Mechanism Design with Set-Theoretic Beliefs*

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

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

March 2, 2012

Abstract

In settings of incomplete information, we put forward

(1) a very conservative —indeed, purely set-theoretic— model of the beliefs (including totally wrong ones) that each player may have about the payoff types of his opponents, and

(2) a new and robust solution concept, based on *mutual* belief of rationality, capable of leveraging such conservative beliefs.

We exemplify the applicability of our new approach for single-good auctions. In particular we show that, under our solution concept, there exists a simple normal-form mechanism, which always sells the good, always has non-negative revenue, and guarantees (up to an arbitrarily small, additive constant) a revenue benchmark that is always greater than or equal to the second-highest valuation, and sometimes much greater. By contrast, we also prove that the same benchmark cannot even be approximated within any positive factor, under classical solution concepts.

*An earlier version of this work has been presented at the Symposium on Foundations of Computer Science, 2011. The present version is under review at the Journal of Economic Theory (special issue dedicated to the interface between economics and computer science).

We would like to thank Gabriel Carroll, Robert Kleinberg, and Ronald Rivest for discussions that motivated us to prove results stronger than the ones we originally had. We also would like to thank Amos Fiat and Anna Karlin for helping us improve the presentation of our results. Many thanks also to Andrés Perea for helping us understand beliefs in economic settings, to Paul Milgrom for helping us clarify the fragility of implementation at equilibria, and to Elchanan Ben-Porath, Sergiu Hart, and Philip Reny for helping us clarify our connections to ex-post equilibria. (Any remaining lack of clarity is solely due to us!) This work is supported in part by ONR Grant No. N00014-09-1-0597.

1 Introduction

We focus on settings of *incomplete* information. Here, a player i knows precisely θ_i, his own (payoff) type, but not θ_{-i}, the type subprofile of his opponents. Accordingly, he may have all kinds of beliefs (even wrong ones) about θ_{-i}. We refer to such beliefs as i's *external beliefs*, and to θ_i as his *internal knowledge*.

For achieving a desired goal, a mechanism designer should in general consider leveraging both the players' internal knowledge and their external beliefs. Mechanisms working in dominant or undominated strategies leverage the former, but not the latter.[1] Mechanisms using Bayesian Nash equilibrium as their underlying solution concept leverage both, but assume that the type profile θ is drawn from a distribution $\mathcal{D}$ that is (commonly) known to the players.

Independent of any additional assumptions (e.g., assumptions about the relationships among the players' individual beliefs), modeling a player i's own belief about the actual type profile θ as a distribution D_i imposes significant constraints. It is true that we focus on settings of incomplete information, and it is also true that modeling uncertainty by distributions is a traditional choice. Yet, we should always be cognizant that this traditional choice constitutes a strong assumption. A distribution D_i is a very *structured* form of incomplete information. In particular, it presupposes that player i can precisely compare any possible pair of type profiles θ' and θ'', and determine —say— that θ' is 3.2718 times more likely than θ''. Often, however, i may not have such structured beliefs. In a single-good auction, i may value the item for sale for 50 and believe that one of his opponents values for more than 100. Such a belief is not a distribution: indeed, i may not know whom such a high-valuing player might be, nor what the probabilities for his valuation being 101, 102, etc. might be. Such belief is not leverageable by Bayesian mechanisms, but it would be nice to be able to leverage it too, somehow.

In sum, classical mechanisms exploit two extremes, (1) the players have no external beliefs and (2) the players' external beliefs consist of probability distributions, but not the vast ground in between. Personally, we consider the first extreme as too pessimistic and the second as too optimistic, and wish to explore a "middle road" to mechanism design.

Our Contributions in a Nutshell We introduce a *set-theoretic* model for the beliefs that a player may have about his opponents. Our model is very conservative. In sharp contrast with the Bayesian setting, we do not even assume that there exist a player i and a pair of type subprofiles for his opponents such that i can tell which of the two subprofiles is more "likely" than the other.

As unstructured as such beliefs may be, we prove that it is possible to design mechanisms that successfully leverage them. Indeed, for single-good auctions we (1) define a new revenue benchmark that is always greater than or equal to the second-highest valuation, and sometimes much higher, and (2) show that this new benchmark can be guaranteed, by a simple mechanism, even when the designer has no information about the players' valuations or their beliefs.

Our mechanism only assumes the players' *mutual* (as opposed to "common") belief of rationality. The exact formalization of our solution concept, *conservative strict implementation*, is of independent interest. Indeed, we have been able to use it for other goals in other settings.

We also prove that our new revenue benchmark cannot even be meaningfully approximated under classical solution concepts, such as implementation in undominated strategies (and thus implementation in dominant strategies), or implementation in ex-post equilibrium. These impossibility results hold even if the designer is allowed to leverage the players' beliefs (e.g., via richer strategy spaces than those traditionally envisaged). Indeed, they hold because of the inability of classical solution concepts to leverage mutual belief of rationality.

Finally, we propose to enlarge the traditional definition of social choice correspondences so as to allow them to depend also on the players' beliefs, and not just on their types. This gives a mechanism designer a richer and meaningful set of "targets", possibly enabling him to "jump over higher bars".

[1] Whenever such mechanisms exist, they achieve their goals no matter what external beliefs the players may have.

Finiteness While our belief model and solution concept are very general, our theorems focus solely on single-good auctions where all valuations are non-negative integers upperbounded by some value V, and all mechanisms provide each player with a finite number of pure strategies.

2 Our Model

2.1 The Conservative-Belief Model

Definition 1. *A **conservative context** C consists of a tuple $(n, \Omega, \Theta, u, \theta, \mathscr{B})$, where*

- *$(n, \Omega, \Theta, u, \theta)$ is a traditional context of incomplete information,*[2] *and*
- *$\mathscr{B}$ is a profile such that, for each player i, (1) $\mathscr{B}_i \subseteq \Theta$ and (2) $t_i = \theta_i$ for all $t \in \mathscr{B}_i$.*

*We refer to $\mathscr{B}$ as the **conservative belief profile**, and say that $\mathscr{B}_i$ is **correct** if $\theta \in \mathscr{B}_i$.*

In a conservative context, $\mathscr{B}_i$ represents all possible candidates for the true type profile in player i's view. (We do not include the players' higher-level beliefs in our contexts because our solution concept prevents such beliefs from affecting a rational play of our mechanism.)

Knowledge and Beliefs Components n, Ω, Θ, and u are common knowledge to everyone. Each player i individually knows θ_i and $\mathscr{B}_i$, is rational, and believes that his opponents are rational. (Any unspecified knowledge and belief of players or mechanism designers can be chosen arbitrarily.)

Important Clarifications

1. *Conservative Beliefs Always Exist.* The conservative-belief profile is a *model* rather than an *assumption*. As usual, a player i knows θ_i, but we make no requirement about his external belief. For instance, he may have no external belief whatsoever. In this case, $\mathscr{B}_i = \Theta_1 \times \cdots \times \Theta_{i-1} \times \{\theta_i\} \times \Theta_{i+1} \times \cdots \times \Theta_n$. On the other extreme, he may have no external uncertainty whatsoever. In this case, $\mathscr{B}_i = \{t\}$ for some type profile t (not necessarily equal to θ).[3]
2. *Players' beliefs can be wrong.* Indeed it may even be the case that $\theta \notin \mathscr{B}_i$ for each player i.
3. *Compatibility with Additional Beliefs.* The profile $\mathscr{B}$ is compatible with the players having *additional* beliefs, even of a probabilistic nature, such as *partial distributions.* For example, a player i may believe that the probability of another player j's valuation being 100 is between 1/3 and 2/3. In no case, however, can these additional beliefs contradict $\mathscr{B}$. For instance, if a player i believes that the true type profile has been drawn from some distribution $\mathcal{D}$, then $\mathscr{B}_i$ should coincide with $\mathcal{D}$'s support.
   Let us stress that our mechanisms leverage $\mathscr{B}$ in order to achieve their goals, but work no matter what additional beliefs (compatible with $\mathscr{B}$) the players might have.
4. *External Beliefs and Payoff Types.* Relative to $\mathscr{B}$, the *external belief* of a player i, $\mathscr{E}_i$, is formally defined to be the set $\{t_{-i} : (\theta_i, t_{-i}) \in \mathscr{B}_i\}$. As a player i's type is a comprehensive description of i in the strategic situation at hand, we are essentially separating i's *payoff* type, θ_i, from his *external-belief* type, $\mathscr{E}_i$.

Conservative Single-Good Auction Contexts A *conservative single-good auction context* is a conservative context $(n, \Omega, \Theta, u, \theta, \mathscr{B})$ where: $\Theta = \{0, 1, \ldots, V\}^n$ for some positive integer V referred to as the *valuation bound*; $\Omega = \{0, 1, \ldots, n\} \times \mathbb{R}^n$;[4] and each utility function u_i is so defined: $u_i(t_i, (a, P))$ equals $t_i - P_i$ if $i = a$, and $-P_i$ otherwise.

[2] That is, $\{1, \ldots, n\}$ is the set of players; Ω the set of outcomes; $\Theta = \Theta_1 \times \cdots \times \Theta_n$ the set of all possible (payoff) type profiles; u the profile of utility functions, each u_i mapping $\Theta_i \times \Omega$ to $\mathbb{R}$, the set of reals; and $\theta \in \Theta$ the profile of true types. If $t_i \in \Theta_i$ and ω a distribution over Ω, then $u_i(t_i, \omega)$ is the expectation induced by ω.

[3] If the context were one of *complete information*, then necessarily $\mathscr{B}_i = \{\theta\}$ for all i.

[4] In an outcome (a, P), a denotes the player getting the good if > 0, or that the good is unallocated if $= 0$; and P denotes the price profile.

If $\omega = (a, P) \in \Omega$, then player i's utility for ω, $u_i(\omega)$, is $u_i(\theta_i, \omega)$; and the *revenue* of ω, $REV(\omega)$, is $\sum_i P_i$. We denote by $\mathscr{C}_n^V$ the set of all conservative single-good auction contexts with n players and valuation bound V, and by $\mathscr{D}_n^V$ the set of all contexts in $\mathscr{C}_n^V$ where the conservative belief of every player is correct.

Remarks

- Working solely in our model, we may drop the term "conservative" or use it for emphasis/clarity only. Further, since all auctions we consider are single-good, we may also omit the term "single-good."
- An auction context C is identified by n, V, θ and $\mathscr{B}$ alone: that is, $C = (n, V, \theta, \mathscr{B})$.
- In an auction context, a player i's true type θ_i —also called i's true *valuation*— represents i's value for the good for sale, and i's conservative belief $\mathscr{B}_i$ is a set of non-negative integer profiles.
- In the discussion below, given an underlying context, θ always represents the true type profile, while a type profile t can be an arbitrary element in a player's conservative belief $\mathscr{B}_i$.

2.2 Conservative-Belief Social Choice Correspondences and Their Advantages

Traditionally, social choice correspondences map *type profiles* to sets of (distributions over) outcomes, but can be naturally extended to map *conservative-belief profiles* to sets of outcomes. This extension strictly enriches the set of "targets" for mechanism design. As noted, each context C implicitly has a conservative-belief profile $\mathscr{B}$, from which the true type profile θ could be easily computed. Thus, for each traditional correspondence f there exists an extended correspondence F such that $f(\theta) = F(\mathscr{B})$, but not vice versa.

The advantage of a meaningful and enlarged "target space" is pretty clear. Very often we do not know how to design mechanisms implementing a given, traditional, social choice correspondence f. Sometimes we can actually prove that designing such mechanisms is impossible (at least for some type of implementation —e.g., in dominant strategies). In these cases, while one can always shop around for new, meaningful, and achievable targets among traditional social choice correspondences, extended social choice correspondences provide access to *additional* targets, which are more tractable, reasonable, but not expressible in terms of θ alone. For instance, in [8] we prove the existence of a very robust mechanism that, in any truly combinatorial auction and without any knowledge about the players' true valuations, generates within a factor of 2 the "maximum revenue that a player could guarantee if he were charged to sell the goods to his competitors by means of take-it-or-leave-it offers."

In this paper, instead of using conservative beliefs for achieving a social choice correspondence "tamer" than classical ones, we use them for (introducing and then) achieving a "tougher" one.

2.3 The Second-Belief Revenue Benchmark

In auction contexts, a *revenue benchmark* F is a function mapping each conservative belief profile $\mathscr{B}$ to a real number. Thus, *de facto*, F is a social choice correspondence: the one mapping each $\mathscr{B}$ to the set of outcomes whose revenue is at least $F(\mathscr{B})$.[5] Let us now define a revenue benchmark for single-good auctions.

