Budget-Balanced Maximization of Social Welfare Resilient to Unrestricted Collusion, Privacy, and Beliefs

Pablo Azar
CSAIL, MIT
Cambridge, MA 02139, USA
azar@csail.mit.edu

Jing Chen
CSAIL, MIT
Cambridge, MA 02139, USA
jingchen@csail.mit.edu

Silvio Micali
CSAIL, MIT
Cambridge, MA 02139, USA
silvio@csail.mit.edu

January 26, 2011

Abstract

Collusion, privacy, and beliefs are forces uniquely capable of affecting the play of a mechanism.

In this paper, we define a class of mechanisms totally resilient to these three forces, and then prove that one such mechanism essentially maximizes social welfare in a budget-balanced way.

Our mechanism works in markets where the players have complete information about each other's utilities (but not about who colludes with whom), meaningfully bypasses classical impossibility results, and enjoys other valuable properties. In particular, it requires a minimal amount of communication.

1 Introduction

The goal of this paper is constructing a market mechanism for budget-balanced maximization of social welfare when the players have complete information about each other's utility functions. Our mechanism bypasses some classical impossibility results, and yet is based on an unusually strong solution concept. Indeed, our mechanism is totally resilient against three forces extremely capable of affecting one's strategic behavior: *collusion*, *privacy,* and *beliefs*.

Collusion can easily prevent a mechanism from achieving its goals. Colluders may be legally prosecuted if caught, yet collusion still occurs and will likely continue to occur. We thus find it important to design mechanisms that continue to work even when the players' ability to collude is essentially unrestricted.

Privacy is a universal desideratum. By definition, privacy-valuing players incur some "utility loss" when revealing certain information about themselves. Thus, though hard to quantify, privacy may affect the players' choices of strategies in a rational execution of a mechanism M, even in a complete-information setting. In fact, although the players know each other's true types, they may worry that *external observers* of M's execution (e.g., M's designer) may deduce information about their true types. We thus find it important to design mechanisms that are "as privacy-preserving as possible".

Beliefs may arise about anything not objectively known by the players and not contradicting their rationality. We thus find it important that a mechanism designed to be "resilient to collusion and privacy" should work no matter what beliefs the players may hold.

1.1 Unrestricted Collusion, Privacy, and Beliefs

Collusion We assume that, as soon as a mechanism is announced, the players have the ability of colluding in a very adversarial way. Namely, they may partition themselves into an *arbitrary number* of coalitions of *arbitrary size*. (For example, there may be two coalitions, each consisting of $n/2$ players; or only the "grand" coalition of all players.) The members of the same coalition $\mathcal{C}$ are assumed capable of *making side payments* to one another and of *perfectly coordinating* their actions —e.g., by entering *binding contracts* with each other.

Accordingly, after selecting a suitable *joint strategy* (i.e., a sub-vector of individual strategies indexed by $\mathcal{C}$), the members of $\mathcal{C}$ can stick to it without any problems. In addition, collusion may be *totally secret:* not only the designer and the "independent" players may have no idea about the existence of collusive players, but each collusive player may only know who colludes with him and nothing more. The only constraint of our collusion model is that each coalition is *rational.* That is, the members of a coalition $\mathcal{C}$ act so as to maximize the sum of their individual utilities. (Via payments between its members, $\mathcal{C}$ can guarantee that any outcome strictly preferred by $\mathcal{C}$ as a whole is also strictly preferred by each of its members.)

Privacy If a designer wishes to exactly implement a given social choice correspondence F, he must assume that the players are willing to divulge the information about their true types implicitly contained in an outcome in $F(\theta)$ where θ is the profile of true types. But, to be "safe", the designer must not assume further restrictions on the players' desire for privacy. That is, he must also assume that they may be reluctant to divulge any information about themselves *additional* to that contained in an outcome in $F(\theta)$. Accordingly, mechanisms based on the assumption that the players have no problem about revealing their entire true types may "overreach", and in practice fail to produce the desired outcomes.

Beliefs We impose no restrictions of the players' beliefs. Indeed, the beliefs of *all* players can be actually specified by an adversary *after* the mechanism is chosen. Accordingly, in our model one cannot expect a mechanism with multiple equilibria to end up in equilibrium. (For such an expectation to be "legitimate" it would be necessary that any combination of equilibrium strategies is itself an equilibrium.)

1.2 Implementation Resilient to Unrestricted Collusion, Privacy, and Beliefs

Intuitively, a mechanism M is resilient to unrestricted collusion, beliefs, and privacy if

- any rational play of M will produce a desired outcome, no matter who colludes with whom and what beliefs the players may have; and
- no external observer of the play can deduce any information about the players' true types beyond that implicit in the produced outcome.

We clarify the second property in our technical sections, but wish to convey the gist of the first right away.

In essence, an extensive-form mechanism M is resilient to collusion and beliefs if it guarantees that (a) at every subgame following the root, each possible independent player/coalition has an individual/joint strictly dominant strategy and (b) assuming that each one of them will choose such a strategy, the player acting at the root (whether he acts so as to maximize his own utility, if he is independent, or the total utility of his coalition, if he is collusive) also has a strictly dominant *action* available to him, and he can compute it without knowing who colludes with whom.[1] In a sense, all rational plays of M are decided at the root. Accordingly, we say that the mechanism *root implements* a given social choice correspondence if, whenever each *agent* (i.e., each independent player or coalition) chooses its strictly dominant strategy, a desired outcome is obtained.

"Mutual" vs. "Common" knowledge of Rationality. Relative to the general notion of implementation in dominant strategies, ours is weaker when all agents are assumed to be independent players (but generally stronger when collusion exists.[2]) For implementation in dominant strategies it is sufficient to rely on "*level-0 rationality*": all agents are rational, but nothing is assumed about their knowledge about the rationality of other agents. For root implementation it is sufficient to rely on *level-1 rationality* (or *mutual* knowledge of rationality): each agent is rational, and knows that all possible agents are rational.[3] Thus,

[1] This solution concept is a simplification of the one put forward in [6] for a more complex setting.

[2] In the sense that the purest form of implementation in dominant strategies is defined —as any equilibrium-based notion— relative only to *individual deviations*, and thus in general offers no protection against collusion.

[3] Actually, root implementation just requires that each agent is rational and that only a specific agent, the "one acting at the root", knows that all possible agents are rational.

root implementation does not rely on *level-2 rationality* —informally, on "knowledge about knowledge of rationality"— let alone on common knowledge of rationality.

Total Participation Resiliency to collusion, privacy and beliefs would not be too meaningful if it did not "encourage player participation." When all players are independent, a mechanism encourages player participation by being *individually rational.* But when some players may collude, individual rationality is no longer sufficient. In particular, we wish to avoid that an independent player experiences negative utility when a coalition $\mathcal{C}$ employs a sub-vector of strategies that would not be chosen by the players in $\mathcal{C}$ if they acted independently. More generally, we demand that root implementation encourage *total* participation. By this we mean that also the utility of every coalition must be non-negative. Encouraging total participation is crucial when a subset of the players, bound to coordinate their actions so as to maximize their collective utility (e.g., a pair of players secretly married to one another), pre-exists the choice of any mechanism.

1.3 Main Properties of Our Mechanism

Informally, we construct a mechanism $\mathcal{M}$ that

(1) is resilient to unrestricted collusion, privacy, beliefs; and
(2) produces a socially optimal outcome with probability arbitrarily close to 1.

We call an outcome *socially optimal* if it maximizes social welfare and is budget-balanced, and we refer to the function mapping a profile of types to the set of corresponding socially optimal outcomes as the *socially optimal correspondence,* denoted by f.

Actually, our $\mathcal{M}$ is always budget-balanced: only social welfare may not be maximum on rare occasions. Almost paradoxically, our $\mathcal{M}$ resiliently maximizes social welfare in a budget-balanced way via its ability to "destroy some of the goods", which it exercises extremely rarely, and its ability to "impose high fines upon deviating players", which it never exercises in a rational play.

Our $\mathcal{M}$ also enjoys other desirable properties.

1.4 Additional Properties of Our Mechanism

Ex-Ante Fairness. Trivially modified, $\mathcal{M}$ gives each player essentially the same expected utility.

Trivial Communication (and Computation) Overhead. We define the *communication overhead* of a mechanism to be the difference between

(a) the number of bits exchanged in a rational execution of $\mathcal{M}$, and

(b) the number of bits required to specify an outcome.

The communication overhead of our $\mathcal{M}$ is $n-1$, that is, less than one bit per player.

Our notion of communication overhead is closely related to (but different from) Fadel and Segal's *communication cost of selfishness* [9]. In a non-Bayesian setting, they use ex-post incentive compatibility as the underlying solution concept and define the communication cost of selfishness as the difference between

(a′) the number of bits exchanged in a rational execution of a mechanism M, and

(b′) the number of bits necessary to describe the players' *true types.*

Note that specifying the true types requires at least as many bits as an outcome, and in general many more. Our mechanism also has quite trivial *computation overhead.*[4]

[4]We define this to be the difference between (a) the sum of the elementary computational steps taken by a mechanism and the players in a rational execution in order to produce the desired outcome, and (b) the number of elementary computational steps required by any algorithm to compute a desired outcome on input the true-type profile θ —volunteered by the players without any incentives. Thus computation overhead measures the additional complexity needed to handle incentives. The computation overhead of our $\mathcal{M}$ is a function linear in each of the relevant variables: namely, n, the number of players, m, the number of goods, and k, the number of bits necessary to describe the value that any player may have for a subset of the goods.

Universal and Incentive-Preserving Approximation The socially optimal correspondence f may be too complex to compute. Fortunately, under our same solution concept and without any significant overhead in communication or computation, our mechanism can also be used to implement *any* social choice correspondence f' that produces budget-balanced outcomes with approximately maximum social welfare.[5]

This property should not be taken for granted. In principle, for different solution concepts, one may be able to find a mechanism implementing (in principle, but not in practice!) a social choice correspondence F that is exponential-time computable, but unable to find any mechanism implementing even *a single* feasible approximation F' of F. At other times, one may be able to implement some feasible approximations F', but not others.[6]

1.5 Bypassing Classical Impossibility Results

Even assuming that the players cannot collude, Green and Laffont [13] and Hurwicz [16] prove that the socially optimal correspondence cannot be implemented in weakly dominant strategies, whether or not the players know each other's utility functions. Assuming that the players can collude, Green and Laffont prove a different impossibility result [14]. That is, even without any concerns about budget balance, maximizing social welfare is impossible via mechanisms that are *coalition incentive compatible* (that is, providing a dominant strategy to each independent player and each coalition).

Dominant strategies and coalition incentive compatibility are excellent ways of respectively guaranteeing (1) resilience to beliefs and (2) resilience to beliefs and collusion. But, as emphasized by our result, they are not the *only* ways, and may actually be too demanding. By showing that the socially optimal correspondence *can* be root-implemented, whether or not the players are collusive, and in a way satisfying privacy and other desirable properties as well, our result suggests that root implementation may be a viable "next alternative" to implementation in dominant strategies, when all players are independent, and to coalition incentive compatible implementation, when the players may be collusive.

Finally, we would like to point out that a result of Myerson and Satterthwaite [31] does not contradict ours. Indeed, they prove the impossibility of implementing markets that are efficient and budget balanced at a Bayesian Nash equilibrium, but in an incomplete-information *Bayesian* setting.

2 Related Work

Perfect Implementation Our work should not be confused with *perfect implementation*, as proposed by [17]. Their work considers the privacy and strategic properties that an already existing (abstract) mechanism M enjoys if executed via a trusted mediator, and shows that they may be exactly preserved by concretely executing M with the players alone, without any trusted mediator. (For example, if no collusion-resilient abstract mechanism implementing a given social choice correspondence is known, then their work does not enable a designer to come up with one.) By contrast, our work is about designing not-yet-existing abstract mechanisms (to be executed by the players alone) so as to enjoy maximum privacy and maximum collusion resiliency.

The VCG mechanism The VCG mechanism [35, 7, 15] maximizes social welfare even in a setting of incomplete information, but is far from achieving our goals. In particular, it is not budget-balanced; it is "privacy-less", that is, asks the players to reveal their true types in their entirety; and, as shown by Ausubel

[5] To the best of our knowledge, this property was put forward and achieved in [6], for revenue-generating mechanisms in combinatorial auctions of incomplete information.

[6] For instance, in a combinatorial auction with sub-modular valuations, let F be the social choice function that allocates the goods so as to maximize social welfare alone (i.e., without any concerns about prices). Then the incentive-compatible mechanism of Dobzinski, Nisan, and Schapira [8] (that actually works for more general types) yields a $\frac{1}{\sqrt{m}}$-approximation F' of F. However, although Lehmann, Lehmann, and Nisan [21] prove that there exists a $\frac{1}{2}$-approximation F'' of F, no incentive compatible mechanism for F'' is known.

and Milgrom [2], it is vulnerable to just two collusive players.[7] In addition, the VCG mechanism is not fair, and its communication overhead is in general very high, because the number of bits required to specify the players' true types may be much more than those sufficient to specify an outcome of maximum social welfare.

The Moore-Repullo Mechanisms Moore and Repullo [27] propose two mechanisms. The first is of extensive form and implements a class of social choice correspondences that include ours at a unique subgame-perfect equilibrium. Accordingly, this mechanism is resilient to beliefs, but it too does not achieve our goals. In particular, it is privacy-less (because it actually requires each player to report an entire type *profile* and an integer, which at equilibrium are θ and 0 respectively), and it is vulnerable to just two colluding players.[8] Moreover, when used to compute a socially optimal outcome in a general market, its communication overhead is n times higher than that of the VCG. Finally, even when all players are independent, it relies on level-n rationality, where n is the number of players, while just level-1 rationality suffices for our mechanism.

Their second mechanism, although described for two players, is generalizable to arbitrarily many ones. However, relative to our goals, it has weaknesses similar to those of their first one.

The Abreu-Matsushima and the Glazer-Perry Mechanisms The Abreu-Matsushima mechanism [1] is a normal-form mechanism that virtually implements essentially all social choice functions F at a unique equilibrium, and is therefore belief-resilient. Their mechanism too, however, does not achieve our goals. To begin with, it is privacy-less; it is always vulnerable to $n-1$ colluding players;[9] and, for some true type profiles, it is even vulnerable to just 3 colluders.[10] In addition, as the mechanism currently stands, when k bits are required to describe a player's value for any possible outcome, it communicates 2^k times more bits than the mechanism of Moore and Repullo.[11] Finally, "for all practical purposes" the mechanism relies on common knowledge of rationality: that is, its complex backwards induction requires level-2^k rationality.

The Glazer-Perry mechanism [11], an extensive-form version of the Abreu-Matsushima mechanism, does not achieve our goals for the same reasons (and requires the same level of rationality).

