. Consider such an equilibrium profile $\gamma'$ of $B'$.

This profile specifies strategies for all agent $i$ for histories that do not belong to $H_\sigma^i$. Finally define $\gamma_i$ as corresponding to a $B$-adaptation of the participation strategy on any history $h \in H_\sigma^i$ and to $\gamma_i'$ otherwise. We claim that the strategy profile $\gamma$ is a subgame-perfect equilibrium. Consider agent $i$ at a node with the history $h \notin H_\sigma^i$ (out of the equilibrium path), then agent $i$ in the game $B$ faces exactly the same decision problem he would have faced in the game $B'$ at the same node and with the same history on previous moves. Since $\gamma'$ is supposed to be an equilibrium, then it is a best response to some out of equilibrium beliefs and $\gamma$ is thus also a best response. Consider agent $i$ at a node with the history $h \notin H_\sigma^i$, consider a deviation by agent $i \in N$: the best outcome he can reach is $\max_{S \supseteq T_\sigma_i} \mathcal{U}_{\sigma(i)}(a(S), t_i(S))$ (given the equilibrium belief that all the agents in $T_\sigma_i$ will finally participate with probability 1 which holds on the equilibrium path) which is smaller than his equilibrium outcome $\mathcal{U}_{\sigma(i)}(a(N), t_i(N))$. QED

\footnote{In general it may fail to be a perfect equilibrium. Nevertheless, if the inequalities $\mathcal{U}_i(a(N), t_i(N)) \geq \mathcal{U}_i(a(S), t_i(S))$ are strict then the strategy profile $\gamma$ is a perfect equilibrium if $\gamma'$ has been selected such that it is a perfect equilibrium.}
6 Links between strong Nash and extensively robust implementation

As a by-product of our analysis, we obtain a link between two distinct robustness implementation criteria: the one we introduce which is non-cooperatively founded and the strong Nash implementation criterium that is based on some ad hoc cooperative flavored constraints. In particular, it gives a straightforward way to check that a revenue is extensively robust implementable or a given surplus profile is extensively robust simply implementable since the strong Nash constraints involve a (finite) set of tractable constraints. For a given surplus profile $e = (e_1, \ldots, e_n)$, denote by $L_e = \{(e'_1, \ldots, e'_n)|e'_i \geq e_i, \forall i \in N\}$ the set of surplus profiles that are higher than $e$. Let $e^\sigma$ denote the higher surplus profile according to the order of the threat $\sigma \in \Sigma(N)$, i.e. $e^\sigma_i = V^*_i(T^\sigma_{\sigma^{-1}(i)})$.

**Proposition 6.1**  (i) The set of extensively robust implementable revenues and the set of strong Nash implementable revenues coincide.

(ii) The set of extensively robust simply implementable surplus profiles and the set of strong Nash implementable surplus profiles coincide and are both equal to $\bigcup_{\sigma \in \Sigma(N)} L_{e^\sigma}$.

The following proof gathers mainly elements from section 4 and 5.

**Proof** It is straightforward to see that both the set of extensively robust implementable revenues and the set of strong Nash implementable revenues are interval of the form $(-\infty, R]$ since any lower revenue can be obtained by an additional uniform transfer to some bidder conditional on his participation. (i) is then a corollary of Theorem 1 and Proposition 4.3.

Basic observation: if a given surplus profile is strong Nash implementable [extensively robust simply implementable] then any higher surplus profile is strong Nash implementable [extensively robust simply implementable] by means of appropriate uniform transfers to the agents.

In particular, it is then sufficient to check that the profiles $e^\sigma$ are strong Nash implementable for any $\sigma \in \Sigma(N)$ (which comes from lemma 4.1) to guarantee than $\bigcup_{\sigma \in \Sigma(N)} L_{e^\sigma}$ is a subset of the set of strong Nash implementable surplus profiles. For the other inclusion, a closer look at the proof of Proposition 4.4 (see remark 4.1)
shows that for any strong Nash implementable surplus profile there exists an order \( \sigma \in \Sigma(N) \) such that this profile is lower than \( e^\sigma \).