Definition 2. *The* ***second-belief*** *benchmark, denoted by* 2^{nd}*, is the revenue benchmark so defined. Relative to given a belief profile* $\mathscr{B}$*, for every player* i *let the* sure maximum price *according to the belief of* i *be* $smp_i \triangleq \min_{t \in \mathscr{B}_i} \max_j t_j$*. Then,* $2^{nd}(\mathscr{B})$ *is the second highest value in* $\{smp_1, \ldots, smp_n\}$*.*

If t were the true valuation profile, then $\max_j t_j$ would be the maximum price that a player is willing to pay for the good. Thus, relative to $\mathscr{B}_i$, smp_i is the maximum price for which player i is *sure* that *some* player (possibly i or a player whose identity is not precisely known to i) is willing to pay for the good.

[5] Notice that we are slightly overloading the notation F here and in the previous subsection. When talking about a generic context F is a social-choice correspondence and $F(\mathscr{B})$ is a set of outcomes, and when talking about generating revenue in single-good auctions F is a revenue benchmark and $F(\mathscr{B})$ is a real number. Such overloading will not cause any ambiguity since in our discussion it is always clear whether we are talking about a general context or single-good auctions.

A Simple Example Consider an auction with three players where $\theta = (100, 80, 60)$ and

$$\mathscr{B}_1 = \{(100, x, y) : x \geq 0, y \geq 0\},\ \mathscr{B}_2 = \{(100, 80, x), (y, 80, 100) : x \geq 0, y \geq 0\},\ \text{and}\ \mathscr{B}_3 = \{(150, 0, 60)\}.$$

Here, the beliefs of players 1 and 2 are correct, but that of player 3 is wrong. Player 1 has no external beliefs: in his eyes, all valuations are possible for his two opponents. Player 2 believes that either player 1 or player 3 has valuation 100, but cannot tell whom. Player 3 has no external uncertainty: in his eyes, $(150, 0, 60)$ is the true valuation profile. According to $\mathscr{B}$, $smp_1 = smp_2 = 100$, $smp_3 = 150$, and thus $2^{nd}(\mathscr{B}) = 100$, which in this specific case happens to be the highest valuation.

Remark Sometimes $2^{nd}(\mathscr{B})$ can exceed the highest valuation, but never when all beliefs are correct. However, since $smp_i \geq \theta_i$ for every player i, it is always the case that "$2^{nd}(\mathscr{B}) \geq 2^{nd}(\theta)$": that is, our benchmark is always greater than or equal to the second highest true valuation. Accordingly, a mechanism designer concerned with generating revenue should try to achieve the second-belief benchmark instead of using the second-price mechanism to generate revenue equal to the second-highest valuation. If he succeeds, *the seller may have something (possibly a lot) to gain and nothing to lose.*

As we prove, however, this more demanding benchmark cannot be achieved via classical solution concepts.

3 Statement and Discussion of Our Results

3.1 The Impossibility of Classically Implementing the Second-Belief Benchmark

Recall that a mechanism M provides each player i with a set of pure strategies, consistently denoted by S_i in this paper, and maps each strategy profile σ to an outcome (or a distribution over outcomes, if M is probabilistic or σ a mixed-strategy profile) denoted by $M(\sigma)$. Also recall that a mechanism is finite if each S_i is finite, and that a game G consists of a context C and a mechanism M: $G = (C, M)$. Finally, when the mechanism M is clear, for any strategy profile σ, we may denote $u_i(M(\sigma))$ by $u_i(\sigma)$ for short.

For our impossibility results, we consider mechanisms that allow the players to "stay home", that is, to opt out of the auction. Otherwise, one could trivially and meaninglessly generate high revenue by forcing the players to participate in an auction mechanism always giving them very negative utility.

Definition 3. *A mechanism M is **reasonable** if it is finite and satisfies the following **opt-out condition**: $\forall$ player i $\exists out_i \in S_i$ such that for (any possible true type θ_i and) any strategy subprofile $s_{-i} \in S_{-i}$,*

$$u_i(M(out_i, s_{-i})) = 0.$$

Remarks

- Having the opt-out condition requiring i's utility to be 0 in expectation, rather than for every outcome in the support of $M(out_i, s_{-i})$, can only make our impossibility results stronger.
- Our impossibility results already hold for auctions with just two players, and when all beliefs are correct. Actually, when the players' beliefs are not correct these results become trivial.[6] Accordingly, we state our impossibility results in terms of $\mathscr{D}_n^V$ instead of $\mathscr{C}_n^V$.
- In our impossibility results we never assume any restrictions on the strategy spaces. In particular, our results also apply to normal-form mechanisms that let the players report their (alleged) conservative beliefs, as it is fair to do so when trying to leverage them.

[6] This is so because, when more than one player's beliefs are not correct, it is trivial to construct contexts for which the second-belief benchmark is much greater than the highest valuation. And no classical notion of implementation can guarantee revenue greater than the highest valuation.

3.1.1 Impossibility of Implementation in Undominated Strategies

Implementation in undominated strategies is a classical notion for settings of incomplete information.[7] We strengthen our first impossibility result by adopting a *weaker* notion of such implementation.[8] Notice that this weaker notion is already *sufficient* from a mechanism designer's point of view.

Definition 4. *A mechanism M **sufficiently** implements a revenue benchmark F for a class $\mathscr{C}$ of auction contexts in undominated strategies if, $\forall$ contexts $C \in \mathscr{C}$ and $\forall$ profiles s of undominated strategies in the game (C, M), denoting by $\mathscr{B}$ the belief profile of C, we have that*

$$REV(M(s)) \geq F(\mathscr{B}).$$

Theorem 1. *$\forall \epsilon \in (1/2, 1]$ and $\forall\ V > \lceil \frac{1}{\epsilon - 1/2} \rceil$, no reasonable mechanism sufficiently implements $\epsilon 2^{nd}$ for $\mathscr{D}_2^V$ in undominated strategies.*

For deterministic mechanisms and purely undominated strategies, our impossibility result holds for arbitrary approximation factors.

Theorem 2. *$\forall \epsilon \in (0, 1]$ and $\forall\ V > \lceil 1/\epsilon \rceil$, no reasonable deterministic mechanism sufficiently implements $\epsilon 2^{nd}$ for $\mathscr{D}_2^V$ in purely undominated strategies.*

We prove Theorem 1 in Section 5. The proof of Theorem 2 is similar (and simpler), and thus omitted.

3.1.2 Impossibility of Implementation in Dominant Strategies

Theorems 1 and 2 immediately yield the following about strictly/weakly/very weakly dominant strategies.[9]

Corollary 1. *$\forall \epsilon \in (1/2, 1]$ and $\forall\ V > \lceil \frac{1}{\epsilon - 1/2} \rceil$, no reasonable mechanism implements $\epsilon 2^{nd}$ for $\mathscr{D}_2^V$ in strictly/weakly dominant strategies or in (**all**) very weakly dominant strategies.*

Corollary 2. *$\forall \epsilon \in (0, 1]$ and $\forall\ V > \lceil 1/\epsilon \rceil$, no reasonable deterministic mechanism implements $\epsilon 2^{nd}$ for $\mathscr{D}_2^V$ in strictly/weakly dominant strategies or in (**all**) very weakly dominant strategies.*

3.1.3 A Crucial Clarification

Note that, in the absence of Theorems 1 and 2, the above two corollaries would be trivial if the players were restricted to bid valuations only. In such a case, in fact, the second-price mechanism is "the only" (weakly) dominant-strategy mechanism for auctions of a single good. And since the revenue it generates is precisely equal to the second-highest valuation, no other dominant-strategy mechanism can generate second-belief revenue. "QED." We thus wish to emphasize again that all our impossibility results hold without any restrictions on strategy spaces, and in particular that a mechanism asking the players to announce conservative beliefs cannot be "simulated" by one asking them to announce only valuations.

By allowing arbitrary strategy spaces, we explicitly allow the designer to leverage each player's external beliefs. However, as Theorems 1 and 2 show, when the designer does not leverage the players' mutual belief of rationality, he cannot hope to even approximate our benchmark.

---

[7] Given a game $G = (C, M)$, a strategy s_i of player i is *weakly dominated* by another (possibly mixed) strategy σ_i if $u_i(\sigma_i, s_{-i}) \geq u_i(s_i, s_{-i})$ for every strategy subprofile s_{-i} of the others, and $u_i(\sigma_i, s'_{-i}) > u_i(s_i, s'_{-i})$ for some strategy subprofile s'_{-i}. A strategy s_i is *undominated* if it is not weakly dominated by any strategy. A strategy s_i is *purely undominated* if it is not weakly dominated by any pure strategy. Thus, to compute his own undominated strategies in a game, a player needs not have any information about his opponents' (payoff) types.

[8] Note that the traditional notion of (full) implementation in undominated strategies —see Jackson [20]— requires not only that every profile of undominated strategies yields an outcome satisfying the desired social choice correspondence, but also that, conversely, for each desired outcome there exists a profile of undominated strategies yielding that outcome. By removing the latter requirement we weaken the notion of implementation and thus strengthen the impossibility result of Theorem 1.

[9] A strategy s_i of player i is *strictly dominant* if for every other strategy s'_i, $u_i(s_i, s_{-i}) > u_i(s'_i, s_{-i})$ for every strategy subprofile s_{-i}. Strategy s_i is *weakly dominant* if for every other strategy s'_i, $u_i(s_i, s_{-i}) \geq u_i(s'_i, s_{-i})$ for every s_{-i}, and the inequality is strict for some s_{-i}. Strategy s_i is *very weakly dominant* if for every other strategy s'_i, $u_i(s_i, s_{-i}) \geq u_i(s'_i, s_{-i})$ for every s_{-i}.

3.1.4 Extra Fragility of Implementation at Some Ex-Post/Very Weakly Dominant Equilibria

A mechanism guaranteeing a given property at *some* equilibria of a given type is certainly more fragile than one guaranteeing it at *all* equilibria of that type. Indeed, one has no control over the equilibrium ultimately selected by the players. But mechanisms implementing $\epsilon 2^{nd}$ at some ex-post or very weakly dominant equilibria have some **extra fragility**. Consider the following mechanism for $\mathscr{C}_2^{100}$.

> **Mechanism** NAIVE. *A strategy of player i has two components: an integer a_i and a set $b_i \subseteq \{0, 1, \ldots, 100\}$. (Allegedly, a_i is player i's true valuation, and b_i his true external belief.) The winner and prices are decided as follows. Let $w = \text{argmax}_i\, a_i$ (ties broken lexicographically), and let $P = \min_{t \in \mathscr{B}'_{-w}} \max_j t_j$ where $\mathscr{B}'_{-w} = \{a_{-w}\} \times b_{-w}$. If $a_w \geq P$, then the good is sold to player w, w pays P, and his opponent pays 0. Else, the good is unsold and both players pay 0.*

According to NAIVE, it is clear that every player announcing his true valuation and true external belief in every context is an ex-post equilibrium. When the players' beliefs are correct, this equilibrium guarantees second-belief revenue. However, consider the context C where

$$\theta = (70, 100),\ \mathscr{B}_1 = \{(70, x) : x \geq 90\},\ \text{and}\ \mathscr{B}_2 = \{(x, 100) : x \geq 60\}.$$

In this context, all beliefs are correct, $2^{nd}(\mathscr{B}) = 90$, the truthful ex-post equilibrium yields the strategy profile $((70, \{x : x \geq 90\}), (100, \{x : x \geq 60\}))$, and it generates revenue 90 as desired. However, it is also clear that $((70, \{x : x \geq 0\}), (100, \{x : x \geq 60\}))$ is an *alternative* Nash equilibrium —corresponding to another ex-post equilibrium— whose revenue is only 70.

In principle —e.g., when two Nash equilibria differ at multiple players, one can argue that a player may be able to establish some belief about which equilibrium is going to be played out by the others, and best respond to his belief. But in the above example, the "truthful" and the "alternative" equilibria differ only at player 1's strategy. Thus, even if player 1 believed that player 2 will play his truthful strategy, it would also be perfectly rational for player 1 to play his own alternative strategy. Viceversa, even if player 2 believed that player 1 will play his alternative strategy, it would also be perfectly rational for player 2 to stick to his own truthful strategy (which coincides with his alternative one in the above example).

Accordingly, *which revenue should we expect from* NAIVE *for context C?* The answer is 90 if player 1 is "generous" towards the seller and 70 otherwise.[10] In Appendix A, we formalize this phenomenon and prove, in Theorems 4 and 5, that such extra fragility is actually *unavoidable* for *any* mechanism implementing (or even approximating) the second-belief benchmark at some ex-post or very weakly dominant equilibria.

Clarification Although very weakly dominant equilibrium and ex-post equilibrium are very related notions, and sometimes one implies the other, there are games where they are not the same (see Appendix ?). Accordingly, our fragility theorems are explicitly stated for both.