Resiliency to Privacy The revenue mechanism of [5] is the closest source of inspiration for our work: it has a unique subgame-perfect equilibrium, fully preserves the privacy of the players, and has low communication overhead. However, it requires level-n rationality and that "who colludes with whom" be common knowledge among the players.[12]

If obtaining an outcome in $F(\theta)$ (where F is the desired social choice correspondence and θ the actual profile of true types) is not the main priority, Talwar and McSherry [25], and Nissim, Smorodinsky and Tennenholtz [32] show that —for some special contexts[13]— it is possible to trade exactness in implementation

[7] Their example was formulated for combinatorial auctions, but can be trivially adapted to general markets as well.

[8] Indeed, when less than $n-1$ players report the same type profile, the reported integers are used in an integer game, in which the player with the highest integer chooses the outcome. Accordingly, two collusive players can jointly deviate from the envisaged unique equilibrium and force the mechanism to enter the integer game, believing that they will win it. Whoever actually wins it, the final outcome will be the one preferred by the winner —or his collusive set— rather than the outcome desired.

[9] Because, when all type profiles reported by $n-1$ players coincide with some θ', the mechanism will, with probability greater than $1-\epsilon$, produce the outcome $F(\theta')$ without punishing the $n-1$ players

[10] That is, when $n>3$ and there is a unique coalition $\mathcal{C}$ of three players, the true-type profile θ may be such that there exists a strong Nash equilibrium (in which neither the independent players nor $\mathcal{C}$ want to deviate) yielding the outcome preferred by $\mathcal{C}$ rather than an outcome in $F(\theta)$.

[11] Such high complexity is necessary for their mechanism to be able to rely on small fines. Indeed, the mechanism requires each player i to announce x type profiles (plus a separate individual type), and keeps i from deviating from the unique equilibrium via a potential fine F that must be roughly greater than the ratio between i's maximum value for an outcome and x, that is, greater than $(1-\epsilon)2^k/x$. Thus F can be small only if x is exponentially large in k.

[12] When the players outside a coalition $\mathcal{C}$ do not know the exact composition of $\mathcal{C}$, a player in $\mathcal{C}$ can, with total impunity, greatly increase $\mathcal{C}$'s joint utility. Interestingly, he (1) causes the individual utility of some of its members to be negative, but (2) makes that of other members extremely high.

[13] In particular, one must be able to add noise to outcomes; the utility functions of players should not be very sensitive to this noise; and the outcome should not be very sensitive to an individual player's actions. Note that neither of these properties hold for our general market context.

for a different notion of privacy. In their model, there is a *curator* who learns the true type of every player and then executes a (possibly approximately) truthful mechanism using these reported types. A player is not concerned with losing privacy to the curator. Instead, he is concerned that someone who knows the true types of all the other players, and the outcome of the mechanism, can deduce some information about his own type.

Resiliency to Collusion When the utility of a coalition consists of the sum of the individual utilities of its members, collusion resiliency has also been studied by Goldberg and Hartline [12]. Like Green and Laffont [14], they use dominant strategies as the underlying solution concept. Their *c-truthful* mechanisms ensure that a coalition of at most c collusive players cannot "collectively gain more than they could by bidding individually." They exemplify their notion for auctions of multiple goods, and prove that, to be c-truthful, a mechanism M must, for any subset of the goods S and player i, fix a price $p_{S,i}$ and offer S to i for that price. (A weaker variant of their notion, *c-truthful with high probability*, has been studied by the same authors.) Laffont and Martimort [20] and Che and Kim [4] consider collusion-resilient mechanisms, under the same utility function for coalitions, based on various solution concepts that are ultimately based on equilibrium. (The latter authors further allow the utility of a coalition to be the weighted sum of the individual utilities of its members.)

Collusion resiliency has also been studied when (1) each coalition prefers an outcome ω to an outcome ω' if and only if each of its members prefers ω to ω'; and (2) players cannot guarantee side-payments to one another. In this model, a mechanism can be considered resilient to collusion if it ensures that any gain for a member of a coalition is accompanied by a loss for another member of the same coalition. Such mechanisms have been constructed under different solution concepts: by [22, 29, 34] using equilibrium, and by [28, 3, 19, 30, 10, 33] using *group (or coalition) strategy-proofness*.

When collusion is unrestricted, and not even rational, *collusion-erasing* mechanisms have been put forward in [26] and exemplified for combinatorial auctions. Such mechanisms essentially ensure at least the same (revenue) performance as when all collusive players spontaneously "leave the player set", so that the auction can be conducted solely with the independent players.

3 Games in Our Model

Mechanism design conceptually partitions any game G into a context $\mathbb{C}$ and a mechanism M, $G = (\mathbb{C}, M)$. Let us thus describe our mechanisms and contexts.

3.1 Mechanisms

All our mechanisms are probabilistic and of extensive form, with the possibility of simultaneous moves. That is, they specify a tree T; an outcome for each terminal node of T; and, for each internal node D of T, whether Nature acts or a subset S^D of the players act. If Nature acts at D, the mechanism specifies a distribution over a set A^D of actions associated with D. Else, the mechanism specifies the set A_i^D of actions available to each i in S^D. The players in S^D select their actions simultaneously. Accordingly,

- A pure strategy s_i for a player i is a function specifying an action in A_i^D for all D such that $i \in S^D$; and
- A pure *joint strategy* $s_{\mathcal{C}}$ for a coalition $\mathcal{C}$ is a sub-vector of pure strategies, one for each player in $\mathcal{C}$.

(I.e., $s_{\mathcal{C}}$ specifies an action in A_i^D for each node D such that $i \in S^D \cap \mathcal{C}$.)

KNOWLEDGE. Once chosen, a mechanism M is common knowledge to everyone.

NOTATION. For any mechanism M we denote by Σ_i the set of pure strategies available to player i, and by Σ the profile of sets of pure strategies. For any strategy profile σ, $M(\sigma)$ and $M[\sigma]$ respectively denote the distribution of outcomes and the distribution of terminal nodes generated by M under σ. (Both distributions

taken over the strategies in σ, if mixed, and the coin tosses of M, if probabilistic.) The support of a distribution D is denoted by $[D]$.

3.2 Collusive Contexts in General

A context $\mathbb{C}$ has six components, $\mathbb{C} = (N, \Omega, \Theta, \theta, u, \mathscr{C})$, where (1) N is a finite set of players; (2) Ω is a set of outcomes; (3) Θ is a profile of finite sets, where each Θ_i is the set of all possible *types* of player i; (4) θ is the profile of true types, where $\theta_i \in \Theta_i$ for all i; (5) u is a function from $N \times \Theta \times \Omega$ to $\mathbb{R}$; and (6) $\mathscr{C}$ is a partition of the players.[14]

KNOWLEDGE. Components N, Ω, Θ, and u are common knowledge to everyone; θ is common knowledge only to the players; and, for each $\mathcal{C} \in \mathscr{C}$, the subset $\mathcal{C}$ itself is common knowledge to the players in $\mathcal{C}$.

NOTATION. We refer to each set of players in $\mathscr{C}$ as a *coalition,* and, for each player i, denote by $\mathscr{C}_i$ the unique coalition in $\mathscr{C}$ containing player i. We call a player i *independent* if $\mathscr{C}_i = \{i\}$, and *collusive* otherwise. We call a context *non-collusive* if each player is independent and this fact (i.e., $\mathscr{C} = \{\{1\}, \ldots, \{n\}\}$) is common knowledge to everyone; and we call a context *collusive* otherwise. For each player i and every subset $\mathcal{C}$ of the players, we denote by u_i and $u_{\mathcal{C}}$ the functions from Ω to $\mathbb{R}$ defined as follows: $\forall \omega \in \Omega$, $u_i(\omega) = u(i, \theta, \omega)$ and $u_{\mathcal{C}}(\omega) = \sum_{i \in \mathcal{C}} u_i(\omega)$. Whenever i is independent, we identify u_i and $u_{\{i\}}$. If $\mathcal{C} \in \mathscr{C}$, we refer to $u_{\mathcal{C}}$ as the utility function of $\mathcal{C}$. When M is clear and σ is a strategy profile of M, we may use the term $u_{\mathcal{C}}(\sigma)$ to denote the expectation of $u_{\mathcal{C}}(M(\sigma))$, taken over all possible sources of randomness: again, the strategies of σ, if mixed, and M itself, if probabilistic.

REMARKS. If i is a collusive player, we do not refer to u_i as i's "utility function", because when a context is collusive in our model only coalitions —rather than players— have utility.

We insist on $\mathscr{C}$ being a partition of the players, and thus on coalitions being disjoint, because otherwise it might be problematic for some players to act so as to maximize the utilities of different coalitions.

3.3 Collusive Market Contexts

Relative to a set G of m goods, initially partitioned among the players, an *n-m-k* market context is a context $\mathbb{C} = (N, \Omega, \Theta, \theta, u, \mathscr{C})$ where

- $N = \{1, \ldots, n\}$;
- $\Omega = \mathcal{A} \times \mathbb{R}^n$, the Cartesian product of the set of all *allocations* and the set of all profiles of *prices*.
  (An allocation A is a partition of G into $n+1$ subsets, $A = A_0, \ldots, A_n$, where A_0 represents the subset of the destroyed goods, and for $i > 0$, A_i represents the set of goods allocated to i. If positive, P_i is the amount of money paid by player i, otherwise $-P_i$ is the amount of money received by i.)
- Each $\Theta_i = \{v_i : 2^G \to \mathbb{Z} \cap (-2^{k-1}, 2^{k-1}) \text{ such that } v_i(\emptyset) = 0\}$.
  (For our markets, we may refer to each possible type v_i as a *valuation.* The k represents the number of bits required to specify the value that a player may have for a subset of the goods: $k-1$ bits for the "magnitude" and one the "sign.")
- $u(i, v, (A, P)) = v_i(A_i) - v_i(E_i) - P_i$.
  (E_i is i's subset of the goods in the initial partition E, that is, "the subset of goods i brings to the market.")

The nature of the goods being irrelevant, an *n-m-k* market context is fully specified by the triple $(E, \theta, \mathscr{C})$.

Notation If $\omega = (A, P)$ is an outcome, then we set $SW(\omega) = SW(A) = \sum_{i=1}^{n} \theta_i(A_i)$, and refer to $SW(\omega)$ and $SW(A)$ as the *social welfare* of ω and A respectively. Accordingly, an outcome ω is *socially optimal* if $SW(\omega) = \max_{A' \in \mathcal{A}} SW(A')$ and $\sum_{i=1}^{n} P_i = 0$.

[14] Our contexts need not specify which beliefs the players hold: an adversary will choose them at the start of an execution.

4 Our Implementation Notions

We define our notions of implementation for general contexts, although we shall later on exemplify them for market contexts only.

4.1 Root Implementation

We have developed several notions of implementation resilient to collusion, privacy, and beliefs, but present only the strongest one capable of supporting our mechanism.

Definition 1. *A mechanism M is* collusively dominant-strategy *(CDS) if, for all collusion structures $\mathscr{C}$ and all coalitions $\mathcal{C} \in \mathscr{C}$, there is a joint strategy $s_{\mathcal{C}}$ such that, for all pure subprofiles of strategies $s'_{\mathcal{C}}$ and $s'_{-\mathcal{C}}$,*

$$u_{\mathcal{C}}(s_{\mathcal{C}}, s'_{-\mathcal{C}}) > u_{\mathcal{C}}(s'_{\mathcal{C}}, s'_{-\mathcal{C}}).$$

(An "empty" mechanism, that is, one whose root coincides with a terminal node, is considered CDS.)

Definition 2. *We say that a mechanism M is* root-solvable *if:*

(1) A single player acts at the root. (Denote such a player by r, and the root by R.)

(2) For each action $x \in A^R_r$, the sub-mechanism rooted at the node reached by x, M^x, is CDS.

(3) For any context $\mathbb{C} = (N, \Omega, \Theta, \theta, u, \mathscr{C})$, there exists a subset A of A^R_r, independent of $\mathscr{C} \setminus \{\mathscr{C}_r\}$,[15] *such that, for all actions $x \in A$ and $x' \in A^R_r$:*

- *$u_{\mathscr{C}_r}(s^x_{\mathscr{C}}) = u_{\mathscr{C}_r}(s^{x'}_{\mathscr{C}})$ whenever $x' \in A$;*
- *$u_{\mathscr{C}_r}(s^x_{\mathscr{C}}) > u_{\mathscr{C}_r}(s^{x'}_{\mathscr{C}})$ whenever $x' \notin A$; and*
- *$u_{\mathcal{C}}(s^x_{\mathscr{C}}) \geq 0$ for all coalitions $\mathcal{C} \in \mathscr{C}$;*

where $s^x_{\mathscr{C}}$ denotes the strategy profile of M in which r plays x at the root and, for all actions $y \in A^R_r$, every coalition in $\mathscr{C}$ plays its CDS strategy in M^y.

If $\mathbb{C}$, M, and A are as above, then we call ROOT the solution concept mapping $G = (\mathbb{C}, M)$ to the set of strategy profiles $\mathsf{ROOT}(G) = \{s^x_{\mathscr{C}} : x \in A\}$; and root-profile any strategy profile in $\mathsf{ROOT}(G)$.

Definition 3. *Let M be a root-solvable mechanism and F a social choice correspondence. We say that M* root-implements *F if for any context $\mathbb{C}$ and any root-profile s of $(\mathbb{C}, M)$, $M(s) \in F(\theta)$.*

Assuming mutual knowledge of rationality, a mechanism M root implementing F is resilient to collusion and beliefs. Indeed, when all coalitions are rational, (a) each coalition $\mathcal{C}$ chooses its collusively dominant strategy at every proper "submechanism"; (b) the only reasonable belief for the coalition $\mathscr{C}_r$ is that each possible coalition will choose its CDS strategy; and (c) based on the latter belief, player r, on behalf of $\mathscr{C}_r$, can only play an action in A at the root. Moreover, because A is independent of $\mathscr{C} \setminus \{\mathscr{C}_r\}$, r can compute A regardless what he, or any possible player in the coalition $\mathscr{C}_r$, may know about the other coalitions. Accordingly, no matter who colludes with whom and what beliefs the players may have, if a mechanism M root implements F, then each rational execution of M generates an outcome in $F(\theta)$.[16]

Finally, Property 3 (third bullet) indeed guarantees that root implementation includes total participation.

[15] That is, the subset A is the same for any two contexts $(N, \Omega, \Theta, \theta, u, \mathscr{C})$ and $(N, \Omega, \Theta, \theta, u, \mathscr{C}')$ such that $\mathscr{C}_r = \mathscr{C}'_r$.

[16] Note that we do not demand "full implementation" in the spirit of Maskin [23]. That is, we do not insist that each possible outcome in $F(\theta)$ be generatable by rationally executing M. We actually could demand and achieve this stronger property, but the corresponding notion of virtual root implementation would become more complex.

