Strategyproofness for “Price Takers” as a Desideratum for Market Design

Eduardo Azevedo and Eric Budish*

May 4, 2011

Abstract

We distinguish between two ways a mechanism can fail to be strategyproof. A mechanism may have manipulations that *persist with market size* (first-order manipulations); and, a mechanism may have manipulations that *vanish with market size* (second-order manipulations). We say that a non-strategyproof mechanism is strategyproof for “price takers” (*SP(p)*) if all of its manipulations vanish with market size. We put “price takers” in quotes because our notion is not limited to mechanisms that explicitly use prices. Our main result is that, given a mechanism with Bayes-Nash or complete information Nash equilibria, there exists a prior free mechanism that is SP(p) and that coincides exactly with the original mechanism in the limit. It coincides approximately in large finite markets, with exponential rate of convergence. Thus, while strategyproofness often severely limits what kinds of mechanisms are possible, for our class of problems $SP(p)$ does not, and hence may be a useful second-best. We illustrate our concepts with examples from single-unit assignment, multi-unit assignment, matching and auctions.

WORKING DRAFT. PRELIMINARY AND INCOMPLETE.

1 Introduction

Strategyproofness – i.e., that playing the game truthfully is a dominant strategy – is perhaps the predominant notion of incentive compatibility in practical market design. There are at least four important reasons why strategyproofness is so heavily emphasized relative to other forms of incentive compatibility, such as Bayes-Nash. First, strategyproof mechanisms are detail free for the designer,

*Harvard Univer ity (azevedo@fa .harvard.edu) and Univer ity of Chicago Booth School of Bu ine (eric.budi h@chicagobooth.edu). For helpful di cu ion we are grateful to Su an Athey, Aaron Bodoh-Creed, Gabriel Carroll, Jeff Ely, Drew Fudenberg, John Hatfield, Richard Holden, Emir Kamenica, Fuhito Kojima, Scott Kominer , Jacob Le hno, Paul Milgrom, Roger Myer on, David Parke , Parag Pathak, Nicola Per ico, Canice Prenderga t, Lar Stole, Glen Weyl, and e pecially Al Roth.

in the sense of Wilson (1987); the designer need not know anything about participants' preferences or beliefs (cf. Bergemann and Morris, 2005). Second, and relatedly, strategyproof mechanisms are strategically simple for participants; participants need not form beliefs about others' preferences or play (Fudenberg and Tirole, 1991; Roth, 2008). Third, with this simplicity comes a measure of fairness; agents' outcomes do not depend on their ability to "game the system" (Friedman, 1991; Pathak and Sonmez, 2008; Abdulkadiroglu et al, 2006). Fourth, Bayesian approaches simply have not yet proved tractable for a number of important market design problems.

However, in a wide variety of economic contexts, impossibility theorems indicate that strategyproofness severely limits what kinds of mechanisms are possible. These range from Gibbard (1973) and Satterthwaite's (1975) dictatorship theorem for general social choice problems, to Hurwicz's (1972) impossibility theorem for general equilibrium settings, to the Green-Laffont (1977) VCG theorem for allocation settings with quasi-linear preferences, to Roth's (1982) impossibility theorem for strategyproof stable matching, to Papai's (2001) dictatorship theorem for multi-unit demand assignment problems, to Abdulkadiroglu, Pathak and Roth's (2009) impossibility theorem for strategyproof and efficient school assignment.

This creates a conundrum for market designers: strategyproofness is the only form of incentive compatibility that the literature finds fully satisfying, yet often there are no good strategyproof mechanisms.

This paper proposes a way out of this conundrum, appropriate for markets that are "large". In most of economics, researchers instinctively understand that markets with a handful of participants should be studied using different tools than markets with thousands or millions of participants. Yet the distinction between strategyproof and non-strategyproof mechanisms ignores market size. Consider the decision of whether to order chicken or fish at a restaurant that you frequent. Suppose you prefer chicken. Suppose further that it is possible that ordering the chicken today will set off a chain reaction of events in the global chicken market, ultimately causing an increase in the restaurant's price of chicken tomorrow. This kind of possibility is in principle enough to make the decision between chicken and fish "not strategyproof." We wager that no market designer would criticize standard restaurant ordering procedures on incentive grounds;[1] yet, limiting attention to strategyproof mechanisms amounts to just that.

Now consider a different kind of restaurant. Again, chicken and fish are two of the choices, but now the ordering process is a bit different. Specifically, you overhear the following exchange:

> Customer: I'll have the Chicken.
>
> Waiter: I'm sorry sir, but we are out of Chicken.
>
> Customer: In that case I'll have the Fish.

[1] Further, we have yet to ee the following propo al: package the chicken entree with a derivative contract tied to global chicken price , o that ordering i truly trategyproof.

> Waiter: Sorry, we only consider your first choice of entree. Since Fish wasn't your first choice, we won't serve you that either.

This restaurant, too, is "not strategyproof"; and in this instance, we imagine that most researchers would indeed find cause to complain about this restaurant's ordering protocol.

We propose a conceptual distinction between two kinds of non-strategyproofness: (i) a mechanism may have profitable manipulations that *persist with market size* ("first-order manipulations"); and (ii) a mechanism may have profitable manipulations that *vanish with market size* ("second-order manipulations"). The restaurant with the strange ordering procedure is first-order manipulable: customers will frequently want to misreport their preferences at a restaurant with those rules, independently of the number of and behavior of patrons of that restaurant (which may be small given those rules!) or of the number of and behavior of agents participating in the global chicken market. The normal restaurant is not first-order manipulable; if we ignore any individual customer's ability to affect the global price of chicken tomorrow, then today's decision between chicken and fish is exactly strategyproof. However, it is second-order manipulable, because of the vanishingly likely possibility that an individual customer affects global prices.

If a mechanism has second-order but not first-order manipulations it is not strategyproof, but it is what we will call *strategyproof for "price takers"*, or $SP(p)$. We put "price takers" in quotations because our notion is not limited to mechanisms that explicitly use prices; indeed, there are many examples of non-price-based mechanisms that are strategyproof for the kinds of agents economists think of as "price takers", and many examples of price-based mechanisms that are not. We argue that, when strategyproof alternatives are unattractive, $SP(p)$ may be a useful second-best incentives criterion; that is, at the very least, market designers should seek to avoid first-order manipulations.

Our first main result is that, in an important class of games that includes many widely studied market-design problems, first-order manipulations can be avoided "for free" in the large market limit. More specifically, we show the following. Suppose there is some mechanism that is both first- and second-order manipulable, and that has Bayes-Nash or complete information Nash equilibria that implement some outcomes as a function of agents' preferences. We show by construction that there exists another mechanism that is $SP(p)$, and that implements exactly the same outcomes as the equilibria of the original mechanism in the limit. That is, in the limit, we can get the attractive aspects of strategyproof design – detail-freeness, strategic simplicity, fairness, tractability – "for free".

We describe our main result in the context of a specific example, the Boston mechanism for school choice. The Boston mechanism resembles the first-order manipulable restaurant described above. Students report their preferences for schools, and then as many students as possible are awarded their first choice; only after as many students as possible have been awarded their first choice are second choices considered. In practice, by the time second choices are considered most of the good schools have already reached capacity; so the system effectively only cares about your first choice, just like the restaurant. Abdulkadiroglu and Sonmez (2003) and Abdulkadiroglu et al (2006) criticized the Boston mechanism on the grounds that it is not strategyproof. These papers proposed that the strategyproof Gale-Shapley deferred acceptance algorithm be used instead; indeed, the Gale-Shapley algorithm was eventually adopted for use in practice (cf. Roth, 2008).

However, as mentioned above, strategyproofness often has costs relative to other forms of incentive compatibility. A second generation of papers on the Boston mechanism argued that the Boston mechanism has a Bayes-Nash equilibrium that yields greater student welfare than does the dominant strategy equilibrium of the Gale-Shapley procedure (Abdulkadiroglu, Che and Yasuda, 2011; Miralles, 2008; Featherstone and Niederle, 2009). These papers argued that the first generation papers mentioned in the previous paragraph were too quick to dismiss the Boston mechanism in favor of strategyproof deferred acceptance.

Bayes-Nash equilibria have their own costs. These second generation papers rely on students having common knowledge of the distribution of other students' preferences; on students being able to coordinate on a specific equilibrium; on students being able to make very precise strategic calculations to determine whether to "risk asking for the chicken"; etc. Our main result says that all of this complexity and non-robustness is unnecessary in a large market. Specifically, in the large market limit, there must exist yet another mechanism that implements the same outcomes as these desirable Bayes-Nash equilibria of the Boston mechanism, but with dominant strategy incentives. Put differently, our result says that even though finite market strategyproofness genuinely limits the scope of what is possible to implement in this environment, strategyproofness for "price takers" does not: the first-order manipulations of the Boston mechanism can be eliminated for free in the large market limit.

Our second main result is about economies away from the limit, i.e. large finite economies. Suppose there exists a mechanism, e.g. the Boston mechanism described above, that has Bayes-Nash equilibria in finite economies. Our second result says that there exists another mechanism that is detail free for the designer, has no first-order manipulations for the participants (i.e., is $SP(p)$), has vanishingly many second-order manipulations, and yields approximately the same outcomes as the Bayes-Nash equilibrium of the original mechanism. Further, we show that both approximations in this new mechanism – the number of second-order manipulations, and the difference in outcomes – vanish exponentially with market size.

We emphasize that there exist mechanisms that do not appear to have any explicit "prices" but that satisfy our notion of $SP(p)$, and also that there exist mechanisms that do have explicit prices but that are not $SP(p)$. Examples of the former include Gale and Shapley's (1962) deferred acceptance algorithm for two-sided matching, and Bogomolnaia and Moulin's (2001) probabilistic serial mechanism for assignment problems. Examples of the latter include the pay-as-bid multi-object auction used until the 1990s by the US Treasury to allocate US government debt (cf. Friedman 1991), and the Bidding Points Auction used by many business schools and law schools to allocate courses to students (cf. Sonmez and Unver, 2010; Budish, 2010). Friedman (1964, 1991) is of special note. He not only criticizes the pay-as-bid auction, but suggests that the US government switch to the uniform price auction instead. The uniform price auction also is not strategyproof. However, uniform price auctions, unlike pay-as-bid auctions, *are* $SP(p)$; this absence of what we call first-order manipulations is presumably what Friedman has in mind when he says that "you do not have to be a specialist" to figure out how to participate in the uniform price auction, because you can just indicate "the maximum amount you are willing to pay for different quantities ... if you bid a higher price [than the market clearing price], you do not lose as you do under the current method."[2]

Related Literature [Especially preliminary]

Our paper is related to a large literature that has studied how market size can ease incentive constraints. An early paper in this tradition is Roberts and Postlewaite (1976) on the Walrasian mechanism, which can be seen as a response to Hurwicz's (1972) critique that the Walrasian mechanism is not strategyproof. Other papers in this tradition include Rustichini, Satterthwaite and Williams (1994) on double auctions with private values, Pesendorfer and Swinkels (2000), Cripps and Swinkels (2006) and Perry and Reny (2006) on double auctions with common-value components, Kojima and Pathak (2009) on deferred acceptance algorithms, and Kojima and Manea (2009) on the Bogomolnaia-Moulin (2001) probabilistic serial mechanism. Each of these papers provides a defense of a *specific mechanism* based on its incentive properties in large markets. Our paper aims to justify strategyproofness for price takers as a *general desideratum* for practical market design. Note that in the context of any of the specific mechanisms named above, our analysis is much less instructive than are previous analyses tailored to the specific mechanism.

