Robust Perfect Revenue From Perfectly Informed Players

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

Avinatan Hassidim
RLE, MIT
Cambridge, MA 02139, USA
avinatanh@gmail.com

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

October 7, 2011

Abstract

Maximizing revenue in the presence of perfectly informed players is a well known goal in mechanism design. Yet, all current mechanisms for this goal are vulnerable to equilibrium selection, collusion, privacy and complexity problems, and therefore far from guaranteeing that maximum revenue will be obtained. In this paper we both clarify and rectify this situation by

- Proving that no *weakly dominant-strategy* mechanism (traditionally considered immune to equilibrium selection) guarantees an arbitrarily small fraction of the maximum possible revenue;

and, more importantly,

- Constructing a new mechanism, of extensive-form and with a *unique* sub-game-perfect equilibrium, which

  (a) guarantees a fraction arbitrarily close to 1 of the maximum possible revenue;

  (b) is provably robust against equilibrium selection, collusion, complexity, and privacy problems; and

  (c) works for any number of players $n > 1$, and without relying on special conditions for the players utilities.

1 Introduction

1.1 Classical Mechanism Design

Contexts and mechanisms. A *context* C describes the players, the outcomes and the players' preferences over the outcomes. A *mechanism* M describes the *strategies* available to the players, and how strategies determine outcomes. Together, a context C and a mechanism M define a *game* G, $G = (C, M)$, in which each rational player will endeavor to choose his own strategy so as to maximize his own utility.

Mechanism design. Mechanism design aims at finding a mechanism M such that, for any context C (or any C in a given class), a desired property $\mathbb{P}$ holds for the outcomes of the game (C, M), when rationally played. The difficulty is that the designer does not exactly know the players' preferences, while $\mathbb{P}$ typically depends on such preferences. In the purest form of mechanism design, *all knowledge about the players lies with the players themselves.* The designer can count only on the players' rationality. And based solely on this fact, he must design M so that it becomes "in the best interest of the players" to satisfy $\mathbb{P}$. That is, he must ensure that $\mathbb{P}$ holds in a rational play of M. But: *What is a rational play?*

The classical interpretation of a rational play. The classical interpretation of a rational play is an *equilibrium*, that is a profile of strategies $\sigma = \sigma_1, \ldots, \sigma_n$ such that no player i has an incentive to deviate from his specified strategy σ_i to an alternative strategy σ'_i. But equilibria are vastly different in their "quality." The weakest form is that of a *Nash equilibrium,* simply stating that i prefers σ_i to any alternative σ'_i only if he *believes that every other player j will stick to his specified σ_j.* That is, Nash equilibrium only guarantees that i prefers $\sigma_1, \ldots, \sigma_i, \ldots, \sigma_n$ to $\sigma_1, \ldots, \sigma'_i, \ldots, \sigma_n$. If σ is a *dominant-strategy equilibrium,* the strongest form of equilibrium, then, for any player i, σ_i is i's best strategy no matter what strategies the other players may choose. More precisely, a dominant-strategy equilibrium σ is called strict (respectively, weak) if, for any player i, any alternative strategy σ'_i, and any strategy sub-profile τ_{-i} for the other players, i's utility when playing σ_i is strictly larger than (respectively, larger than or equal to) his utility when playing σ'_i.

1.2 Our Goal

This paper focuses on a classical context: *quasi-linear utilities with non-negative valuations.* Namely, there are finitely many possible *states*, $\omega_1, \ldots, \omega_k$, including the *null state*, which every player values 0; each player i has non-negative value $v_i(\omega_j)$ for each state ω_j; each outcome consists of a state ω together with a price P_i for each player i; and the utility of each player i for such an outcome is $v_i(\omega) - P_i$. (The *revenue* of an outcome (ω, P) consists of $\sum_i P_i$. The function v_i is i's *valuation.*)

Such context models a great deal of situations. For instance, in an auction of multiple goods, a state ω represents which player wins which items. Accordingly, the utility of player i in an outcome (ω, P) naturally is his value for the items he gets in ω, minus the price he pays. In another example, each state ω represents one of finitely many ways of building a bridge across a given river. Accordingly, and naturally too, each player has different values for each possible bridge. (For instance, a player's value for a given potential bridge may dependent on how distant it would be from his house.) The list of examples could go on and on. In all of them, however, no matter what the mechanisms may be, it is also natural for different subsets of the players to collude —that is, to coordinate their strategies— so as to improve their utilities.

In such a classical context, our goal is equally classical: getting an outcome of maximum revenue when the players have *perfect knowledge.* That is, when each player knows the valuations of all players (as well as who colludes with whom, if collusion exists among the players).

When the players' knowledge is best possible, it is natural to ask whether the best possible revenue can be obtained. Note that, without the ability of imposing arbitrary prices, the best possible revenue that a mechanism can hope to get from rational players is the *maximum social welfare*, that is, $\max_\omega \sum_i v_i(\omega)$. Thus:

Can a mechanism guarantee perfect revenue from perfectly informed players?

1.3 Four Main Obstacles

Plenty of mechanisms have been proposed for our goal. Yet, *none of them achieves it in a robust way.* Four main obstacles stand on their way. Let us explain.

Equilibrium Selection It should be realized that designing a mechanism so as to guarantee a property $\mathbb{P}$ "at a Nash Equilibrium" is a weak guarantee. First, because there may be several Nash equilibria, while $\mathbb{P}$ holds for just some of them. Moreover, even if $\mathbb{P}$ held for all equilibria, $\mathbb{P}$ may not hold at all in a real play. For instance, assume that there exist two equilibria, σ and τ, and that some players believe that σ will be played out, while others believe that τ will. Then, rather than an equilibrium, a mixture of σ and τ will be played out, so that $\mathbb{P}$ may not hold. Of course, this problem worsens as the number of players and/or equilibria grows.

Collusion The problem of collusion in mechanism design is well recognized. The problem occurs for obvious reasons. Any equilibrium, even a dominant-strategy one, only guarantees that no *single* player has incentive to deviate from his strategy. However, two or more players may have all the incentive in the world to *jointly* deviate from their equilibrium strategies. Accordingly, by "guaranteeing" a property $\mathbb{P}$ at equilibrium, a classical mechanism is typically vulnerable to collusion. In a second-price auction, although the mechanism is dominant-strategy, if the players with the highest two valuations for the item on sale collude, then the revenue generated drops from the second-highest to the third-highest valuation. As for a more extreme example, Ausubel and Milgrom [2] show that two sufficiently informed players can totally destroy the economic efficiency of the famous VCG mechanism [26, 9, 12], although it too is dominant-strategy.

Complexity Traditional mechanism design disregards the "complexity of a mechanism." A mechanism of normal-form may require the players to simultaneously announce exponentially long strings. And a mechanism of extensive form may require the players to act over exponentially many rounds. In both cases, therefore, such mechanisms in practice fail to reach their objectives, no matter what their theoretical claims, unless their contexts are extremely "tiny."

Privacy Privacy has been traditionally neglected in mechanism design, and considered a *quite separate* desideratum: nice to have perhaps, but not central for an incentive analysis. Yet, as especially argued by [13], it has a great potential to distort incentives, and thus to derail classical mechanisms from achieving their desired properties. A mechanism typically neglects privacy by requiring the players to reveal a lot of information about themselves. But if the players value privacy (which by definition implies that divulging their secret information causes them to receive a *negative utility*), then the mechanism gives them both positive and negative incentives, and it is no longer clear how these opposing forces will balance out.

1.4 Prior Mechanisms

Let us now discuss the most relevant mechanisms for our design problem. We start with the traditional "at equilibrium" approach, and then proceed to more sophisticated ones.

The Generic Mechanism The following may be the first mechanism that comes to mind for our context.

> HOPE-FOR-THE-BEST: *Each player reports the valuations of all players (including himself). If all reports are the same, then (1) choose the state ω maximizing the sum of the reported valuations and (2) for each player i, choose the price P_i to be his reported value for ω (possibly minus a small discount ϵ to encourage i's participation). If not all reports coincide, then choose the "null outcome" (which all players are assumed to value 0) and price 0 for every player.*

Unfortunately, HOPE-FOR-THE-BEST is extremely vulnerable to equilibrium selection. It is trivial to see that the strategy profile in which each player reports all true valuations is a Nash equilibrium for HOPE-FOR-THE-BEST, indeed, it is the *truthful equilibrium*. It is also trivial to see that in this equilibrium the revenue is the maximum possible (disregarding the negligible quantity $n\epsilon$). Notice too, however, that HOPE-FOR-THE-BEST also has *additional* equilibria, $E_2, E_3, \ldots$, where in E_x all players report all true valuations divided by x. Thus, the truthful equilibrium is E_1, and in each E_x the utility of each player is increased by a factor x, and the money collected is a fraction $1/x$ of the maximum possible revenue. Accordingly,

- In the truthful equilibrium E_1 the designer is "happy", but the players are "sad", while
- in all other equilibria E_x the players are "happy" and the designer is "sad."

This being the case: *which equilibrium E_x is more likely to be selected?* Further, while each E_x at least maximizes social welfare, in plenty of other equilibria both revenue and social welfare are quite poor.[1] Given the multitude of available equilibria and the fact that different equilibria are preferable to different players: *will a play of* HOPE-FOR-THE-BEST *be an equilibrium and generate any revenue at all?*

The JPS Mechanism Jackson, Palfrey, and Srivastava [15] provided a quite different mechanism. Again, their mechanism yields optimal revenue only at the truthful equilibrium τ. But this time τ is a much more meaningful equilibrium: *it is the only Nash equilibrium composed of weakly undominated strategies.* Somewhat counterintuitively, however, their solution too is vulnerable to equilibrium-selection. The point is that, as in HOPE-FOR-THE-BEST, there are plenty of equilibria σ that generate smaller revenue while being more attractive to *all* players. Again too, each such σ consists of reporting all true valuations divided by the same factor x. To be sure, this time each component σ_i is weakly dominated by some other strategy σ'_i. This means that, in all cases (i.e., for all possible subprofiles of strategies for the other players) σ_i provides no more utility to i than σ'_i does, while in at least some cases σ_i provides less utility to i than σ'_i. But in the JPS mechanism this happens in only one case: when *all other players "suicide"* (i.e., when all other players deliberately choose the worst possible strategies for themselves). Thus, as long as a single player does not believe that all others will commit mass suicide, all players prefer σ to the truthful and revenue-maximizing equilibrium τ. Accordingly, *the JPS mechanism too is very vulnerable to equilibrium selection.*

What has happened? At a first glance, the JPS mechanism looks very "robust," because "no one should want to play a weakly dominated strategy." But the problem is that the process of eliminating all *weakly* dominated strategies for yourself and the other players is not well defined. Unlike the iterated elimination of *strictly* dominated strategies, the iterated elimination of weakly dominated strategies depends on the order of elimination. For example, if one eliminates first "suicidal strategies" (in fact, if one eliminates first "suicide" for just another one of the players), then all equilibria become equally reasonable, and the attractive ones from the players' point of view are those generating less revenue.

In addition, *the JPS mechanism is totally vulnerable to collusion.* Indeed, it enables some pairs of players (i, j) to jointly deviate from the truthful equilibrium so as to improve the utility of i without hurting that of j. And when they so deviate its revenue cannot be maximum.

Finally, *the JPS mechanism is totally vulnerable to privacy*, because it relies on the players revealing all their knowledge.

The AM (and GP) Mechanisms Assuming that there are at least 3 players and that some more technical conditions hold, Abreu and Matsushima [1] present a general normal-form mechanism that guarantees that essentially any desired property (including ours) is satisfied in a perfect-knowledge context. Their mechanism is robust against equilibrium selection, because after the iterated elimination of *strictly* dominated strategies, each player is left with a single (and truthful) strategy, and thus the resulting game has a single equilibrium. However, the *AM mechanism is highly vulnerable to collusion, complexity, and privacy problems.* That is,

[1] Let ω be any state such that $v_i(\omega) >> c > 0$ for all players i. And let σ be the strategy profile, where each σ_j consists of reporting that all players have the following valuation v: $v(\omega) = c$ and $v(x) = 0$ for any state $x \neq \omega$. Then, it is easy to see that σ is an equilibrium. Moreover, the revenue of σ is cn, and the social welfare of σ is $\sum_i v_i(\omega)$.

(1) it no longer guarantees its desired property when any two players jointly deviate from their equilibrium strategies; (2) it requires the players to announce a doubly-exponential number of bits even when there are a constant number of outcomes, a constant number of player, and each player can have one of two types; and (3) it relies on the players revealing all the knowledge in their possession.

A variant of the AM mechanism was put forward by Glazer and Perry [11]. The GP mechanism is of extensive form: informally this means that the players act one at a time, over several rounds. The GP mechanism did not suffer from any equilibrium selection problems either, because its corresponding game admits a unique subgame-perfect equilibrium. Essentially, this means that, at each decision node of the game tree, every acting player has a single best action available to him. However, the GP mechanism continues to be vulnerable to collusion, complexity and privacy problems. The vulnerability to collusion and privacy is essentially identical to that of the AM mechanism. Complexity wise the GP mechanisms actually requires exponentially many rounds of communication. A justification of this fact is presented in Appendix B.

1.5 Our Results

Our Impossibility Result The problem of equilibrium selection fully disappears when a mechanism achieves its desired property $\mathbb{P}$ at a strictly dominant-strategy equilibrium, while still "lurks around" for weakly dominant-strategy equilibria. Unfortunately, we prove that neither strong nor weakly dominant-strategy mechanisms exist that can guarantee perfect revenue from perfectly informed players. Worse, our impossibility result holds even even if the mechanism designer were content to generate an *arbitrarily small fraction* of the optimal revenue. In sum, we prove the following.

Thm 1: *No weakly dominant-strategy mechanism guarantees a fraction ϵ of the optimal revenue.*

Our impossibility theorem shows that, in order to guarantee perfect revenue, we must adopt a different solution concept.

Our Possibility Result To enable the design of mechanisms reasonably resilient against equilibrium selection, collusion, complexity, and privacy, the first and third author have developed *rational robustness*, a new solution concept and implementation notion [7]. (Essentially, rational robustness is based on *iterated elimination of distinguishably dominated strategies,* a new notion "in between" iterated elimination of *strictly* and *weakly* dominated strategies.) But in our perfect knowledge context we are able to achieve perfect revenue and perfect resiliency, by means of an extensive-form mechanism, under the classical solution concept of unique subgame-perfect equilibrium (which indeed is a special case of rationally robust implementation).