4.2 Full Privacy

Definition 4. *We say that a mechanism M is* rationally deterministic *relative to a solution concept SC if, for all contexts $\mathbb{C}$ and all strategy profiles $s \in SC(\mathbb{C}, M)$, $M(s)$ consists of a single outcome.*

Definition 5. *Let M be a rationally deterministic mechanism relative to a solution concept SC. We say that M is* fully private *if, for all contexts $\mathbb{C} = (N, \Omega, \Theta, \theta, u, \mathscr{C})$ and $\mathbb{C}' = (N, \Omega, \Theta, \theta', u, \mathscr{C}')$, and all strategy profiles $s \in SC(\mathbb{C}, M)$ and $s' \in SC(\mathbb{C}', M)$,*

$$M(s) = M(s') \Rightarrow M[s] = M[s'].$$

Recall that all components of a context $\mathbb{C}$ are common knowledge to everyone, except for the profile of true types θ and the collusion structure $\mathscr{C}$. Accordingly, the players lose privacy only if someone gains information about θ or $\mathscr{C}$ that he did not already possess.

The very purpose of mechanism design is to come up with a mechanism M that, rationally played in a context $\mathbb{C}$, produces an outcome ω in the set $f(\theta)$, for a given social choice correspondence f. Thus, ω itself is information about θ that might not be available beforehand. As already argued in our introduction, however, for mechanism design to be possible at all, one must postulate that this loss of privacy is acceptable to the players. The designer of M should instead focus on ensuring that a rational play of M does not divulge any more information than that implicitly contained in ω.

Note that an observer of a rational play of M not only learns ω, but also which specific terminal node h associated with ω has been reached by the play. Thus h causes an additional privacy loss for the players only if it enables an observer to further narrow down the set of candidate contexts. That is, h causes an additional privacy loss only if

> "the set of contexts yielding h in a rational play" —that is, $\{\mathbb{C} : \exists s \in SG(\mathbb{C}, M)$ such that $M[s] = h\}$— is smaller than "the set of contexts yielding ω in a rational play."[17]

Accordingly, M does not cause the players any additional privacy loss if, after a rational play, "one cannot use the terminal node to differentiate any two contexts that cannot be differentiated based on the outcome alone." This is the very condition of Definition 5.

Remark Note that the equalities $M(s) = M(s')$ and $M[s] = M[s']$ of Definition 5 can be interpreted as equalities among distributions. Demanding that fully private mechanisms be rationally deterministic is useful for the notion of our next section (and provides for a stronger and yet achievable notion).

4.3 Essential Root Implementation with Full Privacy

Definition 6. *For any constant $\delta \in (0, 1/2)$, we say that a mechanism M $(1-\delta)$-*root implements with full privacy *a social choice correspondence F if*

- *M is root-solvable;*
- *For any context $\mathbb{C}$ and any root-profile s of $(\mathbb{C}, M)$:*
  (1) $M(s)$ assigns at least $1-\delta$ of its total probability mass to a single terminal node, denoted by $M[s]^\delta$; and
  (2) $M(s)^\delta \in F(\theta)$, where $M(s)^\delta$ denotes the outcome associated with $M[s]^\delta$.
- *For all contexts $\mathbb{C} = (N, \Omega, \Theta, \theta, u, \mathscr{C})$ and $\mathbb{C}' = (N, \Omega, \Theta, \theta', u, \mathscr{C}')$, all root-profiles s of $(\mathbb{C}, M)$, and all root-profiles s' of $(\mathbb{C}', M)$,*

$$M(s)^\delta = M(s')^\delta \Rightarrow M[s]^\delta = M[s']^\delta.$$

[17] For example, the intuitive fact that the VCG (with ties broken lexicographically) causes the players to lose privacy, even when all players are guaranteed to be independent, can be argued technically as follows. The solution concept SC is "dominant strategies"; $h = \theta$, that is, h coincides with the profile of true strategies; and ω is the lexicographically first allocation maximizing social welfare. Then the set of contexts yielding h in a rational play has a single element, that is, consists of the actual context $\mathbb{C} = (\{1, \ldots, n\}, \Omega, \Theta, \theta, u, \{\{1\}, \ldots, \{n\}\})$. By contrast, there may be plenty of contexts having ω as their lexicographically first allocation maximizing social welfare, and all of them belong to the set of contexts yielding ω in a rational play.

We say that a social choice correspondence F is essentially root-implementable with full privacy*, if for any constant $\delta \in (0, 1/2)$, there is a mechanism M_δ that $(1-\delta)$-root implements F with full privacy.*

Remarks

- *Existence vs. Uniform Constructibility.* In principle, the required mechanisms M_δ could be totally different from each other, and their existence could be proven non-constructively. One can easily realize, however, that we actually prove the existence of a uniform and polynomial-time algorithm that, on an input consisting of (a) δ and (b) the components of a market context $\mathbb{C}$ that are common knowledge to everyone, outputs (the description of) a mechanism M_δ as desired.
- *Virtual vs. Essential.* Literally following the classical notion of a virtual implementation [1], we could have demanded (informally speaking) that

  > "for each δ, there is a mechanism M_δ that, with probability greater than $1-\delta$, root implements f with full privacy."

  Notice however that the above formulation does not require that M_δ assign at least $1-\delta$ of its total probability mass to a single terminal node.

5 Our Market Result

Our result is formally stated as follows.

Theorem 1. *The socially optimal correspondence is essentially root-implementable with full privacy.*

We prove Theorem 1 by explicitly constructing a mechanism $\mathcal{M}$ that root implements f with full privacy. To develop some intuition, we derive $\mathcal{M}$ via a trial-and-error process, so as to present its underlying ideas one at a time, each one together with its original motivation.

(The reader preferring a drier approach may proceed to Subsection 5.2.)

5.1 "Simple but Wrong" Versions of Our Mechanism

Mechanism *Naive-1*

1. *Each player i, simultaneously with the others, announces a budget-balanced outcome where each price is in $\mathbb{Z} \cap (-2^{k-1}, 2^{k-1})$.*

- *If everyone announces the same outcome (A, P), then (A, P) is the final outcome.*
  Else, the outcome is $(E, 0^n)$ —i.e., the alternative remains the initial one and everyone pays 0.

This mechanism obviously has an equilibrium generating a socially optimal outcome: namely, all players announce the same (A, P) such that A is of maximum social welfare, $P_i = \theta_i(A_i) - \theta_i(E_i)$ for each $i \neq 1$, and $P_1 = -\sum_{i \neq 1} P_i$.[18] However, *Naive-1* does not achieve our goals. In particular, it allows any outcome giving every player non-negative utility to be an equilibrium outcome, and thus is highly vulnerable to belief-mismatch. (In addition, because every player announces an outcome, the communication complexity of *Naive-1* is n times higher than the one required.)

Thus let us consider a different mechanism, whose communication complexity seems as desired. The new mechanism also provides each player a small "discount" ϵ, trying to give the players positive utility whenever possible, encouraging them to "take the right decision when on the fence."

Mechanism *Naive-2*

[18] Notice that the utility of any player other than 1 is 0, and the utility of player 1 is $\theta_1(A_1) - \theta_1(E_1) - P_1 = \sum_i \theta_i(A_i) - \sum_i \theta_i(E_i) \geq 0$. If a player deviates, then the outcome is $(E, 0^n)$ and his utility is 0. Therefore this strategy profile is indeed an equilibrium.

1. *Player 1 announces an outcome* (A, P) *such that* $P_i \in \mathbb{Z} \cap (-2^{k-1}, 2^{k-1})$ *for each* i*, and* $\sum_i P_i = 0$.
- *If* $(A, P) = (E, 0^n)$*, then the mechanism ends and the outcome is* $(E, 0^n)$.
2. *Each player other than player 1 announces YES or NO.*
- *If all players announce YES, then the final outcome is* $(A, (P_1 + (n-1)\epsilon, P_2 - \epsilon, \ldots, P_n - \epsilon))$. *Else, it is* $(E, 0^n)$.

Assume that the initial alternative does not maximize social welfare. Then, since the mechanism is budget balanced, it is easy to see that, when $(n-1)\epsilon < 1$, player 1 maximizes his own utility by announcing a *truthful* outcome, that is, a socially optimal outcome (A, P) such that $P_i = \theta_i(A) - \theta_i(E)$ for each $i \neq 1$. In this case, all other players have utility $\epsilon > 0$ and thus should rationally announce YES. (Player 1's utility is integrally positive, and thus possibly much higher than that of the others, but fairness is not a concern for now.)

When the initial alternative has already maximum social welfare, the mechanism offers player 1 a way out —that is, to announce $(E, 0^n)$— to avoid getting negative utility by paying ϵ to all other players.

Let us explain, however, that *Naive-2* also fails to achieve our goals due to a problem of belief-mismatch.

A Belief-Mismatch Problem Assume that player 1 proposes an outcome (A, P) giving negative utility to each member of a *set* (not a *coalition*, since we are for now analyzing the no-collusion case) of at least two players. Then, if each member of the set believes that another member will say NO, he himself may very well say YES: indeed, a single NO suffices for (A, P) not to be implemented. In turn, if player 1 believes that there will be such a belief-mismatch and all members of the set will announce YES, he has incentives to announce an outcome (A, P) giving himself more utility than an equilibrium outcome. (E.g., A may be player 1's favorite allocation, and $P_2, \ldots, P_n$ may be arbitrarily high.) Accordingly, there is no guarantee about the social welfare achieved. This problem does not go away by "fining" player 1 when a player announces NO.

A False Fix It may appear that the latter problem may simply vanish by modifying the mechanism so that the "simultaneous" Step 2 is replaced by $n-1$ "sequential" steps, where players 2 through n in turn announce YES or NO. This modification, however, would not work when (eventually) considering the possibility of collusion. For instance, assume that player 1 announces an outcome giving negative (individual) utility to independent player $n-1$ and player n. Then, player $n-1$, when his turn comes, may very well announce YES, believing that player n is independent and that he will announce NO later on. However, if player n secretly belongs to a coalition $\mathcal{C}$ whose collective utility is positive, then player n will announce YES.[19] The following fix instead works even in the presence of collusion.

A Better Fix The above belief-mismatch problem is solved by properly using randomness, so that the distribution of the final outcome depends on the precise number of players announcing NO, rather than the mere existence of such players. Namely,

Mechanism *Naive-3*

1. *Player 1 announces an outcome* (A, P) *such that* $P_i \in \mathbb{Z} \cap (-2^{k-1}, 2^{k-1})$ *for each* i*, and* $\sum_i P_i = 0$.
- *If* $(A, P) = (E, 0^n)$ *then the mechanism ends and the outcome is* $(E, 0^n)$.
2. *Every other player announces YES or NO.*
- *If all players announce YES, then the mechanism ends and the outcome is* $(A, (P_1 + (n-1)\epsilon, P_2 - \epsilon, \ldots, P_n - \epsilon))$.
- *Publicly flip a biased coin* c_1 *such that* $\Pr[c_1 = Heads] = 1 - \frac{\epsilon Y}{n}$*, where* Y *is the number of players announcing YES.*
- *If* $c_1 = Heads$ *then the outcome is* $(E, 0^n)$*; otherwise the outcome is* $(A, (P_1 + (n-1)\epsilon, P_2 - \epsilon, \ldots, P_n - \epsilon))$.

[19] Modifications based on "sequentializing" Step 2 do exist, but at least those we are aware of are substantially more complex. In addition, they would require a weaker solution concept.

Naive-3 ensures that a player i's unique best response is announcing YES or NO "truthfully", that is, solely based on the sign of $\theta_i(A_i) - \theta_i(E_i) - P_i$, *no matter what the other players may announce.*

A New Problem When some players announce NO and some players announce YES, *Naive-3* implements (A, P) with small but positive probability. Therefore player 1 has incentives to announce an outcome (A', P') where he gives himself an arbitrarily high utility, e.g. 2^{k-1}, but also gives positive utility to at least another player. By doing so, because 2^{k-1} can be arbitrarily large compared with ϵ, player 1 secures for himself an expected utility arbitrarily higher than his truthful-equilibrium utility. Accordingly, once more, there is no guarantee about the final social welfare.

New Fixes To counter the above problem we need to refine the probabilistic choices of the mechanism, and impose a suitable fine to player 1 when a player announces NO, as shown in Naive-4 below. Here ϵ is an arbitrary constant in $(0, \frac{1}{3n})$, and $B = 2^{k-1}$. Upon termination, (A, P) will be the final outcome.

Mechanism *Naive-4*

1. *Player 1 announces an outcome* (A^*, P^*) *such that* $P_i^* \in \mathbb{Z} \cap (-B, B)$ *for each* i, *and* $\sum_i P_i^* = 0$.

a. If $(A^*, P^*) = (E, 0^n)$, *then set* $(A, P) = (E, 0^n)$ *and HALT.*

2. *Each player* $i \neq 1$ *announces YES or NO.*

b. Let Y *be the number of players announcing YES. If* $Y = n - 1$, *then set* $(A, P) = (A^*, (P_1^* + (n - 1)\epsilon, P_2^* - \epsilon, \ldots, P_n^* - \epsilon))$ *and HALT.*

c. Publicly flip a biased coin c_1 *such that* $\Pr[c_1 = Heads] = 1 - \epsilon$.

d. If $c_1 = Heads$, *then set* $(A, P) = (E, (1, 0, \ldots, 0))$ *and HALT.*

e. If $c_2 = Tails$, *then flip a biased coin* c_2 *such that* $\Pr[c_2 = Heads] = \frac{Y}{nB}$. *If* $c_2 = Heads$, *then set* $(A, P) = (A^*, (P_1^* + (n-1)\epsilon, P_2^* - \epsilon, \ldots, P_n^* - \epsilon))$, *otherwise set* $(A, P) = (E, 0^n)$.

A Collusion Problem With the above changes, it becomes irrational for player 1 to propose any outcome that is not truthful, when it is common knowledge that all players are independent. When collusion exists, player 1 has many different ways to manipulate prices for members of a coalition. Since he does not have complete information about the collusion structure, and may act based on both his knowledge and his beliefs about who is colluding with whom, those manipulations may lead to unexpected outcomes where the desired social welfare is not achieved. For example, assume that player 1 knows that either (`i`) $\{2, 3\}$ is a coalition, or (`ii`) both players 2 and 3 are independent, but has some wrong beliefs. Specifically, he believes that (`i`) is the case, but (`ii`) is the actual truth. Then, according to his beliefs, it is completely rational for player 1 to announce $P_2^* = \theta_2(A_2^*) - \theta_2(E_2) + 1000$ and $P_3^* = \theta_3(A_3^*) - \theta_3(E_3) - 1000$, because according to his belief both players 2 and 3 would announce YES. But if player 1 really does so, then being independent, player 2 will announce NO, and there is no guarantee about the final social welfare.

Final fixes To deal with collusion we further fine tune the randomness used in the mechanism, and, with very small probability, destroy all but one player's goods. These fixes are reflected in the final version of our mechanism, provided in the next subsection.

