Strategy-proofness in the Large*

Eduardo M. Azevedo[†] and Eric Budish[‡]

August 10, 2017

Abstract

We propose a criterion of approximate incentive compatibility, strategy-proofness in the large (SP-L), and argue that it is a useful second-best to exact strategy-proofness (SP) for market design. Conceptually, SP-L requires that an agent who regards a mechanism's "prices" as exogenous to her report – be they traditional prices as in an auction mechanism, or price-like statistics in an assignment or matching mechanism – has a dominant strategy to report truthfully. Mathematically, SP-L weakens SP in two ways: (i) truth-telling is required to be approximately optimal (within epsilon in a large enough market) rather than exactly optimal, and (ii) incentive compatibility is evaluated ex interim, with respect to all full-support i.i.d. probability distributions of play, rather than ex post with respect to all possible realizations of play. This places SP-L in between the traditional notion of approximate strategy-proofness, which evaluates incentives to manipulate ex post, and the traditional notion of approximate Bayes-Nash incentive compatibility, which evaluates incentives to manipulate ex interim with respect to the single common-knowledge probability distribution associated with Bayes-Nash equilibrium.

*First version: October 2011. We thank the editor and four anonymous referees for comments and suggestions that improved the paper considerably. We are grateful to Nabil Al-Najjar, Susan Athey, Aaron Bodoh-Creed, Gabriel Carroll, Sylvain Chassang, Jeff Ely, Alex Frankel, Drew Fudenberg, Matt Gentzkow, Jason Hartline, John Hatfield, Richard Holden, Ehud Kalai, Emir Kamenica, Navin Kartik, Fuhito Kojima, Scott Kominers, Jacob Leshno, Jon Levin, Paul Milgrom, Stephen Morris, Roger Myerson, David Parkes, Parag Pathak, Nicola Persico, Andy Postlewaite, Canice Prendergast, Mark Satterthwaite, Ilya Segal, Eran Shmaya, Lars Stole, Rakesh Vohra, Glen Weyl, Mike Whinston, and especially Al Roth for helpful discussions that have shaped this research. We thank Victor Zhang for excellent research assistance. We are grateful to seminar audiences at Ohio State, the 2011 MFI Conference on Matching and Price Theory, UCLA, Chicago, AMMA 2011, Boston College, the 2011 NBER Conference on Market Design, Duke / UNC, Michigan, Carnegie Mellon / Pittsburgh, Montreal, Berkeley, Northwestern, Rochester, Frontiers of Market Design 2012, ACM EC 2012, Princeton, Maryland, and UPenn.

[†]Wharton, eazevedo@wharton.upenn.edu.

[‡]University of Chicago Booth School of Business, eric.budish@chicagobooth.edu.

1 Introduction

Strategy-proofness (SP), that playing the game truthfully is a dominant strategy, is perhaps the central notion of incentive compatibility in market design. SP is frequently imposed as a design requirement in theoretical analyses, across a broad range of assignment, auction, and matching problems. And, SP has played a central role in several design reforms in practice, including the redesign of school choice mechanisms in several cities, the redesign of the market that matches medical school graduates to residency positions, and efforts to create mechanisms for pairwise kidney exchange (See especially Roth (2008) and Pathak and Sönmez (2008, 2013)). There are several reasons why SP is considered so attractive. First, SP mechanisms are robust: since reporting truthfully is a dominant strategy, equilibrium does not depend on market participants' beliefs about other participants' preferences or information. Second, SP mechanisms are strategically simple: market participants do not have to invest time and effort collecting information about others' preferences or about what equilibrium will be played. Third, with this simplicity comes a measure of fairness: a participant who lacks the information or sophistication to game the mechanism is not disadvantaged relative to sophisticated participants. Fourth, SP mechanisms generate information about true preferences that may be useful to policy makers.[1]

However, SP is restrictive. In a variety of market design contexts, including matching, school choice, course allocation, and combinatorial auctions, impossibility theorems show that SP places meaningful limitations on what other attractive properties a mechanism can hope to satisfy.[2] And, SP is an extremely strong requirement. If there is a single configuration of participants' preferences in which a single participant has a strategic misreport that raises his utility by epsilon, a mechanism is not SP. A natural idea is to look for incentives criteria that are less demanding and less restrictive than SP, while still maintaining some of the

[1] See Wilson (1987) and Bergemann and Morris (2005) on robustness, Fudenberg and Tirole (1991), p. 270 and Roth (2008) on strategic simplicity, Friedman (1991), Pathak and Sönmez (2008) and Abdulkadiroğlu et al. (2006) on fairness, and Roth (2008) and Abdulkadiroğlu et al. (2016)on the advantage of generating preference data.