Informal Thm 2: *There exists an extensive-form mechanism $\mathcal{M}$, guaranteeing a fraction $1-\epsilon$ of the optimal revenue, that*

- *Works for any number of players $n > 1$;*
- *Has a* unique *subgame perfect equilibrium;*
- *Enjoys perfect collusion resilience;*
- *Enjoys perfect privacy;*
- *Does not require any trust from the players; and*
- *Has perfect communication complexity and $n + 1$ rounds.*

Remarks

- Note that $\mathcal{M}$ works with just two players. By contrast, the mechanism of Jackson, Palfrey and Srivasta, that of Abreu and Matsushima, that of Glazer and Perry, and the generic mechanism all require more than 2 players. In essence, because this requirement facilitates (or makes it possible and meaningful) to identify which player deviated from his equilibrium strategies.

- We stress the word "unique" because a game with multiple subgame-perfect equilibria can still suffer from equilibrium-selection problems.
- Perhaps interestingly, in the presence of collusive players, our mechanism $\mathcal{M}$ has multiple ways to be truthful, but *only one* of them is a subgame-perfect equilibrium.
- By saying that $\mathcal{M}$ is perfectly resilient to collusion we mean two things. First, $\mathcal{M}$ guarantees perfect revenue as long as not all players belong to the same coalition, and each coalition acts rationally. (In our setting, a rational coalition maximizes the sum of the individual utilities of its members. Only when the players have imperfect knowledge about each other, one may want to consider weaker models of coalition rationality.) Second, $\mathcal{M}$ achieves perfect revenue no matter how well players belonging to the same collusive set, if any, may coordinate their actions. In particular, $\mathcal{M}$ works even when such players are free to make side-payments to each other and/or to enter into binding contracts with each other.
- By saying that $\mathcal{M}$ is perfectly private we essentially mean that, in any rational play, nothing can be learned about the players' valuations, by the mechanism designer or any observer of the play, except for what is deducible from a perfect-revenue outcome. Of course, our $\mathcal{M}$ can be so "perfect" only because we are dealing with perfectly informed players (so that the only privacy concern is with respect to the "outside world"). But this *is* our setting, and thus one has both the right to demand and the obligation to deliver as a perfect solution as possible.
- By saying that $\mathcal{M}$ does not require any trust from the players we mean several things. First, $\mathcal{M}$ is not a *mediated* mechanism. (Indeed, privacy would be easy to achieve if the players could confide their strategies or their preferences to mediator trusted to announce the right outcome and never to reveal to anyone else any information received from the players). Second, $\mathcal{M}$ does not rely on any complexity assumptions, as needed for running a cryptographically secure protocol. (After all, at least some of such assumptions may turn out to be false.) Third, $\mathcal{M}$ does not rely on the security of some underlying communication channels. Fourth, $\mathcal{M}$ does not rely on the "honesty" of even some of the players.
  Indeed, in our mechanism $\mathcal{M}$ any action of the players becomes public as soon as it is taken.
- By saying that $\mathcal{M}$ has perfect communication complexity we mean that $\mathcal{M}$'s players need to announce, altogether, the same number of bits necessary to describe the desired outcome.
  In addition to "mechanism complexity", one may consider also "player complexity," that is the time required to a player to figure out and thus choose a rational strategy. Here it is worth pointing out that, to play rationally $\mathcal{M}$, a player performs a computation linear in the number of states —and the number of players. This is essentially optimal given the generality of our setting. (Only with respect to a specific choice of states, one can consider whether there exists a compact representation of the states together with a compact representation of the players' utilities for them.)

Comparison with other work

- Note that our notion of collusion resiliency is stronger than that offered by other mechanisms. In particular, *group —or coalition— strategyproofness* [3, 19, 16, 21, 25] rules out collusion, but only under the assumption that the players are not able to make side payments to each other. Without restricting how players might cooperate, *t-truthful mechanisms* [10] offer protection against coalitions of at most t players, but only for single-value games. (In such games, a player i values some outcomes 0, and all other outcomes a fixed value v_i.) Again without restricting cooperation abilities, *collusion neutralization* [20, 6] offers collusion protection in more general games, but their notion too is weaker than the one considered in our paper. (Protection against the coalition of all players has also been considered and achieved, but only in Bayesian settings, where the distributions of player preferences are known to everyone, including the mechanism designer [17, 18, 4, 5].) Finally, a different approach altogether, *collusion leveraging,* has been submitted to this same conference [8].
- Some work on privacy preserving mechanisms has already started. However, the privacy is either limited or gained by adding an additional layer to the mechanism —such as one or more mediators, envelopes,

or encryption— [23, 22, 13]. By contrast, our mechanism $\mathcal{M}$ achieves perfect privacy without relying on any additional infrastructure. Indeed, $\mathcal{M}$ works by asking the players to take only *public actions.*

Easy Variants and Forthcoming Work By hindsight, it is easy to modify our mechanism in various ways —or even the original Abreu-Matsushima or Glazer-Perry mechanisms— so as to keep our perfect robustness against collusion, complexity, and privacy, while gaining some perceived additional advantage, such as reducing the number of steps, achieving *additive* revenue approximation, etc.

There are, however, quite important variants that we would like to point out now, and develop in later versions of this paper. Namely, we can achieve perfect revenue not from just perfectly informed players (as in this paper), but also from players with perfect *distributed* knowledge. For instance, it suffices that for each fact about a player i (i.e., for each $TV_i(\omega)$) there exists an additional player who knows this fact.

2 Preliminaries

Our Contexts We work with reasonably general contexts with quasi-linear utilities, where the players can be collusive and are perfectly informed about each other. More formally,

Definition 1. *A* perfect-knowledge context *C is identified by four quantities, $C = (N, \Omega, TV, \mathbb{C})$, where*

- *N is the finite set of* players*, $N = \{1, \ldots, n\}$*
- *Ω is the finite set of* states*, which includes the* empty state*, $\perp$.*
  The set Ω defines the set of outcomes: namely, $\Omega \times \mathbb{R}^n$. It also defines the set V of valuation profiles v: namely, each valuation v_i is a function from Ω to non-negative reals such that $v_i(\perp) = 0$.
- *TV is the profile of* true *valuations (or types): namely, each valuation TV_i describes player i's actual value for each possible state.*
  Each TV_i defines the utility function *u_i of player i as follows: for each outcome (ω, P), $u_i(\omega, P) = TV_i(\omega) - P_i$. That is, i's utility is his true value for the state minus the price he pays. The profile P is referred to as a* price profile.
- *$\mathbb{C}$ is the* collusion structure*: namely, a partition of N.*
  If S is a subset in $\mathbb{C}$, then S is the maximal subset of players colluding with each other. A collusive set *is a member of $\mathbb{C}$ with cardinality greater than 1. A player i is* independent *if $\{i\} \in \mathbb{C}$. The context is non-collusive if all players are independent, and collusive otherwise.*
  Each independent player tries to maximize his own utility function, and each collusive set tries to maximize the sum of the utilities of its members.

The sets N and Ω are common knowledge to everybody; while the profile TV and the partition $\mathbb{C}$ are only common knowledge to the players.

We stress that the mechanism *designer* has no knowledge about TV (or $\mathbb{C}$)! In other words, we adhere to the classic spirit of mechanism design, where all knowledge lies with just the players.

Our Mechanisms Recall that each mechanism M must specify the players' strategies, including the *opt-out strategy* (for each player i this strategy is always denoted by OUT_i), and the outcome (or distribution over outcomes if M is probabilistic), $M(\sigma)$, associated to each possible strategy profile σ. In addition, we insist that each mechanism M must satisfy the following

OPT-OUT CONDITION: $M(\sigma) = (\perp, (0, \ldots, 0))$ whenever $\sigma_i = \text{OUT}_i$ for some player i.

For conciseness, we refer to a profile of strategies as a *play.* The expected utility of player i in a play σ is $\mathbb{E}[u_i(M(\sigma))]$. As announced, our mechanisms are finite, of extensive form, and *public-action*, that is, each of our M specifies a game tree, where exactly one player acts at each node, knowing all actions played so far.

Unique Subgame Perfect Equilibrium An extensive-form game has a unique subgame perfect equilibrium if:

1. At every decision note of height 1, each player has a (necessarily unique) strictly dominant strategy.
2. By induction: every decision node of height $k-1$, has a unique subgame perfect equilibrium. Assume that in any subgame rooted at node of height $k-1$, the players choose their strategies according to the equilibrium corresponding to this node. Then, for each decision node of height k, each player acting at this node has a strictly dominant strategy.

Social Welfare, Revenue, and Our Goal The *social welfare* and the *revenue* of an outcome (ω, P) are respectively defined to be $\sum_i TV_i(\omega)$ and $\sum_i P_i$.

The *maximum rational revenue* for a context $C = (N, \Omega, TV, \mathbb{C})$ is defined to coincide with the *maximum social welfare* (MSW for short), that is, $\max_{\omega} \sum_i TV_i(\omega)$.

3 Impossibility Result for DST mechanisms

Let us prove that DST mechanisms are incapable of properly leveraging external knowledge: namely, in a perfect-knowledge context, they cannot guarantee even a minuscule fraction of the maximum rational revenue.

Definition 2. *A DST mechanism M guarantees a fraction ϵ of the maximum rational revenue if for any context $C = (N, \Omega, TV)$ we have*

$$(*) \qquad M(TV, \ldots, TV) = (x, P) \text{ implies } \sum P_i \geq \epsilon \cdot MSW.$$

Note that, in proposition $(*)$, each TV is not just the true valuation of a single player, but the profile of all such valuations, because a player's strategy includes his declaration about the others' valuations as well.

Note too that the mechanism is not required to choose the outcome which maximizes the social welfare. Moreover, when not all the players are telling the truth, there is no requirement on the behavior of the mechanism.

Finally note the following immediate corollary of the opt-out condition. Namely,

Non-negative utility property: if M is a DST mechanism and $M(v^1, \ldots, v^n) = (\omega, P)$, then $P_i \leq v_i^i(\omega)$.

Theorem 1. *For any $\epsilon > 0$ no DST mechanism M guarantees a fraction ϵ of the maximum rational revenue.*

Proof. We actually prove our result even for contexts with just two players and only two possible outcomes. Without loss of generality, consider the context (N, Ω, TV) where $N = \{1, 2\}$ and $\Omega = \{\perp, \omega\}$. In such a context, a valuation v_i of a player i coincides with a single number $v_i(\omega)$ (because $v_i(\perp)$ is bound to be 0), and so a strategy v for i coincides with a pair of numbers, $v = (c_1, c_2)$, where c_1 is the declared value for player 1 and c_2 the declared value for player 2.

Our proof is by contradiction. We start by analyzing the behavior of M when the two players make identical and positive (but not necessarily truthful) declarations. More precisely, we prove the following proposition:

$(\star)$ if $c_1, c_2 > 0$, then $M(\ (c_1, c_2)\ ,\ (c_1, c_2)\ ) = (x, (P_1, P_2))$ where

$$\star_1: \quad P_1 + P_2 \geq \epsilon \cdot (c_1 + c_2)$$

$$\star_2: \quad x = \omega$$

To see that proposition $(\star)$ holds, assume the players bid truthfully; that is assume that $c_1 = TV_1(\omega)$ and $c_2 = TV_2(\omega)$. In this case, according to $(*)$ the mechanism must extract a revenue of at least $\epsilon \cdot MSW = \epsilon \cdot (c_1 + c_2)$, and thus $P_1 + P_2 \geq \epsilon \cdot (c_1 + c_2)$, in agreement with inequality $\star_1$.

Now, the hypothesis $c_1 + c_2 > 0$ implies $P_1 + P_2 > 0$. Thus, in light of the non-negative utility property, the state returned by M cannot be $\perp$. Since ω is the only other state, M has to return ω in agreement with equality $\star_2$.

Consider now the declaration $K = (1, 1)$ and let $M(K, K) = (y, Q)$. Then proposition $(\star)$ guarantees that $y = \omega$ and that $Q_1 + Q_2 \geq 2\epsilon$. This implies that $Q_i \geq \epsilon$ for at least a player i. Thus, without loss of generality, we can assume $Q_1 \geq \epsilon$.

Consider now the strategy $\tilde{K} = (\epsilon/2, \epsilon/2)$, and let us analyze the behavior of $M(\tilde{K}, K)$. Let $M(\tilde{K}, K) = (x, P)$.

We start by proving that $x = \omega$. Assume for contradiction purposes that $x = \perp$. Then, when $TV = K$ (and thus player 1 is not truthful), player 2 has an incentive to lie. Indeed, by being truthful, under the current assumption, his utility is 0. However, if player 2 chose the strategy $\tilde{K}$, then according to $(\star)$, the outcome would be (ω, P_1, P_2). In this case, according to the non-negative utility property, since player 2's self-valuation is $\epsilon/2$, $P_2 \leq \epsilon/2$. Thus player 2's utility would be at least $1 - \epsilon/2$. Since this utility is positive, while his utility of being truthful is 0, player 2 has an incentive to lie when $TV = K$ and player 1's strategy is $\tilde{K}$. Therefore we must have $x \neq \perp$, or equivalently $x = \omega$.

Let us now analyze the possible values for P_1 and derive a contradiction in every case.

1. *Case 1:* $P_1 < \epsilon$. In this case, assume that $TV = K$ and compute player 1's utility under the following two strategy profiles: (K, K) and $(\tilde{K}, K)$. In the first case we already know that $M(K, K) = (\omega, Q)$, where $Q_1 \geq \epsilon$. Therefore player 1's utility when being truthful is $1 - Q_1$ which is at most $1 - \epsilon$. On the other hand, under the strategy profile $(\tilde{K}, K)$, player 1's utility is equal to $1 - P_1$ and thus strictly greater than $1 - \epsilon$ by hypothesis. Thus, the context $(\{1, 2\}, \{\perp, \omega\}, K)$ contradicts the dominant-strategy truthfulness of M.

2. *Case 2:* $P_1 > \epsilon/2$. In this case, since $M(\tilde{K}, K) = (\omega, P)$ and $\tilde{K} = (\epsilon/2, \epsilon/2)$, the non-negative utility property implies that $P_1 \leq \epsilon/2$, and thus a contradiction.

In sum, if M guarantees an ϵ fraction of the maximum possible revenue, no price profile exists for $M(\tilde{K}, K)$ that does not contradict the dominant-strategy truthfulness of M. Since we have not assumed any property of M beyond its being DST, this establishes our theorem. *Q.E.D.*

4 Our Mechanism

Notation In the mechanism below,

- ϵ and ϵ^i_j, for $i \in \{2, \ldots, n\}$ and $j \in \{1, \ldots, n\}$, are constants such that
$$\frac{1}{5n} > \epsilon > \epsilon^2_1 > \cdots > \epsilon^2_n > \epsilon^3_1 > \ldots > \epsilon^3_n > \cdots > \epsilon^n_1 > \cdots > \epsilon^n_n > 0.$$
- Numbered steps are taken by the players, while steps marked by letters are taken by the mechanism.
- Sentences between quotation marks are comments, and could be excised if no clarification is needed.
- We denote by n_r the number of outcomes (ω, P) with revenue r. For all such outcomes, we denote by $0 \leq f_r(\omega, P) < n_r$ the rank of the outcome (ω, P) in the lexicographic order that first considers the state and then the price profile (where $P_1, \ldots P_n$ precedes $P'_1 \ldots P'_n$ whenever $P_1 > P'_1$, etc.).