5.2 Our Mechanism

Our mechanism $\mathcal{M}$ uses 6 parameters. Technically, if we wanted to use a heavier notation,

$$\mathcal{M} = \mathcal{M}_{n,m,k,E,\epsilon,\delta}$$

where

- n, m, k and E are the common-knowledge portion of an n-m-k market
  (i.e., n is the number of players; m the number of goods; and k the number of bits to describe a player's possible value of any subset of the goods; and E is the initial partition of the goods, the *endowment*)
- $\epsilon \in (0, \frac{1}{5n})$; and
- $\delta \in (0, \frac{\epsilon}{B})$, where $B = 2^{k-1}$
  (i.e., B is the maximum value that a player may have for a subset of the goods).

Our mechanism $\mathcal{M}$ flips three biased coins c_0, c_1, and c_2, and uses a mechanism $\mathcal{M}'$ as a subroutine. Upon termination, (A, P) will be the final outcome of $\mathcal{M}$. Steps labeled by letters are taken by the mechanism, steps labeled by numbers are taken by the players.

Mechanism $\mathcal{M}$

1. *Player 1 announces an outcome* (A^*, P^*).

a. *If* $P_i^* \notin \mathbb{Z} \cap (-2^k, 2^k)$ *for some player* $i \neq 1$, *or* $\sum_i P_i^* \neq 0$, *then set* $A = E$, $P_1 = 1$, $P_i = 0$ *for each* $i \neq 1$, *and HALT.*

b. *If* $(A^*, P^*) = (E, 0^n)$ *then set* $(A, P) = (E, 0^n)$ *and HALT.*

c. *Flip a biased coin* c_0 *such that* $\Pr[c_0 = Heads] = 1 - \delta$. *If* $c_0 = Tails$ *then go to Mechanism* $\mathcal{M}'$.

2. *Each player* $i \neq 1$ *announces YES or NO.*

d. *Let* Y *be the number of players announcing YES. If* $Y = n - 1$, *then set* $A = A^*$, $P_1 = P_1^* + 2(n-1)\epsilon$, $P_i = P_i^* - 2\epsilon$ *for each* $i \neq 1$, *and HALT.*

e. *Publicly flip a biased coin* c_1 *such that* $\Pr[c_1 = Heads] = 1 - \epsilon$. *If* $c_1 = Heads$, *then set* $A = E$, $P_1 = 2n(n - Y)B + (n-1)\epsilon$, $P_i = -\epsilon$ *for each* $i \neq 1$, *and HALT.*

f. *Flip a biased coin* c_2 *such that* $\Pr[c_2 = Heads] = \frac{Y}{nB}$. *If* $c_2 = Heads$ *then set* $A = A^*$, $P_1 = P_1^* + 2(n-1)\epsilon$, $P_i = P_i^* - 2\epsilon$ *for each* $i \neq 1$, *and HALT.*

g. *Set* $A = E$, $P_1 = (n-1)\epsilon$, $P_i = -\epsilon$ *for each* $i \neq 1$, *and HALT.*

Mechanism $\mathcal{M}'$

h. *Uniformly and randomly choose a player* w.

i. *If* $w = 1$, *then: set* $A_1 = A_1^*$, $P_1 = (n-1)\epsilon$, $A_i = \emptyset$ *and* $P_i = -\epsilon$ *for all* $i \neq 1$; *destroy all goods but* A_1; *and HALT.*

3. *Player* w *announces YES or NO.*

j. *If player* w *announces YES, then: set* $A_w = A_w^*$, $P_w = P_w^* - 2\epsilon$, $P_1 = -P_w^* + n\epsilon$, $A_i = \emptyset$ *for all* $i \neq w$, *and* $P_i = -\epsilon$ *for all* $i \notin \{1, w\}$; *destroy all goods but* A_w; *and HALT.*

k. *Set* $A_w = E_w$, $A_i = \emptyset$ *for all* $i \neq w$, $P_1 = \frac{n^2 B}{\delta} + (n-1)\epsilon$, $P_i = -\epsilon$ *for all* $i \neq 1$; *destroy all goods but* A_w; *and HALT.*

Remarks

- *Budget Balance.* The mechanism may a priori gain money (e.g., from player 1 if $c_0 = Heads$ and some player announces NO), but when the players act rationally, it does not take money from or give money to the players.
- *Almost Ex-Ante Fairness.* In a rational execution, Player 1's utility could be very high (if the difference between the maximum social welfare and the initial one is very high), while all other players' utilities are always between 0 and 2ϵ. In order to make the mechanism ex-ante fair, one could choose a random player to play the role of player 1. This would give every player almost the same expected utility. (The only difference in expected utilities may come via Step c when two players have different values for their respective endowment.)

5.3 Intuitive Analysis of $\mathcal{M}$

In Sections A and B of our appendix, we prove that, for the specific values of δ, $\mathcal{M}$ indeed root-$(1-\delta)$-implements the socially optimal correspondence for any n-m-k market context. In this subsection we give some intuitive explanations. We start by assuming for a moment that all players are independent.

Resiliency against beliefs holds because, for player 1, the only reasonable strategies are those corresponding to announcing some truthful outcome, while each of the other players has a single reasonable strategy: announcing YES or NO truthfully at each of his decision nodes. Accordingly, if the original endowment E already maximizes social welfare, then in any rational play every player's utility equals 0. Otherwise player 1's utility equals the (integral) difference in social welfare between the best allocation and the initial endowments, minus $2(n-1)\epsilon$, while the utility of any other player is between 0 and 2ϵ.

Since, by our choice of ϵ, the players' utilities are always non-negative, it automatically follows that full participation is satisfied. Moreover, since only the player who acts in Step 1 has a very different utility from the others', by choosing this player uniformly at random, ex-ante almost-fairness is satisfied. The communication overhead of $\mathcal{M}$ is small, because in a rational play only $n-1$ YES's are transmitted in addition to the socially optimal outcome (A^*, P^*) announced by player 1. The computation overhead of $\mathcal{M}$ is small because, after player 1 computes (A^*, P^*), the only computation involved is verifying that (A^*, P^*) is budget balanced, checking whether $(A^*, P^*) = (E, 0^n)$, and comparing each $\theta_i(A_i^*)$ with $\theta_i(E_i) + P_i^*$.

Resiliency against privacy holds because the only information leaked about the players' valuations in a rational play is the socially optimal outcome announced by player 1. The rest of the execution of our mechanism (that is, the fact that everybody else announces YES) can be deduced from observing the outcome alone.

Now we remove our assumption about the independence of the players. Resiliency against collusion is essentially due to the propose-and-agree structure of our mechanism, which is already present in Naive-2. This structure, in contrast with prior works, does not give any "additional power" to collusive players. Whether or not player 1 is collusive, his best strategy continues to be proposing an outcome maximizing social welfare. In fact, if player 1 belonged to a coalition with other players, then whatever utility his coalition can derive from an outcome "untruthfully proposed" in their favor can be more than compensated by the money that player 1 can get by proposing some truthful outcome. (Of course there are many ways for player 1 to manipulate his colluders' prices, but all of these are rational and correspond to a socially optimal outcome.) In addition, by proposing some truthful outcome, player 1 is sure that, no matter who colludes with whom, no one will announce NO. Finally, not knowing who else colludes with whom, Player 1 (whether or not he is collusive) is better off announcing not only an outcome with maximum social welfare, but also, for every player i outside his own coalition, a payment equal to $\theta_i(A_i) - \theta_i(E_i)$. Assume for a moment that Player 1 announced $P_2 = \theta_2(A_2) + \theta_3(A_3) - (\theta_2(E_2) + \theta_3(E_3))$, $P_3 = 0$, and $P_j = \theta_j(A_j) - \theta_j(E_j)$ for any other player j. Then, even if players 2 and 3 collude together and they both announce YES in Step 2, player 2 announces NO in Step 3 when $w = 2$, and player 1's utility is strictly less than the utility he would have received by announcing $P_2 = \theta_2(A_2) - \theta_2(E_2)$ and $P_3 = \theta_3(A_3) - \theta_3(E_3)$. Even worse, if players 2 and 3 do not collude together, then player 2 would announce NO if $\theta_3(A_3) - \theta_3(E_3) > 0$, and player 1's coalition risks uselessly

a negative utility. Generalizing, assuming that all players not colluding with 1 are independent, "pricing an alternative with maximum welfare correctly" gives player 1 the same utility as any other socially optimal outcome, no matter who colludes with whom. (Internally to his own coalition, player 1 is free to choose from many pricing schemes without risks and without gains.)

Finally, let us briefly turn our attention to incentive-preserving approximation. Mechanism $\mathcal{M}$ essentially out-sources the evaluation of our social choice correspondence f on the profile of true types θ to player 1. As we said, when the number of goods is large, no one —player 1 included— might in practice be able to compute $f(\theta)$. When this is the case, however, player 1 has all the incentive in the world to evaluate on θ the best feasible approximation f' to f in order to announce his proposed outcome (A^*, P^*). Indeed, the larger the social welfare of (A^*, P^*) is, the larger his utility is. This approach is appealingly simple and effective, and the last two authors have already used in other settings, including settings of incomplete information [6].

6 Conclusions

We have proved that a natural social choice function, which is not implementable in a classical sense, can be meaningfully implemented in a new sense (and with many additional advantages) in settings where the players know each other's individual utility functions. Such settings are theoretically important to establish what is in principle achievable in mechanism design, and have often provided the starting point of further explorations. We believe and hope that this will be the case here too. An increased ability to deal with collusion, privacy, and complexity will be crucial for developing a more robust and comprehensive theory of strategic interaction.

Acknowledgements

We thank Gabriel Carroll, Drew Fudenberg, Stephen Morris, Alessandro Pavan, and Olivier Tercieux for suggesting relevant literature and general comments during oral presentations of this work. We especially thank Gabriel for his detailed comments on an earlier version of this work.

References

[1] D. Abreu and H. Matsushima. Virtual implementation in iteratively undominated strategies: Complete information. *Econometrica*, 60(5):993–1008, 1992.

[2] L.M. Ausubel and P. Milgrom. The Lovely but Lonely Vickrey Auction. *Combinatorial Auctions*, MIT Press, pp. 17-40, 2006.

[3] S. Barbera and M.O. Jackson. Strategy-Proof Exchange. *Econometrica*, Vol.63, No.1, pages 51-87, Jan., 1995.

[4] Y. Che and J. Kim. Robustly Collusion-Proof Implementation. Econometrica, Vol. 74, No. 4, pages 1063-1107, 2006.

[5] J. Chen, A. Hassidim, and S. Micali. Robust Perfect Revenue from Perfectly Informed Players. *Innovations in Computer Science*, pages 94-105, Beijing, 2010.

[6] J. Chen and S. Micali. A New Approach To Auctions And Resilient Mechanism Design. STOC'09: Proceedings of the 41st annual ACM symposium on Theory of computing, pages 503-512. June 2009

[7] E.H. Clarke. Multipart Pricing of Public Goods. *Public Choice*, Vol.11, No.1, pp. 17-33, Sep., 1971.

[8] S. Dobzinski, N. Nisan, and M. Schapira. Approximation algorithms for combinatorial auctions with complement-free bidders. In *Proceedings of the thirty-seventh annual ACM symposium on Theory of computing (STOC '05)*, pages 610-618, 2005.

[9] R. Fadel and I. Segal, The communication cost of selfishness. *Journal of Economic Theory*, Vol. 144, Issue 5, pages 1895-1920, September 2009.

[10] J. Feigenbaum, C. H. Papadimitriou, and S. Shenker. Sharing the Cost of Multicast Transmissions. Journal of Computer and System Sciences, Vol. 63, pages 21-41, 2001.

[11] J. Glazer and M. Perry. Virtual implementation in backwards induction. *Games and Economic Behavior*, 15(1):27 – 32, 1996.

[12] A. Goldberg and J. Hartline. Collusion-Resistant Mechanisms for Single-Parameter Agents. *Symposium on Discrete Algorithms*, pages 620-629, 2005.

[13] J. Green and J. Laffont. Characterization of satisfactory mechanisms for the revelation of preferences for public goods. Econometrica, 45(2), pages 427-438, 1977.

[14] J. Green and J. Laffont. On coailtion incentive compatibility. Review of Economic Studies, Vol. 46, No. 2, pages 243-254, 1979.

[15] T. Groves. Incentives in Teams. *Econometrica*, Vol. 41, No. 4, pp. 617-631, 1973.

[16] L. Hurwicz. On the existence of allocation systems whose manipulative Nash equilibria are Pareto optimal. Unpublished. 1975.

[17] S. Izmalkov, M. Lepinski, and S. Micali. Perfect Implementation. Games and Economic Behavior, to appear. Available at http://people.csail.mit.edu/silvio/Selected Scientific Papers/Mechanism Design/.

[18] M.O. Jackson. A crash course in implementation theory. *Social Choice and Welfare*, 18(4):655–708, 2001.

[19] K. Jain and V. Vazirani. Applications of Approximation Algorithms to Cooperative Games. *Symposium on Theory of Computing*, pages 364-372, 2001.

[20] J. Laffont and D. Martimort. Mechanism Design with Collusion and Correlation. Econometrica, Vol. 68, No. 2, pages 309-342, 2000.

[21] B. Lehmann, D. Lehmann, and N. Nisan. Combinatorial auctions with decreasing marginal utilities. Games and Economic Behavior, Vol. 55, Issue 2, Mini Special Issue: Electronic Market Design, pages 270-296, 2006.

[22] E. Maskin. Implementation and strong Nash-equilibrium. In: J-J. Laffont, ed., Aggregation and revelation of preferences (North-Holland, Amsterdam), 1979.

[23] E. Maskin. Nash equilibrium and welfare optimality. Review of Economic Studies, Vol. 66, pages 23-38, 1999.

[24] E. Maskin and T. Sjostrom. Implementation theory. *Handbook of Social Choice and Welfare*, 1:237–288, 2002.

[25] McSherry, F.; Talwar, K.; , "Mechanism Design via Differential Privacy," Foundations of Computer Science, 2007. FOCS '07. 48th Annual IEEE Symposium on , vol., no., pp.94-103, 21-23 Oct. 2007.

[26] S. Micali and P. Valiant. Resilient Mechanisms for Unrestricted Combinatorial Auctions, 2007. Available at http://people.csail.mit.edu/silvio/Selected Scientific Papers/Mechanism Design.

[27] J. Moore and R. Repullo. Subgame perfect implementation. *Econometrica: Journal of the Econometric Society*, 56(5):1191-1220, 1988.

[28] H. Moulin. Incremental cost sharing: Characterization by coalition strategy-proofness. *Social Choice Welfare*, Vol.16, No.2, pages 279-320, Feb. 1999.

[29] H. Moulin and B. Peleg. Cores of effectivity functions and implementation theory. Journal of Mathematical Economics, Vol. 10, pages 115-145, 1982.

[30] H. Moulin and S. Shenker. Strategyproof Sharing of Submodular Costs: Budget Balance Versus Efficiency. *Economic Theory*, Vol.18, No.3, pages 511-533, Nov. 2001.