[2] In matching problems such as the National Resident Matching Program, SP mechanisms are not stable (Roth, 1982). In multi-unit assignment problems such as course allocation, the only SP and ex-post efficient mechanisms are dictatorships (Papai, 2001; Ehlers and Klaus, 2003; Hatfield, 2009), which perform poorly on measures of fairness and ex-ante welfare (Budish and Cantillon, 2012). In school choice problems, which can be interpreted as a hybrid of an assignment and a matching problem (Abdulkadiroğlu and Sönmez, 2003), there is no mechanism that is both SP and ex-post efficient (Abdulkadiroğlu et al., 2009). In combinatorial auction problems such as the FCC spectrum auctions (Milgrom, 2004; Cramton et al., 2006), the only SP and efficient mechanism is Vickrey-Clarke-Groves (Green and Laffont, 1977; Holmström, 1979), which has a variety of important drawbacks (Ausubel and Milgrom, 2006). Perhaps the earliest such negative result for SP mechanisms is Hurwicz (1972), which shows that SP is incompatible with implementing a Walrasian equilibrium in an exchange economy.

advantages of SP design.

This paper proposes a criterion of approximate strategy-proofness called *strategy-proofness in the large* (SP-L). SP-L weakens SP in two ways. First, whereas SP requires that truthful reporting is optimal in any size economy, SP-L requires that truthful reporting is approximately optimal in a large enough market (within epsilon for large enough n). Second, whereas SP requires that truthful reporting is optimal against any realization of opponent reports, SP-L requires that truthful reporting is optimal only against any full-support, independent and identically distributed (i.i.d.) probability distribution of reports. That is, SP-L examines incentives from the interim perspective rather than ex-post. Because of this interim perspective, SP-L is weaker than the traditional ex-post notion of approximate strategy-proofness; this weakening is important both conceptually and for our results. At the same time, SP-L is importantly stronger than approximate Bayes-Nash incentive compatibility, because SP-L requires that truthful reporting is best against *any* full-support, i.i.d. probability distribution of opponent reports, not just the single probability distribution associated with Bayes-Nash equilibrium play. This strengthening is important because it allows SP-L to approximate, in large markets, the attractive properties such as robustness and strategic simplicity which are the reason why market designers like SP better than Bayes-Nash in the first place.

This combination of approximate incentives in a large market and the interim perspective is powerful for the following reason: it causes each participant to regard the societal distribution of play as exogenous to his own report (more precisely, the distribution of the societal distribution of play). As will become clear, regarding the societal distribution of play as exogenous to one's own play is a generalization of the idea of regarding prices as exogenous, i.e., of price taking. In some settings, such as multi-unit auctions or Walrasian exchange, the two concepts are equivalent. In other settings, such as school choice or two-sided matching, regarding the societal distribution of play as exogenous is equivalent to regarding certain price-like summary statistics of the mechanism as exogenous.

SP-L thus draws a distinction between two ways a mechanism can fail to be SP. If a mechanism is manipulable by participants who can affect prices (or price-like summary statistics), but is not manipulable by participants who regard the societal distribution of play as exogenous, the mechanism is SP-L. If a mechanism is manipulable even by participants who regard the societal distribution of play as exogenous – if even a price taker, or a taker of price-like statistics, wishes to misreport – then the mechanism, in addition to not being SP, is not SP-L. Intuition suggests that these latter violations of SP are especially problematic for practice, because, to manipulate a mechanism, a participant only needs information about aggregate statistics, such as how popular is each school in a school matching mechanism. This is problematic because there are many real-world environments where participants have this kind of information. SP-L rules out mechanisms that violate SP in this particularly serious way.