Mechanism $\mathcal{M}$

(1) *Player* 1 *announces a state* $\omega^\star$ *and a profile* K^1 *of natural numbers.*

"$(\omega^\star, K^1)$ is player 1's proposed outcome, allegedly an outcome of maximum revenue."

(a) *Set* $\omega = \perp$, *and* $P_i = 0$ $\forall i$. *If* $\sum_i K^1_i = 0$, *the mechanism ends right now. Otherwise, proceed to Step 2.*

"Whenever the mechanism ends, ω and P will be, respectively, the final state and price profile."

(2,…, n) *In Step i, $2 \le i \le n$, player i publicly announces a profile Δ^i of natural numbers such that $\Delta^i_i = 0$.*

"By so doing i suggests to raise the current price of j, that is $K^1_j + \sum_{\ell=2}^{i-1} \Delta^\ell_j$, by the amount Δ^i_j."

(b) *For each player i, publicly select bip_i and $P^\star_i$ as follows. Let $R_i = \{j : \Delta^j_i > 0\}$.*
If $R_i \neq \emptyset$, then bip_i is highest player in R_i, and $P^\star_i = K^1_i + \sum_{\ell=2}^{bip_i} \Delta^\ell_i$. Else, $bip_i = 1$ and $P^\star_i = K^1_i$.

"We refer to bip_i as the best informed player about i, and to $P^\star_i$ as the provisional price of i."

($n+1$) *Each player i such that $P^\star_i > 0$ simultaneously announces YES or NO.*
By default, each player i such that $P^\star_i = 0$ announces YES, and player 1 announces YES if $bip_1 = 1$.

"Each player i announces YES or NO to $\omega^\star$ as the final state and to $P^\star_i - \epsilon$ as his own price. (By default player 1 accepts his own price if no one raises it.) At this point the players are done playing, and the mechanism proceeds as follows. If all say YES, the updated proposal $(\omega^\star, P^\star)$ is implemented with probability 1. Else:

- With very high probability the null outcome is chosen, except that the best-informed players of those saying NO are punished.
- With small probability the null outcome is chosen
- With very small probability, proportional to the number of players saying YES, we implement $(\omega^\star, P^\star)$ as if all said YES.

Importantly, as we shall see, all get a small reward at the end for their knowledge."

(c) *Let Y be the number of players announcing YES. If $Y = n$, then reset ω to $\omega^\star$ and each P_i to $P^\star_i - \epsilon$, and go to Step g. If $Y < n$, proceed to Step d.*

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

(e) *If $c_1 = Heads$, reset P_{bip_i} to $P_{bip_i} + 2P^\star_i$ for each player i announcing NO.*

(f) *If $c_1 = Tails$, letting $B = \sum_{i \text{ announces NO}} P^\star_i$, flip a biased coin c_2 such that $\Pr[c_2 = Heads] = \frac{Y}{nB}$.*
If $c_2 = Heads$, reset ω to $\omega^\star$ and each P_i to $P^\star_i - \epsilon$.
If $c_2 = Tails$, ω and P continue to be $\perp$ and 0^n.

(g) *Reset P_1 to $P_1 - \epsilon - 2\epsilon \sum_j K^1_j + \epsilon \frac{f_r(\omega^\star)}{n_r}$ and each other P_i to $P_i - \epsilon - \sum_j \epsilon^i_j \Delta^i_j$.*

"Although players' prices may be negative, we prove that the mechanism never loses money, and that in the unique rational play the utility of every player is non-negative. For clarity, our rewards are proportional to prices and raises."

5 Analysis of Our Mechanism

Mechanism $\mathcal{M}$ induces a game G whose game tree has height $n+1$, and where only players act at each internal node. (The mechanism tosses all its coins at leaf nodes, that are defined to be of height 0.) At each node of height 1 all players act simultaneously, and at every other internal node only a single player acts. Specifically, at each node of height $h \ge 2$ the only acting player is player

$$i_h \triangleq n - h + 2.$$

For each internal node N, we denote by G^N the subgame rooted at N. Recall that a strategy σ_i of player i in G specifies, for each node at which i acts, which action i chooses among all those available to him. By σ^N_i

we denote the restriction of σ_i to subgame G^N. Given a restricted strategy profile σ^N for G^N, the outcome of $\mathcal{M}$ obtained by executing σ^N is denoted by $\mathcal{M}(\sigma^N)$.

For uniformity, we find it sometimes convenient to assume that every player i belongs to a (necessarily unique) collusive set, denoted by $\mathscr{C}_i$. If i is independent, then $\mathscr{C}_i = \{i\}$.

5.1 Statements of Our Lemmas

Lemma 1. *If N is a node of height 1, then G^N has a unique subgame-perfect equilibrium σ^N, where*

- *If i is independent, then σ_i^N consists of announcing YES if and only if $TV_i(\omega^\star) \geq P_i^\star$.*
- *If i belongs to a coalition $\mathscr{C}$, then σ_i^N consists of announcing YES if and only if*

$$bip_i \in \mathscr{C} \quad or \quad \sum_{j\in\mathscr{C}} TV_j(\omega^\star) \geq \sum_{j\in\mathscr{C}} P_j^\star.$$

The proof of this lemma is based on the fact that the probability that an outcome is executed is monotone with the number of players who announce YES. Thus, it is strictly dominant to announce YES, if and only if the player has positive utility from this outcome and price.

Lemma 2. *Let N be a node of height $h \in [2, n]$, $i = i_h$, and $\mathscr{C} = \mathscr{C}_i$. Then G^N has a unique subgame-perfect equilibrium where i acts as follows at node N: For each collusive set $\mathscr{D} \neq \mathscr{C}$,*

1. *if*

$$\sum_{j\in\mathscr{D}} \left(K_j^1 + \sum_{\ell=2}^{i-1} \Delta_j^\ell \right) \geq \sum_{j\in\mathscr{D}} TV_j(\omega^\star)$$

then i announces $\Delta_j^i = 0$ for all $j \in \mathscr{D}$;

2. *if*

$$\sum_{j\in\mathscr{D}} \left(K_j^1 + \sum_{\ell=2}^{i-1} \Delta_j^\ell \right) < \sum_{j\in\mathscr{D}} TV_j(\omega^\star)$$

then letting k be the minimal player in $\mathscr{D}$, i announces $\Delta_j^i = 0$ for all $j \in \mathscr{D} \setminus \{k\}$ and

$$\Delta_k^i = \sum_{j\in\mathscr{D}} \left(TV_j(\omega^\star) - K_j^1 - \sum_{\ell=2}^{i-1} \Delta_j^\ell \right).$$

For his own collusive set $\mathscr{C}$,

1. *if*

$$\sum_{j\in\mathscr{C}} \left(K_j^1 + \sum_{\ell=2}^{i-1} \Delta_j^\ell \right) \geq \sum_{j\in\mathscr{C}} TV_j(\omega^\star) \quad or$$
$$\text{it is the case that } k \in \mathscr{C} \text{ for all } k > i,$$

then i announces $\Delta_j^i = 0$ for all $j \in \mathscr{C}$;

2. *if*

$$\sum_{j\in\mathscr{C}} \left(K_j^1 + \sum_{\ell=2}^{i-1} \Delta_j^\ell \right) < \sum_{j\in\mathscr{C}} TV_j(\omega^\star) \quad and$$
$$\text{there exists player } j > i \text{ such that } j \notin \mathscr{C},$$

then letting k be the minimal player in $\mathscr{C} \setminus \{i\}$, i announces

$$\Delta_k^i = \sum_{j \in \mathscr{C}} \left(TV_j(\omega^\star) - K_j^1 - \sum_{\ell=2}^{i-1} \Delta_j^\ell \right).$$

This lemma is technically involved, but conceptually simple. First, we show that a player i never wants to "overbid," that is raise the price of another player j to more than j's true valuation for the proposed state $\omega^\star$. When j is independent, this holds because we know that j will announce NO to any price above his true valuation, and thus no player after i will want to further raise j's price. Therefore, overbidding on j will cause i to be punished. Care must still be taken to verify the Step-g rewards of i and j will not change this simple analysis. (For example j will not accept a higher price in order to get more reward for volunteering his knowledge about other players.) For coalitions, the argument is more subtle.

After ruling out overbidding, we also show that a player i never wants to "underbid," that is not raise the price of a player j when it is below j's true valuation for the proposed state. Again, this is easier to argue for independent players. Arguing this point for coalitions is the only time that requires exploiting the n^2 reward values $\epsilon_{i,j}$.

Lemma 3. *Let N be the root of the tree (so that $G^N = G$), then G has a unique subgame-perfect equilibrium where player 1 acts as follows at node N:*

1. *player 1 announces $\omega^\star$, the lexicographically first state ω such that $\sum_\ell TV_\ell(\omega) = MSW$;*
2. *for each collusive set $\mathscr{D}$, letting i be the minimal player in $\mathscr{D}$, player 1 announces $K_i^1 = \sum_{j \in \mathscr{D}} TV_j(\omega^\star)$, and $K_j^1 = 0$ for each $j \in \mathscr{D} \setminus \{i\}$.*

The proof of this lemma is also done in two stages. First, given Lemma 2, we prove that it is dominant for player 1 to set the prices correctly (although not exactly truthfully in the case of a coalition). Finally, as the prices are set correctly, choosing the outcome which maximizes the total welfare dominates any other course of action.

Proofs of our lemmas are in Appendix A.

5.2 Our Main Theorem

Theorem 2. *Let σ be the unique subgame perfect equilibrium of G, and let $(\omega, P) = \mathcal{M}(\sigma)$. Then:*

(1) $\sum_i TV_i(\omega) = MSW$, *and*

(2) $\sum_i P_i \geq (1 - 4\epsilon n) MSW$.

Proof. In execution σ, by Lemma 3, player 1 announces $\omega^\star$ such that $\sum_\ell TV_\ell(\omega^\star) = MSW$ and, for each coalition $\mathscr{D}$, also announces $K_\ell^1 = \sum_{j \in \mathscr{D}} TV_j(\omega^\star)$, where ℓ is the minimal player in $\mathscr{D}$. Thus $\sum_i K_i^1 = MSW$.

If $MSW = 0$, then $\sum_i K_i^1 = 0$ and $\mathcal{M}$ ends at Step a, with $\omega = \perp$ and $P_i = 0$ for each player i. Therefore $\sum_i TV_i(\omega) = \sum_i TV_i(\perp) = 0 = MSW$ and $\sum_i P_i = 0 = MSW$.

If $MSW > 0$, then $\sum_i K_i^1 > 0$ and $\mathcal{M}$ ends at Step g. By Lemma 2, for each player $i \neq 1$, i announces $\Delta_k^i = 0$ for each k. Therefore for each player i, $bip_i = 1$. Furthermore, the total price for each coalition $\mathscr{D}$ equals $\mathscr{D}$'s total true valuation for $\omega^\star$: that is, $\sum_{\ell \in \mathscr{D}} P_\ell^\star = \sum_{\ell \in \mathscr{D}} K_\ell^1 = \sum_{\ell \in \mathscr{D}} TV_\ell(\omega^\star)$. By Lemma 1, every player in $\mathscr{D}$ announces YES in Step $n+1$. This implies that, at the end of Step c we have $Y = n$, $\omega = \omega^\star$, and, for each coalition $\mathscr{D}$, $\sum_{\ell \in \mathscr{D}} P_\ell = \sum_{\ell \in \mathscr{D}} P_\ell^\star - |\mathscr{D}|\epsilon = \sum_{\ell \in \mathscr{D}} TV_\ell(\omega^\star) - |\mathscr{D}|\epsilon$. Because $Y = n$, the execution of $\mathcal{M}$ will then proceed directly to Step g, which does not reset the current state. Thus we have that

$$\sum_i TV_i(\omega) = \sum_i TV_i(\omega^\star) = MSW.$$

Because the reward given to each player $i > 1$ in Step g is ϵ, and player 1 gets at most $\epsilon + 2\epsilon MSW$, then the final revenue of the mechanism is

$$\sum_i P_i > \left(\sum_i TV_i(\omega^\star) - n\epsilon \right) - (n-1)\epsilon - \epsilon - 2\epsilon MSW > (1 - 4\epsilon n) MSW,$$

where we parenthesized the prices after step c, and used that MSW is integer and thus $MSW \geq 1$. *Q.E.D.*

References

[1] D. Abreu and H. Matsushima. Virtual Implementation in Iteratively Undominated Strategies: Complete Information. *Econometrica*, Vol. 60, No. 5, pages 993-1008, Sep., 1992.

[2] L.M. Ausubel and P. Milgrom. The Lovely but Lonely Vickrey Auction. *Combinatorial Auctions*, MIT Press, pages 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, Jul., 2006.

[5] Y. Che and J. Kim. Optimal collusion-proof auctions. *Journal of Economic Theory*, Vol.144, Issue 2, pages 565-603, Mar. 2009.

[6] J. Chen and S. Micali. A New Approach to Auctions and Resilient Mechanism Design. *Symposium on Theory of Computing*, pages 503-512, 2009. Full version available at http://people.csail.mit.edu/silvio/Selected Scientific Papers/Mechanism Design/.

[7] J. Chen and S. Micali. Rational Robustness for Mechanism Design. Submitted to STOC'10, full version available at *http://people.csail.mit.edu/silvio/Selected Scientific Papers/Mechanism Design/.*

[8] J. Chen S. Micali, and P. Valiant. Levaraging Collusion in Combinatorial Auctions. Submitted to ICS2010.

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

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

[11] J. Glazer and M. Perry. Virtual Implementation in Backwards Induction. *Games and Economic Behavior*, Vo.15, pages 27-32, 1996.

[12] T. Groves. Incentives in Teams. *Econometrica*, Vol. 41, No. 4, pages 617-631, Jul., 1973.

[13] S. Izmalkov, M. Lepinski, and S. Micali. Perfect Implementation of Normal-Form Mechanisms. *MIT-CSAIL-TR-2008-028*, May 2008. (A preliminary version appeared in FOCS'05, pages 585-595, with title "Rational Secure Computation and Ideal Mechanism Design".)