[31] R. B. Myerson and M. A. Satterthwaite. Efficient mechanisms for bilateral trading. *Journal of Economic Theory*, 29(2):265 – 281, 1983.

[32] K. Nissim , R. Smorodinsky , M. Tennenholtz, "Approximately Optimal Mechanism Design via Differential Privacy". http://arxiv.org/abs/1004.2888

[33] M. Pal and E. Tardos, Group Strategyproof Mechanisms via Primal-Dual Algorithms. *44th Annual IEEE Symposium on Foundations of Computer Science (FOCS'03)*, pages 584-593, 2003.

[34] S. Suh. Implementation with Coalition Formation: A Complete Characterization. Journal of Mathematical Economics, Vol. 26, pages 409-428, 1996.

[35] W. Vickrey. Counterspeculation, Auctions, and Competitive Sealed Tenders. *The Journal of Finance*, Vol. 16, No. 1, pages 8-37, Mar., 1961.

Appendix

It should be realized that our main notions of root implementation, root implementation with full privacy, and essential root implementation with full privacy, not only apply when it is common knowledge that all players are independent, but are actually simpler. Rather than quantifying over all possible contexts $\mathbb{C}$, it suffices to quantify over all possible *non-collusive* contexts C. Accordingly, we do not find any point in restating all our notions, but find it useful to break our proof into two parts: proving first our result for non-collusive contexts, and then for collusive ones.

A Analysis of Our Mechanism Without Collusion

We first analyze our mechanism $\mathcal{M}$ assuming that it is common knowledge that all players are independent.

Theorem 1′. *For any $\delta \in (0, 1/2)$, $\mathcal{M}$ $(1-\delta)$-root implements the socially optimal correspondence f with full privacy, when it is common knowledge that all players are independent.*

Before we prove Theorem 1′, let us clarify some notations. In the remaining part of this section, when we say "any context C", we mean "any non-collusive n-m-k market context $C = (E, \theta, \mathscr{C})$". Moreover, we call an outcome (A^*, P^*) regular if $P_i^* \in \mathbb{Z} \cap (-2^k, 2^k)$ for each $i \neq 1$, $\sum_i P_i^* = 0$, and $(A^*, P^*) \neq (E, 0^n)$. Notice that (A^*, P^*) is regular if and only if $\mathcal{M}$ does not halt in Steps a or b when player 1 announces (A^*, P^*) in Step 1.

To prove Theorem 1′, it suffices to consider $\delta \in (0, \epsilon/B)$, because for any $\delta \geq \epsilon/B$, we can simply take $\delta' \in (0, \epsilon/B)$, and prove that $\mathcal{M}$, configured under δ' instead of δ, $(1-\delta')$-root implements f with full privacy, which implies that $\mathcal{M}$ $(1-\delta)$-root implements f with full privacy.

We start by proving that $\mathcal{M}$ is root-solvable, as defined in Definition 2. Notice that $\mathcal{M}$ clearly has a single player, player 1, acting at the root R, and $A_r^R = \Omega$. The following two lemmas and a corollary prove that for each $x \in \Omega$, $\mathcal{M}^x$ is CDS. (Notice that when it is common knowledge that all players are independent, all coalitions in $\mathscr{C}$ are singletons.)

Lemma 1. *For any context C and any regular outcome $x = (A^*, P^*)$, in the subgame $\mathcal{M}^x$ conditioned on $c_0 = Heads$, for each player $i \neq 1$, it is strictly dominant to announce YES in Step 2 if $\theta_i(A_i^*) - \theta_i(E_i) - P_i^* \geq 0$, and to announce NO otherwise.*

Proof. Notice that for such an x, $\mathcal{M}$ does not halt in Step a or b, and thus c_0 is flipped. Conditioned on $c_0 = Heads$, player $i \neq 1$ has two strategies: to announce YES or to announce NO in Step 2. It is easy to see that the following facts hold:

(1) No matter what strategies the players use, the final outcome is either $(E, (P_1, -\epsilon, \ldots, -\epsilon))$ with some P_1, or $(A^*, (P_1^* + 2(n-1)\epsilon, P_2^* - 2\epsilon, \ldots, P_n^* - 2\epsilon))$.

(2) Player i's utility is ϵ in the former outcome, and $\theta_i(A_i^*) - \theta_i(E_i) - P_i^* + 2\epsilon$ in the latter. Therefore, his expected utility is $\theta_i(A_i^*) - \theta_i(E_i) - P_i^* + \epsilon$ times the probability that the latter outcome is implemented, plus an extra ϵ.

(3) The probability that the latter outcome is implemented is strictly increasing with Y, the number of players announcing YES. Indeed, if $Y = n-1$ then this probability is 1, otherwise this probability is $\frac{\epsilon Y}{nB}$.

(4) $\theta_i(A_i^*) - \theta_i(E_i) - P_i^* \geq 0$ implies that $\theta_i(A_i^*) - \theta_i(E_i) - P_i^* + \epsilon > 0$, and $\theta_i(A_i^*) - \theta_i(E_i) - P_i^* < 0$ implies that $\theta_i(A_i^*) - \theta_i(E_i) - P_i^* + \epsilon < 0$, since the valuations and the prices are all integers and $\epsilon \in (0, 1)$.

Accordingly, when $\theta_i(A_i^*) - \theta_i(E_i) - P_i^* \geq 0$, player i's expected utility in $\mathcal{M}^x$ conditioned on $c_0 = Heads$ is strictly increasing with the probability that $(A^*, (P_1^* + 2(n-1)\epsilon, P_2^* - 2\epsilon, \ldots, P_n^* - 2\epsilon))$ is implemented, and thus strictly increasing with Y, regardless of the other players' strategies. Therefore announcing YES strictly

dominates announcing NO for player i, since the former always increases the value of Y by 1 compared with the latter.

Symmetrically, when $\theta_i(A_i^*)-\theta_i(E_i)-P_i^* < 0$, player i's expected utility in $\mathcal{M}^x$ conditioned on $c_0 = Heads$ is strictly decreasing with Y, regardless of what the other players announce. Thus announcing NO strictly dominates announcing YES for player i.

In sum, Lemma 1 holds. ■

Lemma 2. *For any context C and any regular outcome $x = (A^*, P^*)$, in the subgame $\mathcal{M}^x$ conditioned on $c_0 = Tails$, for each player $i \neq 1$, it is strictly dominant to announce YES in Step 3 when $w = i$ if $\theta_i(A_i^*) - \theta_i(E_i) - P_i^* \geq 0$, and to announce NO otherwise.*

Proof. Again for such an x, $\mathcal{M}$ does not halt in Step a or b. Conditioned on $c_0 = Tails$, the mechanism goes to $\mathcal{M}'$, and player i has two strategies: to announce YES or to announce NO at the decision node where $w = i$. It is easy to see that the following facts hold, no matter what strategies the other players use:

(1) Conditioned on $w \neq i$, $A_i = \emptyset$ and $P_i = -\epsilon$.

(2) Conditioned on $w = i$, $A_i = A_i^*$ and $P_i = P_i^* - 2\epsilon$ if i announces YES, and $A_i = E_i$ and $P_i = -\epsilon$ otherwise.

Accordingly, player i's expected utility in $\mathcal{M}^x$ conditioned on $c_0 = Tails$ is

$$\tfrac{n-1}{n}(0 - \theta_i(E_i) + \epsilon) + \tfrac{1}{n}(\theta_i(A_i^*) - \theta_i(E_i) - P_i^* + 2\epsilon)$$

if he announces YES, and

$$\tfrac{n-1}{n}(0 - \theta_i(E_i) + \epsilon) + \tfrac{1}{n} \cdot \epsilon$$

otherwise. Notice that the difference between these two utilities is $\frac{1}{n}(\theta_i(A_i^*) - \theta_i(E_i) - P_i^* + \epsilon)$. Similar to the proof of Lemma 1, if $\theta_i(A_i^*) - \theta_i(E_i) - P_i^* \geq 0$ then the former utility is strictly greater than the latter, implying that announcing YES strictly dominates announcing NO; while if $\theta_i(A_i^*) - \theta_i(E_i) - P_i^* < 0$ then the latter is strictly greater than the former, implying that announcing NO strictly dominates announcing YES. Therefore Lemma 2 holds. ■

Lemmas 1 and 2 imply the following corollary.

Corollary 1. *For any outcome x, $\mathcal{M}^x$ is CDS. Moreover, for any context C and any regular outcome $x = (A^*, P^*)$, in the subgame $\mathcal{M}^x$, for each player $i \neq 1$, it is strictly dominant to announce YES both in Step 2 and in Step 3 when $w = i$ if $\theta_i(A_i^*) - \theta_i(E_i) - P_i^* \geq 0$, and to announce NO in both steps otherwise.*

Proof. Notice that if x is regular then $\mathcal{M}^x$ is a probabilistic combination of itself conditioned on $c_0 = Heads$ (which occurs with probability $1 - \delta$) and itself conditioned on $c_0 = Tails$ (which occurs with probability δ). Because the outcome of c_0 is independent from the players' strategies, for each player $i \neq 1$, no matter what the other players do, i's expected utility in $\mathcal{M}^x$ is a fixed probabilistic combination of his expected utilities in the corresponding subgames conditioned on the outcomes of c_0. Accordingly, the strategy which consists of playing the strictly dominant strategies in those two subgames is a strictly dominant strategy in $\mathcal{M}^x$.

For any other x, $\mathcal{M}^x$ is "empty", and thus is considered CDS automatically. □

The next two lemmas prove the existence of the subset A of A_r^R, as defined in property (3) of Definition 2. (Notice that $\mathscr{C}_1 = \{1\}$.)

Lemma 3. *For any context C such that $SW(E) = \max_{A' \in \mathcal{A}} SW(A')$, letting $x = (E, (0, \ldots, 0))$ and $S = \{x\}$, we have that $u_1(s^x_{\mathscr{C}}) > u_1(s^{x'}_{\mathscr{C}})$ for any outcome $x' \notin S$, and $u_i(s^x_{\mathscr{C}}) = 0$ for any player i.*

Proof. It is easy to see that $\mathcal{M}^x$ halts in Step b with final outcome $(A, P) = x$. Accordingly, for each player i, $u_i(s^x_{\mathscr{C}}) = \theta_i(A_i) - \theta_i(E_i) - P_i = \theta_i(E_i) - \theta_i(E_i) = 0$. Therefore to prove property (3) of Definition 2, it remains to show that for any outcome $x' = (A^*, P^*) \neq x$, $u_1(s^x_{\mathscr{C}}) > u_1(s^{x'}_{\mathscr{C}})$, or equivalently, $u_1(s^{x'}_{\mathscr{C}}) < 0$.

To do so, first notice that if $P^*_i \notin \mathbb{Z} \cap (-2^k, 2^k)$ for some player $i \neq 1$, or $\sum_i P^*_i \neq 0$, then $\mathcal{M}^{x'}$ halts in Step a with $A_1 = E_1$ and $P_1 = 1$, implying that $u_1(s^{x'}_{\mathscr{C}}) = -1 < 0$. Below we only consider x' such that $P^*_i \in \mathbb{Z} \cap (-2^k, 2^k)$ for all players $i \neq 1$ and $\sum_i P^*_i = 0$, and we distinguish two cases.

Case 1: $\theta_i(A^*_i) - \theta_i(E_i) - P^*_i \geq 0$ for each player $i \neq 1$.

In this case, in the execution of $s^{x'}_{\mathscr{C}}$, according to Corollary 1, conditioned on $c_0 = Heads$, all players announce YES in Step 2, and thus the outcome $(A^*, (P^*_1 + 2(n-1)\epsilon, P^*_2 - 2\epsilon, \ldots, P^*_n - 2\epsilon))$ is implemented with probability 1. Accordingly,

$$\begin{aligned} u_1(s^{x'}_{\mathscr{C}}|c_0 = Heads) &= \theta_1(A^*_1) - \theta_1(E_1) - P^*_1 - 2(n-1)\epsilon = \theta_1(A^*_1) - \theta_1(E_1) + \sum_{i\neq 1} P^*_i - 2(n-1)\epsilon \\ &\leq \theta_1(A^*_1) - \theta_1(E_1) + \sum_{i\neq 1}(\theta_i(A^*_i) - \theta_i(E_i)) - 2(n-1)\epsilon \\ &= \sum_i \theta_i(A^*_i) - \sum_i \theta_i(E_i) - 2(n-1)\epsilon \\ &= SW(A^*) - \max_{A'\in\mathcal{A}} SW(A') - 2(n-1)\epsilon \leq -2(n-1)\epsilon, \end{aligned}$$

where the first inequality is by the hypothesis of Case 1.

On the other hand, conditioned on $c_0 = Tails$ and $w = 1$, we have that $A_1 = A^*_1$ and $P_1 = (n-1)\epsilon$, and thus

$$u_1(s^{x'}_{\mathscr{C}}|c_0 = Tails, w = 1) = \theta_1(A^*_1) - \theta_1(E_1) - (n-1)\epsilon.$$

While according to Corollary 1, for each $i \neq 1$, conditioned on $c_0 = Tails$ and $w = i$, player i announces YES in Step 3, $A_1 = \emptyset$, $P_1 = -P^*_i + n\epsilon$, and thus

$$u_1(s^{x'}_{\mathscr{C}}|c_0 = Tails, w = i) = -\theta_1(E_1) + P^*_i - n\epsilon.$$

In sum,

$$\begin{aligned} u_1(s^{x'}_{\mathscr{C}}|c_0 = Tails) &= \frac{1}{n}(\theta_1(A^*_1) - \theta_1(E_1) - (n-1)\epsilon) + \frac{1}{n}\sum_{i\neq 1}(-\theta_1(E_1) + P^*_i - n\epsilon) \\ &= \frac{\theta_1(A^*_1)}{n} - \theta_1(E_1) - \frac{(n-1)\epsilon}{n} + \frac{\sum_{i\neq 1} P^*_i}{n} - (n-1)\epsilon \\ &\leq \frac{\theta_1(A^*_1)}{n} - \theta_1(E_1) - \frac{(n-1)\epsilon}{n} + \frac{\sum_{i\neq 1}(\theta_i(A^*_i) - \theta_i(E_i))}{n} - (n-1)\epsilon \\ &= \frac{\sum_i \theta_i(A^*_i) - \sum_i \theta_i(E_i)}{n} - \frac{(n-1)\theta_1(E_1)}{n} - \frac{(n-1)\epsilon}{n} - (n-1)\epsilon \\ &< -\frac{(n-1)\theta_1(E_1)}{n} < B, \end{aligned}$$

where again the first inequality is by the hypothesis of Case 1, the second one is because $\sum_i \theta_i(E_i) = \max_{A'\in\mathcal{A}} SW(A') \geq \sum_i \theta_i(A^*_i)$, and the last one is because $\theta_1(E_1) > -B$.