After we present and discuss the formal definition of SP-L, the next part of the paper provides a classification of existing non-SP mechanisms into those that are SP-L and those that are not SP-L. The classification, displayed in Table 1, organizes both the prior theory literature on which non-SP mechanisms have good incentives properties in large markets and the empirical record on when non-SP matters in real-world large markets. In the SP-L column are numerous mechanisms that, while not SP, have been shown theoretically to have approximate incentives for truth telling in large markets. Examples include the Walrasian mechanism (Roberts and Postlewaite, 1976; Jackson and Manelli, 1997), double auctions (Rustichini et al., 1994; Cripps and Swinkels, 2006), multi-unit uniform-price auctions (Swinkels, 2001), the Gale-Shapley deferred acceptance algorithm (Immorlica and Mahdian, 2005; Kojima and Pathak, 2009), and probabilistic serial (Kojima and Manea, 2010). This literature has used a wide variety of definitions of approximate incentive compatibility, as well as a wide variety of analysis techniques. We use a single definition and a single analysis technique (the sufficient conditions for SP-L in Theorem 1) and find that all of these mechanisms are SP-L.[3] Our technique also classifies as SP-L several mechanisms whose large-market incentive properties had not previously been formally studied.

On the other hand, in the non-SP-L column are numerous mechanisms for which there is explicit empirical evidence that real-world market participants strategically misreport their preferences, to the detriment of design objectives such as efficiency or fairness. Examples include multi-unit pay-as-bid auctions (Friedman, 1960, 1991), the Boston mechanism for school choice (Abdulkadiroğlu et al., 2006, 2009), the bidding points auction for course allocation (Sönmez and Ünver, 2010; Budish, 2011), the draft mechanism for course allocation (Budish and Cantillon, 2012), and the priority-match mechanism for two-sided matching (Roth, 2002). This literature has frequently emphasized that the mechanism in question is not SP; our point is that the mechanisms for which there is documentation of important incentives problems in practice not only are not SP, but are not even SP-L. Overall, the

[3]Note as well that the traditional ex-post notion of approximate strategy-proofness is too strong to obtain the classification. For instance, the uniform-price auction is SP-L but is not approximately strategy-proof in an ex-post sense; even in a large economy it is always possible to construct a knife-edge situation where a single player, by shading her demand, can have a large discontinuous influence on the market-clearing price.

Table 1: SP-L and non SP-L mechanisms for some canonical market design problems

Problem	Manipulable in the Large	SP-L
Multi-Unit Auctions	Pay as Bid	Uniform Price
Single-Unit Assignment	Boston Mechanism Adaptive Boston Mechanism	Probabilistic Serial HZ Pseudomarket
Multi-Unit Assignment	Bidding Points Auction HBS Draft	Approximate CEEI Generalized HZ
Matching	Priority Match	Deferred Acceptance
Other		Walrasian Mechanism Double Auctions

Notes: See Supplementary Appendix C for a detailed description of each mechanism in the table as well as a proof of the mechanism's classification as either SP-L or manipulable in the large. Abbreviations: HBS = Harvard Business School; HZ = Hylland and Zeckhauser; CEEI = competitive equilibrium from equal incomes.

classification exercise suggests that the relevant distinction for practice, in markets with a large number of participants, is not "SP vs. not SP", but rather "SP-L vs. not SP-L".

The last part of the paper identifies a precise sense in which, in large markets, SP-L is approximately costless to impose relative to Bayes-Nash incentive compatibility. Formally, we consider social choice functions that take as input both agents' preferences and their beliefs, and that can be implemented by the (limit) Bayes-Nash equilibria of a mechanism, considering agents' beliefs that can be any full-support i.i.d. common prior over payoff types. For example, consider the Boston mechanism for school choice. Abdulkadiroğlu et al. (2011) show, given an arbitrarily set full-support i.i.d. common prior, that the Boston mechanism has a Bayes-Nash equilibrium that yields an ex-ante efficient allocation. This map from preferences and beliefs to equilibrium outcomes is a social choice function that fits our assumptions. Theorem 2 shows, given such a Bayes-Nash implementable social choice function, that there exists an SP-L mechanism that achieves approximately the same outcomes. Thus, while SP is often costly to impose relative to Bayes-Nash incentive compatibility, there is a precise sense in which SP-L is no more restrictive than Bayes-Nash incentive compatibility.