From Proposition 5.1, we obtain that the set of extensively robust simply implementable surplus profiles is a subset of the set of strong Nash implementable surplus profiles. For the other inclusion, it is sufficient (recall the above basic observation) to check that \( e^\sigma \) is extensively robust simply implementable for any \( \sigma \in \Sigma(N) \). For this last step, any \((\alpha, \sigma)\)–optimal mechanism with the modified out of equilibrium transfers as in the proof of the second step of Theorem 1 will work. QED

Note that Proposition 6.1 remains valid if the strong Nash concept is replaced by the more stringent ‘STRONG Nash’ implementation concept introduced in Proposition 5.1: the unique delicate point is to check that the profiles \( e^\sigma \) are STRONG Nash implementable for any \( \sigma \in \Sigma(N) \), which can be tackled with a slight modification of \((\alpha, \sigma)\)–optimal mechanisms where the out of equilibrium transfers \( t_i(S) \) are set low enough, more precisely \( t_i(S) \leq \max_{\alpha \in A} V_i^\alpha \), such that those mechanisms can be shown to be STRONG Nash implementable.\(^{26}\)

7 Full implementation

We have considered until now ‘partial implementation’ concepts, e.g. full participation being an equilibrium. In this section we consider ‘full implementation’ concepts, i.e. full participation being the unique equilibrium outcome. For definition 5, the natural concept for extensively robust full implementation is the following: a revenue \( R \) is ‘extensively robust fully implementable’ if there is a feasible mechanism such that for every PGs any subgame-perfect equilibrium guarantees the expected revenue \( R \) for almost all nature’s move. Under extensively robust implementation criteria, partial and full implementation are not equivalent as can be checked with the ‘Pure Coordination’ game in Table 3 and as it was known to be the case for standard (simultaneous-move) implementation criteria. However, we show that revenue and surplus profiles that are extensively robust implementable are also approximately (“virtually” in the subsequent terminology) extensively robust fully implementable. As a preliminary and in a similar way as in proposition 5.2, we derive a sufficient condition for simple mechanisms that guarantees full implementation. Note that this

\(^{26}\)Such a result is analog to lemma 4.1. The proof is left to the reader.
condition is stronger than the one appearing in Proposition 5.2.

**Proposition 7.1 (A sufficient condition for full implementation)** For a given simple mechanism \((a, t)\), suppose that there exists a permutation \(\sigma \in \Sigma(N)\) and \((pp_i, np_i)_{i \in N} \in (\mathbb{R}^2)^N\) such that, for all set \(S \supset T_{\sigma(i)}\) and for any agent \(i \in N\),

\[
\mathcal{U}_i(a(S), t_i(S)) = pp_i \text{ if } i \in S \quad \text{and} \quad \mathcal{U}_i(a(S), t_i(S)) = np_i \text{ if } i \notin S \text{ and } pp_i > np_i,
\]

then full participation is the unique Nash equilibrium outcome. As a corollary, we obtain that the mechanism \((a, t)\) is extensively robust fully simply implementable.

**Proof** We show by induction on \(i = 1, \ldots, N\), that all the agents in \(T^\sigma_{i+1}\) should participate in a Nash equilibrium outcome. For the initialization, i.e. for \(T^\sigma_1 = \{\sigma(1)\}\), the result comes from the assumption: independently of the participation decision of the remaining agents, agent \(\sigma(1)\) obtains the payoffs \(pp_{\sigma(1)}\) (respectively \(np_{\sigma(1)}\)) if he participates (does not participate) to the mechanism; if agent \(\sigma(1)\)’s expected payoff is strictly less than \(pp_{\sigma(1)}\) then he would strictly benefit to deviate and play a \(B\)-adaptation of the participation strategy (since \(pp_{\sigma(1)} > np_{\sigma(1)}\)); we obtain thus that agent \(\sigma(1)\)’s expected payoff equals \(pp_{\sigma(1)}\) or equivalently that he surely participates in equilibrium. We now move to the induction step. Suppose that all the agents in \(T^\sigma_{i+1}\) should participate in a Nash equilibrium outcome (\(i < N\)). We show then that agent \(\sigma(i+1)\) surely participates in equilibrium. Suppose that he may not participate with some positive probability in equilibrium. From the induction hypothesis, all the agents in \(T^\sigma_{i+1}\) should participate with probability 1 such that agent \(\sigma(i+1)\) obtains the payoffs \(pp_{\sigma(i+1)}\) (respectively \(np_{\sigma(i+1)}\)) if he participates (does not participate) to the mechanism. Then he would strictly benefit to deviate and play a \(B\)-adaptation of the participation strategy (since \(pp_{\sigma(i+1)} > np_{\sigma(i+1)}\)). We obtain thus that agent \(\sigma(i+1)\)’s expected payoff equals \(pp_{\sigma(i+1)}\) or equivalently that he surely participates in equilibrium which would raise a contradiction. So we have completed the induction step.\(^{27}\) QED