3.2 Our New Solution Concept

The inability of achieving the second-belief benchmark via classical notions of implementation encourages us to develop a new one. Intuitively, but erroneously, our notion can be taken to consist of *"two-round elimination of strictly dominated strategies"* (hardly a new solution concept!). The problem is that such elimination is not well defined in a setting of *incomplete* information: without knowing his opponents' payoff types, a player is not capable of figuring out what strategies are left for them after the first round, and thus

[10] Notice that the truthful ex-post equilibrium actually specifies a very weakly dominant strategy for each player in each context, and thus illustrates the lack of robustness for implementation at some very weakly dominant equilibria as well. Such lack of robustness was already pointed out by Saijo, Sjostrom, and Yamato theoretically [24] and by Casona, Saijo, Sjostrom, and Yamato experimentally [6]. In [24] the authors also propose *secure implementation*: essentially, implementation via mechanisms ensuring that (a) each player has a very weakly dominant strategy, and that (b) the desired property holds at all Nash equilibria (and thus all very weakly dominant ones). As we have discussed, therefore, the second-belief revenue benchmark is not securely implementable.

is not capable of figuring out which of his own strategies are dominated in the second round. Therefore we must be more careful.

Sketch of Our Notion Our notion is formally defined in Section 6, but can be summarized as follows.

We say that a normal-form mechanism M *conservatively strictly implements* a social choice correspondence F for a class of contexts $\mathscr{C}$ if, for any context $C \in \mathscr{C}$, denoting by $\mathscr{B}$ the belief profile of C, we have $M(s) \in F(\mathscr{B})$ for any strategy profile s surviving the following two-step elimination procedure:

1. Each player eliminates all of his strictly dominated strategies;
2. Based on his conservative belief $\mathscr{B}_i$, and assuming that everyone completes Step 1, each player i eliminates all his remaining strategies that are *dominated relative to* $\mathscr{B}_i$.

The real novelty of our notion, and the key for meaningfully leveraging set-theoretic beliefs, lie with properly defining "domination relative to $\mathscr{B}_i$" in Step 2. As usual, after Step 1, to determine which of his remaining strategies are dominated, i should know what are the currently surviving strategies of the other players. However, to figure this out, player i must also know what are the true types of the other players —which is precisely a piece of information that he does not have in a setting of incomplete information. We address this concern by breaking down Step 2 into two conceptual sub-steps as follows.

2.1 Each player i, for each type profile t in $\mathscr{B}_i$, computes the profile $S(t)$, where each $S(t)_j$ represents the set of surviving strategies for player j after Step 1, if t were the true type profile.

2.2 Each player i eliminates a Step-1 surviving strategy s_i if and only if there exists another (possibly mixed) Step-1 surviving strategy σ_i that (classically) strictly dominates s_i with respect to $S(t)$ for *each* $t \in \mathscr{B}_i$.

Remark Let us emphasize a subtle point hidden in Step 2.2. Consider the following two ways of defining s_i to be "dominated relative to $\mathscr{B}_i$":

`(i)` for *each* $t \in \mathscr{B}_i$, s_i is strictly dominated with respect to $S(t)$ by *some* σ_i, and

`(ii)` for *each* $t \in \mathscr{B}_i$, s_i is strictly dominated with respect to $S(t)$ by *the same* σ_i.

Although both ways are based on the players' set-theoretic beliefs $\mathscr{B}$, we have adopted the latter one. The reason is that, when he eliminates a strategy s_i dominated according to `(ii)`, player i is sure to have a better strategy to play, namely σ_i, no matter which type profile in $\mathscr{B}_i$ might be the right one. But the same is not true when he eliminates a strategy dominated according to `(i)`.

Example[11] Consider a mechanism M played by two players, where the true type profile is $\theta = (\theta_1, \theta_2)$, and the belief of player 1 is $\mathscr{B}_1 = \{(\theta_1, \theta_2), (\theta_1, \theta_2')\}$. (Since we are going to analyze only player 1's behavior, we do not need to specify $\mathscr{B}_2$ nor the other possible type profiles.) The mechanism gives player 1 the pure strategies a, b, and c, and player 2 the pure strategies d and e. For each type profile in $\mathscr{B}_1$, the players' utilities under M are as follows.

(θ_1, θ_2)

1 \ 2	d	e
a	2,0	2,1
b	-100,0	3,1
c	3,0	-100,1

(θ_1, θ_2')

1 \ 2	d	e
a	2,1	2,0
b	-100,1	3,0
c	3,1	-100,0

Notice that, in Step 1 of our notion, player 1 cannot eliminate any strategy. Player 2 instead would eliminate d (strictly dominated by e) if his true type were θ_2, and e (strictly dominated by d) if his true type were θ_2'. Let us now consider Step 2. If we adopted definition `(i)` in Step 2.2, then player 1 should eliminate strategy a, because it is strictly dominated by b with respect to his candidate type profile (θ_1, θ_2), and by c with respect to his other candidate type profile (θ_1, θ_2'). However, whether player 1 should play b or c in place of a really depends on whether (θ_1, θ_2) or (θ_1, θ_2') is the true type profile. If he makes the wrong choice, then

[11] We thank Paul Valiant for this example.

his loss is huge compared with his possible gain: namely, -100 versus 3. Without any "likelihood" associated with each candidate type profile in his belief $\mathscr{B}_1$, it might be reasonable and safer for player 1 to use a to always get utility 2. (Thus, if M banked on player 1 not choosing a in order to implement its desired social choice correspondence, it may not implement it in a robust sense.)

Mutual Belief of Rationality Implementation in dominant or undominated strategies only requires that every player is rational. Conservative strict implementation instead additionally requires that every player believes that his opponents are rational. However, it does not require "higher-level" beliefs of rationality, let alone common belief. That is,

Conservative strict implementation solely relies on rationality and ***mutual*** *belief of rationality.*

In essence, our notion is only "slightly" weaker than implementation in strictly dominant strategies, yet is defined carefully to explicitly leverage the players' beliefs about others in a robust way.

3.3 The Second-Belief Benchmark is Conservatively Strictly Implementable

Finally, we prove that conservative strict implementation succeeds where classical notions fail. Namely, under our solution concept, we exhibit a mechanism $\mathcal{M}$, the *second–belief mechanism*, that guarantees second-belief revenue, within an arbitrarily small additive value ϵ, in all single-good auction contexts. Our mechanism is uniformly specified for all values ϵ, numbers of players n, and valuation bounds V: $\mathcal{M} = \mathcal{M}_{\epsilon,n,V}$. Formally,

Theorem 3. *For any $\epsilon \in (0, 1]$, n, and V, $\mathcal{M}_{\epsilon,n,V}$ conservatively strictly implements $2^{nd} - \epsilon$ for $\mathscr{C}_n^V$.*

The second-belief mechanism is defined in Section 7 and analyzed in Section 8. In Section 9 we address three concerns raised about our mechanism.

4 Related Work

In Bayesian settings with a common prior, higher revenue benchmarks can be guaranteed, and, more generally, more social choice functions can be implemented, under proper assumptions.[12] These works are not very relevant to ours, since we focus on a non-probabilistic model of incomplete information, and we do not impose any common knowledge assumption about the players' beliefs. Let us instead recall other works, where probabilistic/common-prior assumptions have been substantially relaxed.

Other Models of Incomplete Information Postlewaite and Schmeidler [23] studied *differential information* settings for exchange economies. They model a player's uncertainty as a partition of the set of all possible states of the world, and assume such partitions to be common knowledge. In our case, we do not assume a player to have any knowledge/beliefs about the knowledge/beliefs of another player, and we certainly do not have any common-knowledge requirements. In addition, they further assume that each player has a probabilistic distribution over the state space, and use Bayesian equilibrium as the key solution concept. Their model actually reduces to Harsanyi's incomplete information model [17] if the state space is finite.

Chung and Ely [11] model a player's belief about the state of the world via a *distribution*, but assume that he prefers one outcome ω to another ω' if he locally prefers ω to ω' in every state that is possible according to his belief. In this sense, what matters is the support of the distribution, which is set-theoretic. The authors show that, even when the players only have very small uncertainty about the state of the world, the set of social choice rules implementable at (essentially) undominated Nash equilibria is highly constrained

[12] For instance, Cremer and McLean [12] show that, for ceratin valuation distributions, revenue equal to the highest valuation can be achieved in a single-good auction under Bayesian Nash equilibrium or in weakly dominant strategies. Also, Abreu and Matsushima [2] show that, under some technical conditions, any Bayesian incentive compatible social-choice function can be virtually implemented in iteratively undominated strategies.

compared with that in complete-information settings. Their result is less relevant for settings, like ours, where a player has no uncertainty about his own payoff type. In addition, in our purely set-theoretic model, we have no requirement on how big a player's uncertainty about his opponents can be. Finally, instead of studying implementation at all equilibria (of a given type), we study the fragility of implementation even at some of them.

Artemov, Kunimoto, and Serrano [3] also model the players' beliefs about each other via distributions. But they assume that each player i's belief about the others' payoff types is from a subset Q_i of the set of all possible distributions, and that the Q_i's are common knowledge among the players. By doing so, they assume that the players have some knowledge about each other's first-order belief. They impose no constraint on the players' higher-order beliefs, and assume that no other player knows player i's true first-order belief. Their model is still different from ours. First of all, in our model a player's belief is set-theoretic instead of probabilistic. Second of all, we do not assume that the players have any knowledge about each other. Moreover, their model implicitly assumes that the players' knowledge about each other's first-order belief is *correct* —i.e., player i's true first-order belief is from Q_i, while in our model a player can have arbitrary, perhaps totally wrong, beliefs about others. Finally, the social-choice functions studied in [3] are still defined over the players' payoff types rather than their beliefs.

Our model of external information is also related to other notions in decision theory. In particular, Knight [21] and later Bewley [5] have considered players who have very incomplete information about *their own* type. Specifically, a Knightian player i does not know his own type θ_i, nor the distribution D_i from which θ_i has been drawn. Rather, he knows several distributions, one of which is guaranteed to be D_i. Recently Knightian players have also been studied in mechanism design, in particular, by Lopomo, Rigotti, and Shannon [22] for games with a single player, and by Chiesa, Micali, and Zhu [10] for auctions with multiple players.

Also, Hyafil and Boutilier [19] study regret-minimizing equilibria in games with multiple players having set-theoretic beliefs about each other. But they assume that the players' beliefs come from a common prior, and are always correct. Our model does not make these assumptions.

Impossibility Results Several impossibility results have been proved for implementation in dominant strategies: for instance, for many forms of elections (see Gibbard [13] and Sattherwaite [25]), for maximizing social welfare in a budget-balanced way (see Green and Laffont [16] and Hurwicz [18]), and for maximizing revenue in general settings of quasi-linear utilities (see Chen, Hassidim and Micali [7]). As for mechanisms working in undominated strategies, Jackson [20] shows that the set of social choice correspondences (fully) implementable by bounded mechanisms (which include finite ones) is quite constrained. We note, however, that none of these results imply ours for implementing the second-belief benchmark in either dominant or undominated strategies (indeed, our results do not require full implementation).

Prior-Free Mechanisms Prior-free mechanisms for auctions have also been investigated —in particular, by Baliga and Vohra [4], Segal [26], and Goldberg, Hartline, Karlin, Saks, and Wright [15], although the first two of them do not consider auctions of a single good. The term "prior-free" seems to suggest that this approach be relevant to our set-theoretic setting, but things are quite different. For instance, all cited prior-free mechanisms work in dominant strategies, and we have proved that no dominant-strategy mechanism can even approximate our revenue benchmark. More generally, as for all mechanisms, prior-free ones must be analyzed based on some underlying solution concept, and as long as they use one of the solution concepts we prove inadequate for our benchmark, they would automatically fail to guarantee it.

Our Own Prior Work In [8] we studied mechanisms leveraging only (what we now call) *external correct beliefs*, and, as already mentioned, constructed one such mechanism for truly combinatorial auctions. (This mechanism would also work with incorrect external beliefs, but under a slightly different analysis.) In a later work with Valiant [9], we were able to extend our combinatorial-auction mechanism so as to leverage also, to

a moderate extent, the internal knowledge of the players.[13] In neither of these two prior papers we proved any impossibility results: given that no significant revenue guarantees were known for combinatorial auctions, we were satisfied with achieving new, reasonable benchmarks. Perhaps interestingly, our prior mechanisms were of extensive form, and we still do not know whether equivalent, normal-form ones exist.

5 Proof of Theorem 1

Theorem 1. *$\forall \epsilon \in (1/2, 1]$ and $\forall\ V > \lceil \frac{1}{\epsilon - 1/2} \rceil$, no reasonable mechanism sufficiently implements $\epsilon 2^{nd}$ for $\mathscr{D}_2^V$ in undominated strategies.*

Proof. For sake of contradiction, assume that there exist a value $\epsilon \in (1/2, 1]$, an integer $V > \lceil \frac{1}{\epsilon - 1/2} \rceil$, and a reasonable (probabilistic) mechanism M that sufficiently implements $\epsilon 2^{nd}$ for $\mathscr{D}_2^V$ in undominated strategies. To derive the desired contradiction, letting H be an integer such that

$$V \geq H > \frac{1}{\epsilon - 1/2},$$

we construct two games, G and G', as follows.