Overall, our analysis suggests that in large market settings SP-L approximates the advantages of SP design while being significantly less restrictive. Our hope is that market designers will view SP-L as a practical alternative to SP in settings with a meaningful number of participants and in which SP mechanisms perform poorly. An illustration of this approach is the SP-L course allocation mechanism recently implemented at the Wharton School, replacing a mechanism that was manipulable in the large, in an environment with numerous impossibility theorems for SP.[4]

We emphasize that the idea that market size can ease incentive problems is quite old, with some of the earliest contributions being Roberts and Postlewaite (1976) and Hammond (1979), and rich and active literatures in both implementation theory and market design since these early contributions. We discuss the relevant literatures in detail in the body of the paper. Our paper makes three substantive contributions relative to this literature. First, the criterion of SP-L itself is new. There are many other criteria of approximate incentive compatibility in the literature (see footnote 15). SP-L is perhaps closest to the two criteria it lands in between: the traditional notion of approximate SP (which evaluates incentives to manipulate from the ex-post perspective) and the traditional notion of approximate Bayes-Nash (which evaluates incentives to manipulate from the interim perspective, like SP-L, but with respect only to the single probability distribution associated with equilibrium play). Second, our classification of mechanisms, into those that are SP-L and those that are manipulable even in large markets, is new. We expect few market design researchers will find the classification surprising ex-post; for example, Kojima and Pathak (2009) noted that the Boston mechanism fails their approximate IC criterion (roughly, that truth-telling is approximate Bayes-Nash), and Friedman (1991) informally discusses the incentives difference between pay-as-bid and uniform-price auctions. Rather, our contribution is to produce a classification of mechanisms across all of the canonical market design problems using a single criterion and analytical approach, which is both useful per se and speaks to the applicability of the criterion and analytical methods. Third, the result that SP-L is approximately costless relative to Bayes-Nash is new, and is the first such result, to our knowledge, for any form of approximate strategy-proofness relative to Bayes-Nash; though there are some more restricted settings where SP itself is costless relative to Bayes-Nash (Manelli and Vincent, 2010; Gershkov et al., 2011).

The rest of the paper is organized as follows. Section 2 defines the environment. Section

[4]For further details see Budish et al. (2015) and Budish and Kessler (2016). Notably, while the Wharton administration was concerned about how easy the old mechanism was to manipulate, they were not concerned about the fact that the new mechanism is SP-L but not SP. An excerpt from the student training manual highlights this point: "Doesn't it pay to think strategically? NO! You cannot influence the clearing price (you are only one of 1600 students). So your best 'strategy' is to assume the clearing prices are given. And to tell Course Match [the mechanism] your true preferences so that it can buy you your best schedule, given your preferences, your budget and the given clearing prices." (Wharton, 2013)

3 defines and discusses SP-L. Section 4 presents the classification of non-SP mechanisms. Section 5 presents the theorem that imposing SP-L is approximately costless relative to Bayes-Nash. Section 6 applies the theorem to the Boston mechanism. Section 7 discusses technical extensions. Section 8 discusses related literature. Section 9 concludes. Proofs and other supporting materials are in the appendix.

2 Environment

We work with an abstract mechanism design environment in which mechanisms assign outcomes to agents based on the set of agents' reports. There is a finite set of **payoff types** T and a finite set of **outcomes (or consumption bundles)** X_0**.** The outcome space describes the outcome possibilities for an individual agent. For example, in an auction the elements in X_0 specify both the objects an agent receives and the payment she makes. In school assignment, X_0 is the set of schools to which a student can be assigned. An agent's payoff type determines her preferences over outcomes. For each $t_i \in T$ there is a von Neumann-Morgenstern expected **utility function** $u_{t_i} : X \to [0,1]$, where $X = \Delta X_0$ denotes the set of lotteries over outcomes. Preferences are private values in the sense that an agent's utility depends exclusively on her payoff type and the outcome she receives.

We study mechanisms that are well defined for all possible market sizes, holding fixed X_0 and T. For each market size $n \in \mathbb{N}$, where n denotes the number of agents, an allocation is a vector of n outcomes, one for each agent, and there is a set $Y_n \subseteq (X_0)^n$ of **feasible allocations**. For instance, in an auction, the assumption that X_0 is fixed imposes that the number of potential types of objects is finite, and the sequence $(Y_n)_{\mathbb{N}}$ describes how the supply of each type of object changes as the market grows.