We emphasize that the proof does not rely here on the subgame-perfection refinement we need for partial implementation. All our subsequent results hold indifferentely whether the full implementation criteria relies or not on subgame-perfection. E.g. the definition for extensively robust full implementation can be replaced by the following: a revenue \(R\) is ‘extensively robust (without subgame perfection) fully

\(^{27}\)Note that the argument would not work under the weaker assumption that \(\min_{S \supset T^\sigma_{\sigma(i)} \cup \{i\}} \mathcal{U}_i(a(S), t_i(S)) > \max_{S \supset T^\sigma_{\sigma(i)} \setminus \{i\}} |S| \mathcal{U}_i(a(S), t_i(S))\) for any agent \(i \in N\).
implementable’ if there is a feasible mechanism such that for every PGs any Nash equilibrium guarantees the expected revenue $R$ for almost all nature’s move.

**Definition 8 (virtual implementation)** ° 28 A revenue $R$ is virtually extensively robust fully implementable if any revenue $r < R$ is extensively robust fully implementable.

A surplus profile $e = (e_1, \ldots, e_n)$ is virtually extensively robust simply fully implementable if any surplus profile $e'$ such that $e'_i > e_i$ for any $i \in N$ is extensively robust fully simply implementable.

Any implementable revenue or surplus profile is obviously virtually implementable. Next theorem shows the converse, i.e. there is no fundamental difference in our framework between what can be implemented under a partial implementation criterium and a full implementation criterium.

**Theorem 2** (i) The set of virtually extensively robust fully implementable revenues and the set of extensively robust implementable revenues coincide.

(ii) The set of virtually extensively robust simply fully implementable surplus profiles and the set of extensively robust simply implementable surplus profiles coincide.

**Proof** In the same way as for the proof of proposition 6.1, it is sufficient to show that the surplus profiles corresponding to some $(\alpha, \sigma)$-optimal mechanism can be virtually extensively robust simply fully implemented. The key point is that the simple mechanism $(a^*, t^*)$ derived in the second step of the proof of Theorem 1 can be modified infinitesimally such that the sufficient condition in proposition 7.1 holds. We build $(a, t)$ such that $a$ and $a^*$ coincide while $t_i(S) := t_i^*(S)$ if $i \notin S$ and $t_i(S) := t_i^*(S) - \nu_i$ if $i \in S$. The condition in proposition 7.1 holds with $pp_i - np_i = \nu_i$ for any agent $i$. For any vector $(\nu_i)_{i \in N}$ such that $\nu_i > 0$ we obtain that $(a, t)$ is extensively robust fully simply implementable. QED

Our surplus extraction results are summarized in Table 4.

**Remark 7.1** Finally, we would like to argue that partial implementation concepts under partially-specified games may deserve as much attention as full implementation

---

28°The terminology is reminiscent of Abreu and Matsushima [1]. Our definition does not apply to social choice functions as theirs. On the contrary, it applies to a limited feature of the underlying social choice function: either the principal’s payoff or the agents’ surplus profile.
Table 4: Surplus Extraction according to the class of games for partial and full implementation concepts.

<table>
<thead>
<tr>
<th>Class of Participation Games</th>
<th>Surplus Extraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simultaneous-move</td>
<td>Full: Full Coalition-Proof, Rationalizable a Nash Equilibrium</td>
</tr>
<tr>
<td>Nash, Dominant Strategy</td>
<td>a Nash Equilibrium</td>
</tr>
<tr>
<td>Coalition-Proof, Rationalizable the unique correlated equilibrium</td>
<td></td>
</tr>
<tr>
<td>Partially-Specified</td>
<td>Partial: Partial STRONG Nash Equilibrium, the unique (subgame-perfect) Nash Equilibrium</td>
</tr>
<tr>
<td>Nash</td>
<td>a subgame-perfect Nash Equilibrium, the unique Nash Equilibrium</td>
</tr>
</tbody>
</table>