1. $G = (C, M)$, where $C = (2, V, \theta, \mathscr{B})$ with $\theta = (H, 0)$ and $\mathscr{B}_1 = \mathscr{B}_2 = \{(H, 0)\}$.
   Note: $C \in \mathscr{D}_2^V$ because all beliefs are correct, and $2^{nd}(\mathscr{B}) = H$ because $smp_1 = smp_2 = H$.
2. $G' = (C', M)$, where $C' = (2, V, \theta', \mathscr{B}')$ with $\theta' = (1, 0)$ and $\mathscr{B}'_1 = \mathscr{B}'_2 = \{(1, 0)\}$.
   Note: $C' \in \mathscr{D}_2^V$ and $2^{nd}(\mathscr{B}') = 1$.

After analyzing the (auxiliary) game G', we derive our desired contradiction for G. To clarify the game to which a given quantity refers, we shall use the superscripts G and G'.

Let $UD^{G'} = UD_1^{G'} \times UD_2^{G'}$, where each $UD_i^{G'}$ is player i's set of undominated strategies in G'. Then, by hypothesis:

$$\forall s' \in UD^{G'}, REV(M(s')) \geq \epsilon 2^{nd}(\mathscr{B}') = \epsilon. \tag{1}$$

Denoting as usual by $\Delta(A)$ the set of probabilistic distributions over a set A, we now prove the following statement:

$$\exists \text{ a strategy } \sigma'_1 \in \Delta(UD_1^{G'}) \text{ such that } \forall \text{ strategy } s_2 \text{ of player 2, } u_1^{G'}(M(\sigma'_1, s_2)) \geq 0. \tag{2}$$

Because M satisfies the opt-out condition, player 1 has a strategy out_1 such that $u_1^{G'}(M(out_1, s_2)) = 0\ \forall s_2$. If $out_1 \in UD_1^{G'}$ then Statement 1 follows by taking $\sigma'_1 = out_1$. Otherwise, by the finiteness of M there exists $\sigma'_1 \in \Delta(UD_1^{G'})$ such that out_1 is weakly dominated by σ'_1, which implies $u_1^{G'}(M(\sigma'_1, s_2)) \geq u_1^{G'}(M(out_1, s_2)) = 0\ \forall s_2$, as desired.

Similarly, we have the following statement:

$$\exists \text{ a strategy } \sigma'_2 \in \Delta(UD_2^{G'}) \text{ such that } \forall \text{ strategy } s_1 \text{ of player 1, } u_2^{G'}(M(s_1, \sigma'_2)) \geq 0. \tag{3}$$

Combining Statements 2 and 3, letting ω' be the (possibly probabilistic) outcome $M(\sigma'_1, \sigma'_2)$, and letting p'_i and EP'_i respectively be the probability that player i gets the good and the expected price that i pays according to ω', we have that

$$u_1^{G'}(\omega') = p'_1 - EP'_1 \geq 0 \quad \text{and} \quad u_2^{G'}(\omega') = -EP'_2 \geq 0. \tag{4}$$

Because of Equation 1, and because $\sigma'_i \in \Delta(UD_i^{G'})$ for each i, we have

$$REV(\omega') = EP'_1 + EP'_2 \geq \epsilon. \tag{5}$$

[13] The emphasis of [9] actually was the possibility of leveraging the internal knowledge of coalitions rather than individual ones.

Combining Equations 4 and 5, we have

$$p'_1 \geq EP'_1 \geq \epsilon - EP'_2 \geq \epsilon. \tag{6}$$

Let us now analyze game G. Notice that, under the strategy profile (σ'_1, σ'_2), the (possibly probabilistic) outcome of M is still ω' in game G. Accordingly, following Equation 6 we have that

$$u_1^G(M(\sigma'_1, \sigma'_2)) = u_1^G(\omega') = p'_1 H - EP'_1 \geq p'_1 H - p'_1 \geq \epsilon(H-1),$$

where the second inequality holds further because $H > 1$.

Let $UD^G = UD_1^G \times UD_2^G$, where each UD_i^G is player i's set of undominated strategies in G. We now argue that there exists a strategy $\hat{\sigma}_1 \in \Delta(UD_1^G)$ such that

$$u_1^G(M(\hat{\sigma}_1, \sigma'_2)) \geq \epsilon(H-1). \tag{7}$$

To see why Inequality 7 is true, notice that if $\sigma'_1 \in \Delta(UD_1^G)$ then we can take $\hat{\sigma}_1 = \sigma'_1$. Otherwise, for each strategy s'_1 which is in the support of σ'_1 but not in UD_1^G, there exists $\sigma''_1 \in \Delta(UD_1^G)$ weakly dominating s'_1 in game G (again because M is finite). Thus, we can construct $\hat{\sigma}_1$ from σ'_1 by replacing each such s'_1 with the corresponding σ''_1, and the so constructed $\hat{\sigma}_1$ satisfies $u_1^G(M(\hat{\sigma}_1, \sigma'_2)) \geq u_1^G(M(\sigma'_1, \sigma'_2)) \geq \epsilon(H-1)$, as desired.

Because $\theta_2 = \theta'_2$, we have that player 2's set of undominated strategies is the same in G and G', and so is his utility for each possible outcome. That is,

$$UD_2^G = UD_2^{G'} \quad \text{and} \quad u_2^G(\cdot) = u_2^{G'}(\cdot). \tag{8}$$

Equations 3 and 8 directly imply the following statement:

$$\sigma'_2 \in \Delta(UD_2^G) \quad \text{and} \quad u_2^G(M(\hat{\sigma}_1, \sigma'_2)) \geq 0. \tag{9}$$

Let $\omega = M(\hat{\sigma}_1, \sigma'_2)$, and let p_i and EP_i respectively be the probability that player i gets the good and the expected price that i pays according to ω. Following Equation 7 and the inequality of Statement 9, we have

$$u_1^G(\omega) = p_1 H - EP_1 \geq \epsilon(H-1) \quad \text{and} \quad u_2^G(\omega) = -EP_2 \geq 0. \tag{10}$$

Combining Equation 10 with the facts that $0 \leq p_1 \leq 1$, $1/2 < \epsilon \leq 1$, and $H > \frac{1}{\epsilon - 1/2}$, we have

$$\begin{aligned} REV(\omega) &= EP_1 + EP_2 \leq EP_1 \leq p_1 H - \epsilon(H-1) = H(p_1 - \epsilon + \frac{\epsilon}{H}) \leq H(1 - \epsilon + \frac{1}{H}) \\ &< H(1 - \epsilon + \epsilon - 1/2) = H/2 < \epsilon H. \end{aligned}$$

Accordingly, there exists a strategy profile $\hat{s}$ such that: (1) $\hat{s}_1$ is in the support of $\hat{\sigma}_1$ and $\hat{s}_2$ is in the support of σ'_2, which imply that $\hat{s} \in UD^G$; and (2) $REV(M(\hat{s})) \leq REV(\omega) < \epsilon H = \epsilon 2^{nd}(\mathscr{B})$. That is, we have finally reached the desired contradiction against the hypothesis that M sufficiently implements $\epsilon 2^{nd}$ for $\mathscr{D}_2^V$ in undominated strategies. Thus Theorem 1 holds. ■

6 Conservative Strict Implementation

The following two auxiliary definitions envisage a game with context $C = (n, \Omega, \Theta, u, \theta, \mathscr{B})$ and mechanism M (whose strategy-profile set is denoted by S as usual).

Definition 5. *Let i be a player, t_i a type of i, and $T = T_1 \times \cdots \times T_n$ a set of pure strategy profiles. Then,*

- *We say that a strategy $s_i \in T_i$ is **strictly t_i-T-dominated** by another strategy $\sigma_i \in \Delta(T_i)$, in symbols $s_i <_T^{t_i} \sigma_i$, if for all strategy subprofiles $s_{-i} \in T_{-i}$, $u_i(t_i, M(s_i, s_{-i})) < u_i(t_i, M(\sigma_i, s_{-i}))$.*

- *We denote by $S(t_i)$ the set of pure strategies of i that are not strictly t_i-S-dominated, and, for any type profile t, we set $S(t) = S(t_1) \times \cdots \times S(t_n)$ and $S(t_{-i}) = S(t_1) \times \cdots \times S(t_{i-1}) \times S(t_{i+1}) \times \cdots \times S(t_n)$.*

Accordingly, s_i is strictly dominated by σ_i in the traditional sense if $s_i <_S^{\theta_i} \sigma_i$, and $S(t_i)$ represents the strategies of i that would survive elimination of strictly dominated strategies (in the traditional sense) if his true type were t_i. Also note that, for any $t \in \mathscr{B}_i$, $S(t_i) = S(\theta_i)$, because $t_i = \theta_i$, while $S(t_j)$ and $S(\theta_j)$ may be very different for $j \neq i$. Thus, in general $S(t) \neq S(\theta)$ for $t \neq \theta$.

Definition 6. *A strategy $s_i \in S(\theta_i)$ is **conservatively strictly dominated** if there exists another strategy $\sigma_i \in \Delta(S(\theta_i))$ that strictly θ_i-$S(t)$-dominates s_i for all $t \in \mathscr{B}_i$. Else, s_i is **conservatively strictly rational**.*

We are now ready to formalize our notion of implementation.

Definition 7. *We say that a mechanism M **conservatively strictly implements** a social choice correspondence F for a class of contexts $\mathscr{C}$ if, for all contexts $C \in \mathscr{C}$ and for all profiles s of conservatively strictly rational strategies in (C, M), denoting by $\mathscr{B}$ the belief profile of C, we have that $M(s) \in F(\mathscr{B})$.*

7 The Second-Belief Mechanism

For any $\epsilon \in (0, 1]$, n, and V, the mechanism $\mathcal{M}_{\epsilon,n,V}$ is described below. Note that the mechanism applies to any context in $\mathscr{C}_n^V$, and is of normal form because the players act simultaneously and only once: in Step **1**. Steps **a** through **e** are just "conceptual steps taken by the mechanism". The expression "$X := x$" denotes the operation that sets or resets variable X to value x.

Mechanism $\mathcal{M}_{\epsilon,n,V}$

a: *Set $a := 0$, and $P_i := 0$ for all players i.*

Comment. Upon termination, after all proper resettings, (a, P) will be the final outcome.

1: *Each player i, publicly and simultaneously with the others, announces a pair $(e_i, v_i) \in \{0, 1\} \times \{0, \ldots, V\}$.*

Comment. Allegedly, $v_i = smp_i$, and e_i indicates whether i's announcement is about his internal knowledge (allegedly $e_i = 0$ signifies that $v_i = \theta_i$), or about his external belief.

b: *If $v_i = 0$ for each i, then reset a to be a randomly chosen player, and halt.*

c: *Order the announced n pairs according to $v_1, \ldots, v_n$ decreasingly, breaking ties in favor of those with $e_i = 0$. If there are still ties among some pairs, then break them according to the corresponding players.*

Comment. It does not matter whether the players are ordered lexicographically (increasingly or decreasingly), or according to some other way.

d: *Set a to be the player corresponding to the first pair, and $P_a := \max\{\frac{1}{2}, \max_{j \neq a} v_j\}$.*

e: *For each player i, $P_i := P_i - \delta_i$, where $\delta_i = \frac{\epsilon}{4n}\left[\frac{v_i}{1+v_i} + \frac{1-e_i}{(1+V)^2}\right]$.*

Comment. Each player i receives a reward δ_i.

Remark

- Notice that $\mathcal{M}_{\epsilon,n,V}$ always sells the good.
- *Non-negative Revenue.* Notice that if $\mathcal{M}_{\epsilon,n,V}$ halts in Step **b** then its revenue is 0. Otherwise, its revenue equals the price charged to player a in Step **d** minus the total rewards given to the players in Step **e**. Because for each player i the reward that i receives in Step **e** is $\delta_i < \frac{\epsilon}{4n}(1 + 1) = \frac{\epsilon}{2n} \leq \frac{1}{2n}$, the total rewards given to the players in Step **e** is at most $\frac{1}{2}$. Because the price charged to player a in Step **d** is at least $\frac{1}{2}$, we have that $\mathcal{M}_{\epsilon,n,V}$ always has non-negative revenue.

- *Uniform Construction.* As promised, it is clear that $\mathcal{M}_{\epsilon,n,V}$ is uniformly and efficiently constructible on inputs ϵ, n, and V. In addition, it is very simple. It essentially consists of the second-price mechanism together with carefully designed rewards. In light of our impossibility results about implementing $\epsilon 2^{nd}$ under classical solution concepts, this simplicity suggests that conservative strict implementation can be quite powerful.
- *From Additive to Multiplicative ϵ.* Notice that the reward each player gets in Step **e** is at most $\frac{\epsilon}{2n}$. Thus if a player does not get the good, then his utility is at most $\frac{\epsilon}{2n}$. This is so because we aim at achieving the second belief revenue benchmark up to only an additive ϵ. If we are willing to give up an ϵ fraction of the revenue benchmark, then each player could receive a reward proportional to the second highest bid in the mechanism, so that his utility may still be very high even if he does not get the good. For instance, we can use $\delta_i = \frac{\epsilon \max_{j\neq a} v_j}{4n}\left[\frac{v_i}{1+v_i} + \frac{1-e_i}{(1+V)^2}\right]$.