Definition 1. *Fix a set of outcomes X_0, a set of payoff types T, and a sequence of feasibility constraints $(Y_n)_{\mathbb{N}}$. A* ***mechanism*** *$\{(\Phi^n)_{\mathbb{N}}, A\}$ consists of a finite set of actions A and a sequence of allocation functions*

$$\Phi^n : A^n \to \Delta((X_0)^n), \tag{2.1}$$

each of which satisfies feasibility: for any $n \in \mathbb{N}$ and $a \in A^n$, the support of $\Phi^n(a)$ is contained in the feasible set Y_n. A mechanism is ***direct*** *if $A = T$.*

We assume that mechanisms are **anonymous**, which requires that each agent's outcome depends only on her own action and the distribution of all actions. Formally, a mechanism is anonymous if the allocation function $\Phi^n(\cdot)$ is invariant to permutations for all $n \in \mathbb{N}$. Anonymity is a natural feature of many large-market settings. In Supplementary Appendix B we relax anonymity to **semi-anonymity** (Kalai, 2004). A mechanism is semi-anonymous if agents are divided into a finite set of groups, and an agent's outcome depends only on her own action, her group, and the distribution of actions within each group. Semi-anonymity accommodates applications in which there are asymmetries among classes of participants, such as double auctions in which there are distinct buyers and sellers and school choice problems in which students are grouped into different priority classes, as well as some models of matching markets.

We adopt the following notation. Given a finite set S, the set of probability distributions over S is denoted ΔS, and the set of distributions with full support $\bar{\Delta} S$. Distributions over the set of payoff types will typically be denoted as $\mu \in \Delta T$, and distributions over actions by $m \in \Delta A$. Throughout the analysis we will use the supremum norm on the sets ΔT, ΔA and X. Since the number of payoff types, actions and outcomes is finite, all of these probability spaces are subsets of Euclidean space. Using this representation, we denote the distance between two outcomes $x, x' \in X$ as $\|x - x'\|$, and likewise for distributions over T and A. In particular, we use this topology in the definition of limit mechanisms below.

Given a vector of payoff types $t \in T^n$, we use the notation $\text{emp}[t]$ to denote the empirical distribution of t on T. That is, for each payoff type $\tau \in T$, $\text{emp}[t](\tau)$ is the fraction of coordinates of t that equal τ, and the vector $\text{emp}[t] = (\text{emp}[t](\tau))_{\tau \in T}$. Analogously, given a vector of actions $a \in A^n$, $\text{emp}[a]$ denotes the empirical distribution of a on A.

3 Strategy-proof in the Large

In this section we formally define strategy-proofness in the large (SP-L) and discuss its interpretation and its relationship to previous concepts.

3.1 The Large Market

We begin by defining our notion of the large market. Given a mechanism $\{(\Phi^n)_{\mathbb{N}}, A\}$, define, for each n, the function $\phi^n : A \times \Delta A \to X$ according to

$$\phi^n(a_i, m) = \sum_{a_{-i} \in A^{n-1}} \Phi_i^n(a_i, a_{-i}) \cdot \Pr(a_{-i} | a_{-i} \sim iid(m)), \tag{3.1}$$

where $\Phi_i^n(a_i, a_{-i})$ denotes the marginal distribution of the i^{th} coordinate of $\Phi^n(a)$, i.e., the lottery over outcomes received by agent i when she plays a_i and the other $n-1$ agents play a_{-i}, and $\Pr(a_{-i}|a_{-i} \sim iid(m))$ denotes the probability that the action vector a_{-i} is realized given $n-1$ independent and identically distributed (i.i.d.) draws from the action distribution $m \in \Delta A$. In words, $\phi^n(a_i, m)$ describes what an agent who plays a_i expects to receive, ex interim, if the other $n-1$ agents play i.i.d. according to action distribution m.

We use the interim allocation function ϕ^n to define the large-market limit.

Definition 2. *The* ***large-market limit of mechanism*** $\{(\Phi^n)_{\mathbb{N}}, A\}$ *is the function* $\phi^\infty : A \times \Delta A \to X$ *given by*

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

if this limit exists.