8 Robust implementation against general commitment devices

Going back to Schelling [37], a literature on strategic delegation has emerged (see also, e.g., Fershtman et al. [14]). Recently, this literature has seen a revival of interest with papers formalizing the idea of conditional commitments that lead players to select an action as a function of the commitment choices of his opponents. Various and closely related formalizations appeared under the terminology ‘Program equilibrium’ in Tennenholtz [44], ‘Commitment equilibrium’ in Kalai et al. [22] and
‘Mediated equilibrium’ in Monderer and Tennenholtz [33] and Peleg and Procaccia [36]. Tennenholtz [44] and Kalai et al. [22] obtain some ‘folk theorems’. Monderer and Tennenholtz [33] and Peleg and Procaccia [36] adopt an implementation perspective and investigate the possibility to implement an outcome with a strong Nash mediated equilibrium. We emphasize that our perspective is somehow radically different from those two later papers on implementation since they consider the use of mediated equilibria as a way to extend what can be implemented while we will use commitment devices to check whether what we are trying to implement can be done in a robust way.

The aim of this section is to present informally how our results in section 7 about full implementation extend to general commitment devices à la Kalai et al. [22]. Instead of considering all the extensive versions from a mechanism \((a, t)\), we now consider the set of voluntary commitment games, i.e. the set of all finite simultaneous games where a player chooses a strategy which maps an original strategy in the mechanism \((a, t)\) as a function of the strategies chosen by his opponents and where players have always the possibility to commit to the “neutral devices” which are strategies that guarantee him to play any given strategy in the original mechanism. Furthermore, in the same way as we consider moves from nature in the extensive version, we are also allowing external players to play in the voluntary commitment games. We refer the reader to section 2 in Kalai et al. [22] for a formal definition with two players, a definition which can be extended straightforwardly with more players by considering that commitment devices are fully visible by all players.\(^{29}\) The related partial and full implementation criteria correspond then to the requirement that the desired outcome is the outcome in respectively an equilibrium and all equilibria for any voluntary commitment game. Since the equilibria of the original game remain equilibria in the commitment games (with the use of the corresponding neutral devices), then partial implementation concepts have no bite here contrary to our extensively robust implementation concepts. Next proposition says that the analysis with a full implementation perspective as in section 7 extends to general commitment devices.

**Proposition 8.1** Proposition 7.1 and Theorem 2 extend to the corresponding im-

\(^{29}\)Alternative modeling choices can be made when dealing with more than two players, they seem irrelevant for the present purpose where the analog of proposition 7.1 would still hold.
plementation criteria against general commitment devices.

**Proof** For Theorem 2, we first show that we can not do better with general commitment devices: it comes from the point that the normal form of any PG can be viewed as a voluntary commitment game. Thus if we fail to fully implement a given revenue or a given surplus profile under extensively robust implementation, there exists a subgame-perfect equilibrium in a PG and a fortiori an equilibrium in a voluntary commitment device that fails to implement this revenue or this surplus profile. Second, the fact that we can implement as much with general commitment devices comes from the fact that proposition 7.1 extends straightforwardly such that the optimal design we have constructed to prove Theorem 2 still works when we care about robustness against commitment devices. QED

Robustness against general commitment devices can be interpreted as a robustness criterium against general collusive devices between the agents when monetary transfers are prohibited. An implementation criteria robust to such worst case scenarii are especially salient if we have in mind that a third party can design the commitment device. Main contributions on collusion-proof implementation as Laffont and Martimort [25] and Che and Kim [8] 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 [8] 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 since it excludes monetary transfers. On the other hand, it includes participation decisions. Our results contrast with them since we show that optimal designs that are robust to collusion may raise a strictly lower revenue than standard optimal designs.\(^{30}\)

\(^{30}\)This result also contrasts with the insights of Che and Kim [9] 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 Che and Kim [9] considers a negative-externality-free framework: the sale of a single item in the independent private value framework.
9 Clarifying remarks

9.1 The set of Participation Games

Although, we may think of a PG as resulting from the design of a third party coordinating the agents, the set of PGs in definition 4 may seem excessively large in a mechanism design framework. We give below two possible additional properties such that our analysis in section 4 would not be modified under those additional restrictions as the reader can check in our proofs.

Additional Properties:

5. $B$ 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 up to a permutation of their identities.

In certain environments we may feel uncomfortable since an extensive version as the one in Figure 5 relies on 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 was a motivation in the working paper version [26] to consider also a different class of games with ‘subsequent opportunities and perfect information’ where 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. In other words, agents commit only to participation or to delay their decisions. However, from an optimal design perspective, the analysis remains broadly unchanged: the optimal implementable revenue remains $R^{\text{Partial}}$.