8 Analysis of The Second-Belief Mechanism

Theorem 3. *For any $\epsilon \in (0,1]$, n, and V, $\mathcal{M}_{\epsilon,n,V}$ conservatively strictly implements $2^{nd} - \epsilon$ for $\mathscr{C}_n^V$.*

Proof. Arbitrarily fix $\epsilon \in (0,1]$, n, V, $C = (n, V, \theta, \mathscr{B}) \in \mathscr{C}_n^V$, and a strategy profile s. Denoting $\mathcal{M}_{\epsilon,n,V}$ by $\mathcal{M}$ for short, it suffices for us to prove that, if s is conservatively strictly rational in the game $(C, \mathcal{M})$, then

$$REV(\mathcal{M}(s)) \geq 2^{nd}(\mathscr{B}) - \epsilon. \tag{11}$$

Letting $s_i \triangleq (e_i, v_i)$ for each i, we start by proving three claims.

Claim 1. *$\forall$ player i and $\forall$ type $t_i \in \{0, \ldots, V\}$ of i, if $s_i \in S(t_i)$ then $v_i \geq t_i$.*

Proof of Claim 1. Assume for sake of contradiction that $s_i \in S(t_i)$ and $v_i < t_i$. We shall show that s_i is strictly t_i-S-dominated by $s_i' = (0, t_i)$. By definition, this implies $s_i \notin S(t_i)$, a contradiction. For this purpose, letting s'_{-i} be an arbitrary strategy subprofile of $-i$, it suffices to show that

$$u_i(t_i, (s_i, s'_{-i})) < u_i(t_i, (s_i', s'_{-i})).$$

To do so, let $s_j' = (e_j', v_j')$ for each $j \neq i$. Moreover, in the plays of (s_i, s'_{-i}) and (s_i', s'_{-i}) respectively, let (a, P) and (a', P') be the outcomes, and δ_i and δ_i' the rewards that player i receives in Step **e**.

Because $v_i \geq 0$ by the construction of $\mathcal{M}$ and $v_i < t_i$ by hypothesis, we have that $t_i \geq 1$ and $\mathcal{M}$ does not halt in Step **b** in the play of (s_i', s'_{-i}). Below we shall distinguish two exhaustive cases, according to the play of (s_i, s'_{-i}).

Case 1. $\mathcal{M}$ halts in Step **b** in the play of (s_i, s'_{-i}).

In this case, by the construction of $\mathcal{M}$ we have $v_i = 0$, $v_j' = 0$ for each $j \neq i$, and

$$u_i(t_i, (s_i, s'_{-i})) = \frac{t_i}{n}.$$

Now we consider the play of (s_i', s'_{-i}). Because $t_i \geq 1 > 0 = \max_{j\neq i} v_j'$, we have $a' = i$, $P_i' = \max\{\frac{1}{2}, \max_{j\neq i} v_j'\} - \delta_i' = \frac{1}{2} - \delta_i'$, and $\delta_i' = \frac{\epsilon}{4n}\left[\frac{t_i}{1+t_i} + \frac{1}{(1+V)^2}\right] > 0$. Accordingly,

$$u_i(t_i, (s_i', s'_{-i})) = t_i - P_i' = t_i - \frac{1}{2} + \delta_i' > t_i - \frac{1}{2} \geq \frac{t_i}{n},$$

where the second inequality holds because $t_i \geq 1$ and $n \geq 2$. Therefore $u_i(t_i, (s_i, s'_{-i})) < u_i(t_i, (s_i', s'_{-i}))$ as desired.

Case 2. $\mathcal{M}$ does not halt in Step **b** in the play of (s_i, s'_{-i}).

In this case, by the construction of $\mathcal{M}$ we have

$$\delta_i = \frac{\epsilon}{4n}\left[\frac{v_i}{1+v_i} + \frac{1-e_i}{(1+V)^2}\right] \quad \text{and} \quad \delta'_i = \frac{\epsilon}{4n}\left[\frac{t_i}{1+t_i} + \frac{1}{(1+V)^2}\right].$$

Accordingly,

$$\delta'_i - \delta_i = \frac{\epsilon}{4n}\left[\frac{t_i}{1+t_i} - \frac{v_i}{1+v_i}\right] + \frac{\epsilon}{4n}\left[\frac{1-(1-e_i)}{(1+V)^2}\right] = \frac{\epsilon}{4n}\left[\frac{t_i - v_i}{(1+t_i)(1+v_i)} + \frac{e_i}{(1+V)^2}\right] > 0,$$

where the inequality holds because $v_i < t_i$ by hypothesis and $e_i \geq 0$ by the construction of $\mathcal{M}$. Thus we have

$$\delta'_i > \delta_i.$$

Below we distinguish three exhaustive sub-cases.

Sub-case 2.1. $a' \neq i$.

In this sub-case, we also have $a \neq i$, because $v_i < t_i$. Accordingly, $P_i = -\delta_i$ and $P'_i = -\delta'_i$, and thus $u_i(t_i, (s_i, s'_{-i})) = \delta_i$ and $u_i(t_i, (s'_i, s'_{-i})) = \delta'_i$. Therefore $u_i(t_i, (s_i, s'_{-i})) < u_i(t_i, (s'_i, s'_{-i}))$ as desired.

Sub-case 2.2. $a' = i$ and $a = i$.

In this sub-case, we have $P'_i = \max\{\frac{1}{2}, \max_{j\neq i} v'_j\} - \delta'_i$ and $P_i = \max\{\frac{1}{2}, \max_{j\neq i} v'_j\} - \delta_i$. Because $\delta'_i > \delta_i$, we further have $P_i > P'_i$, which implies $u_i(t_i, (s_i, s'_{-i})) = t_i - P_i < t_i - P'_i = u_i(t_i, (s'_i, s'_{-i}))$ as desired.

Sub-case 2.3. $a' = i$ and $a \neq i$.

In this sub-case, we have $P'_i = \max\{\frac{1}{2}, \max_{j\neq i} v'_j\} - \delta'_i$, $P_i = -\delta_i$, and $t_i \geq \max_{j\neq i} v'_j$. As $t_i \geq 1$ by hypothesis, we further have $t_i \geq \max\{\frac{1}{2}, \max_{j\neq i} v'_j\}$. Accordingly, $u_i(t_i, (s_i, s'_{-i})) = -P_i = \delta_i < \delta'_i \leq (t_i - \max\{\frac{1}{2}, \max_{j\neq i} v'_j\}) + \delta'_i = t_i - P'_i = u_i(t_i, (s'_i, s'_{-i}))$ as desired.

In sum, $u_i(t_i, (s_i, s'_{-i})) < u_i(t_i, (s'_i, s'_{-i}))$ for any s'_{-i}, and thus s_i is strictly t_i-S-dominated by s'_i, contradicting the fact that $s_i \in S(t_i)$. Therefore Claim 1 holds. □

Claim 2. *$\forall$ player i and $\forall$ type $t_i \in \{1, \ldots, V\}$ of i, if $s_i = (1, t_i)$ then $s_i \notin S(t_i)$.*

Proof of Claim 2. By definition, it suffices for us to show that s_i is strictly t_i-S-dominated by strategy $s'_i = (0, t_i)$. For this purpose, arbitrarily fixing a strategy subprofile s'_{-i} of $-i$, it suffices to show that

$$u_i(t_i, (s_i, s'_{-i})) < u_i(t_i, (s'_i, s'_{-i})).$$

To do so, first notice that $\mathcal{M}$ does not halt in Step **b** in either the play of (s_i, s'_{-i}) or the play of (s'_i, s'_{-i}), because $t_i \geq 1$ by hypothesis. The analysis below is very similar to Case 2 of Claim 1. Indeed, in the plays of (s_i, s'_{-i}) and (s'_i, s'_{-i}) respectively, we denote by δ_i and δ'_i the rewards that player i receives in Step **e**, and by (a, P) and (a', P') the final outcomes. Letting $s'_j = (e'_j, v'_j)$ for each player $j \neq i$, we have

$$\delta'_i = \frac{\epsilon}{4n}\left[\frac{t_i}{1+t_i} + \frac{1}{(1+V)^2}\right] > \frac{\epsilon}{4n} \cdot \frac{t_i}{1+t_i} = \delta_i,$$

and we distinguish three cases as before:

- If $a' \neq i$, then $a \neq i$ as well, and we have

$$u_i(t_i, (s_i, s'_{-i})) = -P_i = \delta_i < \delta'_i = -P'_i = u_i(t_i, (s'_i, s'_{-i})).$$

- If $a' = i$ and $a = i$, then $P_i = \max\{\frac{1}{2}, \max_{j\neq i} v'_j\} - \delta_i > \max\{\frac{1}{2}, \max_{j\neq i} v'_j\} - \delta'_i = P'_i$, and we have

$$u_i(t_i, (s_i, s'_{-i})) = t_i - P_i < t_i - P'_i = u_i(t_i, (s'_i, s'_{-i})).$$

- Otherwise, we have that $a' = i$ and $a \neq i$, which implies

$$u_i(t_i, (s_i, s'_{-i})) = -P_i = \delta_i < \delta'_i \leq (t_i - \max\{\frac{1}{2}, \max_{j \neq i} v'_j\}) + \delta'_i = t_i - P'_i = u_i(t_i, (s'_i, s'_{-i})).$$

In sum, s_i is strictly t_i-S-dominated by s'_i, and Claim 2 holds. □

Claim 3. *$\forall$ player i, if s_i is conservatively strictly rational in game $(C, \mathcal{M})$, then $v_i \geq smp_i$.*

Proof of Claim 3. Assume for sake of contradiction that s_i is conservatively strictly rational and $v_i < smp_i$. By definition we have $s_i \in S(\theta_i)$, and thus by Claim 1 we have

$$v_i \geq \theta_i. \tag{12}$$

Let $s'_i = (1, smp_i)$. In order to reach a contradiction it suffices for us to prove the following statement:

$$\forall t \in \mathscr{B}_i, \forall s'_{-i} \in S(t_{-i}), u_i(\theta_i, (s_i, s'_{-i})) < u_i(\theta_i, (s'_i, s'_{-i})). \tag{13}$$

To see why this is sufficient, notice that if $s'_i \in S(\theta_i)$ then by definition Statement 13 implies that s_i is conservatively strictly dominated by s'_i, contradicting the hypothesis that s_i is conservatively strictly rational. If $s'_i \notin S(\theta_i)$, then by definition there exists a strategy $\sigma_i \in \Delta(S_i)$ such that

$$\forall s'_{-i} \in S_{-i}, u_i(\theta_i, (s'_i, s'_{-i})) < u_i(\theta_i, (\sigma_i, s'_{-i})).$$

In other words, s'_i is strictly dominated by σ_i in game $(C, \mathcal{M})$. Because $S(\theta_i)$ is the set of strategies that are not strictly dominated in game $(C, \mathcal{M})$, by well know properties of strict dominance we have that there exists a strategy $\sigma'_i \in \Delta(S(\theta_i))$ such that s'_i is strictly dominated by σ'_i in game $(C, \mathcal{M})$, that is, the following statement holds:

$$\forall s'_{-i} \in S_{-i}, u_i(\theta_i, (s'_i, s'_{-i})) < u_i(\theta_i, (\sigma'_i, s'_{-i})). \tag{14}$$

Because $S(t_{-i}) \subseteq S_{-i}$ for each $t \in \mathscr{B}_i$, Statements 13 and 14 together imply that

$$\forall t \in \mathscr{B}_i, \forall s'_{-i} \in S(t_{-i}), u_i(\theta_i, (s_i, s'_{-i})) < u_i(\theta_i, (\sigma'_i, s'_{-i})),$$

which by definition implies that s_i is conservatively strictly dominated by σ'_i, again contradicting the hypothesis that s_i is conservatively strictly rational.

Below we shall prove Statement 13. Arbitrarily fixing a type profile $t \in \mathscr{B}_i$ and a strategy subprofile $s'_{-i} \in S(t_{-i})$, it suffices to show

$$u_i(\theta_i, (s_i, s'_{-i})) < u_i(\theta_i, (s'_i, s'_{-i})).$$

To do so, let $\star(t) = \text{argmax}_j\, t_j$ with ties broken lexicographically. Because $t \in \mathscr{B}_i$ and $smp_i = \min_{t \in \mathscr{B}_i} \max_j t_j$, we have

$$t_{\star(t)} \geq smp_i.$$

Let $s'_j = (e'_j, v'_j)$ for each $j \neq i$. Because $s'_{\star(t)} \in S(t_{\star(t)})$, by Claim 1 we have that

$$v'_{\star(t)} \geq t_{\star(t)}.$$

Accordingly, the following sequence of inequalities holds:

$$v'_{\star(t)} \geq t_{\star(t)} \geq smp_i > v_i \geq \theta_i \geq 0, \tag{15}$$

where the third inequality holds by hypothesis, and the fourth is just Equation 12. By Sequence 15 we have $v'_{\star(t)} > 0$, thus $\mathcal{M}$ does not halt in Step **b** in the play of (s_i, s'_{-i}) or in the play of (s'_i, s'_{-i}). Below we consider the outcomes of the two plays.