[14] M. Jackson. Implementation in Undominated Strategies: a Look at Bounded Mechanisms. *Review of Economic Studies*, Vol. 59, No. 201, pages 757-775, Oct., 1992.

[15] M.O. Jackson, T. Palfrey and S. Srivastava. Undominated Nash Implementation in Bounded Mechanisms. *Games and Economic Behavior*, 1994.

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

[17] J.J. Laffont and D. Martimort. Collusion under Asymmetric Information. *Econometrica*, Vol.65, No.4, pages 875-911, Jul. 1997.

[18] J.J. Laffont and D. Martimort. Mechanism Design with Collusion and Correlation. *Econometrica*, Vol. 68, Issue 2, pages 309-342, Mar. 2000.

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

[20] S. Micali and P. Valiant. Resilient Mechanisms for Unrestricted Combinatorial Auctions, submitted to FOCS 2007; SODA 2008; and STOC 2009. Available at http://people.csail.mit.edu/silvio/Selected Scientific Papers/Mechanism Design

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

[22] F. McSherry and K. Talwar. Mechanism Design via Differential Privacy.*FOCS07*, 2007.

[23] M. Naor, B. Pinkas and R. Sumner. Privacy Preserving Auctions and Mechanism Design, *EC99*, 1999.

[24] M.J. Osborne and A. Rubinstein. *A Course in Game Theory*, MIT Press, 1994.

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

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

Appendix

A Proofs of Lemmas 1 to 3

Let us restate and prove Lemma 1 in the absence of collusive players.

Lemma 1′. *Let all players be independent, let N be a node of height 1, and for each player i let σ_i^N be the strategy of i in G^N defined according to the following two mutually exclusive cases:*

Case 1: *N is such that $TV_i(\omega^\star) \geq P_i^\star$. In this case, σ_i^N consists of announcing YES.*

Case 2: *N is such that $TV_i(\omega^\star) < P_i^\star$. In this case σ_i^N consists of announcing NO.*

Then, σ^N is the unique subgame-perfect equilibrium of G^N.

Proof. To prove our lemma it suffices to prove that

Each σ_i^N is a (and thus "the") strictly dominant strategy of player i in G^N.

To this end, one needs to consider both cases, but we restrict ourselves to just the first one, because the second can be handled in a totally symmetric way.

Proof for Case 1. Notice that in mechanism $\mathcal{M}$ some players may announce YES in Step $n+1$ by default, depending on their best-informed players and provisional prices. Accordingly to prove the strict dominance of σ_i^N in Case 1 it suffices to consider i such that (a) $P_i^\star > 0$ and (b) either $i \neq 1$ or $i = 1$ but $bip_1 \neq 1$.

Notice that there are only two strategies for i (in Step $n+1$ and thus) in G^N: that is, announcing YES and announcing NO. Since we are working in the hypothesis of Case 1, the first strategy is σ_i^N, and we shall denote the second one by $\bar{\sigma}_i^N$. Thus, all we have to show is that σ_i^N strictly dominates $\bar{\sigma}_i^N$; that is, letting σ_{-i}^N be an arbitrary strategy subprofile for all other players in G^N, all we have to show is that

$$\mathbb{E}[u_i(\mathcal{M}(\sigma_i^N \sqcup \sigma_{-i}^N))] > \mathbb{E}[u_i(\mathcal{M}(\bar{\sigma}_i^N \sqcup \sigma_{-i}^N))].$$

Notice that in any play of G^N the expected utility of player i has three potential components:

1. the reward component that i gets in Step g;
2. the provisional-outcome component, that is, i's value for the provisional state $\omega^\star$ minus i's provisional price $P_i^\star - \epsilon$; and
3. the punishment component that i gets if he is the best informed player of some j announcing NO.

The probability of the reward component is always 1 for any play of G^N. Moreover the amount of the reward component, denoted by U_i^1, is determined by the actions taken in "ancestor nodes" of N, and thus is the same in any play of G^N.[2]

The amount of the provisional-outcome component, denoted by U_i^2, is the same in any play of G^N: namely $U_i^2 = TV_i(\omega^\star) - (P_i^\star - \epsilon)$. However the probability that such component arises depends on the actual play of G^N. Accordingly, we denote this probability by p_i for the play $\sigma_i^N \sqcup \sigma_{-i}^N$, and by $\bar{p}_i$ for the play $\bar{\sigma}_i^N \sqcup \sigma_{-i}^N$.

Finally, the amount of the punishment component only depends on the actions taken by the other players in G^N, and is therefore the same both in play $\sigma_i^N \sqcup \sigma_{-i}^N$ and in play $\bar{\sigma}_i^N \sqcup \sigma_{-i}^N$. We denote this same quantity by U_i^3. Thus $U_i^3 = \sum_{bip_j = i, j \text{ announces NO}} 2P_j^\star$. We denote the probability that i's punishment component arises in play $\sigma_i^N \sqcup \sigma_{-i}^N$ by q_i, and that this component arises in play $\bar{\sigma}_i^N \sqcup \sigma_{-i}^N$ by $\bar{q}_i$.

Accordingly, we have that

$$\mathbb{E}[u_i(\mathcal{M}(\sigma_i^N \sqcup \sigma_{-i}^N))] = U_i^1 + p_i U_i^2 - q_i U_i^3 \quad \text{and} \quad \mathbb{E}[u_i(\mathcal{M}(\bar{\sigma}_i^N \sqcup \sigma_{-i}^N))] = U_i^1 + \bar{p}_i U_i^2 - \bar{q}_i U_i^3$$

so that

$$\mathbb{E}[u_i(\mathcal{M}(\sigma_i^N \sqcup \sigma_{-i}^N))] - \mathbb{E}[u_i(\mathcal{M}(\bar{\sigma}_i^N \sqcup \sigma_{-i}^N))] = U_i^2(p_i - \bar{p}_i) + U_i^3(\bar{q}_i - q_i).$$

[2] Indeed, U_i^1 equals $\epsilon + 2\epsilon \sum_j K_j^1 - \epsilon \frac{f_r(\omega^\star)}{n_r}$ if $i = 1$, and $\epsilon + \sum_j \epsilon_j^i \Delta_j^i$ otherwise.

Therefore to prove our lemma it suffices to show that

$$U_i^2(p_i - \bar{p}_i) + U_i^3(\bar{q}_i - q_i) > 0. \tag{1}$$

Notice that, by hypothesis we have $TV_i(\omega^\star) \geq P_i^\star$, which implies $U_i^2 > 0$. Also notice that $U_i^3 \geq 0$ (i.e., "punishment is never a reward"). To prove Inequality 1 we now distinguish 2 possible scenarios for σ_{-i}^N.

Scenario 1. *All players other than i announce YES.*
In this scenario, letting Y and $\overline{Y}$ be the number of players announcing YES in executions $\sigma_i^N \sqcup \sigma_{-i}^N$ and $\bar{\sigma}_i^N \sqcup \sigma_{-i}^N$ respectively, we have that $Y = n$ and $\overline{Y} = n - 1$ (since i announces YES in σ_i^N and NO in $\bar{\sigma}_i^N$). In turn, this implies $p_i = 1$ and $\bar{p}_i = \epsilon \frac{\overline{Y}}{nP_i^\star} = \frac{\epsilon(n-1)}{nP_i^\star} < 1$ (since $\epsilon < 1$ and $P_i^\star$ is a positive integer). Moreover, in the present scenario, there is no punishment for player i, and thus $U_i^3 = 0$. Accordingly,

$$U_i^2(p_i - \bar{p}_i) + U_i^3(\bar{q}_i - q_i) = U_i^2(1 - \bar{p}_i) > 0,$$

proving Inequality 1.

Scenario 2. *At least a player $j \neq i$ announces NO.*
In this scenario, we have $q_i = \bar{q}_i = 1 - \epsilon$. Moreover, defining Y and $\overline{Y}$ as in the previous scenario, we have that $Y < n$ and $\overline{Y} = Y - 1$, which implies

$$p_i = \epsilon \frac{Y}{n\left(\sum_{j \neq i, j \text{ announces NO}} P_j^\star\right)} \quad \text{and} \quad \bar{p}_i = \epsilon \frac{\overline{Y}}{n\left(P_i^\star + \sum_{j \neq i, j \text{ announces NO}} P_j^\star\right)}.$$

Comparing the fractions expressing p_i and $\bar{p}_i$ reveals that $p_i > \bar{p}_i$, because the first numerator is greater than the second, while the first denominator is smaller than the second. In sum, $p_i - \bar{p}_i > 0$ and $q_i - \bar{q}_i = 0$. Therefore

$$U_i^2(p_i - \bar{p}_i) + U_i^3(\bar{q}_i - q_i) = U_i^2(p_i - \bar{p}_i) > 0,$$

proving again Inequality 1.
Since the above two scenarios are mutually exclusive, Inequality 1 always holds, and thus our lemma is true.
□

Lemma 2′. *Let N be a decision node of height $h \in [2, n]$, then G^N has a unique subgame-perfect equilibrium where at node N player $i = i_h$ acts as follows: For each player $j \neq i$,*

1. *if $K_j^1 + \sum_{\ell=2}^{i-1} \Delta_j^\ell \geq TV_j(\omega^\star)$, then i announces $\Delta_j^i = 0$;*
2. *if $K_j^1 + \sum_{\ell=2}^{i-1} \Delta_j^\ell < TV_j(\omega^\star)$, then i announces $\Delta_j^i = TV_j(\omega^\star) - K_j^1 - \sum_{\ell=2}^{i-1} \Delta_j^\ell$.*

Proof. We proceed by induction on h.

Base Case. When $h = 2$, we have that $i = n$. First we prove implication 1 (called No Overbidding), and then proceed to implication 2 (No Underbidding).

No Overbidding. We begin by giving a high level idea of the proof. Suppose player n overbids on another player j, that is he announces $\Delta_j^n > 0$ such that $K_j^1 + \sum_{\ell=2}^n \Delta_j^\ell > TV_j(\omega^\star)$. Consider an alternative strategy $\hat{\sigma}_n$ for player n in which he announces

$$\widehat{\Delta}_j^n = \min(0, TV_j(\omega^\star) - (K_j^1 + \sum_{\ell=2}^{n-1} \Delta_j^\ell))$$

and keeps the rest of his declarations.

The difference in utility between the strategies comes from three terms:

1. The initial strategy gives player n a larger reward. However, this reward is upper bounded by $\epsilon_j^n \Delta_j^n$, which is much less than Δ_j^n
2. With probability $1-\epsilon$ player n will be punished by $2P_j^\star$, which is greater than $2\Delta_j^n$
3. The probability that $c_2 = Tails$ changes. This change can either increase or decrease player n's expected utility, depending on whether he prefers $\perp$ without paying, or $\omega^\star$ for $P_n^\star$. We show that if this decreases player n's utility, than it does so by at most $1 < \Delta_j^n$.

Summing the contributions shows that $\hat{\sigma}$ gives player n a larger expected utility.

We now give the details. By contradiction, assume that there exist a decision node N of height 2, a player $j \neq n$, and a restricted strategy profile σ^N of subgame G^N such that: (1) σ_n^N is a subgame-perfect strategy of player n in G^N; (2) for each proper subgame G^M of G^N and each player ℓ, σ_ℓ^M is the unique subgame-perfect strategy of ℓ at G^M; and (3) at node N, $K_j^1 + \sum_{\ell=2}^{n-1} \Delta_j^\ell \geq TV_j(\omega^\star)$, and σ_n^N consists of player n announcing $\Delta_j^n > 0$. Consider the following alternative strategy $\widehat{\sigma}_n$ for n.

Strategy $\widehat{\sigma}_n$

Step n.	Run σ_n and compute Δ^n as σ_n does.
	For each player $\ell \neq j$, announce $\widehat{\Delta}_\ell^n = \Delta_\ell^n$.
	If $K_j^1 + \sum_{\ell=2}^{n-1} \Delta_j^\ell \geq TV_j(\omega^\star)$, then announce $\widehat{\Delta}_j^n = 0$.
	If $K_j^1 + \sum_{\ell=2}^{n-1} \Delta_j^\ell < TV_j(\omega^\star)$, then announce $\widehat{\Delta}_j^n = TV_j(\omega^\star) - K_j^1 - \sum_{\ell=2}^{n-1} \Delta_j^\ell$.
Step n+1.	If $P_n^\star = 0$, announce nothing.
	If $P_n^\star > 0$ and $TV_n(\omega^\star) \geq P_n^\star$, announce YES.
	Otherwise, announce NO.

Notice that $\widehat{\sigma}_n^N$ consists of n announcing $\widehat{\Delta}_j^n = 0$. To emphasize the difference between execution σ^N and execution $\widehat{\sigma}_n^N \sqcup \sigma_{-n}^N$, we write σ^N as $\sigma_n^N \sqcup \sigma_{-n}^N$. We prove that $\mathbb{E}[u_n(\mathcal{M}(\sigma_n^N \sqcup \sigma_{-n}^N))] < \mathbb{E}[u_n(\mathcal{M}(\widehat{\sigma}_n^N \sqcup \sigma_{-n}^N))]$, which implies that σ_n^N is not a subgame-perfect strategy of n at G^N, contradicting our hypothesis about σ_n^N. Because the two executions $\sigma_n^N \sqcup \sigma_{-n}^N$ and $\widehat{\sigma}_n^N \sqcup \sigma_{-n}^N$ are restricted to G^N, for each variable whose value does not change in G^N, we use the same notation in both executions —$\omega^\star$, K^1, and Δ^ℓ for each $\ell \neq n$—, without causing any ambiguity. For the other variables, we use different notations in the two executions —Δ^n and $\widehat{\Delta}^n$, bip_ℓ and $\widehat{bip}_\ell$, $P^\star$ and $\widehat{P}^\star$, etc, for $\sigma_n^N \sqcup \sigma_{-n}^N$ and $\widehat{\sigma}_n^N \sqcup \sigma_{-n}^N$ respectively—, and it should be clear from the context which execution a notation belongs to.

We have the following observations:

O_1: in Step n+1 of both σ_n^N and $\widehat{\sigma}_n^N$, n announces YES or NO consistently with Lemma 1.

O_2: for each player $\ell \neq n$, in both executions $\sigma_n^N \sqcup \sigma_{-n}^N$ and $\widehat{\sigma}_n^N \sqcup \sigma_{-n}^N$, ℓ announces YES or NO in Step n+1 consistently with Lemma 1.

O_3: In execution $\sigma_n^N \sqcup \sigma_{-n}^N$, $bip_j = n$, $P_j^\star = K_j^1 + \sum_{\ell=2}^{n} \Delta_j^\ell$, and j announces NO.