In words, $\phi^\infty(a_i, m)$ equals the lottery that an agent who plays a_i receives, in the limit as the number of agents grows large, when the other agents play i.i.d. according to the probability distribution m.[5]

It is easy to construct examples of mechanisms that do not have limits. For instance, if a mechanism is a uniform-price auction when n is even and is a pay-as-bid auction when n is odd, then the mechanism does not have a limit. The main results in this paper, Theorems 1 and 2, are valid regardless of the existence of the limit. Nevertheless, the limit is useful to understand the definition of SP-L, and is useful in the analysis of specific mechanisms.

The Role of Randomness in the Large Market The randomness in how we let the market grow large is important for the following reason: from the interim perspective, as the market grows large in our way, the *distribution of the empirical distribution of play* becomes exogenous to any particular agent's own play. We state this claim formally in the Appendix as Lemma A.1. Intuitively, if a fair coin is tossed n times the distribution of the number of heads is stochastic, and the influence of the i^{th} coin toss on this distribution vanishes to zero as $n \to \infty$; whereas if the market grew large in a deterministic fashion one player's decision between heads or tails could be pivotal as to whether the number of heads is greater than or less than 50%.

We interpret treating the societal distribution of play as exogenous to one's own report as a generalized version of price taking. Suppose that a mechanism has prices that are a

[5]The randomness in how we take the large-market limit is in contrast with early approaches to large-market analysis, such as Debreu and Scarf's (1963) replicator economy and Aumann's (1966) continuum economy. It is more closely related to the random economy method used in Immorlica and Mahdian's (2005) and Kojima and Pathak's (2009) studies of large matching markets.

function of the empirical distribution of play. For example, in a uniform-price auction, price is determined based on where reported demand equals reported supply. In our large market, because the distribution of the empirical distribution of play is exogenous to each agent, the distribution of prices is exogenous to each agent. Now suppose that a mechanism does not have prices, but has price-like statistics that are functions of the empirical distribution of play and sufficient statistics for the outcomes received by agents who played each action. For example, in Bogomolnaia and Moulin's (2001) assignment mechanism, the empirical distribution of reports determines statistics called "run-out times", which describe at what time in their algorithm each object exhausts its capacity. In our large market, each agent regards the distribution of these price-like statistics as exogenous to their own report.

3.2 Definition of SP-L

A mechanism is strategy-proof (SP) if it is optimal for each agent to report truthfully, in any size market, given any realization of opponent reports.

Definition 3. *The direct mechanism $\{(\Phi^n)_{\mathbb{N}}, T\}$ is* ***strategy-proof (SP)*** *if, for all n, all $t_i, t_i' \in T$, and all $t_{-i} \in T^{n-1}$*

$$u_{t_i}[\Phi_i^n(t_i, t_{-i})] \geq u_{t_i}[\Phi_i^n(t_i', t_{-i})].$$

We say that a mechanism is strategy-proof in the large (SP-L) if, for any full-support i.i.d. distribution of opponent reports, reporting truthfully is approximately optimal in large markets. For mechanisms that have a limit, this is equivalent to, for any full-support i.i.d. distribution of opponent reports, reporting truthfully being optimal in the limit.

Definition 4. *The direct mechanism $\{(\Phi^n)_{\mathbb{N}}, T\}$ is* ***strategy-proof in the large (SP-L)*** *if, for any $m \in \bar{\Delta}T$ and $\epsilon > 0$ there exists n_0 such that, for all $n \geq n_0$ and all $t_i, t_i' \in T$,*

$$u_{t_i}[\phi^n(t_i, m)] \geq u_{t_i}[\phi^n(t_i', m)] - \epsilon.$$

If the mechanism has a limit, this is equivalent to, for any $m \in \bar{\Delta}T$ and all $t_i, t_i' \in T$,

$$u_{t_i}[\phi^\infty(t_i, m)] \geq u_{t_i}[\phi^\infty(t_i', m)]. \tag{3.2}$$

Otherwise, the mechanism is ***manipulable in the large****.*

SP-L weakens SP in two ways. First, while SP requires that truthful reporting is optimal in any size market, SP-L requires that truthful reporting is only approximately optimal in a large enough market. Second, SP evaluates what report is best based on the (ex-post) realization of reports, whereas SP-L evaluates based on the (ex-interim) probability distribution of reports. A mechanism can be SP-L even if it has the property that, given $\epsilon > 0$, in any size market n one can find a payoff type t_i and realization of opponent reports t_{-i} for which t_i has a misreport worth more than ϵ. What SP-L rules out is that there is a full-support i.i.d. probability distribution of opponent reports with this property. Implicitly, SP-L takes a view on what information participants have in a large market when they decide how to play – they may have a (possibly incorrect) sense of the distribution of opponent preferences, but they do not know the exact realization of opponent preferences.