Let (a, P) and (a', P') be the final outcomes of (s_i, s'_{-i}) and (s'_i, s'_{-i}) respectively. By Sequence 15 we have $v'_{\star(t)} > v_i$, thus $\star(t) \neq i$. If $v'_{\star(t)} > smp_i$, then by the construction of $\mathcal{M}$ we have that $(e'_{\star(t)}, v'_{\star(t)})$ is ordered before $(1, smp_i)$, and thus is also ordered before (e_i, v_i). If $v'_{\star(t)} = smp_i$, then by Sequence 15 we have $v'_{\star(t)} = t_{\star(t)}$. Also by Sequence 15 we have $t_{\star(t)} \geq 1$. Thus by Claim 2 we have $e'_{\star(t)} = 0$, which implies that $(e'_{\star(t)}, v'_{\star(t)})$ is ordered before $(1, smp_i)$, and thus is also ordered before (e_i, v_i). Accordingly, no matter what $v'_{\star(t)}$ is, we always have

$$a \neq i \quad \text{and} \quad a' \neq i,$$

therefore the utilities of player i only depend on his rewards in Step **e** in both plays.

Let δ_i and δ'_i be the rewards that player i receives in Step **e**, in the plays of (s_i, s'_{-i}) and (s'_i, s'_{-i}) respectively. We have

$$\begin{aligned}
\delta'_i - \delta_i &= \frac{\epsilon}{4n} \cdot \frac{smp_i}{1 + smp_i} - \frac{\epsilon}{4n}\left[\frac{v_i}{1+v_i} + \frac{1-e_i}{(1+V)^2}\right] \\
&= \frac{\epsilon}{4n}\left[\frac{smp_i - v_i}{(1+smp_i)(1+v_i)} - \frac{1-e_i}{(1+V)^2}\right] \geq \frac{\epsilon}{4n}\left[\frac{1}{(1+smp_i)(1+v_i)} - \frac{1}{(1+V)^2}\right] \\
&> \frac{\epsilon}{4n}\left[\frac{1}{(1+smp_i)^2} - \frac{1}{(1+V)^2}\right] \geq \frac{\epsilon}{4n}\left[\frac{1}{(1+V)^2} - \frac{1}{(1+V)^2}\right] = 0,
\end{aligned}$$

where the first inequality holds because $v_i < smp_i$ and $e_i \geq 0$, the second because $v_i < smp_i$, and the last because $smp_i \leq V$. Accordingly we have

$$\delta'_i > \delta_i,$$

which implies

$$u_i(\theta_i, (s_i, s'_{-i})) = \delta_i < \delta'_i = u_i(\theta_i, (s'_i, s'_{-i}))$$

as we wanted to show. Therefore Claim 3 holds. □

Now we are ready to prove that if s is conservatively strictly rational then Inequality 11 holds, which implies Theorem 3. Because s is conservatively strictly rational, by Claim 3 we have that

$$v_i \geq smp_i \text{ for each } i. \tag{16}$$

If $\mathcal{M}$ halts in Step **b**, then $v_i = 0$ for each i, which together with Equation 16 implies that $smp_i = 0$ for each i, and thus $2^{nd}(\mathscr{B}) = 0$. Accordingly,

$$REV(\mathcal{M}(s)) = 0 = 2^{nd}(\mathscr{B}) > 2^{nd}(\mathscr{B}) - \epsilon.$$

Otherwise, by Equation 16 we have that the second highest value in $\{v_1, \ldots, v_n\}$ is greater than or equal to the second highest value in $\{smp_1, \ldots, smp_n\}$, which is precisely $2^{nd}(\mathscr{B})$. By the construction of $\mathcal{M}$ we have that for each reward δ_i in Step **e**,

$$\delta_i = \frac{\epsilon}{4n}\left[\frac{v_i}{1+v_i} + \frac{1-e_i}{(1+V)^2}\right] < \frac{\epsilon}{4n} \cdot (1+1) = \frac{\epsilon}{2n}.$$

Letting (a, P) be the outcome of s, we have: (1) $P_a = \max\{\frac{1}{2}, \max_{j \neq a} v_j\} - \delta_a$, (2) $\forall i \neq a$, $P_i = -\delta_i$, and (3) $\max_{j \neq a} v_j$ is the second highest value in $\{v_1, \ldots, v_n\}$, which implies $\max\{\frac{1}{2}, \max_{j \neq a} v_j\} \geq 2^{nd}(\mathscr{B})$. Accordingly,

$$REV(\mathcal{M}(s)) = P_a + \sum_{i \neq a} P_i \geq 2^{nd}(\mathscr{B}) - \delta_a - \sum_{i \neq a} \delta_i > 2^{nd}(\mathscr{B}) - n \cdot \frac{\epsilon}{2n} > 2^{nd}(\mathscr{B}) - \epsilon.$$

Therefore Theorem 3 holds. *Q.E.D.*

Remark. If a player's belief is not correct, then according to mechanism $\mathcal{M}$ his utility may be negative and he may be "shocked" when seeing the final outcome. But when the game is played he believes that his utility will be non-negative and thus behaves as specified by our solution concept, in particular by Claim 3.

9 Three Concerns About the Second-Belief Mechanism "in Practice"

A concern raised about the second-belief mechanism is that "ϵ rewards" may not be enough motivation for the players to participate. When the relevant players opt to "stay at home", the second-belief benchmark cannot be guaranteed, and thus the second-price mechanism might in practice generate higher revenue.

Let us have a closer look. First, it should be appreciated that any rational player prefers a positive utility, no matter how small, to 0 utility, which is the utility he would receive if he opted out of the auction, both in the second-belief and the second-price mechanism. (Saying otherwise requires an alternative notion of rationality.[14]) Second, as we have already observed, conservative beliefs are implicit in any context, whether or not a designer tries to leverage them. Accordingly, to compare properly the second-belief and the second-price mechanism, one should consider the same, underlying, conservative belief profile $\mathscr{B}$. Consider a player i who does not believe that his valuation is the highest. Then i concludes that he will receive "ϵ utility" under the second-belief mechanism, and 0 utility under the second-price one. Therefore, according to any reasonable (traditional or not) notion of rationality, if i chooses to opt out in the second-belief mechanism, he should also opt out in the second-price mechanism. In neither mechanism, therefore, can player i be relied upon to achieve the corresponding revenue benchmark. Consider now a player i who believes that he might be the one with the highest valuation. Then, in either mechanism, it is dominant for him to participate in the auction. (In particular, in the second-belief mechanism, opting out is strictly dominated by $(0, \theta_i)$, which always has positive utility.) Accordingly, if i chooses to participate the second-price mechanism, he should also participate the second-belief one.

Another (related) concern was raised for the case in which the players only have very unprecise external beliefs. In this case, while the revenue generated by the second-price mechanism is equal to the second-highest valuation, $2^{nd}(\theta)$, the one generated by the second-belief mechanism is "$2^{nd}(\theta) - \epsilon$." Again, such a concern is based on an "unfair" comparison. The second-belief mechanism works no matter what beliefs the seller may have about the quality of the players' conservative beliefs, and insists on guaranteeing *strictly positive utilities* to the players (when they play conservatively and not all players have value 0). By contrast, the second-price mechanism only guarantees that the players' utilities are ≥ 0, and thus cannot guarantee the participation of players who believe that they do not have the highest valuation. Accordingly, for the seller to gain an extra ϵ in revenue by adopting the second-price mechanism instead of the second-belief one, it is necessary that he has enough information about the players: namely, *he must be sure that each player believes that he might be the one with the highest valuation.* In absence of this information, to guarantee the participation of all players, the second-price mechanism must be modified so as to provide some form of "ϵ rewards" as well, and thus will miss its target revenue in its purest form. To be sure, the second-price mechanism can be perturbed so that all players with non-zero valuations get strictly positive utilities and it is strictly dominant for them to participate. But then the revenue of the seller becomes "$2^{nd}(\theta) - \epsilon$" as well.

A third concern raised is that the second-belief mechanism may miss its benchmark because its players may prefer decreasing their opponents' utilities to increasing their own ones. Indeed, if (1) the player with the highest valuation is player i, (2) i believes that he is the player with the highest valuation, (3) i believes that $\theta_i \geq 2^{nd}(\mathscr{B})$, and (4) i further believes that $2^{nd}(\mathscr{B}) > 2^{nd}(\theta)$, then, when all other players act rationally, by sufficiently underbidding his own valuation —e.g., by bidding $(0, 0)$— player i will cause another player

[14] To be sure, such alternative notions exist: in particular, ϵ-Nash equilibrium. Note however that *any* mechanism which, like ours, achieves a revenue benchmark —at least in some contexts— close to the highest true valuation, *must* rely on the traditional notion of rationality, instead of any ϵ-alternative. This is so because, when the revenue benchmark equals the highest valuation minus ϵ, by definition the sum of the players' utilities must be at most ϵ. Therefore any ϵ-alternative notion of rationality will make the players indifferent between participating and opting out. And when players opt out, the mechanism cannot guarantee its desired benchmark.

to receive negative utility. However, let us emphasize that, while leveraging the players' external beliefs, we continue to use the *classical utility function* for single-good auctions: namely, the utility of every player equals his true valuation minus the price he pays if he wins the good, and 0 minus the price he pays otherwise. Under such a classical utility function, the second-belief mechanism achieves its benchmark at every rational play. The concern about a player having a different type of preference is therefore out of the model.

10 Future Directions

We believe that much work can be done in leveraging the players' set-theoretic beliefs.

Indeed, in a very recent work with Rafael Pass, we exhibit mechanisms that guarantee even higher revenue benchmarks (based on the players' set-theoretic higher-order beliefs), under different solution concepts.

Beyond single-good auctions, we plan to investigate what social choice correspondences can be implemented by leveraging the players' set-theoretic beliefs in other strategic settings.

Finally, we should investigate models where some of the players' beliefs are set-theoretic, and some are probabilistic, but without assuming the correctness of such beliefs, let alone their being common knowledge.

Appendix

A Extra Fragility of Implementation at Some Ex-Post/Very Weakly Dominant Equilibria

The notion of ex-post equilibrium, originally defined for Bayesian settings, is naturally extended to our set-theoretic setting.

Definition 8. *Given a class of contexts $\mathscr{C}$ and a mechanism M, an **ex-post equilibrium** $\boldsymbol{s}$ is a profile of functions, where each $\boldsymbol{s}_i$ maps player i's conservative beliefs to his (possibly mixed) strategies, such that: $\forall$ context $C \in \mathscr{C}$, denoting by $\mathscr{B}$ the belief profile of C and by $\boldsymbol{s}(\mathscr{B})$ the strategy profile $(\boldsymbol{s}_1(\mathscr{B}_1), \ldots, \boldsymbol{s}_n(\mathscr{B}_n))$, $\boldsymbol{s}(\mathscr{B})$ is a Nash equilibrium of the game (C, M).*

*If the range of each $\boldsymbol{s}_i$ only consists of pure strategies, then we say that $\boldsymbol{s}$ is a **pure** ex-post equilibrium.*

Note that ex-post equilibrium and very weakly dominant equilibrium are different notions.[15]

Definition 9. *A mechanism M **implements at some ex-post equilibrium** a social choice correspondence F for a class of contexts $\mathscr{C}$ if, $\exists$ an ex-post equilibrium $\boldsymbol{s}$ such that, $\forall$ context $C \in \mathscr{C}$, denoting by $\mathscr{B}$ the belief profile of C, we have that $M(\boldsymbol{s}(\mathscr{B})) \in F(\mathscr{B})$.*

*If $\boldsymbol{s}$ is the ex-post equilibrium required above, then we further say that M implements F **at** $\boldsymbol{s}$.*

Definition 10. *A mechanism M implementing a social choice correspondence F at some ex-post equilibrium for a class of contexts $\mathscr{C}$ is **fragile** if, $\forall$ ex-post equilibrium $\boldsymbol{s}$ at which M implements F, $\exists$ another ex-post equilibrium $\boldsymbol{s}'$ satisfying the following two properties:*

(1) $\exists$ a player i and a conservative belief $\mathscr{B}_i$ of i such that $\boldsymbol{s}'$ and $\boldsymbol{s}$ differ only at $\mathscr{B}_i$;[16] and

(2) $\forall$ context $C = (n, \Omega, \Theta, u, \theta', \mathscr{B}')$ in $\mathscr{C}$ such that $\mathscr{B}'_i = \mathscr{B}_i$, we have that $M(\boldsymbol{s}'(\mathscr{B}')) \notin F(\mathscr{B}')$.

(The notion of *implementation at some very weakly dominant equilibrium* and the corresponding notion of fragility are similarly defined.)