O_4: In execution $\widehat{\sigma}_n^N \sqcup \sigma_{-n}^N$, $\widehat{bip}_j \neq n$.

O_1 is by hypothesis about σ_n^N and by construction of $\widehat{\sigma}_n^N$; O_2 is by hypothesis about σ_{-n}^N; O_3 is by construction of $\mathcal{M}$, O_2, and the fact that $P_j^\star = (K_j^1 + \sum_{\ell=2}^{n-1} \Delta_j^\ell) + \Delta_j^n \geq TV_j(\omega^\star) + \Delta_j^n > TV_j(\omega^\star)$; and O_4 is by construction of $\mathcal{M}$ and the fact that $\widehat{\Delta}_j^n = 0$.

We now compare the reward that player n gets in Step g in the two executions. In execution $\sigma_n^N \sqcup \sigma_{-n}^N$ this is $\epsilon + \sum_\ell \epsilon_\ell^n \Delta_\ell^n$; while in execution $\widehat{\sigma}_n^N \sqcup \sigma_{-n}^N$ this is $\epsilon + \sum_\ell \epsilon_\ell^n \widehat{\Delta}_\ell^n$. By construction of $\widehat{\sigma}_n^N$, $\widehat{\Delta}_\ell^n = \Delta_\ell^n$ for all $\ell \neq j$, and thus

$$\left(\epsilon + \sum_\ell \epsilon_\ell^n \widehat{\Delta}_\ell^n\right) - \left(\epsilon + \sum_\ell \epsilon_\ell^n \Delta_\ell^n\right) = \epsilon_j^n (\widehat{\Delta}_j^n - \Delta_j^n) = -\epsilon_j^n \Delta_j^n. \tag{2}$$

Therefore it suffices to show that before Step g, $\mathbb{E}[u_n(\mathcal{M}(\widehat{\sigma}_i \sqcup \tau_{-i}))] - \mathbb{E}[u_i(\mathcal{M}(\sigma_i \sqcup \tau_{-i}))] > \epsilon_j^n \Delta_j^n$.

To do so, notice that for each player $\ell \neq j$ (in particular for $\ell = n$), $bip_\ell = \widehat{bip}_\ell$, $P^\star_\ell = \widehat{P}^\star_\ell$, and ℓ announces the same content in Step n+1 in the two executions. The only player whose best informed player, provisional price, and announcement in Step n+1 may be different in the two executions is player j, due to player n's announcement in Step n about him.

We define the following variables:

- $u_n^1 = TV_n(\omega^\star) - P_n^\star + \epsilon$.
  This is the utility that player n gets in both executions due to the provisional outcome $(\omega^\star, P^\star)$ being implemented, either because everybody announces YES, or because $c_1 = $ Tails and $c_2 = $ Heads.
- $p_n^1 = \sum_{\ell \neq j:\ bip_\ell = n,\ \ell \text{ announces NO}} 2P^\star_\ell$.
  This is the punishment that n pays to the mechanism in both executions due to players other than j announcing NO, when $c_1 = $ Heads.

We distinguish two cases, according to u_n^1.

Case 1: $u_n^1 > 0$.

In this case, player n announces YES in Step n+1, either because Lemma 1 or because $P_n^\star = 0$. We distinguish the following two subcases.

Subcase 1.1: $K_j^1 + \sum_{\ell=2}^{n-1} \Delta_j^\ell > TV_j(\omega^\star)$.

In this subcase, player j announces NO in both executions, despite the value of bip_j and $\widehat{bip}_j$. Accordingly, we have the following facts: (1) $Y = \widehat{Y}$, and actually for each player ℓ, his announcements in Step n+1 in the two executions are the same; (2) $B - \widehat{B} = P_j^\star - \widehat{P}_j^\star = \Delta_j^n > 0$ which is equivalent to say $B > \widehat{B}$; and (3) c_1 and c_2 are flipped in both executions. According to fact (3), before Step g

$$\mathbb{E}[u_n(\mathcal{M}(\sigma_n^N \sqcup \sigma_{-n}^N))] = -(1-\epsilon)(p_n^1 + 2P_j^\star) + \frac{\epsilon Y}{nB} u_n^1,$$

and

$$\mathbb{E}[u_n(\mathcal{M}(\widehat{\sigma}_n^N \sqcup \sigma_{-n}^N))] = -(1-\epsilon)p_n^1 + \frac{\epsilon \widehat{Y}}{n\widehat{B}} u_n^1.$$

Substract the first equation above from the second one, we have that before Step g

$$\begin{aligned} & \mathbb{E}[u_n(\mathcal{M}(\widehat{\sigma}_n^N \sqcup \sigma_{-n}^N))] - \mathbb{E}[u_n(\mathcal{M}(\sigma_n^N \sqcup \sigma_{-n}^N))] \\ = \ & (1-\epsilon)2P_j^\star + (\frac{\epsilon \widehat{Y}}{n\widehat{B}} - \frac{\epsilon Y}{nB}) u_n^1 > (1-\epsilon)\Delta_j^n > \epsilon \Delta_j^n > \epsilon_j^n \Delta_j^n, \end{aligned}$$

where the first inequality is because facts (1) and (2) we have deduced in this subcase, the fact that $u_n^1 > 0$, and that $P_j^\star = K_j^1 + \sum_{\ell=2}^{n-1} \Delta_j^\ell + \Delta_j^n > TV_j(\omega^\star) + \Delta_j^n \geq \Delta_j^n$; the second one is because $\Delta_j^n > 0$ and $1 - \epsilon > \epsilon > 0$ (since $0 < \epsilon < 1/5$); and the last one is because $\epsilon > \epsilon_j^n$. We are done in this subcase.

Subcase 1.2: $K_j^1 + \sum_{\ell=2}^{n-1} \Delta_j^\ell = TV_j(\omega^\star)$.

In this subcase, in execution $\widehat{\sigma}_n^N \sqcup \sigma_{-n}^N$, j announces YES and $\widehat{P}_j^\star = K_j^1 + \sum_{\ell=2}^{n-1} \Delta_j^\ell = TV_j(\omega^\star)$. Combining with O_3, we have that: (1) $\widehat{Y} = Y + 1$; (2) $B - \widehat{B} = P_j^\star \geq \Delta_j^n \geq 1$.

If $\widehat{Y} = n$, then before Step g,

$$\mathbb{E}[u_n(\mathcal{M}(\widehat{\sigma}_n^N \sqcup \sigma_{-n}^N))] = u_n^1 \quad \text{and} \quad \mathbb{E}[u_n(\mathcal{M}(\sigma_n^N \sqcup \sigma_{-n}^N))] = -(1-\epsilon)2P_j^\star + \frac{\epsilon Y}{nB} u_n^1$$

(notice that $p_n^1 = 0$ since $Y = n-1$ and j is the only player who announces NO). Accordingly, we have that

$$\mathbb{E}[u_n(\mathcal{M}(\widehat{\sigma}_n^N \sqcup \sigma_{-n}^N))] - \mathbb{E}[u_n(\mathcal{M}(\sigma_n^N \sqcup \sigma_{-n}^N))] = (1-\epsilon)2P_j^\star + (1 - \frac{\epsilon Y}{nB})u_n^1 > (1-\epsilon)\Delta_j^n > \epsilon_j^n \Delta_j^n,$$

where the first inequality is because $Y < n$, $\epsilon < 1$, $B \geq 1$, and $P_j^\star \geq \Delta_j^n$.

If $\widehat{Y} < n$, then before Step g,

$$\mathbb{E}[u_n(\mathcal{M}(\widehat{\sigma}_n^N \sqcup \sigma_{-n}^N))] = -(1-\epsilon)p_n^1 + \frac{\epsilon \widehat{Y}}{n\widehat{B}} u_n^1$$

and

$$\mathbb{E}[u_n(\mathcal{M}(\sigma_n^N \sqcup \sigma_{-n}^N))] = -(1-\epsilon)(p_n^1 + 2P_j^\star) + \frac{\epsilon Y}{nB} u_n^1.$$

Accordingly, we have that

$$\mathbb{E}[u_n(\mathcal{M}(\widehat{\sigma}_n^N \sqcup \sigma_{-n}^N))] - \mathbb{E}[u_n(\mathcal{M}(\sigma_n^N \sqcup \sigma_{-n}^N))] = (1-\epsilon)2P_j^\star + (\frac{\epsilon \widehat{Y}}{n\widehat{B}} - \frac{\epsilon Y}{nB})u_n^1 > (1-\epsilon)\Delta_j^n > \epsilon_j^n \Delta_j^n,$$

where the first inequality is because $\widehat{Y} > Y$, $\widehat{B} < B$, and $P_j^\star \geq \Delta_j^n$.

Therefore $\mathbb{E}[u_n(\mathcal{M}(\widehat{\sigma}_n^N \sqcup \sigma_{-n}^N))] - \mathbb{E}[u_n(\mathcal{M}(\sigma_n^N \sqcup \sigma_{-n}^N))] > \epsilon_j^n \Delta_j^n$ before Step g in this subcase.

To summarize, in Case 1 we have $\mathbb{E}[u_n(\mathcal{M}(\widehat{\sigma}_n^N \sqcup \sigma_{-n}^N))] - \mathbb{E}[u_n(\mathcal{M}(\sigma_n^N \sqcup \sigma_{-n}^N))] > \epsilon_j^n \Delta_j^n$ before Step g.

Case 2: $u_n^1 < 0$.

In this case, $P_n^\star > 0$, player n announces NO in Step n+1 in both executions, and thus before Step g,

$$\mathbb{E}[u_n(\mathcal{M}(\widehat{\sigma}_n^N \sqcup \sigma_{-n}^N))] = -(1-\epsilon)p_n^1 + \frac{\epsilon \widehat{Y}}{n\widehat{B}} u_n^1,$$

and

$$\mathbb{E}[u_n(\mathcal{M}(\sigma_n^N \sqcup \sigma_{-n}^N))] = -(1-\epsilon)(p_n^1 + 2P_j^\star) + \frac{\epsilon Y}{nB} u_n^1.$$

Subtracting the second equation from the first one, we have that before Step g,

$$\mathbb{E}[u_n(\mathcal{M}(\widehat{\sigma}_n^N \sqcup \sigma_{-n}^N))] - \mathbb{E}[u_n(\mathcal{M}(\sigma_n^N \sqcup \sigma_{-n}^N))] = (1-\epsilon)2P_j^\star + \frac{\epsilon \widehat{Y}}{n\widehat{B}} u_n^1 - \frac{\epsilon Y}{nB} u_n^1.$$

Because $0 > u_n^1 = TV_n(\omega^\star) - P_n^\star + \epsilon > -P_n^\star$, $B \geq P_n^\star$, and $Y < n$, we have that $0 > \frac{\epsilon Y}{nB} u_n^1 > -\epsilon$. Similarly, $0 > \frac{\epsilon \widehat{Y}}{n\widehat{B}} u_n^1 > -\epsilon$ also. Thus $\frac{\epsilon \widehat{Y}}{\widehat{B} n} u_n^1 - \frac{\epsilon Y}{Bn} u_n^1 > -\epsilon$, and

$$\begin{aligned} & \mathbb{E}[u_n(\mathcal{M}(\widehat{\sigma}_n^N \sqcup \sigma_{-n}^N))] - \mathbb{E}[u_n(\mathcal{M}(\sigma_n^N \sqcup \sigma_{-n}^N))] \\ > \ & (1-\epsilon)2P_j^\star - \epsilon \geq (1-\epsilon)\Delta_j^n - \epsilon = \epsilon \Delta_j^n + (1-2\epsilon)\Delta_j^n - \epsilon \\ > \ & \epsilon_j^n \Delta_j^n + 1 - 2\epsilon - \epsilon > \epsilon_j^n \Delta_j^n, \end{aligned}$$

where the second inequality is because $P_j^\star \geq \Delta_j^n$, the third one is because $\Delta_j^n \geq 1$ and $\epsilon > \epsilon_j^n$, and the last one is because $\epsilon < 1/5$. We are done in Case 2 also.

In sum, in both Case 1 and Case 2 we have that $\mathbb{E}[u_n(\mathcal{M}(\widehat{\sigma}_n^N \sqcup \sigma_{-n}^N))] - \mathbb{E}[u_n(\mathcal{M}(\sigma_n^N \sqcup \sigma_{-n}^N))] > \epsilon_j^n \Delta_j^n$ before Step g, which together with Equation 2 implies that $\mathbb{E}[u_n(\mathcal{M}(\widehat{\sigma}_n^N \sqcup \sigma_{-n}^N))] > \mathbb{E}[u_n(\mathcal{M}(\sigma_n^N \sqcup \sigma_{-n}^N))]$. Implication 1 follows.

No Underbidding.

Again we begin by giving a high level idea of the proof. Suppose player n underbids on another player j, that is he announces Δ_j^n such that $K_j^1 + \sum_{\ell=2}^{n-1} \Delta_j^\ell + \Delta_j^n < TV_j(\omega^\star)$. Consider an alternative strategy $\hat{\sigma}_n$ for player n in which he announces

$$\hat{\Delta}_j^n = TV_j(\omega^\star) - (K_j^1 + \sum_{\ell=2}^{n-1} \Delta_j^\ell)$$

and keeps the rest of his declarations.

The difference in utility between the strategies comes only from the reward, which is bigger in the latter strategy.

Now we prove Implication 2 for $h = 2$ (and thus $i = n$), that is if $K_j^1 + \sum_{\ell=2}^{n-1} \Delta_j^\ell < TV_j(\omega^\star)$, then σ_n^N consists of n announcing $\Delta_j^n = TV_j(\omega^\star) - K_j^1 - \sum_{\ell=2}^{n-1} \Delta_j^\ell$ at Step n. We proceed by contradiction, and assume that σ_n^N is a subgame-perfect strategy of n at G^N, and consists of n announcing $\Delta_j^n \neq TV_j(\omega^\star) - K_j^1 - \sum_{\ell=2}^{n-1} \Delta_j^\ell$. Consider the same alternative strategy $\widehat{\sigma}_n^N$ as defined before, we are going to show that $\mathbb{E}[u_n(\mathcal{M}(\sigma_n^N \sqcup \sigma_{-n}^N))] < \mathbb{E}[u_n(\mathcal{M}(\widehat{\sigma}_n^N \sqcup \sigma_{-n}^N))]$. Using the same rule to refer to variables in execution $\sigma_n^N \sqcup \sigma_{-n}^N$ and execution $\widehat{\sigma}_n^N \sqcup \sigma_{-n}^N$, we distinguish two cases.

Case 1: σ_n^N consists of n announcing $\Delta_j^n > TV_j(\omega^\star) - K_j^1 - \sum_{\ell=2}^{n-1} \Delta_j^\ell$.