Technically, our paper is most closely related to Kalai (2004). Kalai (2004; Theorem 1) shows that Bayes-Nash equilibria are approximately ex-post Nash in a class of large anonymous games. In words, if a large number of agents with private information about their types play some BNE, then ex-post – i.e. after seeing each agent's chosen action – agents will have vanishingly little incentive to revise their play. The difference between our Theorem 2 and Kalai's Theorem 1 is that Kalai shows that a given BNE is approximately ex-post Nash, whereas we use the BNE of a given mechanism to create a new mechanism that is approximately strategyproof. In our new mechanism players need not have common knowledge of the prior, or of what equilibrium is being played, nor need they be strategically sophisticated in any way.

[2] The debate about whether to favor uniform price or pay-a -bid auction continue to thi day. In the context of de ign di cu ion concerning the US Trea ury' Troubled A et Relief Program (TARP), Au ubel and Cramton (2008) criticize the trategic complexity of pay-a -bid auction a follow : "Bidder hate pay-a -bid auction , a they look fooli h (or unemployed) after elling at unnece arily low price ." See al o Au ubel and Cramton (1996) for an analy i that how that the uniform-price auction propo ed by Friedman i not trategyproof.

Third, our paper is related to the literature on the role of strategyproofness in practical market design. Wilson (1987) famously argued that practical market designs should aim to be detail free, and Bergemann and Morris (2005) formalized the sense in which strategyproof mechanisms are robust in the sense of Wilson. Several recent papers have argued that strategyproofness can be viewed as a design objective and not just as a constraint: papers on this theme include Abdulkadiroglu et al (2006, 2009), Pathak and Sonmez (2008), and Roth (2008). Our paper contributes to this literature by showing that strategyproofness is approximately attainable in large markets. Also, the distinction we draw between first- and second-order manipulations highlights that many mechanisms in practice are manipulable in a preventable way.

Last, our paper is conceptually related to Pathak and Sonmez (2011), who also seek a way to say something more useful about non-strategyproof mechanisms than simply that they are not strategyproof. We view the two papers as complementary. An advantage of the Pathak-Sonmez (2011) approach is that it empirically organizes several recent policy changes in the design of school choice systems. An advantage of our approach is that it yields an explicit design desideratum, namely that mechanisms be strategyproof for price takers.

Organization of the paper The rest of this paper is organized as follows. Section 2 describes the environment and some key assumptions. Section 3 defines strategyproof for "price takers" and related concepts. Section 4 goes through several examples. Section 5 presents the main theoretical results. Section 6 concludes.

2 En ironment

Mechanisms Many papers ask what kinds of mechanisms are possible given a single, commonly known probability distribution, and/or a single market size n. In this paper, we are interested in mechanisms that are well defined for a range of market sizes and that are "detail free" for the designer, in the sense that the probability distribution over agents' types does not enter into the mechanism description (cf. Wilson, 1987). For these reasons we define a mechanism as follows:

Definition 1. *A* ***mechanism*** *is a finite action space A, a finite set of possible (sure) outcomes X_0 for the agents, with $X = \Delta X_0$ the set of lotteries of over outcomes, and a sequence of functions*

$$\Phi^n : A^n \to X^n$$

that maps a vector of n actions $\bar{a}$ into a vector of n random allocations $\Phi^n(\bar{a})$. A particular function in the sequence, for a particular size n, is called an ***n-mechanism****.*

An outcome might be the consumption bundle an agent receives, a match partner, a social decision, etc.. Since our interest is in mechanisms that induce truthful reporting, it is important that there be a well-defined notion of "truthful".

Definition 2. *Associated with each mechanism is a finite type space* Θ*, vNM utility functions* $u_\theta : X_0 \to \mathbb{R}_+$ *for each* $\theta \in \Theta$*, and an onto function* $\tau : \Theta \to A$*. For each* $\theta \in \Theta$*, action* $\tau(\theta)$ *is called* θ*'s* ***truthful play****.*

Note that our setup implicitly assumes that agents have private values, in the sense that their preferences depend only on their own type.

To illustrate our terminology, consider the Boston mechanism for school choice, described above. In that mechanism the action space A is the set of ordinal preference rankings over schools. The appropriate type space Θ is the discretized set of von-Neumann Morgenstern utility functions over schools, which is a larger space. The truthful play of type θ is simply the ordinal preference ranking associated with that type's vNM utility function.

The assumption that τ is onto means that each action is the truthful play of some type. This allows us to speak of an agent "playing θ"; we use this to mean that the agent plays action $\tau(\theta)$. Whenever we use a type $\theta \in \Theta$ as an argument of a mechanism, we mean the associated action $\tau(\theta) \in A$.

Two key assumptions of our analysis are that mechanisms are anonymous and equicontinuous. We define these in turn, following the terminology of Kalai (2004).

Definition 3. *For every vector of actions* $\bar{a}$*, define the* ***empirical distribution*** *of* $\bar{a}$ *on the action set* A *by the vector*

$$emp_a[\bar{a}] = \frac{(\text{the number of coordinates of } \bar{a} \text{ with } a_i = a)}{(\text{the number of coordinates of } \bar{a})}$$

A mechanism $(\Phi^n)_{\mathbb{N}}$ *is* ***anonymous*** *if, for every* n*, and every* $\bar{a}$ *and* $\bar{a}'$ *with* $emp(\bar{a}) = emp(\bar{a}')$ *and* $a_i = a'_j$*, we have* $\Phi^n_i(a_i, a_{-i}) = \Phi^n_j(a'_j, a'_{-j})$*.*

Assumption 1. *Mechanisms are anonymous.*

In an anonymous mechanism, two players who take the same action get the same (lottery over) outcomes. For this reason, anonymity is sometimes called "equal treatment of equals", after Aristotle's famous dictum (cf. Moulin 1995, Thomson 2008). Another implication of anonymity is that each player's outcome depends only on the distribution of others' actions, not on exactly who played what. This is a natural feature of many large-market settings. Examples of anonymous mechanisms include the Walrasian mechanism, most well-known single-object, combinatorial- and double- auction formats, and most of the mechanisms that have been proposed for single- and multi-unit assignment problems.

[3] We believe that all of the re ult in thi preliminary analy i can be obtained if we relax

Definition 4. *Mechanism* $(\Phi^n)_{n\in\mathbb{N}}$ *is* ***(Kalai equi)continuous in actions*** *if, for every* $\epsilon > 0$*, there exists a* $\delta > 0$ *s.t.: for every* n*, for every* $\theta \in \Theta$*, and any action profiles* $\bar{a}, \bar{a}'$ *with* $a_i = a'_i$ *and* $\left|emp[a_{-i}] - emp[a'_{-i}]\right|_{sup} < \delta$*,* $u_{\theta_i}[\Phi^n_i(\bar{a})] - u_{\theta_i}[\Phi^n_i(\bar{a}')] < \epsilon$*.*

Assumption 2. *Mechanisms are continuous in actions.*

Equicontinuity in actions requires that if the empirical distribution of others' actions changes by a small amount, then the payoff to any particular action changes by a small amount. Note that the amount by which each individual agent's play affects the overall distribution of actions grows small with market size; specifically, the maximum effect a single agent can have on this measure in a market with n participants is $\frac{1}{n}$. Below, we will define analogous notions of continuity with respect to reports and with respect to strategies.

2.1 Standard Equilibrium Concepts

Suppose there are n agents, and consider a measure m on the set of types, i.e., $m \in \Delta\Theta$. Let:

$$\phi^n(\theta_i|m) = \Phi^n_i(\theta_i, \theta_{-i}) \tag{1}$$

where θ_{-i} is an $n-1$ vector of types distributed iid according to m. The object $\phi^n(\theta_i|m)$ is a random bundle in X that describes what a generic agent i can expect to receive from the n-mechanism $\Phi^n(\cdot)$ when he reports θ_i and the other $n-1$ agents report according to m. Recall that we equate reporting type $\theta_i \in \Theta$ with playing action $\tau(\theta_i) \in A$.

A **strategy** is a map $\sigma : \Theta \to \Delta\Theta$, assigning for each type a probability distribution over reports (which in turn induces a probability distribution over actions, via $\tau(\cdot)$). Given a probability distribution $m \in \Delta\Theta$ over types and a strategy $\sigma(\cdot)$, denote by $\sigma(m)$ the induced distribution over reports.

Definition 5. *A* ***μ-Bayes-Nash equilibrium*** *(μ-BNE) of n-mechanism* Φ^n *is a strategy* $\sigma^*_\mu(\cdot)$ *such that for all* $\theta_i, \hat{\theta}_i \in \Theta$

$$u_{\theta_i}[\phi^n(\sigma^*_\mu(\theta_i)|\sigma^*_\mu(\mu))] \geq u_{\theta_i}[\phi^n(\hat{\theta}_i|\sigma^*_\mu(\mu))]$$

This is the standard definition of Bayes Nash Equilibrium in our notation. In words, the strategy σ^*_μ is a BNE if each agent's expected utility from playing according to σ^*_μ is higher than that from any other action, given that the other agents' types are distributed according to μ and that they also play according to σ^*_μ. Notice that there is no guarantee that $\sigma^*_\mu(\theta)$ is the best strategy for an agent of type θ if the other agents play differently, which could occur, e.g., if

anonymity to emi-anonymity (Kalai, 2004). Semi-anonymity accommodate many additional etting in which there are a ymmetrie among t cla e of participant , e.g. in certain kind of two- ided matching market (cf. Kojima and Pathak, 2009). Another example with a ym-metrie among t cla e of participant i   chool choice (e.g., Abdulkadiroglu, Pathak and Roth, 2009), in which agent are treated differently depending on whether they have "walk zone priority", " ibling priority", etc.

the other agents make systematic mistakes, or play a different equilibrium, or if their types have a different distribution than μ. Put differently, even though mechanisms in our analysis are always detail free for the designer, the agents themselves will need to know the prior μ in order to play BNEs.[4] Another equilibrium concept with similar issues to worry about is complete information Nash equilibrium.

Definition 6. *A* ***$\bar{\theta}$-complete information Nash equilibrium*** *($\bar{\theta}$-NE) of n-mechanism Φ^n is a strategy $\sigma^*_{\bar{\theta}}(\cdot)$ such that for all $\theta_i \in \bar{\theta}$ and $\hat{\theta}_i \in \Theta$:*

$$u_{\theta_i}[\Phi^n_i(\sigma^*_{\bar{\theta}}(\theta_i), \sigma^*_{\bar{\theta}}(\theta_{-i}))] \geq u_{\theta_i}[\Phi^n_i(\sigma^*_{\bar{\theta}}(\hat{\theta}_i), \sigma^*_{\bar{\theta}}(\theta_{-i}))]$$

The informational requirements for complete information Nash equilibria are arguably even more severe than for Bayes Nash. Agents must know both the equilibrium strategy $\sigma^*_{\bar{\theta}}(\cdot)$ and the precise realization of the distribution of the other players' types.