9.2 Non-convexity of the set of optimal mechanisms

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 \((a, t)\).\(^{31}\) Thus the set of implementable mechanisms 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 \(m_\sigma\) 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, neither the set of strong Nash implementable mechanisms nor the set of extensively robust implementable mechanisms are convex in general, except in externality-free frameworks. As an example of such a non-convexity, we can come back to the simple example of section 3. In any (strict) mixture of the two mechanisms from Table 2, each agents would make a loss with regards to the allocation if they both do not participate and the coalitional constraint \(CP(\{1, 2\})\) breaks down. Indeed, in this example, simple mechanisms that are symmetric do not implement a better revenue than \(v\) the payoff under full non-participation.

9.3 Beyond quasi-linear utilities

Our analysis can be extended in a straightforward way -up to some cumbersome notation- if agents have general von Neumann-Morgenstern utility functions \(U_i(a, t_i)\), where \(U_i(a, t_i)\) is strictly decreasing in \(t_i\) and the image of \(U_i(a, .)\) is assumed to be \(\mathbb{R}\), for any \(a \in A\) and \(i \in N\). Let \(U_i^{-1}(a, u)\) denote the unique solution of the implicit equation \(U_i(a, x) = u\) according to \(x\). The harshest feasible threat that the principal can inflict on \(i\) given that the agents in \(S\) have accepted the mechanism is now given by \(a_i^*(S) \in \text{Arg} \min_{a \in A(S)} U_i(a, 0)\). The transfer that you can extract from agent \(i\) under the threat \(a_i^*(S)\) and if you implement the allocation \(a\) is now given by: \(U_i^{-1}[a, U_i(a_i^*(S), 0)]\). In particular, it can be shown that the optimal revenues under both the extensively robust implementable mechanism design program and the strong Nash implementable program are equal to:

\(^{31}\)The implicit space structure according to which linearity applies is the following. For two mechanisms, \((a, t)\) and \((a', t')\) and a real number \(\lambda \in [0, 1]\), the mechanism \(\lambda \cdot (a, t) + (1 - \lambda) \cdot (a', t')\) is the mechanism that implements the mechanism \((a, t)\) (respectively \((a', t')\)) with probability \(\lambda\) (resp. \((1 - \lambda))\).
$$R_{\text{Partial}}^* = R_{\text{strong}}^* = \max_{(\alpha, \sigma) \in A \times \Sigma(N)} \left\{ V_0^\alpha + \sum_{i=1}^n U_i^{-1}(\alpha, U_i[a_i^{\sigma}(T_{\sigma^{-1}(i)}), 0]) \right\}. $$

Note that this program is slightly harder to solve since it not separable anymore in the choice of the final allocation $\alpha$ and the choice of the order of the threats $\sigma$.

10 Additional Comments: applications & extensions

As illustrated by the example in section 3, an important class of applications where our ‘coalitional constraints’ are binding are auctions with negative externalities as in Jehiel et al. [19, 20, 21]. See also Aseff and Chade [2] and Figueroa and Skreta [15] for the characterization of the optimal design in such setups with externalities. This final section briefly discusses other general applications where our implementation criteria may provide new benchmarks.

10.1 The use of veto-power in mechanism design

The way the mechanism design toolbox is used does not always stand in line with the standard framework. In particular, it is not rare to see the modeler imposing veto-power constraints that have no clear non-cooperative foundation. E.g., in the recent mechanism design literature on collusion as in Che and Kim [8], one agent (or a third party) proposes a collusion 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 to the collusive mechanism while the status of such a constraint is not clear since the refusal of an agent to join a collusive device should not prevent the remaining agents to collude on their own. 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 illustrate those differences in a simple stylized example under complete information: a symmetric triopoly under Cournot competition. Each firm has a constant null marginal cost and a maximum capacity $q_{\text{max}} = 0.5$. 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 various models 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.

10.2 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 [30] and Skreta [43] 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 [17] 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 Theorem 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.

10.3 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 [16], Segal and Whinston [40], Segal [39] and Winter [47]. Nevertheless, in those papers, discrimination comes from both the bi-
lateral nature of the contractual relationship and from severe coordination failures (unique implementation).

Genicot and Ray [16] and Segal and Whinston [40] 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 [40], 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 [16], 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.

10.4 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. Nevertheless, we emphasize that all our analysis is directly relevant and can be easily applied to environments where the timing is such that participation decisions occur before agents are learning their types as in a vast area of the mechanism design literature.

References