In this case, notice that $\Delta_j^n > 0$ and $\widehat{\Delta}_j^n > 0$, $bip_j = \widehat{bip}_j = n$, j announces NO in execution $\sigma_n^N \sqcup \sigma_{-n}^N$, while announcing YES in execution $\widehat{\sigma}_n^N \sqcup \sigma_{-n}^N$. The remaining analysis is very similar to Subcase 1.2 and Case 2 in the proof of Implication 1 above, and thus ignored here.

Case 2: σ_n^N consists of n announcing $\Delta_j^n < TV_j(\omega^\star) - K_j^1 - \sum_{\ell=2}^{n-1} \Delta_j^\ell$.

In this case, in executions $\sigma_n^N \sqcup \sigma_{-n}^N$ and $\widehat{\sigma}_n^N \sqcup \sigma_{-n}^N$, the rewards that player n receives in Step g are $\epsilon + \sum_\ell \epsilon_\ell^n \Delta_\ell^n$ and $\epsilon + \sum_\ell \epsilon_\ell^n \widehat{\Delta}_\ell^n$ respectively. Since $\Delta_\ell^n = \widehat{\Delta}_\ell^n$ for all $\ell \neq j$ and $\Delta_j^n < TV_j(\omega^\star) - K_j^1 - \sum_{\ell=2}^{n-1} \Delta_j^\ell = \widehat{\Delta}_j^n$, we have that

$$\epsilon + \sum_\ell \epsilon_\ell^n \Delta_\ell^n < \epsilon + \sum_\ell \epsilon_\ell^n \widehat{\Delta}_\ell^n,$$

that is player n receives strictly less reward in execution $\sigma_n^N \sqcup \sigma_{-n}^N$ than in execution $\widehat{\sigma}_n^N \sqcup \sigma_{-n}^N$. Therefore it suffices to show that before Step g, $\mathbb{E}[u_n(\mathcal{M}(\sigma_n^N \sqcup \sigma_{-n}^N))] \leq \mathbb{E}[u_n(\mathcal{M}(\widehat{\sigma}_n^N \sqcup \sigma_{-n}^N))]$.

Since j announces YES in both executions, we have that $Y = \widehat{Y}$ and $B = \widehat{B}$. Thus before Step g, either $Y = n$ and $\mathbb{E}[u_n(\mathcal{M}(\sigma_n^N \sqcup \sigma_{-n}^N))] = \mathbb{E}[u_n(\mathcal{M}(\widehat{\sigma}_n^N \sqcup \sigma_{-n}^N))] = u_n^1$, or $Y < n$ and $\mathbb{E}[u_n(\mathcal{M}(\sigma_n^N \sqcup \sigma_{-n}^N))] = \mathbb{E}[u_n(\mathcal{M}(\widehat{\sigma}_n^N \sqcup \sigma_{-n}^N))] = -(1-\epsilon)p_n^1 + \frac{\epsilon Y}{nB} u_n^1$. That is, $\mathbb{E}[u_n(\mathcal{M}(\sigma_n^N \sqcup \sigma_{-n}^N))] = \mathbb{E}[u_n(\mathcal{M}(\widehat{\sigma}_n^N \sqcup \sigma_{-n}^N))]$ before Step g, despite the value of Y.

In sum, we have that $\mathbb{E}[u_n(\mathcal{M}(\sigma_n^N \sqcup \sigma_{-n}^N))] < \mathbb{E}[u_n(\mathcal{M}(\widehat{\sigma}_n^N \sqcup \sigma_{-n}^N))]$ in Case 2.

Combining Case 1 and Case 2, Implication 2 follows for $h = 2$.

Inductive Step. Now we prove that Lemma 2' is true for each node N of height $h \in [3, n]$, given that it is true for each node M of height $< h$. The proof is very similar to the base of the induction, and thus some of the cases are omitted.

No Overbidding. We first prove, by contradiction, Implication 1. Assume that there exist a decision node N of height h, a player $j \neq i$, and a restricted strategy profile σ^N for subgame G^N such that: (1) σ_i^N is a subgame-perfect strategy of player i in G^N; (2) for each proper subgame G^M of G^N and each player ℓ, σ_ℓ^M is the unique subgame-perfect strategy of ℓ at G^M; and (3) at node N, $K_j^1 + \sum_{\ell=2}^{i-1} \Delta_j^\ell \geq TV_j(\omega^\star)$, and σ_i^N consists of player i announcing $\Delta_j^i > 0$. Consider the following alternative strategy $\widehat{\sigma}_i$ for i (essentially $\widehat{\sigma}_i$ is the same as $\widehat{\sigma}_n$ defined before, except that every reference to player n now is to player i).

Strategy $\widehat{\sigma}_i$

Step i.	Run σ_i and compute Δ^i as σ_i does.
	For each player $\ell \neq j$, announce $\widehat{\Delta}_\ell^i = \Delta_\ell^i$.
	If $K_j^1 + \sum_{\ell=2}^{i-1} \Delta_j^\ell \geq TV_j(\omega^\star)$, then announce $\widehat{\Delta}_j^i = 0$.
	If $K_j^1 + \sum_{\ell=2}^{i-1} \Delta_j^\ell < TV_j(\omega^\star)$, then announce $\widehat{\Delta}_j^i = TV_j(\omega^\star) - K_j^1 - \sum_{\ell=2}^{i-1} \Delta_j^\ell$.
Step n+1.	If $P_i^\star = 0$, announce nothing.
	If $P_i^\star > 0$ and $TV_i(\omega^\star) \geq P_i^\star$, announce YES.
	Otherwise, announce NO.

We prove that $\mathbb{E}[u_i(\mathcal{M}(\sigma_i^N \sqcup \sigma_{-i}^N))] < \mathbb{E}[u_i(\mathcal{M}(\widehat{\sigma}_i^N \sqcup \sigma_{-i}^N))]$, which implies that σ_i^N is not a subgame-perfect strategy of i at G^N, contradicting our hypothesis about σ_i^N.

To do so, we define the following variables:

- $$u_i^1 = \begin{cases} TV_i(\omega^\star) - (K_i^1 + \sum_{\ell=2}^{i-1} \Delta_i^\ell) + \epsilon & \text{if} \quad K_i^1 + \sum_{\ell=2}^{i-1} \Delta_i^\ell > TV_i(\omega^\star); \\ \epsilon & \text{if} \quad K_i^1 + \sum_{\ell=2}^{i-1} \Delta_i^\ell \leq TV_i(\omega^\star). \end{cases}$$

  This is the utility that player i gets in both executions due to the provisional outcome $(\omega^\star, P^\star)$ being implemented. Notice that this definition is different from its counterpart u_n^1 defined before. This is because players $\ell > i$ will announce their Δ^ℓ after Step i, and $P_i^\star$ has not been defined yet in Step i. However, according to our induction assumption, the value of $P_i^\star$ has been fully determined by the values of $K_i^1 + \sum_{\ell=2}^{i-1} \Delta_i^\ell$ and $TV_i(\omega^\star)$:
  if $K_i^1 + \sum_{\ell=2}^{i-1} \Delta_i^\ell > TV_i(\omega^\star)$, then for each player $\ell > i$, σ_ℓ^N consists of ℓ announcing $\Delta_i^\ell = 0$ at step ℓ, and thus $P_i^\star = K_i^1 + \sum_{\ell=2}^{i-1} \Delta_i^\ell$ and $u_i^1 = TV_i(\omega^\star) - P_i^\star + \epsilon = TV_i(\omega^\star) - (K_i^1 + \sum_{\ell=2}^{i-1} \Delta_i^\ell) + \epsilon$;
  if $K_i^1 + \sum_{\ell=2}^{i-1} \Delta_i^\ell \leq TV_i(\omega^\star)$, then σ_{i+1}^N consists of player $i+1$ announcing $\Delta_i^{i+1} = TV_i(\omega^\star) - (K_i^1 + \sum_{\ell=2}^{i} \Delta_i^\ell) = TV_i(\omega^\star) - (K_i^1 + \sum_{\ell=2}^{i-1} \Delta_i^\ell)$ at Step $i+1$, and σ_ℓ^N consists of player ℓ announcing $\Delta_i^\ell = 0$ at Step ℓ for each $\ell > i+1$, which implies that $P_i^\star = TV_i(\omega^\star)$ and $u_i^1 = TV_i(\omega^\star) - P_i^\star + \epsilon = \epsilon$.

- $$p_i^1 = \sum_{\ell \neq j:\ \Delta_\ell^i > 0,\ K_\ell^1 + \sum_{k=2}^{i} \Delta_\ell^k > TV_\ell(\omega^\star)} 2\left(K_\ell^1 + \sum_{k=2}^{i} \Delta_\ell^k\right).$$

  This is the punishment that i pays to the mechanism in both executions due to players other than j announcing NO in Step n+1, when c_1 = Heads. Notice that this definition is different from its counterpart p_n^1. But according to our induction assumption, and by the fact that $\Delta_\ell^i = \widehat{\Delta}_\ell^i$ for all $\ell \neq j$, the punishment that player i pays after Step n+1 in both executions has been fully determined by the players' announcement till Step i:
  For each player $\ell \neq j$ such that $\Delta_\ell^i > 0$ and $K_\ell^1 + \sum_{k=2}^{i} \Delta_\ell^k > TV_\ell(\omega^\star)$, for each player $k > i$, σ_k^N consists of announcing $\Delta_\ell^k = 0$. Thus $bip_\ell = \widehat{bip}_\ell = i$, $P_\ell^\star = \widehat{P}_\ell^\star = K_\ell^1 + \sum_{k=2}^{i} \Delta_\ell^k$, σ_ℓ^N consists of ℓ announcing NO in Step n+1, and i is punished by $2P_\ell^\star$ in both executions.

For every other player ℓ, either $\Delta_\ell^i = 0$ and $bip_\ell = \widehat{bip}_\ell \neq i$, or $\Delta_\ell^i > 0$ and $K_\ell^1 + \sum_{k=2}^{i} \Delta_\ell^k \leq TV_\ell(\omega^\star)$. No matter which is the case, in both executions i is not punished due to the announcement of player ℓ in Step n+1.

Once u_i^1 and p_i^1 are defined, the remaining analysis for proving $\mathbb{E}[u_i(\mathcal{M}(\sigma_i^N \sqcup \sigma_{-i}^N))] < \mathbb{E}[u_i(\mathcal{M}(\widehat{\sigma}_i^N \sqcup \sigma_{-i}^N))]$ is almost the same as the analysis for Implication 1 when $h = 2$ and $i = n$, and is ignored here.

No Underbidding. Again, this is very similar to the base case.

To prove Implication 2 for $h \in [3, n]$, notice that if σ_i^N consists of i announcing $\Delta_j^i > TV_j(\omega^\star) - (K_j^1 + \sum_{\ell=2}^{i-1} \Delta_j^\ell)$, then $\mathbb{E}[u_i(\mathcal{M}(\sigma_i^N \sqcup \sigma_{-i}^N))] < \mathbb{E}[u_i(\mathcal{M}(\widehat{\sigma}_i^N \sqcup \sigma_{-i}^N))]$ for the same reasons as in Implication 1, with $\widehat{\sigma}_i^N$ defined as before. If σ_i^N consists of i announcing $\Delta_j^i < TV_j(\omega^\star) - (K_j^1 + \sum_{\ell=2}^{i-1} \Delta_j^\ell)$, as in Implication 2 when $h = 2$, we have that $\mathbb{E}[u_i(\mathcal{M}(\sigma_i^N \sqcup \sigma_{-i}^N))] = \mathbb{E}[u_i(\mathcal{M}(\widehat{\sigma}_i^N \sqcup \sigma_{-i}^N))]$ before Step g, but i receives strictly larger reward in Step g in execution $\widehat{\sigma}_i^N \sqcup \sigma_{-i}^N$, which implies $\mathbb{E}[u_i(\mathcal{M}(\sigma_i^N \sqcup \sigma_{-i}^N))] < \mathbb{E}[u_i(\mathcal{M}(\widehat{\sigma}_i^N \sqcup \sigma_{-i}^N))]$.

To summarize, we have that for all $h \in [2, n]$, Lemma 2' holds. $\square$

Lemma 2. *Let N be a node of height $h \in [2, n]$, $i = i_h$, and $\mathscr{C} = \mathscr{C}_i$. Then G^N has a unique subgame-perfect equilibrium where i acts as follows at node N: For each collusive set $\mathscr{D} \neq \mathscr{C}$,*

1. *if*
$$\sum_{j \in \mathscr{D}} \left(K_j^1 + \sum_{\ell=2}^{i-1} \Delta_j^\ell \right) \geq \sum_{j \in \mathscr{D}} TV_j(\omega^\star)$$
then i announces $\Delta_j^i = 0$ for all $j \in \mathscr{D}$;
2. *if*
$$\sum_{j \in \mathscr{D}} \left(K_j^1 + \sum_{\ell=2}^{i-1} \Delta_j^\ell \right) < \sum_{j \in \mathscr{D}} TV_j(\omega^\star)$$
then letting k be the minimal player in $\mathscr{D}$, i announces $\Delta_j^i = 0$ for all $j \in \mathscr{D} \setminus \{k\}$ and
$$\Delta_k^i = \sum_{j \in \mathscr{D}} \left(TV_j(\omega^\star) - K_j^1 - \sum_{\ell=2}^{i-1} \Delta_j^\ell \right).$$

For his own collusive set $\mathscr{C}$,

1. *if*
$$\sum_{j \in \mathscr{C}} \left(K_j^1 + \sum_{\ell=2}^{i-1} \Delta_j^\ell \right) \geq \sum_{j \in \mathscr{C}} TV_j(\omega^\star) \quad \textit{or}$$
it is the case that $k \in \mathscr{C}$ for all $k > i$,

then i announces $\Delta_j^i = 0$ for all $j \in \mathscr{C}$;
2. *if*
$$\sum_{j \in \mathscr{C}} \left(K_j^1 + \sum_{\ell=2}^{i-1} \Delta_j^\ell \right) < \sum_{j \in \mathscr{C}} TV_j(\omega^\star) \quad \textit{and}$$
there exists player $j > i$ such that $j \notin \mathscr{C}$,

then letting k be the minimal player in $\mathscr{C} \setminus \{i\}$, i announces
$$\Delta_k^i = \sum_{j \in \mathscr{C}} \left(TV_j(\omega^\star) - K_j^1 - \sum_{\ell=2}^{i-1} \Delta_j^\ell \right).$$

Proof. For each collusive set $\mathscr{D} \neq \mathscr{C}$, the reasoning is almost the same as in Lemma 2', with the following new points: For each player $j \in \mathscr{D}$ and each player $\ell \in \mathscr{C}$ such that $\ell > i$, we have that $\epsilon_j^i > \epsilon_j^\ell$, and thus letting i raise j's price gives $\mathscr{C}$'s members more reward than letting ℓ do so. Therefore given that ℓ (if exists) will raise $\mathscr{D}$'s members' price, it is still preferable for i to do so, whenever he has a chance (that is, whenever $\sum_{j\in\mathscr{D}}\left(K_j^1 + \sum_{\ell=2}^{i-1}\Delta_j^\ell\right) < \sum_{j\in\mathscr{D}} TV_j(\omega^\star)$), but he should never raise so much that he gets punished due to $\mathscr{D}$'s members announcing NO. Finally, i prefers to raise price for the minimal player k in $\mathscr{D}$, because doing so gives him the most reward ($\epsilon_k^i > \epsilon_{k'}^i$ for each $k' > k$), and will not affect $\mathscr{D}$'s members announcement in Step n+1, by Lemma 1.