Part of the appeal of strategyproof mechanisms is that these informational requirements are no longer concerns. Specifically:

Definition 7. *An n-mechanism Φ^n is* ***strategyproof*** *(SP) if, for all $\theta_i, \hat{\theta}_i \in \Theta$, and all $\theta_{-i} \in \Theta^{n-1}$:*

$$u_{\theta_i}[\Phi^n_i(\theta_i, \theta_{-i})] \geq u_{\theta_i}[\Phi^n_i(\hat{\theta}_i, \theta_{-i})]$$

Definitions (5)-(7) provide equilibrium concepts defined for a particular market size and, in the cases of BNE and NE, a particular distribution of opponent types. What does it mean for a mechanism to have equilibria for any market size and any prior?

Definition 8. *The mechanism $(\Phi^n)_{\mathbb{N}}$ has Bayes-Nash equilibria if, for any n and any $\mu \in \Delta\Theta$, Φ^n has a μ-BNE. The mechanism $(\Phi^n)_{\mathbb{N}}$ has complete information Nash equilibria if, for any n and any $\bar{\theta} \in \Theta^n$, Φ^n has a $\bar{\theta}$-NE. The mechanism $\{\Phi^n\}_n$ is strategyproof if, for any n, Φ^n is strategyproof.*

When researchers report that the Boston mechanism has attractive Bayes-Nash equilibria, or that the Generalized Second Price Auction has attractive complete information Nash equilibria, or that the Random Serial Dictatorship is strategyproof, they are using these terms in the more universal sense of Definition 8.

[4] By the tandard revelation principle (cf. Fudenberg and Tirole, 1991; Section 7.2), for any mechani m with a Baye -Na h equilibrium in which agent mi report their preference , there exi t a direct-revelation mechani m in which telling the truth i a Baye -Na h equilibrium. Thi direct-revelation mechani m, however, i no longer detail free for the de igner; the map between type and outcome will have to vary with the prior. For in tance, in the direct-revelation mechani m ver ion of the fir t-price ealed bid auction, the amount by which the center will hade each type' bid mu t vary with the prior in order for truthful reporting to be a BNE.

2.2 Continuity of Equilibria

For our main results in Section 5, we will need slightly different continuity assumptions from Kalai (2004). Specifically, our Theorems will be for mechanisms whose *equilibria* are equicontinuous, in the following sense:

Definition 9. *A family of Bayes-Nash Equilibria σ_μ^n of mechanism $(\Phi^n)_{\mathbb{N}}$ is* ***(equi)continuous in reports*** *if, for every $\epsilon > 0$, there exists a $\delta > 0$ s.t.: for every n, μ, θ_i, $\hat{\theta}_i$ and any θ_{-i}, θ'_{-i} with $\left|emp[\theta_{-i}] - emp[\theta'_{-i}]\right|_{sup} < \delta$, we have $u_{\theta_i}[\Phi_i^n(\sigma_\mu^n(\hat{\theta}_i), \sigma_\mu^n(\theta_{-i}))] - u_{\theta_i}[\Phi_i^n(\sigma_\mu^n(\hat{\theta}_i), \sigma_\mu^n(\theta'_{-i}))] < \epsilon$.*

The family is ***(equi)continuous in strategies*** *if, for every $\epsilon > 0$, there exists a $\delta > 0$ s.t.: for every n, θ_i, $\hat{\theta}_i$, θ_{-i}, and any priors μ, μ' with $|\mu - \mu'|_{sup} < \delta$, we have $u_{\theta_i}[\Phi_i^n(\sigma_\mu^n(\hat{\theta}_i), \sigma_\mu^n(\theta_{-i}))] - u_{\theta_i}[\Phi_i^n(\sigma_{\mu'}^n(\hat{\theta}_i), \sigma_{\mu'}^n(\theta_{-i}))] < \epsilon$.*

In words, if a family of Bayes-Nash Equilibria is continuous in reports, a small change in the realized distribution of other players' types has just a small affect on a particular player's equilibrium payoff. Continuity in reports is similar to Kalai's (2004) continuity in actions, but for the direct-revelation mechanism versions of the Bayes-Nash equilibria of $(\Phi_n)_{\mathbb{N}}$. If the family is continuous in strategies, then a small change in the strategy all players use in equilibrium – e.g., due to a small change in the prior – has a small affect on a particular player's equilibrium payoff.

Our Theorems 1 and 2 are for mechanisms that have equilibria satisfying these conditions, though what we really use is equicontinuity locally around the actual or conjectured play. Though strong, one possible interpretation of the conditions is the following: if a mechanism is *not* equicontinuous in the above sense, then the analyst's prediction of equilibrium outcomes is highly sensitive to small changes in information about the environment. This itself is arguably an undesirable feature of a mechanism.

The analogous definitions for complete information Nash equilibria are:

Definition 10. *A family of complete information Nash equilibria $\sigma_{\bar{\theta}}^n$ of mechanism $(\Phi^n)_{\mathbb{N}}$ is* ***(equi)continuous in reports*** *if, for every $\epsilon > 0$, there exists a $\delta > 0$ s.t.: for every n, θ_i, $\hat{\theta}_i$, and any θ_{-i}, θ'_{-i} with $\left|emp[\theta_{-i}] - emp[\theta'_{-i}]\right|_{sup} < \delta$, we have $u_{\theta_i}[\Phi_i^n(\sigma_{\bar{\theta}}^n(\hat{\theta}_i), \sigma_{\bar{\theta}}^n(\theta_{-i}))] - u_{\theta_i}[\Phi_i^n(\sigma_{\bar{\theta}}^n(\hat{\theta}_i), \sigma_{\bar{\theta}}^n(\theta'_{-i}))] < \epsilon$.*

The family is ***(equi)continuous in strategies*** *if, for every $\epsilon > 0$, there exists a $\delta > 0$ s.t.: for every n, θ_i, $\hat{\theta}_i$, and any θ_{-i}, θ'_{-i} with $\left|emp[\theta_{-i}] - emp[\theta'_{-i}]\right|_{sup} < \delta$, we have $u_{\theta_i}[\Phi_i^n(\sigma_{\bar{\theta}}^n(\hat{\theta}_i), \sigma_{\bar{\theta}}^n(\theta_{-i}))] - u_{\theta_i}[\Phi_i^n(\sigma_{\bar{\theta}'}^n(\hat{\theta}_i), \sigma_{\bar{\theta}'}^n(\theta_{-i}))] < \epsilon$.*

Our Theorems 3 and 4 are for mechanisms that have equilibria satisfying these conditions.

3 Strategyproofness for "Price Takers"

Suppose that Φ^n is not strategyproof. This means that there exist θ_i, $\hat{\theta}_i$, θ_{-i} such that

$$u_{\theta_i}[\Phi_i^n(\hat{\theta}_i, \theta_{-i})] > u_{\theta_i}[\Phi_i^n(\theta_i, \theta_{-i})] \tag{2}$$

i.e., an agent of type θ_i would like to misreport as $\hat{\theta}_i$. In this section we distinguish between two such kinds of manipulations. Some manipulations persist with market size, whereas others vanish with market size. To formalize this distinction, we first need the concept of a limit mechanism.

3.1 Limit Mechanisms

In an anonymous mechanism, each agent's outcome is a function of his own report and the distribution of all agents' reports. As a market grows larger, each individual agent's ability to influence the aggregate distribution of all reports grows smaller, with this influence converging to zero in the limit. This motivates our definition of a limit mechanism:

Definition 11. *A **limit mechanism** is a function:*

$$\phi^{\infty} : \Theta \times \Delta\Theta \to X$$

*that indicates an agent's outcome as a function of his report $\theta \in \Theta$ and the distribution of all agents' reports $m \in \Delta\Theta$. The limit mechanism ϕ^{∞} is the **limit of mechanism** Φ^n if, for all θ, m:*

$$\phi^{\infty}(\theta, m) = \lim_{n\to\infty} \phi^n(\theta|m)$$

where ϕ^n is as defined in (1).

A feature of our method of taking the limit is that each ϕ^n in the sequence is random, in the sense that the types of the agent's $n-1$ opponents are stochastic (drawn from distribution m). In the limit this randomness vanishes due to the law of large numbers, and so ϕ^{∞} itself is not random in this way. This is in contrast with, e.g., Debreu and Scarf's (1963) replicator economy, or with the approach pioneered by Aumann (1964) that looks directly at a continuum economy. This randomness seems intuitively appealing in the context of the study of large anonymous mechanisms.

Assumption 3. *Mechanisms have limits.*

Most (if not all) practical market design mechanisms we are aware of have limits, but we note that it is very easy to construct examples of mechanisms that do not. For instance, if a mechanism acts differently depending on whether n is even or odd it will not have a limit.

3.2 A Distinction Between Two Kinds of Non-Strategyproofness

We propose the following distinction between kinds of manipulations.

Definition 12. *Suppose, as in (2), that $u_{\theta_i}[\Phi_i^n(\hat{\theta}_i, \theta_{-i})] > u_{\theta_i}[\Phi_i^n(\theta_i, \theta_{-i})]$ for suitable n, θ_i, $\hat{\theta}_i$, and θ_{-i}, and let $m = emp[\theta_i, \theta_{-i}]$. This manipulation is said to **persist with market size** if:*

$$\phi^{\infty}(\hat{\theta}_i, m) > \phi^{\infty}(\theta_i, m) \tag{3}$$

If inequality (3) does not obtain, then we say that the manipulation ***vanishes with market size****.*

We sometimes refer to manipulations that persist with market size as "first-order manipulations" and to manipulations that vanish with market size as "second-order manipulations". The desiderata that we propose as a second-best alternative to strategyproofness is that mechanisms not have first-order manipulations:

Definition 13. *Mechanism* $(\Phi^n)_{\mathbb{N}}$ *is* ***strategyproof for "price takers"****, or* $\mathbf{SP(p)}$*, if, for all* $\theta_i, \hat{\theta}_i, m$*:*

$$\phi^{\infty}(\theta_i, m) \geq \phi^{\infty}(\hat{\theta}_i, m).$$

Equivalently, a mechanism is $SP(p)$ *if all finite-economy manipulations vanish with market size. If a mechanism has manipulations that persist with market size then it is* ***manipulable by "price takers"****, or* $\mathbf{Manip(p)}$*.*

We use the terminology "price takers" because the agents in our limit economy are like the price-taking agents familiar in other areas of economics. Further, the most familiar examples of mechanisms that are $SP(p)$ are those that use prices – e.g., the Walrasian mechanism or double auctions. However, there are many mechanisms that do not explicitly use prices that are nevertheless $SP(p)$. There also are numerous mechanisms that do explicitly use prices that nevertheless are manipulable even for price takers. In the following section we describe a class of mechanisms for which checking whether the mechanism is $SP(p)$ or $Manip(p)$ is especially simple.