[15] If for each player i and each strategy s_i there exists $\mathscr{B}_i$ such that $\mathbf{s}_i(\mathscr{B}_i) = s_i$, then for each $\mathscr{B}$ the strategy profile $\mathbf{s}(\mathscr{B})$ is also a very weakly dominant equilibrium. But otherwise not.

[16] That is, $\mathbf{s}_i(\mathscr{B}_i) \neq \mathbf{s}'_i(\mathscr{B}_i)$; $\mathbf{s}_i(\mathscr{B}'_i) = \mathbf{s}'_i(\mathscr{B}'_i)$ for all $\mathscr{B}'_i \neq \mathscr{B}_i$; and $\mathbf{s}_j = \mathbf{s}'_j$ for all $j \neq i$.

Theorem 4. *$\forall \epsilon \in (0,1]$ and $\forall\ V > 6/\epsilon^2$, any reasonable deterministic mechanism implementing $\epsilon 2^{nd}$ for $\mathscr{D}_2^V$ at some pure ex-post/very weakly dominant equilibrium is fragile.*

Theorem 5. *$\forall \epsilon \in (\frac{1}{2},1]$ and $\forall\ V > \lceil \frac{1}{\epsilon - 1/2} \rceil$, any reasonable mechanism implementing $\epsilon 2^{nd}$ for $\mathscr{D}_2^V$ at some ex-post/very weakly dominant equilibrium is fragile.*

We prove Theorems 4 and 5 for ex-post equilibrium only, as the proof for very weakly dominant equilibrium is almost the same. (Notice that the lower-bound of V in Theorem 4 depends on ϵ^2 rather than being linear in ϵ as in the other impossibility results. This is not a crucial difference, and is only because in the proof of Theorem 4 we require that player 2's true valuation is > 1, instead of ≥ 0 as typically assumed for auctions.)

A.1 Proof of Theorem 4

Let ϵ be a value in $(0,1]$, V an integer greater than $6/\epsilon^2$, and M a reasonable deterministic mechanism implementing $\epsilon 2^{nd}$ for $\mathscr{D}_2^V$ at some pure ex-post equilibrium $\boldsymbol{s}$. Further, let

- θ_2^* be a positive integer such that $3\theta_2^*/\epsilon^2 < V$, and
- $\mathscr{B}_2^* = \{(x, \theta_2^*) : x \geq \lceil 3\theta_2^*/\epsilon^2 \rceil\}$.

Notice that the desired θ_2^* exists because $V > 6/\epsilon^2$. Notice also that there exists some context in $\mathscr{D}_2^V$ with player 2's conservative belief being $\mathscr{B}_2^*$.

We prove that there exists another pure ex-post equilibrium $\boldsymbol{s}'$ such that:

(1) $\boldsymbol{s}$ and $\boldsymbol{s}'$ differ only at the conservative belief $\mathscr{B}_2^*$ of player 2; and

(2) for every context $C = (n, V, \theta, \mathscr{B}) \in \mathscr{D}_2^V$ with $\mathscr{B}_2 = \mathscr{B}_2^*$, $REV(M(\boldsymbol{s}'(\mathscr{B}))) < \epsilon 2^{nd}(\mathscr{B})$.

To do so, we analyze two (classes of) related games G and G'. Again to clarify the game to which a given quantity refers, we shall use the superscripts G and G'.

$$\boxed{G = (C, M), \text{ where } C = (2, V, \theta, \mathscr{B}) \text{ is an arbitrary context in } \mathscr{D}_2^V \text{ with } \mathscr{B}_2 = \mathscr{B}_2^*.}$$

In G we have that $smp_1 \geq \theta_1$ and $smp_2 = \lceil 3\theta_2^*/\epsilon^2 \rceil$. Because C has correct conservative beliefs, we have

$$\theta_1 \geq \lceil 3\theta_2^*/\epsilon^2 \rceil, \qquad \theta_2 = \theta_2^*, \qquad \text{and} \qquad \theta \in \mathscr{B}_1.$$

Letting $(a, P) = M(\boldsymbol{s}(\mathscr{B}))$, we claim (and prove later) that the following (in)equalities hold:

(g) $\quad 2^{nd}(\mathscr{B}) \geq 3\theta_2^*/\epsilon^2, \qquad REV(a, P) \geq 3\theta_2^*/\epsilon, \qquad a = 1, \qquad 3\theta_2^*/\epsilon \leq P_1 \leq \theta_1, \qquad \text{and} \qquad P_2 \leq 0.$

$$\boxed{G' = (C', M), \text{ where } C' = (2, V, \theta', \mathscr{B}') \text{ with } \theta' = \theta,\ \mathscr{B}_1' = \mathscr{B}_1, \text{ and } \mathscr{B}_2' = \{(x, \theta_2) : x \geq \lceil 2\theta_2^*/\epsilon \rceil\}.}$$

Notice that $C' \in \mathscr{D}_2^V$. In G' we have that $smp_1' \geq \theta_1' \geq \lceil 3\theta_2^*/\epsilon^2 \rceil$ and $smp_2' = \lceil 2\theta_2^*/\epsilon \rceil$. Letting $(a', P') = M(\boldsymbol{s}(\mathscr{B}'))$, we claim that the following (in)equalities hold:

(g′) $\quad 2^{nd}(\mathscr{B}') \geq 2\theta_2^*/\epsilon, \qquad REV(a', P') \geq 2\theta_2^*, \qquad a = 1, \qquad 2\theta_2^* \leq P_1' \leq \theta_1', \qquad \text{and} \qquad P_2' \leq 0.$

Let us explain why all 5 (in)equalities in (g) hold. The first follows because $2^{nd}(\mathscr{B}) = smp_2$. The second follows from our hypothesis of M and $\boldsymbol{s}$. Now notice that, because M satisfies the opt-out condition and $\boldsymbol{s}(\mathscr{B})$ is an equilibrium of G, we have $u_1^G(a, P) \geq 0$ and $u_2^G(a, P) \geq 0$. Accordingly, because there is only a single good, at most one player can pay a positive price. That is, the revenue must come from a single player, and must be at least $3\theta_2^*/\epsilon$. Because this player has to get non-negative utility, he must be player 1, and must get the good. Thus the other (in)equalities of (g) hold. The (in)equalities of (g′) hold for similar reasons.

We construct the desired pure ex-post equilibrium $\boldsymbol{s}'$ as follows: $\boldsymbol{s}_2'(\mathscr{B}_2^*) = \boldsymbol{s}_2(\mathscr{B}_2')$, and $\boldsymbol{s}'$ coincides with $\boldsymbol{s}$ everywhere else. By construction, for any context $C'' \in \mathscr{D}_2^V$ with conservative belief profile $\mathscr{B}''$ such that $\mathscr{B}_2'' \neq \mathscr{B}_2^*$, $\boldsymbol{s}'(\mathscr{B}'') = \boldsymbol{s}(\mathscr{B}'')$, and thus $\boldsymbol{s}'(\mathscr{B}'')$ is a Nash equilibrium of (C'', M). Because context C is a generic context in $\mathscr{D}_2^V$ with player 2's conservative belief being $\mathscr{B}_2^*$, it remains for us to prove that $\boldsymbol{s}'$ satisfies the following properties:

(A) $\boldsymbol{s}'(\mathscr{B})$ is a Nash equilibrium of G; and

(B) $REV(M(\boldsymbol{s}'(\mathscr{B}))) < 3\theta_2^*/\epsilon$.

Proof of Property A Because $\boldsymbol{s}'_1 = \boldsymbol{s}_1$, to establish Property A, it suffices to show that $(\boldsymbol{s}_1(\mathscr{B}_1), \boldsymbol{s}'_2(\mathscr{B}_2))$ is a Nash equilibrium of G.

By construction, player 1's true type is the same in C and C', and the same is true for player 2's true type. Therefore we have that

$$u_1^{G'}(\cdot) = u_1^G(\cdot) \quad \text{and} \quad u_2^{G'}(\cdot) = u_2^G(\cdot).$$

Accordingly, $(\boldsymbol{s}_1(\mathscr{B}_1), \boldsymbol{s}'_2(\mathscr{B}_2))$ is a Nash equilibrium of G if and only if it is a Nash equilibrium of G'. Because, again by construction, player 1's conservative belief is the same in C and C', we further have

$$\boldsymbol{s}_1(\mathscr{B}'_1) = \boldsymbol{s}_1(\mathscr{B}_1),$$

and thus

$$(\boldsymbol{s}_1(\mathscr{B}_1), \boldsymbol{s}'_2(\mathscr{B}_2)) = (\boldsymbol{s}_1(\mathscr{B}'_1), \boldsymbol{s}'_2(\mathscr{B}^*_2)) = (\boldsymbol{s}_1(\mathscr{B}'_1), \boldsymbol{s}_2(\mathscr{B}'_2)) = \boldsymbol{s}(\mathscr{B}').$$

Because $\boldsymbol{s}(\mathscr{B}')$ is a Nash equilibrium of G', $(\boldsymbol{s}_1(\mathscr{B}_1), \boldsymbol{s}'_2(\mathscr{B}_2))$ is a Nash equilibrium of G' and thus of G, and Property A holds. □

Proof of Property B Because $\boldsymbol{s}'(\mathscr{B}) = (\boldsymbol{s}_1(\mathscr{B}_1), \boldsymbol{s}'_2(\mathscr{B}_2)) = \boldsymbol{s}(\mathscr{B}')$, to establish Property B, it suffices to prove that

$$REV(M(\boldsymbol{s}(\mathscr{B}'))) < 3\theta_2^*/\epsilon. \tag{17}$$

Because the last inequality in (g′) states that at the strategy profile $\boldsymbol{s}(\mathscr{B}')$ the revenue is at most the price paid by player 1, P'_1, it actually suffices to show that

$$P'_1 < 3\theta_2^*/\epsilon. \tag{18}$$

We prove Inequality 18 by contradiction. Assume $P'_1 \geq 3\theta_2^*/\epsilon$, and consider the auxiliary game G''.

> $G'' = (C'', M)$, where $C'' = (2, V, \theta'', \mathscr{B}'')$ with
> $\theta'' = (\lceil 2\theta_2^*/\epsilon \rceil, \theta_2)$, $\mathscr{B}''_1 = \{(\theta''_1, x) : (\theta_1, x) \in \mathscr{B}_1\}$, and $\mathscr{B}''_2 = \mathscr{B}'_2$.

Notice that $C'' \in \mathscr{D}_2^V$. In G'' we have that $smp''_1 \geq smp''_2 = \lceil 2\theta_2^*/\epsilon \rceil$. Letting $(a'', P'') = M(\boldsymbol{s}(\mathscr{B}''))$, similar to what we have seen before, the following (in)equalities hold:

(g″) $\quad 2^{nd}(\mathscr{B}'') \geq 2\theta_2^*/\epsilon, \quad REV(a'', P'') \geq 2\theta_2^*, \quad a'' = 1, \quad 2\theta_2^* \leq P''_1 \leq \theta''_1 = \lceil 2\theta_2^*/\epsilon \rceil, \quad \text{and} \quad P''_2 \leq 0.$

Let us now prove that, in game G', player 1 can profitably deviate from the equilibrium $\boldsymbol{s}(\mathscr{B}')$ by playing $\boldsymbol{s}_1(\mathscr{B}''_1)$ instead of $\boldsymbol{s}_1(\mathscr{B}'_1)$. Indeed, $\mathscr{B}''_2 = \mathscr{B}'_2$ implies that $\boldsymbol{s}_2(\mathscr{B}''_2) = \boldsymbol{s}_2(\mathscr{B}'_2)$, and thus $M(\boldsymbol{s}_1(\mathscr{B}''_1), \boldsymbol{s}_2(\mathscr{B}'_2)) = M(\boldsymbol{s}(\mathscr{B}'')) = (a'', P'')$. According to (g″), in the outcome (a'', P'') player 1 still wins the good, but pays $\leq \lceil 2\theta_2^*/\epsilon \rceil$ rather than $\geq 3\theta_2^*/\epsilon$.

Because θ_2^* is an integer > 1, we have that $3\theta_2^*/\epsilon - 2\theta_2^*/\epsilon > 1$. Therefore $3\theta_2^*/\epsilon > \lceil 2\theta_2^*/\epsilon \rceil$, implying that player 1 can indeed benefit by deviating from $\boldsymbol{s}_1(\mathscr{B}'_1)$ to $\boldsymbol{s}_1(\mathscr{B}''_1)$ in equilibrium $\boldsymbol{s}(\mathscr{B}')$. Thus, assuming that Inequality 18 is false contradicts the fact that $\boldsymbol{s}(\mathscr{B}')$ in an equilibrium of G'. The contradiction establishes Inequality 17 and thus Property B. □

Therefore Theorem 4 holds. ■

A.2 Proof of Theorem 5

Let ϵ be a value in $(1/2, 1]$, V an integer greater than $\lceil \frac{1}{\epsilon - 1/2} \rceil$, and M a reasonable mechanism implementing $\epsilon 2^{nd}$ for $\mathscr{D}_2^V$ at some ex-post equilibrium $\boldsymbol{s}$. To prove that M is fragile, let H be an integer such that

$$V \geq H > \frac{1}{\epsilon - 1/2}.$$

Similar to previous proofs, we are going to consider different contexts and thus different games, and we use superscripts to clarify the game to which a given quantity refers.