Accordingly,

$$\begin{aligned} u_1(s^{x'}_{\mathscr{C}}) &= (1-\delta)u_1(s^{x'}_{\mathscr{C}}|c_0 = Heads) + \delta u_1(s^{x'}_{\mathscr{C}}|c_0 = Tails) < -2(1-\delta)(n-1)\epsilon + \delta B \\ &< -2(1-\delta)(n-1)\epsilon + \epsilon < -\frac{3}{2}\epsilon + \epsilon < 0, \end{aligned}$$

where the inequalities are because $0 < \delta < \epsilon/B < \frac{1}{5nB} < 1/4$ and $n \geq 2$.

Case 2: $\theta_i(A_i^*) - \theta_i(E_i) - P_i^* < 0$ for some player $i \neq 1$.

In this case, at least one player, i.e., player i, announces NO in both Step 2 and Step 3 when he is player w. Accordingly, conditioned on $c_0 = Heads$, we have that:

- $Y < n-1$;
- $A_1 = E_1$ and $P_1 = 2n(n-Y)B + (n-1)\epsilon$ with probability $1-\epsilon$;
- $A_1 = A_1^*$ and $P_1 = P_1^* + 2(n-1)\epsilon$ with probability $\frac{\epsilon Y}{nB}$; and
- $A_1 = E_1$ and $P_1 = (n-1)\epsilon$ with probability $\epsilon(1 - \frac{Y}{nB})$.

Thus

$$\begin{aligned} & u_1(s_{\mathscr{C}}^{x'} | c_0 = Heads) \\ = \; & -(1-\epsilon)(2n(n-Y)B + (n-1)\epsilon) + \frac{\epsilon Y}{nB}(\theta_1(A_1^*) - \theta_1(E_1) - P_1^* - 2(n-1)\epsilon) - \epsilon(1 - \frac{Y}{nB})(n-1)\epsilon \\ < \; & -(1-\epsilon)B + \frac{\epsilon Y}{nB} \cdot 4B < -(1-\epsilon)B + 4\epsilon < 0, \end{aligned}$$

where the first inequality is because $Y < n-1$, $\theta_1(A_1^*)$ and $\theta_1(E_1)$ are in $(-B, B)$, and $P_1^* \in (-2B, 2B)$, and the last one is because $B \geq 1$ and $\epsilon < 1/5$.

Conditioned on $c_0 = Tails$ and $w = i$, we have that $A_1 = \emptyset$ and $P_1 = \frac{n^2 B}{\delta} + (n-1)\epsilon$. Recall from Case 1 the utility of player 1 conditioned on $w = 1$ and that conditioned on w being some player who announces YES in Step 3, we have that

$$\begin{aligned} u_1(s_{\mathscr{C}}^{x'} | c_0 = Tails) \; = \; & \frac{1}{n}(\theta_1(A_1^*) - \theta_1(E_1) - (n-1)\epsilon) + \frac{1}{n} \sum_{j:j \text{ announces YES}} (-\theta_1(E_1) + P_j^* - n\epsilon) \\ & + \frac{1}{n} \sum_{j:j \text{ announces NO}} (-\theta_1(E_1) - \frac{n^2 B}{\delta} - (n-1)\epsilon) \\ < \; & \frac{2B}{n} + \frac{3(n-2)B}{n} - \frac{\frac{n^2 B}{\delta} - B}{n} \\ = \; & \frac{B}{n}(3n - 3 - \frac{n^2}{\delta}) < 0. \end{aligned}$$

In sum, $u_1(s_{\mathscr{C}}^{x'}) = (1-\delta)u_1(s_{\mathscr{C}}^{x'} | c_0 = Heads) + \delta u_1(s_{\mathscr{C}}^{x'} | c_0 = Tails) < 0$ in this case.

Combining the two cases, we have that $u_1(s_{\mathscr{C}}^{x'}) < 0$ for any $x' \neq x$, and thus Lemma 3 holds. ■

We use the following definition to simplify the statement and the proof of the next lemma.

Definition 7. *Given a context C, an outcome (A, P) is* truthful *if $SW(A) = \max_{A' \in \mathcal{A}} SW(A')$, $\sum_i P_i = 0$, and $P_i = \theta_i(A_i) - \theta_i(E_i)$ for each $i \neq 1$.*

It is easy to see that the set of truthful outcomes is independent of $\mathscr{C} \setminus \{\mathscr{C}_1\}$.

Lemma 4. *For any context C such that $SW(E) < \max_{A' \in \mathcal{A}} SW(A')$, letting S be the set of truthful outcomes, we have that for any $x \in S$ and $x' \in \Omega$: (1) $u_1(s_{\mathscr{C}}^x) = u_1(s_{\mathscr{C}}^{x'})$ whenever $x' \in S$; (2) $u_1(s_{\mathscr{C}}^x) > u_1(s_{\mathscr{C}}^{x'})$ whenever $x' \notin S$; and (3) $u_i(s_{\mathscr{C}}^x) \geq 0$ for any player i.*

Proof. Write x as (A^*, P^*), below we consider the execution of $s_{\mathscr{C}}^x$, in order to compute $u_i(s_{\mathscr{C}}^x)$ for each player i. By the definition of truthful outcomes, $\mathcal{M}$ does not halt in Step a. Because $SW(E) < \max_{A' \in \mathcal{A}} SW(A')$, we have that $x \neq (E, (0, \ldots, 0))$, and thus $\mathcal{M}$ does not halt in Step b.

By Corollary 1, every player $i \neq 1$ announces YES in both Step 2 and Step 3 when $w = i$. Accordingly, conditioned on $c_0 = Heads$, we have that $Y = n-1$, $A = A^*$, $P_1 = P_1^* + 2(n-1)\epsilon$, and $P_i = P_i^* - 2\epsilon$ for

each $i \neq 1$. Therefore

$$\begin{aligned} u_1(s^x_{\mathscr{C}}|c_0 = Heads) &= \theta_1(A^*_1) - \theta_1(E_1) - P^*_1 - 2(n-1)\epsilon = \theta_1(A^*_1) - \theta_1(E_1) + \sum_{i\neq 1} P^*_i - 2(n-1)\epsilon \\ &= \theta_1(A^*_1) - \theta_1(E_1) + \sum_{i\neq 1}(\theta_i(A^*_i) - \theta_i(E_i)) - 2(n-1)\epsilon \\ &= \sum_i \theta_i(A^*_i) - \sum_i \theta_i(E_i) - 2(n-1)\epsilon = \max_{A'\in\mathcal{A}} \sum_i SW(A') - SW(E) - 2(n-1)\epsilon, \end{aligned}$$

and for each player $i \neq 1$,

$$u_i(s^x_{\mathscr{C}}|c_0 = Heads) = \theta_i(A^*_i) - \theta_i(E_i) - P^*_i + 2\epsilon = 2\epsilon.$$

Conditioned on $c_0 = Tails$, similarly to what we have seen in the proof of Lemma 3, we have that

$$\begin{aligned} u_1(s^x_{\mathscr{C}}|c_0 = Tails) &= \frac{1}{n}(\theta_1(A^*_1) - \theta_1(E_1) - (n-1)\epsilon) + \frac{1}{n}\sum_{i\neq 1}(-\theta_1(E_1) + P^*_i - n\epsilon) \\ &= \frac{\theta_1(A^*_1)}{n} - \theta_1(E_1) - \frac{(n-1)\epsilon}{n} + \frac{\sum_{i\neq 1} P^*_i}{n} - (n-1)\epsilon \\ &= \frac{\theta_1(A^*_1)}{n} - \theta_1(E_1) - \frac{(n-1)\epsilon}{n} + \frac{\sum_{i\neq 1}(\theta_i(A^*_i) - \theta_i(E_i))}{n} - (n-1)\epsilon \\ &= \frac{\sum_i \theta_i(A^*_i) - \sum_i \theta_i(E_i)}{n} - \frac{(n-1)\theta_1(E_1)}{n} - \frac{(n-1)\epsilon}{n} - (n-1)\epsilon \\ &= \frac{\max_{A'\in\mathcal{A}} SW(A') - SW(E)}{n} - \frac{(n-1)\theta_1(E_1)}{n} - \frac{(n^2-1)\epsilon}{n}; \end{aligned}$$

and similar to what we have seen in the proof of Lemma 2, we have that for each $i \neq 1$,

$$u_i(s^x_{\mathscr{C}}|c_0 = Tails) = \frac{n-1}{n}(-\theta_i(E_i) + \epsilon) + \frac{1}{n}(\theta_i(A^*_i) - \theta_i(E_i) - P^*_i + 2\epsilon) = -\frac{(n-1)\theta_i(E_i)}{n} + \frac{(n+1)\epsilon}{n}.$$

In sum, we have that

$$\begin{aligned} u_1(s^x_{\mathscr{C}}) &= (1-\delta)u_1(s^x_{\mathscr{C}}|c_0 = Heads) + \delta u_1(s^x_{\mathscr{C}}|c_0 = Tails) \\ &= (1-\delta)\left(\max_{A'\in\mathcal{A}} SW(A') - SW(E) - 2(n-1)\epsilon\right) \\ &\quad +\delta\left(\frac{\max_{A'\in\mathcal{A}} SW(A') - SW(E)}{n} - \frac{(n-1)\theta_1(E_1)}{n} - \frac{(n^2-1)\epsilon}{n}\right), \end{aligned}$$

and that for each $i \neq 1$,

$$u_i(s^x_{\mathscr{C}}) = (1-\delta)2\epsilon + \delta\left(-\frac{(n-1)\theta_i(E_i)}{n} + \frac{(n+1)\epsilon}{n}\right).$$

We can derive two conclusions from the above two equations. On one hand, it is easy to see that $u_1(s^x_{\mathscr{C}})$ only depends on the fact that (A^*, P^*) is truthful, and nothing else. Since x can be any truthful outcome, we have that for any $x' \in S$, $u_1(s^{x'}_{\mathscr{C}}) = u_1(s^x_{\mathscr{C}})$, and property (1) of Lemma 4 holds. On the other hand, because E does not maximize social welfare, we have that $\max_{A'\in\mathcal{A}} SW(A') - SW(E) \geq 1$. Combining with the fact that $\epsilon \in (0, 1/(5n))$, $\delta \in (0, \epsilon/B)$, and that a player's valuation on any subset of the goods is in $(-B, B)$, we have that

$$\begin{aligned} u_1(s^x_{\mathscr{C}}) &> (1-\delta)(1-2(n-1)\epsilon) - \delta\left(\frac{(n-1)\theta_1(E_1)}{n} + \frac{(n^2-1)\epsilon}{n}\right) \\ &> (1-\delta)(1-2(n-1)\epsilon) - \frac{\epsilon}{B}(B + (n+1)\epsilon) > 1 - \delta - 2(n-1)\epsilon - 2\epsilon \\ &> 1 - (2n+1)\epsilon > 0, \end{aligned}$$

and that for each $i \neq 1$,

$$u_i(s^x_{\mathscr{C}}) > 2(1-\delta)\epsilon - \frac{\epsilon}{B} \cdot \frac{n-1}{n} \cdot B = (1 - 2\delta + \frac{1}{n})\epsilon > 0.$$

Accordingly, property (3) of Lemma 4 holds.

It remains to show that $u_1(s^x_{\mathscr{C}}) > u_1(s^{x'}_{\mathscr{C}})$ whenever $x' \notin S$. To do so, writing x' as $(\widehat{A}, \widehat{P})$, we consider the execution of $s^{x'}_{\mathscr{C}}$. If x' causes $\mathcal{M}$ to halt in Step a, then $u_1(s^{x'}_{\mathscr{C}}) = -1 < 0 < u_1(s^x_{\mathscr{C}})$. If $x' = (E, (0, \ldots, 0))$ and $\mathcal{M}$ halts in Step b, then $u_1(s^{x'}_{\mathscr{C}}) = 0 < u_1(s^x_{\mathscr{C}})$. Below we only consider x' such that $\mathcal{M}$ does not halt in Step a or b, and we distinguish three cases.

Case 1: $\theta_i(\widehat{A}_i) - \theta_i(E_i) - \widehat{P}_i < 0$ for some player $i \neq 1$.

In this case, the analysis is very similar to Case 2 of Lemma 3, that is, player i announces NO in both Step 2 and Step 3 when $w = i$, and player 1 is punished heavily both when $c_0 = Heads$ and when $c_0 = Tails$ and $w = i$. Using similar formulas, we have that $u_1(s^{x'}_{\mathscr{C}}) < 0$ in this case, and thus $< u_1(s^x_{\mathscr{C}})$.

Case 2: $\theta_i(\widehat{A}_i) - \theta_i(E_i) - \widehat{P}_i \geq 0$ for each $i \neq 1$, and the inequality is strict for some player.

In this case, by Corollary 1, all players announce YES in both Step 2 and Step 3. Similar to Case 1 of Lemma 3, we have that

$$\begin{aligned}
u_1(s^{x'}_{\mathscr{C}}|c_0 = Heads) &= \theta_1(\widehat{A}_1) - \theta_1(E_1) - \widehat{P}_1 - 2(n-1)\epsilon = \theta_1(\widehat{A}_1) - \theta_1(E_1) + \sum_{i\neq 1} \widehat{P}_i - 2(n-1)\epsilon \\
&< \theta_1(\widehat{A}_1) - \theta_1(E_1) + \sum_{i\neq 1}(\theta_i(\widehat{A}_i) - \theta_i(E_i)) - 2(n-1)\epsilon \\
&= \sum_i \theta_i(\widehat{A}_i) - \sum_i \theta_i(E_i) - 2(n-1)\epsilon \\
&\leq \max_{A' \in \mathcal{A}} SW(A') - SW(E) - 2(n-1)\epsilon = u_1(s^x_{\mathscr{C}}|c_0 = Heads),
\end{aligned}$$

and

$$\begin{aligned}
u_1(s^{x'}_{\mathscr{C}}|c_0 = Tails) &= \frac{1}{n}(\theta_1(\widehat{A}_1) - \theta_1(E_1) - (n-1)\epsilon) + \frac{1}{n}\sum_{i\neq 1}(-\theta_1(E_1) + \widehat{P}_i - n\epsilon) \\
&= \frac{\theta_1(\widehat{A}_1)}{n} - \theta_1(E_1) - \frac{(n-1)\epsilon}{n} + \frac{\sum_{i\neq 1}\widehat{P}_i}{n} - (n-1)\epsilon \\
&< \frac{\theta_1(\widehat{A}_1)}{n} - \theta_1(E_1) - \frac{(n-1)\epsilon}{n} + \frac{\sum_{i\neq 1}(\theta_i(\widehat{A}_i) - \theta_i(E_i))}{n} - (n-1)\epsilon \\
&= \frac{\sum_i \theta_i(\widehat{A}_i) - \sum_i \theta_i(E_i)}{n} - \frac{(n-1)\theta_1(E_1)}{n} - \frac{(n-1)\epsilon}{n} - (n-1)\epsilon \\
&\leq \frac{\max_{A' \in \mathcal{A}} SW(A') - SW(E)}{n} - \frac{(n-1)\theta_1(E_1)}{n} - \frac{(n^2-1)\epsilon}{n} \\
&= u_1(s^x_{\mathscr{C}}|c_0 = Tails).
\end{aligned}$$

Accordingly, we have that $u_1(s^{x'}_{\mathscr{C}}) = (1-\delta)u_1(s^{x'}_{\mathscr{C}}|c_0 = Heads) + \delta u_1(s^{x'}_{\mathscr{C}}|c_0 = Tails) < (1-\delta)u_1(s^x_{\mathscr{C}}|c_0 = Heads) + \delta u_1(s^x_{\mathscr{C}}|c_0 = Tails) = u_1(s^x_{\mathscr{C}})$.