3.3 Mechanisms with p-Based Representations

In general a mechanism is a sequence of functions that maps a vector of actions to a vector of outcomes. A number of the mechanisms we are interested in have a more compact representation:

Definition 14. *Mechanism* $(\Phi^n)_{\mathbb{N}}$ *has a* **p***-****based representation*** *if there exists a compact set* P*, an onto p-function* $p : \Delta\Theta \to P$*, and an outcome function* $x : \Theta \times P \to X$ *such that, for all* n *and all* $\bar{\theta} \in \Theta^n$*:*

$$\big(x(\theta_1, p(emp[\bar{\theta}])), \ldots, x(\theta_n, p(emp[\bar{\theta}]))\big) = \Phi^n(\bar{\theta}).$$

In words, when a mechanism has a p-based representation, each agent's outcome is a function of his own report and a set of statistics, p. The statistics themselves are a function of the distribution of all reports. The statistics could be a set of prices, as in a Walrasian setting. Or, in algorithms such as the Boston Mechanism for school choice or the HBS Draft mechanism for course allocation, the statistics are algorithm "sell-out times" (cf. Budish and Cantillon, 2011).

For this class of mechanisms, our concept of strategyproof for price takers takes an especially simple form:

Remark 1 (Checking SP(p)). *A mechanism with a p-based representation is $SP(p)$ iff, for all $\theta, \hat{\theta} \in \Theta$ and all $p \in P$, $u_\theta[x(\theta, p)] \geq u_\theta[x(\hat{\theta}, p)]$.*

Proof. By the assumption that mechanisms are equicontinuous in actions and have limits, $x(\theta, p(m))$ is continuous in m and, for all θ, m, we have: $x(\theta, p(m)) = \phi^\infty(\theta, m)$.

[If] If $u_\theta[x(\theta, p)] \geq u_\theta[x(\hat{\theta}, p)]$ for all θ, $\hat{\theta}$, p, then for any $m \in \Delta\Theta$, we have $\phi^\infty(\theta, m) = x(\theta, p(m)) \geq_{u_\theta} x(\hat{\theta}, p(m)) = \phi^\infty(\hat{\theta}, m)$, hence by Definition 13 the mechanism is $SP(p)$.

[Only If] Suppose $u_\theta[x(\theta, p)] < u_\theta[x(\hat{\theta}, p)]$. Since $p(\cdot)$ is onto there exists m such that $p(m) = p$. Hence $\phi^\infty(\theta, m) = x(\theta, p(m)) <_{u_\theta} x(\hat{\theta}, p(m)) = \phi^\infty(\hat{\theta}, m)$, hence by Definition 13 the mechanism is $Manip(p)$.

□

That is, to check whether a mechanism is $SP(p)$, we just need to study the outcome function. If, for any fixed prices p, reporting truthfully selects each type of agent's most preferred outcome, then the mechanism is $SP(p)$. If not, it is not. We call this set of potential outcomes the agent's opportunity set:

Definition 15. *In a mechanism with a p-based representation, the* ***opportunity set at*** **p** *is the set $\{x(\theta', p)\}_{\theta' \in \Theta}$.*

The opportunity set allows us to ask what outcomes an agent could achieve by varying his report, if somehow the aggregate distribution of reports were held fixed. The finite economy manipulations of $SP(p)$ mechanisms involve agents' misreporting so as to advantageously change their opportunity set:

Remark 2 (Nature of Manipulations). *If a mechanism has a p-based representation and is $SP(p)$, then any manipulations in finite markets take the following form: there exists $\theta_i, \hat{\theta}_i, \theta_{-i}$, with $p = p(emp[\theta_i, \theta_{-i}])$ and $\hat{p} = p(emp[\hat{\theta}_i, \theta_{-i}])$, such that*

$$u_{\theta_i}[x(\hat{\theta}_i, \hat{p})] \geq u_{\theta_i}[x(\theta_i, p)].$$

But from Remark 1 we also have that

$$u_{\theta_i}[x(\theta_i, \hat{p})] \geq u_{\theta_i}[x(\hat{\theta}_i, \hat{p})].$$

Putting these two inequalities together we have the following interpretation: by misreporting as $\hat{\theta}_i$ instead of θ_i, the agent changes the mechanism's statistics ("prices") from p to $\hat{p}$, and this change in statistics is sufficiently advantageous for the agent that it more than compensates for the fact that, at these new statistics $\hat{p}$, the agent is allocated a bundle based on false preferences $\hat{\theta}_i$ and not his true preferences θ_i. If he could obtain $x(\theta_i, \hat{p})$ he would be better off still.

Our last observation on mechanisms with p-based representation relates to fairness, one of the advantages the literature has attributed to strategyproof mechanisms over other forms of implementation. Perhaps the most widely used notion of fairness in the economics literature on distributive justice is Foley's

(1967) *envy freeness*: an allocation is envy free if each agent weakly prefers his own allocation to any other agent's allocation. We have the following simple observation:

Remark 3 (SP(p) and Envy Freeness)**.** *If a mechanism has a p-based representation and is $SP(p)$, then it is envy free for truthful players. That is, for all types θ_i and θ_j and any realization of the statistics p, we have:*

$$u_{\theta_i}[x(\theta_i, p)] \geq u_{\theta_i}[x(\theta_j, p)]$$

An agent who misreports, even if this misreporting is profitable, is not guaranteed an envy free allocation. That is because it is possible, in the terminology of Remark 2, to have

$$u_{\theta_i}[x(\theta_i, \hat{p})] > u_{\theta_i}[x(\hat{\theta}_i, \hat{p})] > u_{\theta_i}[x(\theta_i, p)]$$

in which case type θ_i profits from misreporting as type $\hat{\theta}_i$, but will envy any other agent who reports θ_i.

In practice, Remark 3 enables a market administrator to advise participants that "Only reporting truthfully guarantees that you will prefer your allocation to any other agents' allocation." This is not as powerful a piece of strategic advice as that enabled by a strategyproof mechanism, but it may nevertheless be useful in practice. In particular, if a mechanism has a p-based representation and is $SP(p)$, then truthful reporting may be a "safe" strategy in the informal sense of Roth (2008). See further discussion on this point in the conclusion.

4 Examples

In this section we illustrate the concepts of Section 3 with a series of examples. Our first example is uniform-price and pay-as-bid auctions, two mechanisms best known for their use in the allocation of government securities. The former is an example of a price-based mechanism that is $SP(p)$, the latter an example of a price-based mechanism that is $Manip(p)$.

Example 1 (Multi-Unit Auctions)**.** *There are qn units of a homogeneous good, with $q \in \mathbb{Z}_+$.*

To simplify notation, we assume that agents' preferences take the form of linear utility functions, up to a capacity limit. Specifically, each agent i's type consists of a per-unit value v_i and a maximum capacity q_i, with $V = \{1, \ldots, \bar{v}\}$ the set of possible values, $Q = \{1, \ldots, \bar{q}\}$ the set of possible capacity limits, and $\Theta = V \times Q$. The set of possible actions is just $A = \Theta$, with τ the identify function.

For both Uniform-Price and Pay-as-Bid auctions, price is calculated as a function of the reports $\bar{\theta}$ as follows:

$$p^*(\bar{\theta}) = \max_p \sum_{i=1}^{n} q_i \cdot \mathbf{1}\{v_i \geq p\} \geq qn$$

i.e., p^* is the highest price at which demand weakly exceeds supply. Allocations of the good are equivalent across the two mechanisms: an agent who reports (v_i, q_i) is allocated q_i units if $v_i > p^*$, is allocated 0 units if $v_i < p^*$, and is rationed if $v_i = p^*$. In the limit, if the measure of agents' reports is $d\theta$, then price can be calculated as[5]

$$p^* = \max_p \int q_i \cdot \mathbf{1}\{v_i \geq p\}\ d\theta \geq q$$

Notice that in the limit mechanism each agent regards price as exogenous to her own report, because they cannot affect $d\theta$. It is easy to see that the Uniform-Price Auction is $SP(p)$ whereas the Pay-as-Bid auction is $Manip(p)$. In particular, in the Pay-as-Bid auction an agent of type (v_i, q_i) with $v_i > p^* + 1$ wishes to misreport as $(\hat{v}_i = p^* + 1, \hat{q}_i = q_i)$ to get the same quantity at a strictly lower price than if he reports truthfully.

This example is consistent with Milton Friedman's (1991) observation that "you do not have to be a specialist" to participate in the uniform price auction, because you can just indicate "the maximum amount you are willing to pay for different quantitites ... if you bid a higher price [than the market clearing price], you do not lose as you do under the current [pay-as-bid] method." Friedman seems to be talking about the absence of what we call first-order manipulations, and seems to be less concerned by the second-order manipulability of the uniform-price auction.

Our next example is the Boston mechanism for school choice, a mechanism that does not explicitly have prices in the description (though it does have a p-based representation, as we will see below). As mentioned in the introduction, this mechanism was criticized by Abdulkadiroglu and Sonmez (2003) and Abdulkadiroglu et al (2006) for not being strategyproof. We show something stronger, which is that it is not even $SP(p)$.

Example 2 (Boston Mechanism)**.** *Let X_0 be the set of schools, each with capacity $q = \left\lceil \frac{n}{|X_0|} \right\rceil$.*

Agents' types take the form of von-Neumann Morgenstern utility functions over the set of schools, i.e., functions of the form $u_\theta : X_0 \to \{0, 1, \ldots, \bar{u}\}$ for large integer $\bar{u}$. The set of actions A is the set of ordinal preferences over X_0, with $\tau(\theta)$ denoting the true ordinal preferences of type θ.

The Boston mechanism awards as many students as possible their reported first choice school; then, awards as many students as possible their reported second choice school; etc. To keep the description concise we focus just on the first choices. Let d_j^1 denote the number of students who report that school $j \in X_0$ is their first choice: these students receive school j with probability $\min(1, \frac{q}{d_j^1})$. Let $p_j = \min(1, \frac{q}{d_j^1})$.

[5] The notation $\mathbf{1}\{statement\}$ denote the indicator function which return 1 if the tatement i true and 0 if the tatement i fal e.

In the limit, if the measure of agents' reports is $d\hat{\theta}$, then the probability that a student who ranks j first gets it can be calculated as

$$p_j^* = \min(1, \frac{1}{|X_0| \int \mathbf{1}\{j \text{ is first choice}\} d\hat{\theta}})$$

Notice that in the limit mechanism each agent regards the p_j^*'s as exogenous to their own report. Agent θ will wish to misreport her first choice school if there exists $j' \neq j$ such that $u_\theta(j) p_j^* < u_\theta(j') p_{j'}^*$. Therefore the mechanism is $Manip(p)$.

Our next example contrasts two mechanisms for the course-allocation problem that are based on articficial currency markets. The widely-used Bidding Points Auction, studied by Sonmez and Unver (2010) is based on the strategyproof Vickrey auction. However, because it uses fake money and not real money it is not strategyproof; in fact we show that it is $Manip(p)$. A new mechanism introduced by Budish (2010) is based on the general equilibrium theory idea of Competitive Equilibrium from Equal Incomes (CEEI). It too is not strategyproof, but it is $SP(p)$.