Let $\mathscr{B}_2^* = \{(H, 0)\}$. Notice that there exist some contexts in $\mathscr{D}_2^V$ with player 2's conservative belief being $\mathscr{B}_2^*$ —indeed, these are contexts where player 1's true valuation is H, player 2's true valuation is 0, and player 2 believes that player 1's true valuation is H (that is, player 1's external belief are the only undetermined part). Our goal is to show that there exists another ex-post equilibrium $\boldsymbol{s}'$ such that:

(1) $\boldsymbol{s}'$ and $\boldsymbol{s}$ differ only at the conservative belief $\mathscr{B}_2^*$ of player 2; and

(2) for every context $C = (n, V, \theta, \mathscr{B}) \in \mathscr{D}_2^V$ with $\mathscr{B}_2 = \mathscr{B}_2^*$, $REV(M(\boldsymbol{s}'(\mathscr{B}))) < \epsilon 2^{nd}(\mathscr{B})$.

To do so, we analyze two (classes of) related games, G and G', as follows.

$$\boxed{G = (C, M), \text{ where } C = (2, V, \theta, \mathscr{B}) \text{ is an arbitrary context in } \mathscr{D}_2^V \text{ with } \mathscr{B}_2 = \mathscr{B}_2^*.}$$

In G we have that $\theta = (H, 0)$, $\theta \in \mathscr{B}_1$, $smp_1 = smp_2 = H$, and thus $2^{nd}(\mathscr{B}) = H$ no matter what $\mathscr{B}_1$ is.

$$\boxed{\begin{array}{l} G' = (C', M), \text{ where } C' = (2, V, \theta', \mathscr{B}') \text{ with} \\ \theta' = (1, 0), \mathscr{B}'_1 = \{(1, x) : (H, x) \in \mathscr{B}_1\}, \text{ and } \mathscr{B}'_2 = \{(x, 0) : x \geq 1\}. \end{array}}$$

Notice that $C' \in \mathscr{D}_2^V$ and $2^{nd}(\mathscr{B}') = 1$.

Let us now analyze game G'. Let $\omega' = M(\boldsymbol{s}(\mathscr{B}'))$, and p'_i and EP'_i respectively be the probability that player i gets the good and the expected price that player i pays according to ω'. By the opt-out condition and our hypothesis, we have

$$u_1^{G'}(\omega') = p'_1 - EP'_1 \geq 0, \quad u_2^{G'}(\omega') = -EP'_2 \geq 0, \quad \text{and} \quad EP'_1 + EP'_2 \geq \epsilon 2^{nd}(\mathscr{B}') = \epsilon.$$

Combining these three inequalities, we have

$$p'_1 \geq EP'_1 \geq \epsilon - EP'_2 \geq \epsilon. \tag{19}$$

We construct the desired ex-post equilibrium $\boldsymbol{s}'$ as follows:

$$\boldsymbol{s}'_2(\mathscr{B}_2^*) = \boldsymbol{s}_2(\mathscr{B}'_2),$$

and $\boldsymbol{s}'$ coincides with $\boldsymbol{s}$ everywhere else. By construction, for any context $C'' \in \mathscr{D}_2^V$ with conservative belief profile $\mathscr{B}''$ such that $\mathscr{B}''_2 \neq \mathscr{B}_2^*$, $\boldsymbol{s}'(\mathscr{B}'') = \boldsymbol{s}(\mathscr{B}'')$, and thus $\boldsymbol{s}'(\mathscr{B}'')$ is a Nash equilibrium of the game (C'', M). Because C is a generic context in $\mathscr{D}_2^V$ with player 2's conservative belief being $\mathscr{B}_2^*$, it remains for us to prove that $\boldsymbol{s}'$ satisfies the following properties:

(A) $\boldsymbol{s}'(\mathscr{B})$ is a Nash equilibrium of G; and

(B) $REV(M(\boldsymbol{s}'(\mathscr{B}))) < \epsilon H$.

Proof of Property A. We do so by introducing another (auxiliary) game G''.

$$\boxed{G'' = (C'', M), \text{ where } C'' = (2, V, \theta'', \mathscr{B}'') \text{ with } \theta'' = \theta, \mathscr{B}''_1 = \mathscr{B}_1, \text{ and } \mathscr{B}''_2 = \mathscr{B}'_2.}$$

Notice that $C'' \in \mathscr{D}_2^V$, and that C'' differs from C only at player 2's belief and from C' only at player 1's true valuation (of course $\mathscr{B}''_1$ has to be consistent with θ''_1 which is H, and thus differs from $\mathscr{B}'_1$, but player 1's external belief does not change).

Because $\boldsymbol{s}'_1 = \boldsymbol{s}_1$, $\mathscr{B}_2 = \mathscr{B}_2^*$, $\mathscr{B}_1 = \mathscr{B}''_1$, $\boldsymbol{s}'_2(\mathscr{B}_2^*) = \boldsymbol{s}_2(\mathscr{B}'_2)$, and $\mathscr{B}'_2 = \mathscr{B}''_2$, we have that

$$\boldsymbol{s}'(\mathscr{B}) = (\boldsymbol{s}'_1(\mathscr{B}_1), \boldsymbol{s}'_2(\mathscr{B}_2)) = (\boldsymbol{s}_1(\mathscr{B}_1), \boldsymbol{s}'_2(\mathscr{B}_2^*)) = (\boldsymbol{s}_1(\mathscr{B}''_1), \boldsymbol{s}_2(\mathscr{B}'_2)) = (\boldsymbol{s}_1(\mathscr{B}''_1), \boldsymbol{s}_2(\mathscr{B}''_2)) = \boldsymbol{s}(\mathscr{B}'').$$

Because $\boldsymbol{s}(\mathscr{B}'')$ is a Nash equilibrium of G'' by the definition of $\boldsymbol{s}$, $\boldsymbol{s}'(\mathscr{B})$ is also a Nash equilibrium of G''. Because G and G'' have the same true valuation profile, $\boldsymbol{s}'(\mathscr{B})$ is a Nash equilibrium of G, and Property A holds.

Proof of Property B. Notice that in game G,

$$u_1^G(M(\boldsymbol{s}(\mathscr{B}'))) = u_1^G(\omega') = p_1'H - EP_1' \geq p_1'H - p_1' \geq \epsilon(H-1),$$

where the inequalities hold by Equation 19.

Because $\boldsymbol{s}'(\mathscr{B}) = (\boldsymbol{s}_1(\mathscr{B}_1), \boldsymbol{s}_2'(\mathscr{B}_2^*)) = (\boldsymbol{s}_1(\mathscr{B}_1), \boldsymbol{s}_2(\mathscr{B}_2'))$ is a Nash equilibrium of G, we have that

$$u_1^G(M(\boldsymbol{s}_1(\mathscr{B}_1), \boldsymbol{s}_2(\mathscr{B}_2'))) \geq u_1^G(M(\boldsymbol{s}(\mathscr{B}'))) \geq \epsilon(H-1),$$

and

$$u_2^G(M(\boldsymbol{s}_1(\mathscr{B}_1), \boldsymbol{s}_2(\mathscr{B}_2'))) \geq u_2^G(M(\boldsymbol{s}_1(\mathscr{B}_1), out_2)) = 0.$$

Let $\omega'' = M(\boldsymbol{s}_1(\mathscr{B}_1), \boldsymbol{s}_2(\mathscr{B}_2'))$, and p_i'' and EP_i'' respectively be the probability that player i gets the good and the expected price that player i pays according to ω''. Combining with the above two lines of equations, we have

$$u_1^G(\omega'') = p_1''H - EP_1'' \geq \epsilon(H-1) \quad \text{and} \quad u_2^G(\omega'') = -EP_2'' \geq 0.$$

Combining with the facts that $0 \leq p_1'' \leq 1$, $1/2 < \epsilon \leq 1$, and $H > \frac{1}{\epsilon - 1/2}$, we have

$$\begin{aligned} REV(M(\boldsymbol{s}'(\mathscr{B}))) &= REV(M(\boldsymbol{s}_1(\mathscr{B}_1), \boldsymbol{s}_2(\mathscr{B}_2'))) = EP_1'' + EP_2'' \leq EP_1'' \leq p_1''H - \epsilon(H-1) = H(p_1'' - \epsilon + \frac{\epsilon}{H}) \\ &\leq H(1 - \epsilon + \frac{1}{H}) < H(1 - \epsilon + \epsilon - 1/2) = H/2 < \epsilon H. \end{aligned}$$

Therefore Property B holds, and so does Theorem 5. ■

References

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

[2] D. Abreu and H. Matsushima. Virtual Implementation in Iteratively Undominated Strategies: Incomplete Information. Mimeo, 1992.

[3] G. Artemov, T. Kunimoto, and R. Serrano. Robust Virtual Implementation with Incomplete Information: Towards a Reinterpretation of the Wilson Doctrine. Working paper, 2007.

[4] S. Baliga and R. Vohra. Market Research and Market Design. *Advances in Theoretical Economics*, 3(1), Article 5, 2003.

[5] T. F. Bewley. Knightian decision theory Part I. *Decisions in Economics and Finance*, Vol. 25, No. 2, pp. 79-110, 2002.

[6] T. N. Casona, T. Saijob, T. Sjostrom, and T. Yamatoe. Secure Implementation Experiments: Do Strategy-Proof Mechanisms Really Work? *Games and Economic Behavior*, 57(2): 206-235, 2006.

[7] J. Chen, A. Hassidim, and S. Micali. Robust Perfect Revenue from Perfectly Informed Players. *Innovations in Computer Science (ICS)*, pp. 94-105, 2010.

[8] J. Chen and S. Micali. A New Approach to Auctions and Resilient Mechanism Design. *41st Symposium on Theory of Computing (STOC)*, pp. 503-512, 2009.

[9] J. Chen, S. Micali, and P. Valiant. Robustly Leveraging Collusion in Combinatorial Auctions. *Innovations in Computer Science (ICS)*, pp. 81-93, 2010.

[10] A. Chiesa, S. Micali, and Z. A. Zhu. Mechanism Design with Approximate Valuations. To appear at *Innovations in Theoretical Computer Science (ITCS)*, 2012.

[11] K.S. Chung and J.C. Ely. Implementation with Near-Complete Information. *Econometrica*, 71(3): 857-871, 2003.

[12] J. Cremer and R.P. McLean. Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions. *Econometrica*, Vol.56, No.6, pp. 1247-1257, Nov., 1988.

[13] A. Gibbard. Manipulation of Voting Schemes: A General Result. *Econometrica*, 41(4): 587-602, 1973.

[14] J. Glazer and M. Perry. Virtual Implementation in Backwards Induction. *Games and Economic Behavior*, Vol.15, pp. 27-32, 1996.

[15] A. Goldberg, J. Hartline, A. Karlin, M. Saks, and A. Wright. Competitive Auctions. *Games and Economic Behavior*, 55(2): 242-269, 2006.

[16] J. Green and J. Laffont. Characterization of Satisfactory Mechanisms for the Revelation of Preferences for Public Goods. *Econometrica*, 45(2): 427-438, 1977.

[17] J. Harsanyi. Games with Incomplete Information Played by "Bayesian" Players, I-III. Part I. The Basic Model. *Management Science*, 14(3) Theory Series: 159-182, 1967.

[18] L. Hurwicz. On the Existence of Allocation Systems Whose Manipulative Nash Equilibria Are Pareto Optimal. Unpublished. 1975.

[19] N. Hyafil and C. Boutilier. Regret Minimizing Equilibria and Mechanisms for Games with Strict Type Uncertainty. *Proceedings of the Twentieth Annual Conference on Uncertainty in Artificial Intelligence*, pp. 268-277, 2004.

[20] M. Jackson. Implementation in Undominated Strategies: A Look at Bounded Mechanisms. *The Review of Economic Studies*, 59(4): 757-775, 1992.

[21] F. H. Knight. Risk, Uncertainty and Profit. Houghton Mifflin, Boston, MA. 1921.

[22] G. Lopomo, L. Rigotti, and C. Shannon. Uncertainty in Mechanism Design. Revise and resubmit at *Review of Economic Studies*, 2009.

[23] A. Postlewaite and D. Schmeidler. Implementation in Differential Information Economies. *Journal of Economic Theory*, 39(1): 14-33, 1986.

[24] T. Saijo, T. Sjostrom, and T. Yamato. Secure Implementation: Strategy-Proof Mechanisms Reconsidered. Unpublished. 2003.

[25] M. Satterthwaite. Strategy-Proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions. *Journal of Economic Theory*, 10(2): 187-217, 1975.

[26] I. Segal. Optimal Pricing Mechanisms with Unknown Demand. *American Economic Review*, 93(3): 509-529, 2003.