For collusive set $\mathscr{C}$, we have the following intuition: For players in $\mathscr{C}$, if $\sum_{j\in\mathscr{C}}\left(K_j^1 + \sum_{\ell=2}^{i-1}\Delta_j^\ell\right) < \sum_{j\in\mathscr{C}} TV_j(\omega^\star)$ and all players $j > i$ are in $\mathscr{C}$, then nobody after i will raise price for $\mathscr{C}$'s members, and i prefers not raising either, since by raising his colluder ℓ's price by 1, he gets reward at most ϵ more, but ℓ loses utility 1, and the sum of their utilities decreases. While if $\sum_{j\in\mathscr{C}}\left(K_j^1 + \sum_{\ell=2}^{i-1}\Delta_j^\ell\right) < \sum_{j\in\mathscr{C}} TV_j(\omega^\star)$ and there exists a player $j > i$ not in $\mathscr{C}$, then j will raise price for $\mathscr{C}$'s members anyway, in which case i prefers doing so by himself, because this will not hurt his colluders' utilities but gives i himself more reward. Moreover, i prefers raising price for $\mathscr{C}$'s members by himself, instead of letting some colluder $i' > i$ do so, because $\epsilon_\ell^i > \epsilon_{\ell'}^{i'}$ for each ℓ and ℓ', and raising price by i gives $\mathscr{C}$'s members more reward. Of course, i will never raise so much that $\mathscr{C}$'s members jointly get negative utility. Finally, i prefers to raise price for the minimal player k in $\mathscr{C}\setminus\{i\}$, because he can not raise price for himself, and raising price for k gives him the most reward (since $\epsilon_k^i > \epsilon_{k'}^i$ for any $k' > k$).

The detailed proof is complicated and highly repetitive, and thus ignored here. □

Lemma 3′. *Let N be root of the tree (so that $G^N = G$), then G has a unique subgame-perfect equilibrium where player 1 acts as follows at node N:*

1. *player 1 announces $\omega^\star$, the lexicographically first state ω such that $\sum_\ell TV_\ell(\omega) = MSW$;*
2. *player 1 announces $K_i^1 = TV_i(\omega^\star)$ for each player i.*

Proof. We first prove Implication 2, that is, (actually no matter what $\omega^\star$ is,) player 1 announces $K_i^1 = TV_i(\omega^\star)$ for each player i. We proceed by contradiction. Assume there exist a subgame perfect equilibrium σ such that at node N, σ_1 consists of player 1 announcing K^1 such that $K_i^1 \neq TV_i(\omega^\star)$ for some player i. Consider the following alternative strategy $\widehat{\sigma}_1$.

Strategy $\widehat{\sigma}_i$

Step 1.	Announce $\widehat{\omega}^\star = \omega^\star$.
	Announce $\widehat{K}_\ell^1 = TV_\ell(\omega^\star)$ for each player ℓ.
Step n+1.	If $P_1^\star = 0$ or $bip_1 = 1$, announce nothing.
	If $bip_1 \neq 1$, $P_1^\star > 0$, and $TV_1(\omega^\star) \geq P_1^\star$, announce YES.
	If $bip_1 \neq 1$, $P_1^\star > 0$, and $TV_1(\omega^\star) < P_1^\star$, announce NO.

To emphasize the difference between execution σ and execution $\widehat{\sigma}_1 \sqcup \sigma_{-1}$, we write σ as $\sigma_1 \sqcup \sigma_{-1}$. We prove that $\mathbb{E}[u_1(\mathcal{M}(\sigma_1 \sqcup \sigma_{-1}))] < \mathbb{E}[u_1(\mathcal{M}(\widehat{\sigma}_1 \sqcup \sigma_{-1}))]$, which implies that σ_1 is not a subgame-perfect strategy of player 1 in G, contradicting our hypothesis about σ_1.

To do so, notice that the two executions $\sigma_1 \sqcup \sigma_{-1}$ and $\widehat{\sigma}_1 \sqcup \sigma_{-1}$ differ from Step 1. Accordingly, for every variable in the mechanism, we use different notations to refer to it in the two executions ($\omega^\star$ and $\widehat{\omega}^\star$, K^1 and $\widehat{K}^1$, Δ^ℓ and $\widehat{\Delta}^\ell$, $P^\star$ and $\widehat{P}^\star$, etc). It should be clear from the context to which execution a notation belongs.

We have the following three observations:

O_1: in both executions, in Step n+1, every player ℓ announces YES or NO consistently with Lemma 1′.

O_2: every player $\ell \neq 1$ announces Δ^ℓ and $\widehat{\Delta}^\ell$ in Step ℓ in both executions consistently with Lemma 2′. That is, for each player $k \neq \ell$, ℓ announces Δ_k^ℓ (respectively, $\widehat{\Delta}_k^\ell$) to be 0 if $K_k^1 + \sum_{j=2}^{\ell-1}\Delta_k^j \geq TV_k(\omega^\star)$ (respectively, if $\widehat{K}_k^1+\sum_{j=2}^{\ell-1}\widehat{\Delta}_k^j \geq TV_k(\widehat{\omega}^\star)$), and $TV_k(\omega^\star)-(K_k^1+\sum_{j=2}^{\ell-1}\Delta_k^j)$ (respectively, $TV_k(\widehat{\omega}^\star)-(\widehat{K}_k^1+\sum_{j=2}^{\ell-1}\widehat{\Delta}_k^j)$) otherwise.

O_3: in execution $\widehat{\sigma}_1 \sqcup \sigma_{-1}$, every player $\ell \neq 1$ announces $\widehat{\Delta}_k^\ell = 0$ for each $k \neq \ell$; and for every player ℓ (including player 1 himself), $\widehat{bip}_\ell = 1$, player ℓ announces YES in Step n+1, and player 1 is never punished.

O_4: in execution $\sigma_1 \sqcup \sigma_{-1}$, $K_1^1 + \sum_{k=2}^n \Delta_1^k \geq TV_1(\omega^\star)$.

Here O_1 is by our hypothesis of $\sigma_1 \sqcup \sigma_{-1}$, by construction of $\widehat{\sigma}_1$, and by Lemma 1′; O_2 is by our hypothesis of σ_{-1} and by Lemma 2′; O_3 is by O_2 and O_1, and by construction of $\widehat{\sigma}_1$; and O_4 is directly implied by O_2.

We distinguish 4 exhaustive cases according to σ_1.

Case 1: there exists $j \neq 1$ such that $K_j^1 > TV_j(\omega^\star)$.

Without loss of generality, assume that j is the only such player. In this case, in execution $\sigma_1 \sqcup \tau_{-1}$, $\sum_l K_l^1 \geq K_j^1 > 0$, and the mechanism $\mathcal{M}$ does not end at Step a. Therefore $bip_j = 1$ by O_2 and the mechanism, and $P_j^\star = K_j^1 > TV_j(\omega^\star)$. According to O_1, j announces NO in Step $n+1$, and player 1 is punished by $2K_j^1$ when $c_1 =$ Heads. Because $P_1^\star \geq TV_1(\omega^\star)$ by O_4, the utility that player 1 can get in Step f (when $c_1 =$ Tails and $c_2 =$ Heads) is at most ϵ. Further because the reward that player 1 can get in Step g is less than $\epsilon + 2\epsilon \sum_\ell K_\ell^1$, we have that

$$\mathbb{E}[u_1(\mathcal{M}(\sigma_1 \sqcup \tau_{-1}))] < -(1-\epsilon)2K_j^1 + \frac{\epsilon Y}{Bn}\cdot\epsilon + \epsilon + 2\epsilon \sum_\ell K_\ell^1 < -(1-3\epsilon)2K_j^1 + 2\epsilon + 2\epsilon \sum_{\ell\neq j} K_\ell^1 < 2\epsilon \sum_{\ell\neq j} K_\ell^1,$$

where the second inequality is because $Y \leq n-1 < n$, $B \geq K_j^1 \geq 1$, and $\epsilon < 1$; and the last one is because $K_j^1 \geq 1$ and $\epsilon < 1/5$.

In execution $\widehat{\sigma}_1 \sqcup \tau_{-1}$, if $\sum_l \widehat{K}_l^1 = 0$, then $\mathcal{M}$ ends at Step a with $\omega = \perp$ and $P_1 = 0$, therefore

$$\mathbb{E}[u_1(\mathcal{M}(\widehat{\sigma}_1 \sqcup \tau_{-1}))] = 0.$$

But then $\sum_{\ell\neq j} K_\ell^1 = 0$, and thus $\mathbb{E}[u_1(\mathcal{M}(\sigma_1 \sqcup \tau_{-1}))] < 0$.

If $\sum_l \widehat{K}_l^1 > 0$, then by O_3, $\widehat{Y} = n$ and $\widehat{P}_1^\star = \widehat{K}_1^1 = TV_1(\omega^\star)$. Accordingly, $\mathbb{E}[u_1(\mathcal{M}(\widehat{\sigma}_1 \sqcup \tau_{-1}))] = TV_1(\omega^\star) - \widehat{P}_1^\star + \epsilon = \epsilon$ before Step g, and player 1 gets reward at least $2\epsilon \sum_\ell \widehat{K}_\ell^1$ in Step g. By assumption, $\widehat{K}_\ell^1 = TV_\ell(\omega^\star) \geq K_\ell^1$ for each $\ell \neq j$, which implies

$$\mathbb{E}[u_1(\mathcal{M}(\widehat{\sigma}_1 \sqcup \tau_{-1}))] \geq \epsilon + 2\epsilon \sum_\ell \widehat{K}_\ell^1 \geq \epsilon + 2\epsilon \sum_{\ell\neq j} K_\ell^1.$$

Therefore in Case 1 we have that $\mathbb{E}[u_1(\mathcal{M}(\sigma_1 \sqcup \tau_{-1}))] < \mathbb{E}[u_1(\mathcal{M}(\widehat{\sigma}_1 \sqcup \tau_{-1}))]$.

Case 2: $K_l^1 \leq TV_l(\omega^\star)$ for all $l \neq 1$, but there exists $j \neq 1$ such that $K_j^1 < TV_j(\omega^\star)$.

In this case, in execution $\sigma_1 \sqcup \tau_{-1}$, by O_2, for each player $l \neq 1$, no matter whom bip_l is, we have that $P_l^\star = TV_l(\omega^\star)$ and l announces YES in Step $n+1$.

If $K_1^1 > TV_1(\omega^\star)$, then by O_2, $\Delta_1^l = 0$ for each player $l \neq 1$, $bip_1 = 1$, and $P_1^\star = K_1^1$. By the mechanism, player 1 announces YES "by default", therefore $Y = n$, and

$$\mathbb{E}[u_1(\mathcal{M}(\sigma_1 \sqcup \tau_{-1}))] = TV_1(\omega^\star) - P_1^\star + \epsilon = TV_1(\omega^\star) - K_1^1 + \epsilon$$

before Step g. Since the reward that player 1 gets in Step g is less than $\epsilon + 2\epsilon \sum_\ell K^1_\ell$, we have that

$$\begin{aligned}
& \mathbb{E}[u_1(\mathcal{M}(\sigma_1 \sqcup \tau_{-1}))] \\
< \ & TV_1(\omega^\star) - K^1_1 + 2\epsilon + 2\epsilon \sum_\ell K^1_\ell \\
= \ & -(1-2\epsilon)(K^1_1 - TV_1(\omega^\star)) + 2\epsilon + 2\epsilon \sum_{\ell \neq 1} K^1_\ell + 2\epsilon TV_1(\omega^\star) \\
\leq \ & -1 + 4\epsilon + 2\epsilon \sum_{\ell \neq 1} K^1_\ell + 2\epsilon TV_1(\omega^\star) \\
< \ & 2\epsilon \sum_{\ell \neq 1} K^1_\ell + 2\epsilon TV_1(\omega^\star),
\end{aligned}$$

where the second inequality is because $K^1_1 - TV_1(\omega^\star) \geq 1$, and the last inequality is because $\epsilon < 1/4$. But

$$\mathbb{E}[u_1(\mathcal{M}(\widehat{\sigma}_1 \sqcup \tau_{-1}))] \geq \epsilon + 2\epsilon \sum_\ell \widehat{K}^1_\ell \geq \epsilon + 2\epsilon \sum_{\ell \neq 1} K^1_\ell + 2\epsilon TV_1(\omega^\star),$$

where the first inequality is as in Case 1. Thus $\mathbb{E}[u_1(\mathcal{M}(\sigma_1 \sqcup \tau_{-1}))] < \mathbb{E}[u_1(\mathcal{M}(\widehat{\sigma}_1 \sqcup \tau_{-1}))]$.

If $K^1_1 \leq TV_1(\omega^\star)$, then $\mathbb{E}[u_1(\mathcal{M}(\sigma_1 \sqcup \tau_{-1}))] = 0$ if $\sum_l K^1_l = 0$. If $\sum_l K^1_l > 0$, then by O_2, no matter whom bip_1 is, we have that $P^\star_1 = TV_1(\omega^\star)$, and player 1 always announces YES in execution $\sigma_1 \sqcup \tau_{-1}$. Therefore $\mathbb{E}[u_1(\mathcal{M}(\sigma_1 \sqcup \tau_{-1}))] = \epsilon + \epsilon + 2\epsilon \sum_\ell K^1_\ell - \epsilon \frac{f_r(\omega^\star)}{n_r}$. That is,

$$\mathbb{E}[u_1(\mathcal{M}(\sigma_1 \sqcup \tau_{-1}))] \leq 2\epsilon + 2\epsilon \sum_\ell K^1_\ell - \epsilon \frac{f_r(\omega^\star)}{n_r}$$

no matter what $\sum_l K^1_l$ is. While in execution $\widehat{\sigma}_1 \sqcup \tau_{-1}$, by O_3,

$$\mathbb{E}[u_1(\mathcal{M}(\widehat{\sigma}_1 \sqcup \tau_{-1}))] = 2\epsilon + 2\epsilon \sum_\ell \widehat{K}^1_\ell - \epsilon \frac{f_r(\omega^\star)}{n_r}.$$

Since $\widehat{K}^1_l = TV_l(\omega^\star) \geq K^1_l$ for all $l \neq j$ and $\widehat{K}^1_j = TV_j(\omega^\star) > K^1_j$, we have that $\mathbb{E}[u_1(\mathcal{M}(\sigma_1 \sqcup \tau_{-1}))] < \mathbb{E}[u_1(\mathcal{M}(\widehat{\sigma}_1 \sqcup \tau_{-1}))]$.