Example 3 (Course Allocation)**.** *Let C be the set of courses, and X_0 the powerset of C (for convenience we ignore scheduling constraints). Each course has capacity $q = \lceil \kappa n \rceil$, with $\kappa \in [0, 1]$. Students have additive-separable preferences over courses, normalized so that their utility from consuming one seat in each course is $b \in \mathbb{Z}_+$. Specifically, student i's type is described by a vector $v_i \in \{0, 1, \ldots, b\}^{|C|}$ where component v_{ij} indicates i's utility from course j, and $\sum_j v_{ij} = b$. Her utility from schedule $x \in X_0$ is $\sum_j v_{ij} \mathbf{1}\{c \in x\}$. The set of possible actions is equal to the set of types.*

In both the Bidding Points Auction and CEEI, the constant b will play the role of students' "budgets" of an artificial currency with no outside use. Suppose the agents submit type profile v. In the Bidding Points Auction (cf. Sonmez and Unver, 2010), prices are calculated as in a real-money multi-unit Vickrey auction, as follows:

$$p_j^* = \max_p \sum_{i=1}^{n} \mathbf{1}[v_{ij} \geq p] \geq q + 1 \tag{4}$$

Then, the q highest bidders for j are awarded a seat in the course. In the limit, the formula becomes

$$p_j^* = \max_p \int \mathbf{1}[v_{ij} \geq p] \; dv \geq \kappa$$

It is easy to see that, despite the resemblance to the strategyproof Vickrey auction, the BPA is $Manip(p)$. The reason is that agent i's most preferred affordable bundle at price vector p^* is

$$x_i^* = \arg \max_{x \in \{0,1\}^{|C|}} \left(\sum v_{ij} x_{ij} : \sum p_j^* x_{ij} \leq b \right) \tag{5}$$

but the BPA instead awards agent i the bundle

$$x_i^{BPA} = \arg \max_{x \in \{0,1\}^{|C|}} \left(\sum v_{ij} x_{ij} - \sum p_j^* x_{ij} \right). \tag{6}$$

Whenever $x_i^* \neq x_i^{BPA}$, agent i profits from misreporting. The CEEI-based mechanism proposed by Budish (2010) calculates prices in such a way that it is then able to award agents bundles according to (5). This ensures that the mechanism is $SP(p)$.

There are numerous other examples. For single-unit assignment problems such as in Example 2, Hylland and Zeckhauser's (1979) pseudomarket mechanism is an example of a price-based mechanism that is $SP(p)$, while Bogomolnaia and Moulin's (2001) probabilistic serial mechanism is an example of a mechanism that does not explicitly use prices in the original description but that is $SP(p)$ (cf. Kojima and Manea, 2009). For multi-unit assignment problems such as in Example 3, Harvard Business School's draft mechanism is an example of a non-price mechanism that is $Manip(p)$ (cf. Budish and Cantillon, 2011), whereas the generalization of Hylland and Zeckhauser's mechanism to multi-unit demand is $SP(p)$ (Budish, Che, Kojima and Milgrom, 2011).

The concepts can also be applied to two-sided matching mechanisms, if we generalize the class of mechanisms considered to be the class of semi-anonymous mechanisms (Kalai, 2004), and not just anonymous mechanisms. Then, techniques in Kojima and Pathak (2009) or Azevedo and Leshno (2011) can be used to show that Gale and Shapley's deferred acceptance algorithm is $SP(p)$ in semi-anonymous environments. It is also easy to see that the priority-match algorithm, criticized by Roth (2002) and others, is $Manip(p)$.

The following table summarizes this informal discussion.

Table 1. Which Non-SP Market Designs are SP(p)?

Problem	**Manipulable for Price Takers**	**Strategyproof for Price Takers**
Single-unit Assignment	Boston Mechanism	Prob Serial, HZ Pseudomarket
Multi-unit Assignment	Bidding Points Auction	Approximate CEEI
	HBS Draft Mechanism	Generalized HZ Pseudomarket
Multi-unit Auctions	Pay-as-Bid Auctions	Uniform-Price Auctions
Two-Sided Matching	Priority-Match Algorithm	Deferred Acceptance Algorithm

5 Main Results

Our main result is that, in our class of mechanisms, strategyproofness is in a certain sense "free" in the limit and "approximately free" in large finite economies, relative to Bayes-Nash incentive compatibility and complete information Nash equilibrium. Thus, whereas strategyproofness often severely limits what kinds of mechanisms are possible, there is an important sense in which $SP(p)$ does not.

We give the results for Bayes-Nash incentive compatibility in Sections 5.1-5.2, and then give the analogous results for complete information Nash equilibrium in Section 5.3.

5.1 Limit Result for Bayes-Nash Equilibria

Our limit result is the following:

Theorem 1 (Limit Result for BNE)**.** *Consider a mechanism $(\Phi^n)_{\mathbb{N}}$ with Bayes-Nash equilibria σ^n_μ continuous in reports and strategies, and with $\lim_{n\to\infty}\sigma^n_\mu = \sigma^\infty_\mu$. There exists another mechanism $(F^n)_{\mathbb{N}}$ that is strategyproof for price takers and gives agents the same utilities as the original mechanism $(\Phi_n)_{n\in\mathbb{N}}$ in the limit. That is, for any type $\theta \in \Theta$ and any prior $\mu \in \Delta\Theta$, $u_\theta[f^\infty(\theta|\mu)] = u_\theta[\phi^\infty(\sigma^\infty_\mu(\theta)|\sigma^\infty_\mu(\mu))]$.*

Proof. We construct the mechanism $(F^n)_{\mathbb{N}}$ as follows. Suppose in a market of size n agents report $\bar{\theta}$. Let $m = emp[\bar{\theta}]$ denote the empirical distribution of $\bar{\theta}$ in $\Delta\Theta$. Let:

$$F^n(\bar{\theta}) = \Phi^n(\sigma^n_m(\bar{\theta})) \tag{7}$$

In words, F^n plays action $\sigma^n_m(\theta_i)$ for agent i who reports θ_i, where m is not the true distribution of agents' types μ (which is not known to the mechanism) but rather the *empirical distribution of agents' reports*. (In the Bayes-Nash equilibria of Φ^n agent θ_i plays $\sigma^n_\mu(\theta_i)$).

As in (1) above, the object $f^n(\theta_i|m)$ denotes an agents' outcome under F^n when he reports θ_i and the $n-1$ other agents' reports are distributed iid according to some $m \in \Delta\Theta$.

Lemma 1. *For each $m \in \Delta\Theta$ and each $\theta_i, \hat{\theta}_i \in \Theta$,*

$$\lim_{n\to\infty}[u_{\theta_i}[f^n(\hat{\theta}_i|m)] - u_{\theta_i}[\phi^n(\sigma^n_m(\hat{\theta}_i)|\sigma^n_m(m))]] = 0$$

Proof of Lemma. By the construction of $(F^n)_{n\in\mathbb{N}}$ we have

$$f^n(\hat{\theta}_i|m) = \Phi^n_1(\sigma^n_{\hat{m}}(\hat{\theta}_i), \sigma^n_{\hat{m}}(\theta_{-i})),$$

where θ_{-i} is an $n-1$ vector of types drawn iid according to m, and then $\hat{m} = emp[\hat{\theta}_i, \theta_{-i}]$. Recall that

$$\phi^n(\sigma^n_m(\hat{\theta}_i)|\sigma^n_m(m)) = \Phi^n_1(\sigma^n_m(\hat{\theta}_i), \sigma^n_m(\theta_{-i})).$$

Hence, we need to show that

$$u_{\theta_i}[\Phi^n_1(\sigma^n_{\hat{m}}(\hat{\theta}_i), \sigma^n_{\hat{m}}(\theta_{-i}))] - u_{\theta_i}[\Phi^n_1(\sigma^n_m(\hat{\theta}_i), \sigma^n_m(\theta_{-i}))] \tag{8}$$

gets small as n gets large. Without loss of generality, normalize utility functions such that the range of each u_θ is $[0,1]$. Let $B^n_\delta = \{\theta_{-i} \in \Theta^{n-1}$:

$\Big|emp[\hat{\theta}_i, \theta_{-i}] - m\Big|_{sup} < \delta\}$ denote the set of realizations of θ_{-i} such that the realized empirical distribution is close to the distribution, m, from which each of the $n-1$ elements of θ_{-i} are drawn iid. Fix $\epsilon > 0$. By equicontinuity in strategies, we can choose $\delta > 0$ such that, if the realization of θ_{-i} is in B_δ^n, then the realized value of (8) is less than $\frac{\epsilon}{2}$. By the law of large numbers, we have that the probability that θ_{-i} is in B_δ^n converges to 1 as $n \to \infty$; in particular we can take n_0 large enough such that, for all $n \geq n_0$, the probability that the realization of θ_{-i} is in B_δ^n is at least $1-\frac{\epsilon}{2}$. In the other $\frac{\epsilon}{2}$ of cases, the realization of (8) is at most 1, because of our utility normalization. Hence

$$u_{\theta_i}[\Phi_1^n(\sigma_{\hat{m}}^n(\hat{\theta}_i), \sigma_{\hat{m}}^n(\theta_{-i}))] - u_{\theta_i}[\Phi_1^n(\sigma_m^n(\hat{\theta}_i), \sigma_m^n(\theta_{-i}))] \leq (1 - \frac{\epsilon}{2})\frac{\epsilon}{2} + \frac{\epsilon}{2} \leq \epsilon$$

which completes the proof of the lemma.

□

Given the Lemma on the limit of our mechanism $(F^n)_{\mathbb{N}}$, the next step is to show that in this limit it is a dominant strategy for each type to report truthfully. This requires that for any type θ_i, any alternate report $\hat{\theta}_i$, and any distribution m, we have that:

$$u_{\theta_i}[f^\infty(\theta_i, m)] \geq u_{\theta_i}[f^\infty(\hat{\theta}_i, m)]$$

which given the Lemma is equivalent to:

$$u_{\theta_i}[\phi^\infty(\sigma_m^\infty(\theta_i), m)] \geq u_{\theta_i}[\phi^\infty(\sigma_m^\infty(\hat{\theta}_i), m)] \tag{9}$$

which obtains because σ_m^∞ is the limit of the sequence of Bayes-Nash equilibrium strategies $(\sigma_m^n)_{\mathbb{N}}$. That is, if agents' report according to m, and so σ_m^∞ is the limit Bayes-Nash equilibrium that is "activated" by our new mechanism $(F^n)_{\mathbb{N}}$, then each type θ_i wants to report his own type truthfully, so that the mechanism plays the correct BNE response on his behalf. Since this obtains for any $m \in \Delta\Theta$, we have dominant strategy incentives in the limit, as required.

The remaining step is to show that truthful play of our new mechanism coincides with Bayes-Nash equilibrium play of the original mechanism in the limit. This follows from the Lemma, setting m equal to the true prior, μ.