Case 3: $\theta_i(\widehat{A}_i) - \theta_i(E_i) = \widehat{P}_i$ for each player $i \neq 1$, and $SW(\widehat{A}) < \max_{A' \in \mathcal{A}} SW(A')$.

This case is very similar to Case 2, and all players announce YES in both Step 2 and Step 3. As before, the utility of player 1 solely depends on the difference between the social welfare of the outcome announced in Step 1 and that of E, and is strictly increasing with this difference. Since the social welfare of E is fixed, the utility of player 1 is strictly increasing with the social welfare of the outcome announced in Step 1. Because in the execution of $s^{x'}_{\mathscr{C}}$ player 1 announces $(\widehat{A}, \widehat{P})$ which does not maximize social welfare by hypothesis, and in the execution of $s^x_{\mathscr{C}}$ player 1 announces (A^*, P^*) which maximizes social welfare, $u_1(s^{x'}_{\mathscr{C}}) < u_1(s^x_{\mathscr{C}})$.

Because the above three cases have exhausted all possibilities for x' being not truthful, property (3) of Lemma 4 holds, and so does Lemma 4. ■

The above lemmas and corollary have proved that $\mathcal{M}$ is root-solvable. The next lemma proves that $\mathcal{M}$ root-implements the socially optimal correspondence f with probability $\geq 1-\delta$, that is, the second property of Definition 6.

Lemma 5. *For any context C and any root-profile s of $(C,\mathcal{M})$, $\mathcal{M}(s)$ assigns at least $1-\delta$ of its total probability mass to a single terminal node, and $\mathcal{M}(s)^\delta \in f(\theta)$.*

Proof. If C is such that $SW(E) = \max_{A'\in\mathcal{A}} SW(A')$, then by Lemma 3, the only root-profile of $(C,\mathcal{M})$ is $s^x_{\mathscr{C}}$ where $x = (E,(0,\ldots,0))$. In the execution of $s^x_{\mathscr{C}}$, $\mathcal{M}$ halts in Step b, assigning probability 1 to the corresponding terminal node, with $\mathcal{M}(s^x_{\mathscr{C}})^\delta = (E,(0,\ldots,0)) \in f(\theta)$.

Otherwise, by Lemma 4, the set of root-profiles of $(C,\mathcal{M})$ is $S = \{s^x_{\mathscr{C}} : x \text{ is truthful}\}$. For any truthful outcome $x = (A^*,P^*)$, in the execution of $s^x_{\mathscr{C}}$, $\mathcal{M}$ does not halt in Step a or b, and c_0 is flipped. Conditioned on $c_0 = Heads$, which occurs with probability $1-\delta$, every player announces YES in Step 2, and $\mathcal{M}$ halts in Step d. Therefore $\mathcal{M}$ assigns probability $1-\delta$ to the terminal node corresponding to Step d, with final outcome $\mathcal{M}(s^x_{\mathscr{C}})^\delta = (A^*,(P^*_1+2(n-1)\epsilon, P^*_2-2\epsilon,\ldots,P^*_n-2\epsilon))$. Because $SW(A^*) = \max_{A'\in\mathcal{A}} SW(A')$ and $P^*_1+2(n-1)\epsilon+\sum_{i\neq 1}(P^*_i-2\epsilon) = \sum_i P^*_i + 2(n-1)\epsilon - 2(n-1)\epsilon = \sum_i P^*_i = 0$, we have that $\mathcal{M}(s^x_C)^\delta \in f(\theta)$.

In sum, Lemma 5 holds. ■

Finally we discuss the privacy of $\mathcal{M}$, and prove the third and last property of Definition 6.

Lemma 6. *For all contexts $C = (E,\theta,\mathscr{C})$ and $C' = (E,\theta',\mathscr{C}')$, all root-profiles s of $(C,\mathcal{M})$, and all root-profiles s' of $(C',\mathcal{M})$, $\mathcal{M}(s)^\delta = \mathcal{M}(s')^\delta \Rightarrow \mathcal{M}[s]^\delta = \mathcal{M}[s']^\delta$.*

Proof. Following the proof of Lemma 5, we have that $\mathcal{M}(s)^\delta = (E,(0,\ldots,0))$ if and only if $\sum_i \theta_i(E_i) = \max_{A'\in\mathcal{A}} \sum_i \theta_i(A'_i)$, and $\mathcal{M}(s')^\delta = (E,(0,\ldots,0))$ if and only if $\sum_i \theta'_i(E_i) = \max_{A'\in\mathcal{A}} \sum_i \theta'_i(A'_i)$. Accordingly, if $\mathcal{M}(s)^\delta = \mathcal{M}(s')^\delta = (E,(0,\ldots,0))$, then both $\mathcal{M}[s]^\delta$ and $\mathcal{M}[s']^\delta$ are such that player 1 announces $(E,(0,\ldots,0))$ in Step 1, and $\mathcal{M}$ halts in Step b. Therefore we have that $\mathcal{M}[s]^\delta = \mathcal{M}[s']^\delta$.

Again following the proof of Lemma 5, we have that for any outcome $(A,P) \neq (E,(0,\ldots,0))$, $\mathcal{M}(s)^\delta = (A,P)$ if and only if $\sum_i \theta_i(E_i) < \max_{A'\in\mathcal{A}} \sum_i \theta_i(A'_i)$, and $(A,P) = (A^*,(P^*_1+2(n-1)\epsilon, P^*_2-2\epsilon,\ldots,P^*_n-2\epsilon))$ for some outcome (A^*,P^*) which is truthful for C; and the same thing can be said for $\mathcal{M}(s')^\delta$. Accordingly, if $\mathcal{M}(s)^\delta = \mathcal{M}(s')^\delta = (A^*,(P^*_1+2(n-1)\epsilon, P^*_2-2\epsilon,\ldots,P^*_n-2\epsilon))$ for some (A^*,P^*), then both $\mathcal{M}[s]^\delta$ and $\mathcal{M}[s']^\delta$ are such that player 1 announces (A^*,P^*) in Step 1, $c_0 = Heads$, every player announces YES in Step 2, and $\mathcal{M}$ halts in Step d. Therefore we again have that $\mathcal{M}[s]^\delta = \mathcal{M}[s']^\delta$.

In sum, Lemma 6 holds. ■

Combining all the lemmas and the corollary above, we can conclude that Theorem 1′ holds. *Q.E.D.*

B Analysis of Our Mechanism With Collusion

We now analyze our mechanism with the existence of collusion.

Theorem 1″. *For any $\delta \in (0,1/2)$, $\mathcal{M}$ $(1-\delta)$-root implements the socially optimal correspondence f with full privacy.*

The proof of Theorem 1″ is very similar to that of Theorem 1′, and thus most repeated details have been omitted. In the remaining part of this section, when we say "any context $\mathbb{C}$", we mean "any n-m-k market context $\mathbb{C} = (E,\theta,\mathscr{C})$". Notice that the coalition acting at the root is $\mathscr{C}_1$. Again it suffices to consider $\delta \in (0,\epsilon/B)$, and we use the notion of regular outcomes to simplify the analysis. Recall that an outcome (A^*,P^*) is regular if $P^*_i \in \mathbb{Z} \cap (-2^k,2^k)$ for each $i \neq 1$, $\sum_i P^*_i = 0$, and $(A^*,P^*) \neq (E,0^n)$.

Lemma 7. *For any context $\mathbb{C}$ and any regular outcome $x = (A^*,P^*)$, in the subgame $\mathcal{M}^x$ conditioned on $c_0 = Heads$, for any coalition $\mathcal{C} \not\ni 1$, it is strictly dominant for $\mathcal{C}$'s members to all announce YES in Step 2 if $\sum_{i\in\mathcal{C}}(\theta_i(A^*_i) - \theta_i(E_i) - P^*_i) \geq 0$, and to all announce NO otherwise.*

Proof. Similar to that of Lemma 1. ■

Lemma 8. *For any context $\mathbb{C}$ and any regular outcome $x = (A^*, P^*)$, in the subgame $\mathcal{M}^x$ conditioned on $c_0 = Tails$, for any coalition $\mathcal{C} \not\ni 1$, the following strategy is strictly dominant: for any player $i \in \mathcal{C}$, when $w = i$, i announces YES in Step 3 if $\theta_i(A_i^*) - \theta_i(E_i) - P_i^* \geq 0$, and announces NO otherwise.*

Proof. Similar to that of Lemma 2. ■

Notice that according to Lemma 8, when $w = i$, the coalition $\mathcal{C}$ decides whether i should announce YES or NO solely depending on player i's valuation on the goods he gets and player i's price, instead of $\mathcal{C}$'s members' total valuation and total price. This is because according to $\mathcal{M}$, when $c_0 = Tails$, only player w can get some goods —A_w^* if he announces YES and E_w otherwise, and all of the goods not allocated to player w are destroyed.

The following lemma is the only one which is completely new to this section, and thus needs careful proof.

Lemma 9. *For any context $\mathbb{C}$ and any regular outcome $x = (A^*, P^*)$, in the subgame $\mathcal{M}^x$, it is strictly dominant for $\mathscr{C}_1$'s members other than player 1 to all announce YES in both Step 2 and Step 3 when $w \in \mathscr{C}_1$.*

Proof. First of all, for any such outcome x, $\mathcal{M}$ does not halt in Step a or b, and thus $\mathcal{M}^x$ is not "empty". If player 1 is independent, that is, $\mathscr{C}_1 = \{1\}$, then $\mathscr{C}_1$'s strategy set in $\mathcal{M}^x$ is empty, as player 1 is never asked to take actions in this subgame. Below we focus on the case when $\mathscr{C}_1 \neq \{1\}$. The key point is that, no matter what the outcome of c_0 is, whenever some player announces NO, player 1 is charged a big fine. The more the players who announce NO, the bigger the fine is, and the increase of the fine cancels out any possible gain that player 1 and his colluders may get from the remaining part of the final outcome, and leaves them with negative utility. Let us be more formal.

Notice that for each player $i \neq 1$, i has the following actions available to him in $\mathcal{M}^x$: to announce YES or NO in Step 2 conditioned on $c_0 = Heads$, and to announce YES or NO in Step 3 conditioned on $c_0 = Tails$ and $w = i$. Therefore each player $i \neq 1$ has 4 strategies, and the coalition $\mathscr{C}_1$ has $4^{|\mathscr{C}_1|-1}$ strategies in $\mathcal{M}^x$. Let $s_{\mathscr{C}_1}$ be the strategy of $\mathscr{C}_1$ such that everybody in $\mathscr{C}_1 \setminus \{1\}$ announces YES in both Step 2 and Step 3. For any other strategy $s'_{\mathscr{C}_1}$ of $\mathscr{C}_1$ and any strategy subprofile $t_{-\mathscr{C}_1}$ for players in $-\mathscr{C}_1$ in $\mathcal{M}^x$, we are going to prove that

$$u^x_{\mathscr{C}_1}(s_{\mathscr{C}_1} \sqcup t_{-\mathscr{C}_1}) > u^x_{\mathscr{C}_1}(s'_{\mathscr{C}_1} \sqcup t_{-\mathscr{C}_1}),$$

where $u^x_{\mathscr{C}_1}$ is the expected utility of $\mathscr{C}_1$ in $\mathcal{M}^x$. We distinguish two cases.

Case 1. $s'_{\mathscr{C}_1}$ is such that every player in $\mathscr{C}_1 \setminus \{1\}$ announces YES in Step 2.

In this case, $u^x_{\mathscr{C}_1}(s_{\mathscr{C}_1} \sqcup t_{-\mathscr{C}_1} | c_0 = Heads) = u^x_{\mathscr{C}_1}(s'_{\mathscr{C}_1} \sqcup t_{-\mathscr{C}_1} | c_0 = Heads)$, and it suffices to show that

$$u^x_{\mathscr{C}_1}(s_{\mathscr{C}_1} \sqcup t_{-\mathscr{C}_1} | c_0 = Tails) > u^x_{\mathscr{C}_1}(s'_{\mathscr{C}_1} \sqcup t_{-\mathscr{C}_1} | c_0 = Tails).$$

Because $s'_{\mathscr{C}_1}$ is different from $s_{\mathscr{C}_1}$, it must be different from the latter conditioned on $c_0 = Tails$, and thus must have some player announcing NO in Step 3. Letting $D \subseteq \mathscr{C}_1 \setminus \{1\}$ be the set of such players, we have that

$$u^x_{\mathscr{C}_1}(s_{\mathscr{C}_1} \sqcup t_{-\mathscr{C}_1} | c_0 = Tails) = \frac{1}{n} \sum_{i \in D} u^x_{\mathscr{C}_1}(s_{\mathscr{C}_1} \sqcup t_{-\mathscr{C}_1} | c_0 = Tails, w = i) + \frac{1}{n} \sum_{i \notin D} u^x_{\mathscr{C}_1}(s_{\mathscr{C}_1} \sqcup t_{-\mathscr{C}_1} | c_0 = Tails, w = i),$$

and that

$$u^x_{\mathscr{C}_1}(s'_{\mathscr{C}_1} \sqcup t_{-\mathscr{C}_1} | c_0 = Tails) = \frac{1}{n} \sum_{i \in D} u^x_{\mathscr{C}_1}(s'_{\mathscr{C}_1} \sqcup t_{-\mathscr{C}_1} | c_0 = Tails, w = i) + \frac{1}{n} \sum_{i \notin D} u^x_{\mathscr{C}_1}(s'_{\mathscr{C}_1} \sqcup t_{-\mathscr{C}_1} | c_0 = Tails, w = i).$$

For each player $i \notin D$, conditioned on $c_0 = Tails$ and $w = i$, i's strategies are the same in $s_{\mathscr{C}_1} \sqcup t_{-\mathscr{C}_1}$ and in $s'_{\mathscr{C}_1} \sqcup t_{-\mathscr{C}_1}$, so are the final outcomes. Thus

$$\frac{1}{n} \sum_{i \notin D} u^x_{\mathscr{C}_1}(s_{\mathscr{C}_1} \sqcup t_{-\mathscr{C}_1} | c_0 = Tails, w = i) = \frac{1}{n} \sum_{i \notin D} u^x_{\mathscr{C}_1}(s'_{\mathscr{C}_1} \sqcup t_{-\mathscr{C}_1} | c_0 = Tails, w = i),$$

and it suffices to show that for each player $i \in D$,

$$u^x_{\mathscr{C}_1}(s_{\mathscr{C}_1} \sqcup t_{-\mathscr{C}_1} | c_0 = Tails, w = i) > u^x_{\mathscr{C}_1}(s'_{\mathscr{C}_1} \sqcup t_{-\mathscr{C}_1} | c_0 = Tails, w = i).$$