In sum, in Case 2 we have $\mathbb{E}[u_1(\mathcal{M}(\sigma_1 \sqcup \tau_{-1}))] < \mathbb{E}[u_1(\mathcal{M}(\widehat{\sigma}_1 \sqcup \tau_{-1}))]$.

Case 3: $K^1_j = TV_j(\omega^\star)$ for all $j \neq 1$, but $K^1_1 > TV_1(\omega^\star)$.

In this case, similar to Case 2 when $K^1_1 > TV_1(\omega^\star)$, we have that

$$\mathbb{E}[u_1(\mathcal{M}(\sigma_1 \sqcup \tau_{-1}))] < 2\epsilon \sum_{\ell \neq 1} K^1_\ell + 2\epsilon TV_1(\omega^\star) < \mathbb{E}[u_1(\mathcal{M}(\widehat{\sigma}_1 \sqcup \tau_{-1}))].$$

Case 4: $K^1_j = TV_j(\omega^\star)$ for all $j \neq 1$, but $K^1_1 < TV_1(\omega^\star)$.

In this case, similar to Case 2 when $K^1_1 \leq TV_1(\omega^\star)$, we have that

$$\mathbb{E}[u_1(\mathcal{M}(\sigma_1 \sqcup \tau_{-1}))] \leq 2\epsilon + 2\epsilon \sum_\ell K^1_\ell - \epsilon \frac{f_r(\omega^\star)}{n_r} < \mathbb{E}[u_1(\mathcal{M}(\widehat{\sigma}_1 \sqcup \tau_{-1}))].$$

To summarize, we have that $\mathbb{E}[u_1(\mathcal{M}(\sigma_1 \sqcup \tau_{-1}))] < \mathbb{E}[u_1(\mathcal{M}(\widehat{\sigma}_1 \sqcup \tau_{-1}))]$ in all four cases. Therefore Implication 2 holds.

We now prove Implication 1, that is, player 1 announces $\omega^\star$ to be the lexicographically first state ω such that $\sum_\ell TV_\ell(\omega) = MSW$. We again proceed by contradiction. Assume there exists a subgame perfect equilibrium

σ such that at node N, σ_1 consistes of player 1 announcing $\omega^\star$ either with $\sum_\ell TV_\ell(\omega^\star) < MSW$, or with $\sum_\ell TV_\ell(\omega^\star) = MSW$ but $\omega^\star$ not being the lexicographically first such state.

By Implication 2, we have that $K^1_\ell = TV_\ell(\omega^\star)$ for each player ℓ. Consider the following alternative strategy $\widehat{\sigma}_1$ for player 1.

Strategy $\widehat{\sigma}_i$

Step 1.	Announce $\widehat{\omega}^\star = \mathrm{argmax}_{o\in\Omega} \sum_l TV_l(o)$, with ties broken lexicographically. Announce $K^1_l = TV_l(\widehat{\omega}^\star)$ for each player l.
Step $n+1$.	If $P^\star_1 = 0$ or $bip_1 = 1$, announce nothing. If $bip_1 \neq 1$, $P^\star_1 > 0$, and $TV_1(\omega^\star) \geq P^\star_1$, announce YES. If $bip_1 \neq 1$, $P^\star_1 > 0$, and $TV_1(\omega^\star) < P^\star_1$, announce NO.

We prove that $\mathbb{E}[u_1(\mathcal{M}(\sigma_1 \sqcup \sigma_{-1}))] < \mathbb{E}[u_1(\mathcal{M}(\widehat{\sigma}_1 \sqcup \sigma_{-1}))]$, which implies that σ_1 is not a subgame-perfect strategy of player 1 in G, contradicting our hypothesis about σ_1.

Similar to the proof of Lemma $2'$, for each variable in the mechanism, we refer to it using different notations in the two executions $\sigma_1 \sqcup \sigma_{-1}$ and $\widehat{\sigma}_1 \sqcup \sigma_{-1}$ (K^1 and $\widehat{K}^1$, $\omega^\star$ and $\widehat{\omega}^\star$, $P^\star$ and $\widehat{P}^\star$, etc). It should be clear from the context to which execution a notation belongs.

By Lemmas $1'$ and $2'$, we have that for each player $l \neq 1$, l announces $\Delta^l_k = \widehat{\Delta}^l_k = 0$ for each $k \neq l$. Therefore $bip_l = \widehat{bip}_l = 1$ for each player l, and l announces YES in Step $n+1$ in both executions. Similar to Case 2 of Implication 2, we have that

$$\mathbb{E}[u_1(\mathcal{M}(\widehat{\sigma}_1 \sqcup \sigma_{-1}))] = 2\epsilon + 2\epsilon \sum_\ell \widehat{K}^1_\ell - \epsilon\frac{f_r(\widehat{\omega}^\star)}{n_r}.$$

If $\mathbb{E}[u_1(\mathcal{M}(\sigma_1 \sqcup \sigma_{-1}))] < MSW$, then

$$\mathbb{E}[u_1(\mathcal{M}(\sigma_1 \sqcup \sigma_{-1}))] \leq 2\epsilon + 2\epsilon \sum_\ell K^1_\ell - \epsilon\frac{f_r(\omega^\star)}{n_r} \leq 2\epsilon + 2\epsilon \sum_\ell \widehat{K}^1_\ell - 2\epsilon - \epsilon\frac{f_r(\omega^\star)}{n_r} \leq 2\epsilon \sum_\ell \widehat{K}^1_\ell < \mathbb{E}[u_1(\mathcal{M}(\widehat{\sigma}_1 \sqcup \sigma_{-1}))].$$

If $\sum_\ell TV_\ell(\omega^\star) = MSW$, then

$$\mathbb{E}[u_1(\mathcal{M}(\sigma_1 \sqcup \sigma_{-1}))] = 2\epsilon + 2\epsilon \sum_\ell K^1_\ell - \epsilon\frac{f_r(\omega^\star)}{n_r} = 2\epsilon + 2\epsilon \sum_\ell \widehat{K}^1_\ell - \epsilon\frac{f_r(\omega^\star)}{n_r} < \mathbb{E}[u_1(\mathcal{M}(\widehat{\sigma}_1 \sqcup \sigma_{-1}))],$$

since $f_r(\omega^\star) > f_r(\widehat{\omega}^\star)$ by hypothesis.

In sum, we have that $\mathbb{E}[u_1(\mathcal{M}(\sigma_1 \sqcup \sigma_{-1}))] < \mathbb{E}[u_1(\mathcal{M}(\widehat{\sigma}_1 \sqcup \sigma_{-1}))]$. Therefore Implication 1 holds $\square$

Proving Lemma 3 based on Lemmas 1 and 2 is almost the same as proving Lemma $3'$ based on Lemmas $1'$ and $2'$, and thus ignored here.

B The Complexity of the GP Mechanism

In this appendix we give an example in which the GP mechanism requires an exponential number of rounds, although the number outcomes is two, there are five players, each player can have at most two types, and the mechanism is only required to succeed with constant probability (The same example also requires a double-exponential amount of communication in the AM mechanism).

Context Sketch We give an informal description of the context of the AM [1] and GP [11, 24] mechanisms. Let n denote the number of players. Each player p_i has a set T_i of all possible types for this player. The set T_i is finite, and is known to the designer. In the example we present $|T_i| = 2$ for every i. There is also a set of outcomes A, and for $t_i \in T_i$, and $\alpha \in A$ we denote by $t_i(\alpha)$ how much player i values the outcome α if his type is t_i. The GP mechanism requires that if $s, t \in T_i$, then there exists an outcome α such that $s(\alpha) \neq t(\alpha)$. In addition, each player has a true type $\Theta_i \in T_i$, which is unknown to the designer.

The designer is also given as input a *Social Choice Function*, which is a function from the type of all the players to outcomes

$$SCC:\ T_1 \times T_2 \times \ldots \times T_n \to A.$$

The goal of the designer is to guarantee that with probability at least $1-\epsilon$ the output of the mechanism is $SCC(\Theta_1, \ldots, \Theta_n)$.

Sketch of the GP Mechanism Following Abreu and Matsushima, Glazer and Perry design rely on a function f from types of the players to probability distribution over outcomes, such that for every two types $t_i, \tilde{t}_i \in T_i$, we have that if the player's true valuation is t_i, then he strictly prefers $f(t_i)$ to $f(\tilde{t}_i)$. The GP mechanism then proceeds by performing $k+1$ rounds, where in each round all the players speak, one after the other. In the first k rounds, every player declares the types of all the players, while in the last round each player declares only his own type. Finally, the mechanism chooses the outcome in the following manner:

1. Choosing the outcome:
   (a) With probability ϵ: round $k+1$ is chosen. Pick a random player i, and let t_i denote his declaration in round $k+1$. The outcome is $f(t_i)$ up to fines (which will be described later)
   (b) With probability $1-\epsilon$: pick a random round $1 \leq j \leq k$. If at least $n-1$ players declared the same vector of types $(t_1, \ldots, t_n)$ in this round, select the outcome $SCC((t_1, \ldots, t_n))$. Else, pick an arbitrary outcome O.
2. Choosing the fines:
   (a) Let $t_1, \ldots, t_n$ be the vector of types declared by the players in the final round.
   (b) Let p be the last player not to declare $t_1, \ldots t_n$, in the previous k rounds. That player pays a small fine δ.

Analysis of their Mechanism In order for this scheme to work, one must be careful about the choice of parameters. To define these parameters, for $t_i, \tilde{t}_i \in T_i$, let $L_f(t_i, t_j) = U(t_i, f(t_i)) - U(t_i, f(t_j))$ be the advantage that a player whose true type is t_i gains from reporting t_i and not t_j, where in computing A_{dv} we take the minimum over all players i and pairs of types. In addition, let $A_{dv} = \min_{\Theta_i, \Theta_j} L_f(\Theta_i, \Theta_j)$ be the minimum such advantage. Also, let B denote the maximal difference in valuation between the default outcome O, and the outcome dictated by the social choice $SCC(\Theta_1, \ldots \Theta_n)$. We require

$$\epsilon A_{adv}/n \geq \delta \geq B\frac{1}{k}$$

Given these constraints, the proof of Glazer Perry uses a backward induction argument to prove the existence of a unique subgame perfect equilibrium. In the last round, it is better for a player i to report his true valuation regardless of anything which happened in the previous rounds, as the small probability that the result will be $f(\Theta_i)$ and not $f(t_i)$ for some $t_i \neq \Theta_i$ is enough to overcome any fine which the player will need to bear. Given that we assume that all the players report the truth in the last round, a backwards induction shows that no player has incentive to deviate from the truth. Indeed, let i denote the last player to deviate from the truth, and suppose he does so in round j. In this case, consider the case in which the player does not deviate from the truth in that round. By telling the truth, player i can decrease the fine he pays, but also changes the output of the mechanism if round j is chosen in step 2. While decreasing the fine is always good for player i, he may either gain or loose by the new outcome. To guarantee that he doesn't loose too much by the new outcome, the GP mechanism relies on the low probability of this event, which is proportional to $1/k$. This requires k to be large.

Our Example Finally, we can present our example. There are two possible outcomes, α, β. We have $T_1 = T_2 = T_3 = T_4$, and $T_1 = \{t_\alpha, t_\beta\}$, where

$$t_\alpha(\alpha) = 1, \quad t_\alpha(\beta) = 0$$

and similarly

$$t_\beta(\alpha) = 0, \quad t_\beta(\beta) = 1.$$

That is, each one of the first four players has a slight preference for one of the alternatives. The fifth player is similar, except that he has a strong preference, that is $T_5 = \{\hat{t}_\alpha, \hat{t}_\beta\}$, where

$$\hat{t}_\alpha(\alpha) = 2^n, \quad \hat{t}_\alpha(\beta) = 0$$

and similarly

$$\hat{t}_\beta(\alpha) = 0, \quad \hat{t}_\beta(\beta) = 2^n.$$

Note that this preference only requires $O(n)$ bits to describe; still the number of rounds will be exponential.

Finally, the required *SCC* picks the alternative which is worse for most of the players. That is, if a majority of the players have type t_α or $\hat{t}_\alpha$ the SCC requires to output β, and vice versa.

Analysis of Our Example We begin by showing that the fine has to be less than ϵ. We do so by considering player 1. Let f_1 denote the function the mechanism uses to decide on the outcome in the $k+1$ round, given that player 1 was picked at random in the last round. Any such function must give $A_{dv} \leq 1$ (this can be obtained by $f_1(t_\alpha) = \alpha$, and $f_1(t_\beta) = \beta$, which is essentially giving player 1 his preferred outcome). However, since the proof requires that the player is truthful at the last round even if this makes him pay a fine, we must require that the fine $\delta < \epsilon/5$, his expected utility gain from being truthful.

Consider O, the default outcome outputted by the mechanism in case that round j is chosen at random, and there are no $n-1 = 4$ players who declared the same vector of valuations in that round. Since the types of the players and the SCC function are symmetric, we can assume wlog that $O = \alpha$. Suppose that the true types of the players are $\Theta_1 = \Theta_2 = \Theta_3 = \Theta_4 = t_\alpha$, and $\Theta_5 = \hat{t}_\alpha$. Following the induction proof, we can safely assume that since the fine is small, all the players will be truthful in round $k+1$. But what will happen in round k?

Consider the case in which the first three players declared the truth, and the fourth player lied. We want to argue that in this case the fifth player should declare the truth. If the fifth player declares the truth, and the k'th round is chosen, the outcome of the mechanism will be β. In this case, player 5 will not be punished, and player 4 will be.

On the other hand, if he chooses to lie, and the mechanism chooses round k, the outcome will be α. However, whether round k is being picked or not, player 5 will have to pay a fine of $\delta < \epsilon/5$. The expected difference in the expected utility between these options is $\frac{2^n}{k} - \delta$. To make this negative (so that the truthful option is chosen), we require $\delta > \frac{2^n}{k}$, which gives $k > 5 \cdot 2^n/\epsilon > 2^n$, as required.