□

The key idea in the proof is (7), which indicates how to construct the $SP(p)$ mechanism $(F^n)_{\mathbb{N}}$ given the Bayes-Nash mechanism $(\Phi^n)_{\mathbb{N}}$. In words, the empirical distribution of agents' reports, $m \in \Delta\Theta$, "activates" the Bayes-Nash equilibrium strategy $\sigma_m(\cdot)$. An agent who reports type θ thus plays what her Bayes-Nash equilibrium action *would have been* if the true prior were m, not μ. By construction, the distribution of opponents' actions will be very close to what it would have been were the true prior m, and then everyone played according to $\sigma_m(\cdot)$. Thus, even if m is very different from the true prior μ, our agent remains happy to have told the truth – it does not matter that the true prior is μ, because the other plays are behaving "as if" the true prior is m.

This "trick" is the reason why our mechanism in (7) provides dominant-strategy incentives as opposed to just Bayes-Nash incentives. Furthermore, this trick allows the mechanism to be prior free for the participants; they need not actually know the true μ to play the mechanism, they just need report their true type.

Our mechanism $(F^n)_{\mathbb{N}}$ is importantly different from a traditional Bayes-Nash direct revelation mechanism (cf. Fudenberg and Tirole, 1991; Section 7.2). In a traditional Bayes-Nash DRM, the mechanism needs to know the prior μ. It then announces a BNE strategy $\sigma_\mu(\cdot)$, and plays $\sigma_\mu(\theta)$ on behalf of an agent who reports θ. Our mechanism *infers* a prior from the empirical distribution of approximate strategyproofness and with respect to matching the outcome of the original Bayesian mechanism from which our mechanism is derived.

Theorem 2 (Large Finite Markets Result for BNE)**.** *Consider a mechanism* $(\Phi^n)_{\mathbb{N}}$ *with Bayes-Nash equilibria* σ_μ^n *continuous in reports and strategies. There exists another mechanism* $(F^n)_{\mathbb{N}}$ *with the following property: For any* $\epsilon > 0$*, there exist constants* $\alpha > 0$ *(*$= \alpha((\Phi^n)_{\mathbb{N}}, \epsilon)$*) and* $0 < \beta < 1$ *(*$= \beta((\Phi^n)_{\mathbb{N}}, \epsilon)$*), such that the mechanism* $(F^n)_{\mathbb{N}}$*:*

1. *is* $(\epsilon, \alpha\beta^n)$*-strategyproof in markets of size* n
2. *has outcomes under truthful play that* $(\epsilon, \alpha\beta^n)$ *approximate the Bayes-Nash equilibrium outcomes under the original mechanism. Specifically, for all* θ_i*, and for* θ_{-i} *drawn iid according to the true prior* μ*, we have* $\Pr[u_{\theta_i}[F_i^n(\theta_i, \theta_{-i})] - u_{\theta_i}[\Phi_i^n(\sigma_\mu^n(\theta_i), \sigma_\mu^n(\theta_{-i}))] > \epsilon] < \alpha\beta^n$*.*

The proof of Theorem 2 is contained in Appendix A. The first step of the proof shows that, for any belief about the distribution of opponents' reports m, truth telling under our mechanism approximately maximizes *expected* utility. The reason this is true for the true belief, μ, is the continuity in strategies assumption: if players report truthfully, the empirical distribution $\hat{\mu}$ will be "close" to the true distribution μ, and hence $\sigma_{\hat{\mu}}(\cdot)$ will be close in terms of outcomes to $\sigma_\mu(\cdot)$. The reason this then can be extended to any belief $m \neq \mu$ is continuity combined with the method of mechanism construction: if the expectation is m, then the realization $\hat{m}$ will be close to m, and hence the "activated" strategy $\sigma_{\hat{m}}(\cdot)$ will be close to the expected strategy $\sigma_m(\cdot)$.

The second step of the proof then uses the expected utility analysis and law of large numbers techniques due to Kalai (2004) to show that truth telling also approximately maximizes *realized* utility, with high probability.

Remark 4 (Small Cost of Optimizing)**.** *Suppose that there is a cost* $c > 0$ *associated with calculating an optimal strategy, that truthful reporting is costless, and (wlog) that the range of each* u_θ *is* $[0, 1]$*. Then Theorem 2 implies that in a large enough finite market, for any conjecture* $m \in \Delta\Theta$ *about how the other agents will play, reporting truthfully and hence avoiding the cost c is expected utility maximizing in our mechanism. Simply set* $\epsilon = \frac{c}{2}$ *and then choose* n *large enough that* $\alpha\beta^n \leq \frac{c}{2}$*.*

5.3 Results for Complete Information Nash Equilibria

[Especially preliminary]

In this section we provide results analogous to Theorems 1 and 2 but for complete information Nash equilibria, instead of Bayes-Nash equilibria. It is important to keep in mind that assuming that a mechanism has complete information Nash equilibria is very different from assuming that participants actually have complete information.

Theorem 3 (Limit Result for NE)**.** *Consider a mechanism* $(\Phi^n)_{\mathbb{N}}$ *with complete information Nash equilibria* σ_θ^n *continuous in reports and strategies. There exists* *another mechanism $(F^n)_{\mathbb{N}}$ that is strategyproof for price takers and gives agents the same utilities as the original mechanism $(\Phi_n)_{n\in\mathbb{N}}$ in the limit.*

Theorem 4 (Large Finite Markets Result for NE)**.** *Consider a mechanism $(\Phi^n)_{\mathbb{N}}$ with complete information Nash equilibria $\sigma^n_{\bar{\theta}}$ continuous in reports and strategies. There exists another mechanism $(F^n)_{\mathbb{N}}$ with the following property: For any $\epsilon > 0$, there exist constants $\alpha > 0$ $(= \alpha((\Phi^n)_{\mathbb{N}}, \epsilon))$ and $0 < \beta < 1$ $(= \beta((\Phi^n)_{\mathbb{N}}, \epsilon))$, such that the mechanism $(F^n)_{\mathbb{N}}$:*

1. *is $(\epsilon, \alpha\beta^n)$-strategyproof in markets of size n*
2. *has outcomes under truthful play that exactly coincide with the complete information Nash equilibrium outcomes under the original mechanism.*

The mechanism we use to prove these results is especially simple:

$$F^n(\bar{\theta}) = \Phi^n(\sigma^n_{\bar{\theta}}(\bar{\theta})) \tag{10}$$

In words, the mechanism computes a complete-information Nash equilibrium of the game induced by the players' reports, and then executes these actions on behalf of each player. Note, somewhat subtly, that it is *not* actually a Nash equilibrium for each player to report their preferences truthfully to this mechanism in finite markets. The reason is that, by changing one's report from say θ_i to $\hat{\theta}_i$, one changes the profile of reported types from say $\bar{\theta}$ to $\hat{\bar{\theta}}$, and this in turn changes the strategy that is activated from $\sigma^n_{\bar{\theta}}(\cdot)$ to $\sigma^n_{\hat{\bar{\theta}}}(\cdot)$. Thus, i changing his *report* can have the effect of the mechanism changing j's *action*, from $\sigma^n_{\bar{\theta}}(\theta_j)$ to $\sigma^n_{\hat{\bar{\theta}}}(\theta_j)$.

The reason our mechanism constructed according to (10) provides approximate dominant strategy incentives in large finite markets, and exact incentives in the limit, is that, by the continuity in strategies assumption, as $emp[\bar{\theta}] - emp[\hat{\bar{\theta}}]$ gets small the effect of a change from $\sigma^n_{\bar{\theta}}(\cdot)$ to $\sigma^n_{\hat{\bar{\theta}}}(\cdot)$ on i's utility gets small. As the market gets larger, i's ability to affect the empirical distribution vanishes to zero. Thus in a large market, for any fixed set of reports by i's opponents $\hat{\theta}_{-i}$, our mechanism will activate a strategy "close" to $\sigma^n_{(\theta_i,\hat{\theta}_{-i})}(\cdot)$ no matter how i reports. Since θ_i is exactly a best response to $\hat{\theta}_{-i}$ if the strategy is fixed to be exactly $\sigma^n_{(\theta_i,\hat{\theta}_{-i})}(\cdot)$, it is approximately a best response for i to tell the truth when the strategy is always close to $\sigma^n_{(\theta_i,\hat{\theta}_{-i})}(\cdot)$, even though the strategy varies slightly with i's own report. Since this argument holds for any $\hat{\theta}_{-i}$, our mechanism transforms exact complete-information Nash equilibria into a mechanism that provides approximate dominant strategy incentives.

Note as well that, if agents tell the truth in finite markets, this mechanism produces outcomes that are identical to the outcomes under the NE of the original mechanism. By contrast with BNE our mechanism only approximates the finite market outcomes (cf. how Theorem 4(2) is stronger than Theorem 2(2)).

An Empirical Application A recent paper by Budish and Cantillon (2011) applies the mechanism (10) empirically, in the context of course allocation at Harvard Business School. The original mechanism $(\Phi^n)_{\mathbb{N}}$ is HBS's draft mechanism, in which students take turns choosing courses one at a time over a series of rounds. They call the constructed mechanism $(F^n)_{\mathbb{N}}$ the "proxy draft", to denote that students submit their preferences to the mechanism, which then acts as a strategic proxy on each student's behalf. They find evidence that the proxy draft improves welfare relative to the original mechanism, in part because it prevents strategic mistakes. That is, the "robustness" of dominant strategy equilibria has a payoff in terms of welfare.

6 Discussion

A goal of this paper is to persuade market designers that strategyproofness for "price takers" ($SP(p)$) is a useful desideratum for practice, especially in the numerous problem contexts where strategyproof mechanisms are known to be unattractive. Our main formal results in support of this argument are theorems which suggest that, in a class of mechanisms, $SP(p)$ is "free" in the limit relative to other forms of incentive compatibility, and approximately free in large finite markets. We conclude the paper with a few informal arguments in support of $SP(p)$.

Empirical Evidence on Manip(p) and SP(p) Mechanisms There are several empirical studies of mechanisms which are $Manip(p)$ according to our definition, and which have been shown to have important incentives problems in practice. These include Jegadeesh (1993) and others on the 1991 pay-as-bid auction scandals, Abdulkadiroglu et al (2006, 2009) on the Boston mechanism for school choice, Budish and Cantillon (2011) on Harvard Business School's course-allocation draft mechanism, Krishna and Unver (2008) and Budish (2010) on the Bidding Points Auction, Edelman and Ostrovsky (2007) on pay-as-bid keyword auctions, Cramton et al (2011) and Kim et al (2011) on a proposed Medicare auction for durable equipment, Roth (2002) and others on non-stable matching algorithms such as the priority match, and potentially others. By contrast, to the best of our knowledge, there are no empirical examples of market designs that are $SP(p)$ but which have been shown to be harmfully manipulated in large finite markets.

To the extent that this pattern is indeed true, it suggests that perhaps the relevant distinction for practice, in contexts with a large number of participants, is not SP vs. non-SP, but rather $SP(p)$ vs. $Manip(p)$. Or, more conservatively, SP vs. $SP(p)$ vs. $Manip(p)$.