For any such player i, conditioned on $c_0 = Tails$ and $w = i$, in $s_{\mathscr{C}_1} \sqcup t_{-\mathscr{C}_1}$ the final outcome is such that $A_i = A_i^*$, $P_i = P_i^* - 2\epsilon$, $P_1 = -P_i^* + n\epsilon$, $A_j = \emptyset$ for any $j \neq i$, and $P_j = -\epsilon$ for any $j \notin \{1, i\}$. Accordingly,

$$\begin{aligned} & u^x_{\mathscr{C}_1}(s_{\mathscr{C}_1} \sqcup t_{-\mathscr{C}_1} | c_0 = Tails, w = i) \\ = \; & (-\theta_1(E_1) + P_i^* - n\epsilon) + (\theta_i(A_i^*) - \theta_i(E_i) - P_i^* + 2\epsilon) + \sum_{j \in \mathscr{C}_1 \setminus \{1,i\}} (-\theta_j(E_j) + \epsilon) \\ = \; & \theta_i(A_i^*) - \sum_{j \in \mathscr{C}_1} \theta_j(E_j) - (n - |\mathscr{C}_1|)\epsilon. \end{aligned}$$

On the other hand, in $s'_{\mathscr{C}_1} \sqcup t_{-\mathscr{C}_1}$ the final outcome is such that $A_i = E_i$, $A_j = \emptyset$ for any $j \neq i$, $P_1 = \frac{n^2 B}{\delta} + (n-1)\epsilon$, $P_j = -\epsilon$ for any $j \neq 1$. Accordingly,

$$\begin{aligned} & u^x_{\mathscr{C}_1}(s'_{\mathscr{C}_1} \sqcup t_{-\mathscr{C}_1} | c_0 = Tails, w = i) \\ = \; & \left(-\theta_1(E_1) - \frac{n^2 B}{\delta} - (n-1)\epsilon\right) + (\theta_i(E_i) - \theta_i(E_i) + \epsilon) + \sum_{j \in \mathscr{C}_1 \setminus \{1,i\}} (-\theta_j(E_j) + \epsilon) \\ = \; & \theta_i(E_i) - \sum_{j \in \mathscr{C}_1} \theta_j(E_j) - \frac{n^2 B}{\delta} - (n - |\mathscr{C}_1|)\epsilon. \end{aligned}$$

Because $\theta_i(A_i^*) > -B > B - \frac{n^2 B}{\delta} > \theta_i(E_i) - \frac{n^2 B}{\delta}$, we have that $u^x_{\mathscr{C}_1}(s_{\mathscr{C}_1} \sqcup t_{-\mathscr{C}_1} | c_0 = Tails, w = i) > u^x_{\mathscr{C}_1}(s'_{\mathscr{C}_1} \sqcup t_{-\mathscr{C}_1} | c_0 = Tails, w = i)$ for each $i \in D$, as desired.

Case 2. $s'_{\mathscr{C}_1}$ is such that some player $i \in \mathscr{C}_1 \setminus \{1\}$ announces NO in Step 2.

Following Case 1, we have that $u^x_{\mathscr{C}_1}(s_{\mathscr{C}_1} \sqcup t_{-\mathscr{C}_1} | c_0 = Tails) \geq u^x_{\mathscr{C}_1}(s'_{\mathscr{C}_1} \sqcup t_{-\mathscr{C}_1} | c_0 = Tails)$, and thus it suffices to show that

$$u^x_{\mathscr{C}_1}(s_{\mathscr{C}_1} \sqcup t_{-\mathscr{C}_1} | c_0 = Heads) > u^x_{\mathscr{C}_1}(s'_{\mathscr{C}_1} \sqcup t_{-\mathscr{C}_1} | c_0 = Heads).$$

According to $s'_{\mathscr{C}_1} \sqcup t_{-\mathscr{C}_1}$, conditioned on $c_0 = Heads$, letting Y' be the number of players announcing YES in Step 2, we have that $Y' \leq n - 2$, and thus the final outcome is:

- $(E, (2n(n - Y')B + (n-1)\epsilon, -\epsilon, \ldots, -\epsilon))$ with probability $1 - \epsilon$;
- $(A^*, (P_1^* + 2(n-1)\epsilon, P_2^* - 2\epsilon, \ldots, P_n^* - 2\epsilon))$ with probability $\frac{\epsilon Y'}{nB}$; and
- $(E, ((n-1)\epsilon, -\epsilon, \ldots, -\epsilon))$ with probability $\epsilon(1 - \frac{Y'}{nB})$.

Therefore we have that

$$
\begin{aligned}
& u^x_{\mathscr{C}_1}(s'_{\mathscr{C}_1} \sqcup t_{-\mathscr{C}_1} | c_0 = Heads) \\
= \;& (1-\epsilon)[-2n(n-Y')B - (n-1)\epsilon + (|\mathscr{C}_1|-1)\epsilon] \\
& + \frac{\epsilon Y'}{nB} \cdot \left[\sum_{j \in \mathscr{C}_1} \theta_j(A^*_j) - \sum_{j \in \mathscr{C}_1} \theta_j(E_j) - P^*_1 - 2(n-1)\epsilon - \sum_{j \in \mathscr{C}_1 \setminus \{1\}} P^*_j + 2(|\mathscr{C}_1|-1)\epsilon \right] \\
& + \epsilon(1 - \frac{Y'}{nB})[-(n-1)\epsilon + (|\mathscr{C}_1|-1)\epsilon] \\
< \;& (1-\epsilon)[-2n(n-Y')B - (n-|\mathscr{C}_1|)\epsilon] + \frac{\epsilon Y'}{nB} \cdot \left[2|\mathscr{C}_1|B + \sum_{j \notin \mathscr{C}_1} P^*_j - 2(n-|\mathscr{C}_1|)\epsilon \right] \\
& - \epsilon(1 - \frac{Y'}{nB})(n-|\mathscr{C}_1|)\epsilon \\
< \;& -2n(1-\epsilon)(n-Y')B + \frac{\epsilon Y'}{nB} \cdot 2nB < -4n(1-\epsilon)B + 2n\epsilon < -3nB.
\end{aligned}
$$

Now we compute $u^x_{\mathscr{C}_1}(s_{\mathscr{C}_1} \sqcup t_{-\mathscr{C}_1} | c_0 = Heads)$. If $t_{-\mathscr{C}_1}$ is such that some player in $-\mathscr{C}_1$ announces NO in Step 2, then letting Y be the number of players announcing YES, we have that $n-1 > Y > Y'$ (the second inequality is because every player in $\mathscr{C}_1 \setminus \{1\}$ announcing YES according to $s_{\mathscr{C}_1}$, but some of them announce NO according to $s'_{\mathscr{C}_1}$). Similar to the formula above, we have that

$$
\begin{aligned}
& u^x_{\mathscr{C}_1}(s_{\mathscr{C}_1} \sqcup t_{-\mathscr{C}_1} | c_0 = Heads) \\
= \;& (1-\epsilon)[-2n(n-Y)B - (n-1)\epsilon + (|\mathscr{C}_1|-1)\epsilon] \\
& + \frac{\epsilon Y}{nB} \cdot \left[\sum_{j \in \mathscr{C}_1} \theta_j(A^*_j) - \sum_{j \in \mathscr{C}_1} \theta_j(E_j) - P^*_1 - 2(n-1)\epsilon - \sum_{j \in \mathscr{C}_1 \setminus \{1\}} P^*_j + 2(|\mathscr{C}_1|-1)\epsilon \right] \\
& + \epsilon(1 - \frac{Y}{nB})[-(n-1)\epsilon + (|\mathscr{C}_1|-1)\epsilon].
\end{aligned}
$$

Therefore

$$
\begin{aligned}
& u^x_{\mathscr{C}_1}(s_{\mathscr{C}_1} \sqcup t_{-\mathscr{C}_1} | c_0 = Heads) - u^x_{\mathscr{C}_1}(s'_{\mathscr{C}_1} \sqcup t_{-\mathscr{C}_1} | c_0 = Heads) \\
= \;& (1-\epsilon)2n(Y-Y')B + \frac{\epsilon(Y-Y')}{nB} \cdot \left[\sum_{j \in \mathscr{C}_1} \theta_j(A^*_j) - \sum_{j \in \mathscr{C}_1} \theta_j(E_j) - P^*_1 - \sum_{j \in \mathscr{C}_1 \setminus \{1\}} P^*_j - 2(n-|\mathscr{C}_1|)\epsilon \right] \\
& + \frac{\epsilon(Y-Y')}{nB}(n-|\mathscr{C}_1|)\epsilon \\
= \;& 2n(1-\epsilon)(Y-Y')B + \frac{\epsilon(Y-Y')}{nB} \cdot \left[\sum_{j \in \mathscr{C}_1} \theta_j(A^*_j) - \sum_{j \in \mathscr{C}_1} \theta_j(E_j) + \sum_{j \notin \mathscr{C}_1} P^*_j - (n-|\mathscr{C}_1|)\epsilon \right] \\
> \;& 2n(1-\epsilon)(Y-Y')B - \frac{\epsilon(Y-Y')}{nB} \cdot [2nB + (n-|\mathscr{C}_1|)\epsilon] \\
> \;& 2n(1-\epsilon)(Y-Y')B - 2\epsilon(Y-Y') - \epsilon > 2n(1-\epsilon) - 2n\epsilon > 0,
\end{aligned}
$$

and thus $u^x_{\mathscr{C}_1}(s_{\mathscr{C}_1} \sqcup t_{-\mathscr{C}_1} | c_0 = Heads) > u^x_{\mathscr{C}_1}(s'_{\mathscr{C}_1} \sqcup t_{-\mathscr{C}_1} | c_0 = Heads)$ as desired.
On the other hand, if $t_{-\mathscr{C}_1}$ is such that every player in $-\mathscr{C}_1$ announces YES in Step 2, then letting Y be the number of players announcing YES, we have that $Y = n-1$ and the final outcome is $(A^*, (P^*_1 +$

$2(n-1)\epsilon, P_2^* - 2\epsilon, \ldots, P_n^* - 2\epsilon)$), which implies that

$$\begin{aligned} & u^x_{\mathscr{C}_1}(s_{\mathscr{C}_1} \sqcup t_{-\mathscr{C}_1} | c_0 = Heads) \\ = & \sum_{j \in \mathscr{C}_1} \theta_j(A_j^*) - \sum_{j \in \mathscr{C}_1} \theta_j(E_j) - P_1^* - \sum_{j \in \mathscr{C}_1 \setminus \{1\}} P_j^* - 2(n - |\mathscr{C}_1|)\epsilon \\ = & \sum_{j \in \mathscr{C}_1} \theta_j(A_j^*) - \sum_{j \in \mathscr{C}_1} \theta_j(E_j) + \sum_{j \notin \mathscr{C}_1} P_j^* - 2(n - |\mathscr{C}_1|)\epsilon \\ > & -2nB - 2(n - |\mathscr{C}_1|)\epsilon > -3nB, \end{aligned}$$

and thus $u^x_{\mathscr{C}_1}(s_{\mathscr{C}_1} \sqcup t_{-\mathscr{C}_1} | c_0 = Heads) > u^x_{\mathscr{C}_1}(s'_{\mathscr{C}_1} \sqcup t_{-\mathscr{C}_1} | c_0 = Heads)$ as desired.

Combining the two cases, we have $u^x_{\mathscr{C}_1}(s_{\mathscr{C}_1} \sqcup t_{-\mathscr{C}_1}) > u^x_{\mathscr{C}_1}(s'_{\mathscr{C}_1} \sqcup t_{-\mathscr{C}_1})$ always, and Lemma 9 holds. ■

The above three lemmas together imply that for any outcome x, $\mathcal{M}^x$ is CDS. Moreover, these lemmas have also specified the structure of $s^x_{\mathscr{C}}$. The next two lemmas prove property (3) of Definition 2, concluding the proof that $\mathcal{M}$ is root-solvable.

Lemma 10. *For any context $\mathbb{C}$ such that $SW(E) = \max_{A' \in \mathcal{A}} SW(A')$, letting $x = (E, (0, \ldots, 0))$ and $S = \{x\}$, we have that $u_{\mathscr{C}_1}(s^x_{\mathscr{C}}) > u_{\mathscr{C}_1}(s^{x'}_{\mathscr{C}})$ for any outcome $x' \notin S$, and $u_{\mathcal{C}}(s^x_{\mathscr{C}}) = 0$ for any coalition $\mathcal{C} \in \mathscr{C}$.*

Proof. Similar to that of Lemma 3. ■

The next lemma uses the following definition.

Definition 8. *Given a context $\mathbb{C}$, an outcome (A, P) is* semi-truthful *if $SW(A) = \max_{A' \in \mathcal{A}} SW(A')$, $\sum_i P_i = 0$, $P_i = \theta_i(A_i) - \theta_i(E_i)$ for each $i \notin \mathscr{C}_1$, and $P_i \in \mathbb{Z} \cap (-2^k, 2^k)$ for each $i \in \mathscr{C}_1 \setminus \{1\}$.*

Notice that the set of semi-truthful outcomes is independent of $\mathscr{C} \setminus \{\mathscr{C}_1\}$. Indeed, if $\mathbb{C} = (E, \theta, \mathscr{C})$ and $\mathbb{C}' = (E, \theta, \mathscr{C}')$ are two contexts such that $\mathscr{C}_1 = \mathscr{C}'_1$, then the set of semi-truthful outcomes with respect to $\mathbb{C}$ is the same as that with respect to $\mathbb{C}'$.

Lemma 11. *For any context $\mathbb{C}$ such that $SW(E) < \max_{A' \in \mathcal{A}} SW(A')$, letting S be the set of semi-truthful outcomes, we have that for any $x \in S$ and $x' \in \Omega$: (1) $u_{\mathscr{C}_1}(s^x_{\mathscr{C}}) = u_{\mathscr{C}_1}(s^{x'}_{\mathscr{C}})$ whenever $x' \in S$; (2) $u_{\mathscr{C}_1}(s^x_{\mathscr{C}}) > u_{\mathscr{C}_1}(s^{x'}_{\mathscr{C}})$ whenever $x' \notin S$; and (3) $u_{\mathcal{C}}(s^x_{\mathscr{C}}) \geq 0$ for any coalition $\mathcal{C} \in \mathscr{C}$.*

Proof. Similar to that of Lemma 4. ■

Now we can conclude that $\mathcal{M}$ is root-solvable. The proofs of the second and the third properties of Definition 6 are very similar to that of Lemmas 5 and 6, and thus omitted. *Q.E.D.*