These two weakenings place SP-L between two commonly used notions of incentive compatibility. SP-L is weaker than the standard ex-post notion of asymptotic strategy-proofness, which requires that reporting truthfully is approximately optimal, in a large enough market, for any realization of opponent reports.[6] This distinction is important for the classification below; nearly all of the mechanisms that are classified as SP-L would fail this stronger criterion (e.g., uniform-price auctions, deferred acceptance), with the lone exception being the probabilistic serial mechanism. At the same time, SP-L is stronger than approximate Bayes-Nash incentive compatibility, which requires that truthful reporting is approximately optimal against the true probability distribution of opponent reports, which itself is assumed to be common knowledge. In contrast, SP-L requires truthful reporting to be approximately optimal for any full-support i.i.d. probability distribution of opponent reports. This distinction is what allows SP-L mechanisms to maintain, at least approximately, some of the attractive features of SP design such as robustness, strategic simplicity, and fairness to unsophisticated agents (cf. Section 7.3).

Finally, the definition of the limit gives a useful way to think about SP-L as a generalization of price-taking. As discussed above, in our large market limit agents regard the distribution of the aggregate distribution of play as exogenous to their report. A mechanism fails to be SP-L if even an agent who regards the aggregate distribution of play as exogenous may wish to misreport. In the case of mechanisms with prices that are a function of the

[6]For example, Liu and Pycia (2011) define a mechanism as asymptotically strategy-proof if, given $\epsilon > 0$, there exists n_0 such that for all $n \geq n_0$, types t_i, t_i', and a vector of $n-1$ types t_{-i},

$$u_{t_i}[\Phi_i^n(t_i, t_{-i})] \geq u_{t_i}[\Phi_i^n(t_i', t_{-i})] - \epsilon.$$

A similar definition is in Hatfield, Kojima and Kominers (2015).

distribution of play, a mechanism fails to be SP-L if even an agent who regards prices as exogenous wishes to misreport.

3.3 Clarifying Example: Multi-Unit Auctions

We illustrate several of the key definitions of this section using the example of multi-unit auctions for identical objects, such as government bond auctions. The two most common multi-unit auction formats are uniform-price auctions and pay-as-bid auctions. While neither mechanism is SP (Ausubel and Cramton, 2002), Milton Friedman famously argued in favor of the uniform-price auction on incentives grounds (Friedman, 1960, 1991). We will see that the uniform-price auction is SP-L whereas the pay-as-bid auction is manipulable in the large. The example also illustrates the definition of the large-market limit, the role of the full-support requirement, and the contrast between SP-L and the traditional notion of approximate strategy-proofness based on ex-post realizations of others' play rather than interim distributions of others' play. The example is not needed for the remainder of the analysis, so some readers may prefer to skip it.

Example 1. (**Multi-Unit Auctions**). There are kn units of a homogeneous good. To simplify notation, we assume that agents assign a constant per-unit value to the good, up to a capacity limit. Specifically, each agent i's type $t_i = (v_i, q_i)$ consists of a per-unit value v_i and a maximum capacity q_i. The set of possible values is $V = \{1, \dots, \bar{v}\}$, the set of possible capacity limits is $Q = \{0, 1, \dots, \bar{q}\}$ with $1 < k < \bar{q}$, and $T = V \times Q$. The set of outcomes is $X_0 = (\{1, 2, \cdots, \bar{v}\} \times \{1, 2, \cdots, \bar{q}\}) \cup \{0\}$, with an outcome consisting either of a per-unit payment and an allotted quantity, or 0 to denote that the agent receives no units and makes no payment.