Several Arguments for Strategyproof Design are also Arguments for SP(p) Design In traditional mechanism design, incentives are viewed as a constraint, not an objective. A number of recent papers in the market design literature have suggested, either formally or informally, that strategyproofness be viewed as an explicit design objective.

One such argument is that strategyproof mechanisms are fair, in the sense that they do not penalize participants who are strategically unsophisticated (cf. Abdulkadiroglu et al, 2006; Pathak and Sonmez, 2008). Our Remark 3 suggests that an important class of $SP(p)$ mechanisms are fair, in the sense that agents who report their preferences truthfully are guaranteed to prefer their allocation to any other agent's allocation. Sophisticated agents able to find profitable second-order manipulations may be able to improve their welfare by misreporting, but even so unsophisticated agents who tell the truth will not envy them.

Another such argument is that strategyproof mechanisms eliminate any unmodeled costs of calculating an optimal response; e.g., Roth (2008) argues that good markets are "sufficiently simple to participate in". For any such cost $\epsilon > 0$, our Theorems 2 and 4 indicate that under our $SP(p)$ mechanism, in a large enough market, an agent is better off simply telling the truth than incurring the ϵ cost and then optimizing accordingly.

Last, Roth (2008) argues that good markets make it "safe to participate straightforwardly", as opposed to "engaging in costly and risky strategic behavior." Though "safe" is never defined formally, some mechanisms that are strategyproof are described as safe, as well as some other mechanisms, like deferred acceptance, that are not strategyproof but that are $SP(p)$. We propose that $SP(p)$ mechanisms, particularly those with a p-based representation and hence envy-free for truth-tellers, are "safe" in the informal sense of Roth (2008).

Two Notes of Caution Our arguments above suggest that $SP(p)$ may be viewed as a kind of necessary condition for good design in large anonymous settings. However we conclude with a note of caution about viewing it as sufficient. Though a useful approximation in many markets of interest, the assumption of price-taking behavior is never exactly correct. Even wheat farmers are atomic.

That said, we anticipate that market designers will often be faced with the following choice: use a mechanism that is attractive under truthful play and $SP(p)$, but not strategyproof; or, use a mechanism that is less attractive under truthful play but exactly strategyproof? Our paper also cautions against ignoring the former in favor of the latter, which is what researchers do whenever they limit attention to strategyproof mechanisms.

A Omitted Proofs

Proof of Theorem 2. For ease of exposition fix arbitrary $\theta_i \in \Delta\Theta$, $n \in \mathbb{N}$, $m \in \Delta\Theta$, but note that the analysis holds uniformly for any such values.

Step 1. *Bound variation in expected utility under original mechanism* $(\Phi^n)_{\mathbb{N}}$ *due to variation in* $\sigma_m(\cdot)$.

We know that if i's $n-1$ opponents draw their types iid according to m, and then play the original mechanism Φ^n using strategy $\sigma^n_m(\cdot)$, then i's Bayesian best response is to play $\sigma^n_m(\theta_i)$. That is, any other play $\sigma^n_m(\hat{\theta}_i)$ generates weakly less utility in expectation, or:

$$E_{\theta_{-i}}[u_{\theta_i}[\Phi^n_i(\sigma^n_m(\hat{\theta}_i), \sigma^n_m(\theta_{-i}))]] - E_{\theta_{-i}}[u_{\theta_i}[\Phi^n_i(\sigma^n_m(\theta_i), \sigma^n_m(\theta_{-i}))]] \leq 0 \tag{11}$$

The first step in our argument is to use equicontinuity of strategies to show that (11) approximately holds if we replace $\sigma^n_m(\cdot)$ everywhere with $\sigma^n_{\tilde{m}}(\cdot)$, and $\tilde{m}$ is "close" to m. Specifically, for any $\epsilon_1 > 0$, there exists a $\delta_1 > 0$ such that, if $|\tilde{m} - m|_{sup} < \delta_1$, then

$$|u_{\theta_i}[\Phi^n_i(\sigma^n_m(\theta_i), \sigma^n_m(\theta_{-i}))] - u_{\theta_i}[\Phi^n_i(\sigma^n_{\tilde{m}}(\theta_i), \sigma^n_{\tilde{m}}(\theta_{-i}))]| < \epsilon_1 \tag{12}$$

and

$$|u_{\theta_i}[\Phi^n_i(\sigma^n_m(\hat{\theta}_i), \sigma^n_m(\theta_{-i}))] - u_{\theta_i}[\Phi^n_i(\sigma^n_{\tilde{m}}(\hat{\theta}_i), \sigma^n_{\tilde{m}}(\theta_{-i}))]| < \epsilon_1 \tag{13}$$

for any θ_{-i}. Moreover, this constant δ_1 is uniform over θ_i, n, m. Combining equations (11)-(13) we have that, for any strategy profile $\tilde{m}$ satisfying $|\tilde{m} - m|_{sup} < \delta_1$, any deviation from θ_i can generate at most $2\epsilon_1$ more utility in expectation:

$$E_{\theta_{-i}}[u_{\theta_i}[\Phi^n_i(\sigma^n_{\tilde{m}}(\hat{\theta}_i), \sigma^n_{\tilde{m}}(\theta_{-i}))]] - E_{\theta_{-i}}[u_{\theta_i}[\Phi^n_i(\sigma^n_{\tilde{m}}(\theta_i), \sigma^n_{\tilde{m}}(\theta_{-i}))]] < 2\epsilon_1. \tag{14}$$

We will argue below that such strategy profiles $\tilde{m}$ are likely under our mechanism.

Step 2. *Bound variation in realized utility under new mechanism* $(F^n)_{\mathbb{N}}$ *due to variation in* θ_{-i}.

Recall that our new mechanism $(F^n)_{\mathbb{N}}$ is constructed from the original mechanism $(\Phi^n)_{\mathbb{N}}$ according to

$$F^n_i(\theta_i, \theta_{-i}) = \Phi^n_i(\sigma_{emp[\theta]}(\theta_i), \sigma_{emp[\theta]}(\theta_{-i})).$$

Our analysis from Step 1 concerned variation in i's expected utility due to variation in what strategy $\sigma_{emp[\theta]}(\cdot)$ is "activated". The present step concerns variation in i's realized utility due to variation in what profile θ_{-i} is realized.

By equicontinuity of reports and strategies, there exists a region of θ_{-i} near $emp[\theta_{-i}] = m$ where the payoff to reporting θ_i varies very little with θ_{-i}. Specifically, for any $\epsilon_2 > 0$, there exists a $\delta_2 > 0$ such that, if $|emp[\theta^1_{-i}] - m|_{sup} < \frac{\delta_2}{2}$ and $|emp[\theta^2_{-i}] - m|_{sup} < \frac{\delta_2}{2}$, and hence $|emp[\theta^1_{-i}] - emp[\theta^2_{-i}]|_{sup} < \delta_2$, then

$$|u_{\theta_i}[\Phi^n_i(\sigma_{emp[\theta_i,\theta^1_{-i}]}(\theta_i), \sigma_{emp[\theta_i,\theta^1_{-i}]}(\theta^1_{-i}))] - u_{\theta_i}[\Phi^n_i(\sigma_{emp[\theta_i,\theta^2_{-i}]}(\theta_i), \sigma_{emp[\theta_i,\theta^2_{-i}]}(\theta^2_{-i}))]| < \epsilon_2 \tag{15}$$

As the constant δ_2 is uniform over θ_i, n, m, the same holds for any potential deviation strategy $\hat{\theta}_i$. Specifically,

$$|u_{\theta_i}[\Phi^n_i(\sigma_{emp[\hat{\theta}_i,\theta^1_{-i}]}(\hat{\theta}_i), \sigma_{emp[\hat{\theta}_i,\theta^1_{-i}]}(\theta^1_{-i}))] - u_{\theta_i}[\Phi^n_i(\sigma_{emp[\hat{\theta}_i,\theta^2_{-i}]}(\hat{\theta}_i), \sigma_{emp[\hat{\theta}_i,\theta^2_{-i}]}(\theta^2_{-i}))]| < \epsilon_2 \tag{16}$$

Hence, within this region, the difference between the *realized* utility from playing θ_i ($\hat{\theta}_i$) and the *expected* utility from playing θ_i ($\hat{\theta}_i$) is at most ϵ_2 as well. Specifically, conditioning on θ_{-i} s.t. $|emp[\theta_{-i}] - m| < \frac{\delta_2}{2}$, we have :

$$u_{\theta_i}[\Phi^n_i(\sigma^n_{emp[\hat{\theta}_i,\theta_{-i}]}(\hat{\theta}_i), \sigma^n_{emp[\hat{\theta}_i,\theta_{-i}]}(\theta_{-i}))]] - E_{\theta_{-i}}[u_{\theta_i}[\Phi^n_i(\sigma^n_{emp[\hat{\theta}_i,\theta_{-i}]}(\hat{\theta}_i), \sigma^n_{emp[\hat{\theta}_i,\theta_{-i}]}(\theta_{-i}))]] < \epsilon_2 \tag{17}$$

and

$$E_{\theta_{-i}}[u_{\theta_i}[\Phi^n_i(\sigma^n_{emp[\theta_i,\theta_{-i}]}(\theta_i), \sigma^n_{emp[\theta_i,\theta_{-i}]}(\theta_{-i}))]] - u_{\theta_i}[\Phi^n_i(\sigma^n_{emp[\theta_i,\theta_{-i}]}(\theta_i), \sigma^n_{emp[\theta_i,\theta_{-i}]}(\theta_{-i}))]] < \epsilon_2 \tag{18}$$

In words, (17) says that the realized payoff from playing $\hat{\theta}_i$ goes up by at most ϵ_2 relative to the expected payoff, conditional on being in the region of θ_{-i} where $|emp[\theta_{-i}] - m|_{sup} < \frac{\delta_2}{2}$. The next equation (18) says that the realized payoff from playing θ_i goes *down* by at most ϵ_2 relative to the expected payoff.

Step 3. *Combine the bounds from Steps 1 and 2 to get an overall bound on the gain from misreporting, for a region of θ_{-i}.*

If we condition as well on $|emp[\theta_i, \theta_{-i}] - m| < \delta_1$ and $|emp[\hat{\theta}_i, \theta_{-i}] - m| < \delta_1$, then we can combine (17), (18) and (14) to obtain the following bound on the difference in realized utilities:

$$u_{\theta_i}[\Phi^n_i(\sigma^n_{emp[\hat{\theta}_i,\theta_{-i}]}(\hat{\theta}_i), \sigma^n_{emp[\hat{\theta}_i,\theta_{-i}]}(\theta_{-i}))]] - u_{\theta_i}[\Phi^n_i(\sigma^n_{emp[\theta_i,\theta_{-i}]}(\theta_i), \sigma^n_{emp[\theta_i,\theta_{-i}]}(\theta_{-i}))]] < 2\epsilon_1 + 2\epsilon_2$$

Hence, by the construction of our new mechanism,

$$u_{\theta_i}[F^n_i(\hat{\theta}_i, \theta_{-i})] - u_{\theta_i}[F^n_i(\theta_i, \theta_{-i})] < 2\epsilon_1 + 2\epsilon_2 \tag{19}$$

for all realizations of θ_{-i} satisfying:

1. $|emp(\theta'_i, \theta_{-i}) - m|_{sup} < \delta_1$ for all $\theta'_i \in \Theta$
2. $|emp[\theta_{-i}] - m|_{sup} < \frac{\delta_2}{2}$

with δ_1 and δ_2 constants that are uniform over n, m.

Step 4. *Bound the probability with which θ_{-i} is in the appropriate region.*

Hence, we can limit i's gain from misreporting to $\epsilon \equiv 2\epsilon_1 + 2\epsilon_2 > 0$, with probability of at least the probability that events (1) and (2) occur given that θ_{-i} is drawn iid according to m. By Kalai's Lemma 5, the probability that (1) is not satisfied is less than $2|\Theta|e^{-2[(n\delta_1-1)/(n-1)]^2(n-1)}$ and by Kalai's Lemma 4 the probability that (2) is not satisfied is less than $2|\Theta|e^{-2(\frac{\delta_2}{2})^2(n-1)}$. Suppose that $2|\Theta|e^{-2(\frac{\delta_2}{2})^2(n-1)} > 2|\Theta|e^{-2[(n\delta_1-1)/(n-1)]^2(n-1)}$, the reverse case being similar in what follows. Then $4|\Theta|e^{-2(\frac{\delta_2}{2})^2(n-1)}$ is a weak upper bound on the probability that either (1) or (2) is not satisfied, so setting $\alpha = 4|\Theta|e^{2(\frac{\delta_2}{2})^2}$ and $\beta = e^{-2(\frac{\delta_2}{2})^2}$ completes the argument that $(\bar{\Phi}_n)_{\mathbb{N}}$ is $(\epsilon, \alpha\beta^n)$-strategyproof, as required by part (1) of the Theorem statement.

Step 5. *Complete the proof by showing that the analysis implies that $(F_n)_{\mathbb{N}}$ is close to $(\Phi_n)_{\mathbb{N}}$ in utility terms as well.*

The last step in the proof is to show that, for all θ_i, and for θ_{-i} drawn iid according to the true prior μ, we have $\Pr[u_{\theta_i}[F_i^n(\theta_i, \theta_{-i})] - u_{\theta_i}[\Phi_i^n(\sigma_\mu^n(\theta_i), \sigma_\mu^n(\theta_{-i}))] > \epsilon] < \alpha\beta^n$. Rewrite the bracketed expression as

$$u_{\theta_i}[\Phi_i^n(\sigma_{emp(\theta)}^n(\theta_i), \sigma_{emp(\theta)}^n(\theta_{-i}))] - u_{\theta_i}[\Phi_i^n(\sigma_\mu^n(\theta_i), \sigma_\mu^n(\theta_{-i}))] \qquad (20)$$

By equicontinuity in strategies, there exists $\delta > 0$ such that (20) is less than ϵ whenever $|emp(\theta) - \mu|_{sup} < \epsilon$. We can use the same δ_1 as in Step 1, hence (20) is less than $\epsilon_1 < \epsilon$ whenever event (i) in Step 3 occurs for $m = \mu$. By Step 4 this occurs with probability of at least $1 - \alpha\beta^n$.

□

References

[1] Abdulkadiroğlu, Atila, Yeon-Koo Che and Yosuke Yasuda, (2011). "Resolving Conflicting Preferences in School Choice: the Boston Mechanism Reconsidered." American Economic Review, forthcoming.

[2] Abdulkadiroğlu, Atila and Tayfun Sönmez (2003). "School Choice: A Mechanism Design Approach." American Economic Review, 93(3), 729-47.

[3] Abdulkadiroğlu, Atila, Parag Pathak, Tayfun Sönmez and Alvin E. Roth (2006). "Changing the Boston School Choice Mechanism: Strategyproofness as Equal Access". Mimeo.

[4] Abdulkadiroğlu, Atila, Parag Pathak, and Alvin E. Roth (2009). "Strategyproofness versus Efficiency in Matching with Indifferences: Redesigning the NYC High School Match." *American Economic Review,* 99(5), 1954-78.

[5] Aumann, Robert J., 1964. “Existence of Competitive Equilibria in Markets with a Continuum of Traders.” *Econometrica*, 32: 39-50.

[6] Ausubel, Larry and Peter Cramton (1996). “Demand Reduction and Inefficiency in Multi-Unit Auctions.” Mimeo.

[7] Azevedo, Eduardo and Jacob Leshno (2011). “The College Admissions Problem with a Continuum of Students.”

[8] Bergemann, Dirk and Stephen Morris (2005). "Robust Mechanism Design." Econometrica, 73(6), 1771-1813.

[9] Bogomolnaia, Anna and Hervé Moulin (2001). “A New Solution to the Random Assignment Problem.” *Journal of Economic Theory*, 100, 295-328.

[10] Budish, Eric (2010). “The Combinatorial Assignment Problem: Approximate Competitive Equilibrium from Equal Incomes.” Mimeo.

[11] Budish, Eric and Estelle Cantillon (2011). “The Multi-Unit Assignment Problem: Theory and Evidence from Course Allocation at Harvard.” Mimeo.

[12] Cramton, Peter, Sean Ellermeyer and Brett Katzman (2011). “Designed to Fail: the Medicare Auction for Durable Medical Equipment.” Mimeo.

[13] Cripps, Martin and Jeroen Swinkels (2006). “Depth and Efficiency of Large Double Auctions.” Econometrica, 74, 47-92.

[14] Debreu, George and Herbert Scarf (1963). “A Limit Theorem on the Core of an Economy.” *Econometrica*

[15] Edelman, Benjamin, and Michael Ostrovsky (2007). “Strategic Bidder Behavior in Sponsored Search Auctions.” Decision Support Systems, 43(1), 192-8.

[16] Edelman, Benjamin, Michael Ostrovsky and Michael Schwarz (2007). "Internet Advertising and the Generalized Second-Price Auction: Selling Billions of Dollars Worth of Keywords." American Economic Review, 97(1), 242-59.

[17] Featherstone, Clayton, and Muriel Niederle. 2008. “Ex Ante Eciency in School Choice Mechanisms: An Experimental Investigation.” NBER WP#14618.

[18] Foley, Duncan (1967). “Resource Allocation and the Public Sector.” *Yale Economic Essays*, 7, 45-98.

[19] Fudenberg, Drew and Jean Tirole (1991). Game Theory. MIT Press.

[20] Friedman, Milton (1964). “Comment on Collusion in the Auction Market for Treasury Bills.” Journal of Political Economy, 72(5), 513-4.

[21] Friedman, Milton (1991). "How to Sell Government Securities." *Wall Street Journal*, August 28.

[22] Gale, David and Lloyd Shapley (1962). "College Admissions and the Stability of Marriage." *American Mathematical Monthly*, 69(1), 9-15.

[23] Gibbard, A., 1973. Manipulation of voting schemes: a general result. Econometrica 41, 587–601.

[24] Green, J. and J.J. Laffont (1977). "Characterization of Satisfactory Mechanisms for the Revelation of Preferences for Public Goods." *Econometrica*, Vol 45(2), 427-38.

[25] Hurwicz, Leonid (1972). "On Informationally Decentralized Systems." in Decision and Organization: A Volume in Honor of Jacob Marschak, ed. by C. B. McGuire and Roy Radner. Amsterdam: North-Holland.

[26] Hylland, Aanund and Richard Zeckhauser (1979). "The Efficient Allocation of Individuals to Positions." *Journal of Political Economy*, 87(2), 293-314

[27] Jegadeesh, Narasimhan (1993). "Treasury Auction Bids and the Salomon Squeeze." Journal of Finance, 48(4), 1403-1419.

[28] Kalai, Ehud (2004). "Large Robust Games." Econometrica, 72(6), 1631-1665.

[29] Kim, Caroline, Brian Merlob, Kathryn Peters, Charles R. Plott, Andre Pradhana, and Yuanjun Zhang (2010), "An Evaluation of the Proposed Procurement Auction for the Purchase of Medicare Equipment: Experimental Tests of the Auction Architecture," Working Paper, California Institute of Technology, October.

[30] Kojima, Fuhito and Parag Pathak (2009). "Incentives and Stability in Large Two-Sided Matching Markets." *American Economic Review*, 99, 608-27.

[31] Kojima, Fuhito and Mihai Manea (2009). "Incentives in the Probabilistic Serial Mechanism." Journal of Economic Theory.

[32] Krishna, Aradhna and Utku Unver (2008). "Improving the Efficiency of Course Bidding at Business Schools: Field and Laboratory Studies." Marketing Science, 27, 262-82.

[33] Miralles, Antonio. 2008. "School Choice: The Case for the Boston Mechanism." Mimeo.

[34] Pápai, Szilvia (2001). "Strategyproof and Nonbossy Multiple Assignments." *Journal of Public Economic Theory*, 3(3): 257-71.

[35] Pathak, Parag and Tayfun Sönmez (2008). "Leveling the Playing Field: Sincere and Sophisticated Players in the Boston Mechanism." American Economic Review, 98(4), 1636-52.

[36] Pathak, Parag and Tayfun Sönmez (2011). "School Admissions Reform in Chicago and England: Comparing Mechanisms by their Vulnerability to Manipulation." Mimeo.

[37] Perry, Motty and Philip J. Reny (2006). "Toward a Strategic Foundation for Rational Expectations Equilibrium." *Econometrica*, 74, 1231-69.

[38] Pesendorfer, Wolfgang and Jeroen Swinkels (2000). "Efficiency and Information Aggregation in Auctions," American Economic Review, 90, 499-525.

[39] Roberts, John and Andrew Postlewaite (1976). "Incentives for Price-Taking Behavior in Large Exchange Economies." *Econometrica*, 44(1), 115-27.

[40] Roth, Alvin E. (1982). The economics of matching: stability and incentives. Math Oper Res 7:617–628

[41] Roth, Alvin E. (2002). "The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics." *Econometrica*, 70(4): 1341-1378

[42] Roth, Alvin E. (2008). "What Have We Learned from Market Design?" Hahn Lecture, Economic Journal, 118 (March), 285-310.

[43] Rustichini, Aldo, Mark A. Satterthwaite and Steven R. Williams (1994). "Convergence to Efficiency in a Simple Market with Incomplete Information." *Econometrica*, 62(5), 1041-63

[44] Satterthwaite, M.A., 1975. Strategy-proofness and Arrow's conditions: existence and correspondence theorems for voting procedures and social welfare functions. Journal of Economic Theory 10, 187–217.

[45] Sönmez, Tayfun and Utku Ünver (2010). "Course Bidding at Business Schools." *International Economic Review*, 51(1), 99-123.

[46] Wilson, Robert (1987). "Game-Theoretic Analyses of Trading Processes," in Advances in Economic Theory: Fifth World Congress, ed. by Timothy Bewley. Cambridge, UK: Cambridge University Press, Chap. 2, 33-70.