We first describe the uniform-price auction. Bids consist of a per-unit value and a maximum capacity, so the action set $A = T$. Given a vector of n bidders' reports t, denote the demand for the object at a price of p as $D(p; t) = \sum_{i=1}^{n} q_i \cdot 1\{v_i \geq p\}$, where $1\{\cdot\}$ is the indicator function. The market-clearing price $p^*(t)$ is the highest price at which demand exceeds supply. That is,

$$p^*(t) = \max \left\{ p \in V : \frac{D(p; t)}{n} \geq k \right\} \tag{3.3}$$

if $D(1, t) \geq k$ and $p^*(t) = 0$ otherwise. The uniform-price auction allocates each bidder i her demanded quantity at $p^*(t)$, with the exception that bids with $v_i = p^*(t)$ are rationed with equal probability. Formally, $\Phi_i^n(t)$ allots each bidder the following number of units of the good,

Reported Value	Expected Number of Units
$v_i < p^*(t)$	0
$v_i = p^*(t)$	$\bar{r} \cdot q_i$
$v_i > p^*(t)$	q_i

at a price per unit of $p^*(t)$, and the rationing probability $\bar{r}$ set so that the market clears.[7]

We now analyze the large-market limit of the uniform-price auction. Let $\rho^*(m)$ denote the price that clears supply and *average demand* given bid distribution m. That is,

$$\rho^*(m) = \max\{p \in V : \mathbb{E}[D(p;t_i)|t_i \sim m] \geq k\} \tag{3.4}$$

if $\mathbb{E}[D(1;t_i)|t_i \sim m] \geq k$ and 0 otherwise.

Generically, expected demand at price $\rho^*(m)$ strictly exceeds supply, that is,

$$\mathbb{E}[D(\rho^*(m);t_i)|t_i \sim m] > k.$$

In this generic case, as the market grows large, the realized price as defined in (3.3) will be equal to $\rho^*(m)$ with probability converging to one. Therefore, the limit mechanism allocates each bidder their demand at $\rho^*(m)$, again with the exception that bidders with value exactly equal to $\rho^*(m)$ are rationed, and with all winning bidders paying $\rho^*(m)$ per unit. Formally, $\phi^\infty(t_i, m)$ gives player i

Reported Value	Expected Number of Units
$v_i < \rho^*(m)$	0
$v_i = \rho^*(m)$	$\bar{r} \cdot q_i$
$v_i > \rho^*(m)$	q_i

at a per unit price of $\rho^*(m)$, and the rationing probability $\bar{r}$ is set so that the market clears on average.[8] Note that, in this generic case, the price in the limit is deterministic and is

[7]Since preferences are linear up to the capacity limit, the exact form of the rationing is immaterial in the analysis below. The rationing constant is

$$\bar{r} = \frac{kn - D(p^*(t)+1;t)}{D(p^*(t);t) - D(p^*(t)+1;t)}.$$

[8]That is, $\bar{r}$ satisfies

$$\bar{r} = \frac{k - E[D(\rho^*(m)+1;t_i')|t_i' \sim m]}{E[D(\rho^*(m);t_i')|t_i' \sim m] - E[D(\rho^*(m)+1;t_i')|t_i' \sim m]}.$$

exogenous from the perspective of each individual bidder.

In addition to the generic case, there is a knife-edge case in which expected demand at $\rho^*(m)$ is exactly equal to supply. That is, $\mathbb{E}[D(\rho^*(m);t_i)|t_i \sim m] = k$ and $\rho^*(m) > 0$. In this case, focusing for now on m with full support, the price is stochastic even in the large-market limit. Given large n, the realized per-capita demand at price $\rho^*(m)$ will be weakly greater than per-capita supply k with probability of about $\frac{1}{2}$, and will be strictly smaller than per-capita supply k with probability of about $\frac{1}{2}$.[9] Therefore, the price in the limit will be $\rho^*(m)$ with probability of $\frac{1}{2}$, and $\rho^*(m) - 1$ with probability of $\frac{1}{2}$. $\phi^\infty(t_i, m)$ assigns to player i the following expected number of units,

Reported Value	Expected Number of Units
$v_i < \rho^*(m)$	0
$v_i \geq \rho^*(m)$	q_i

and prices are $\rho^*(m)$ or $\rho^*(m) - 1$ with equal probability. Note that bids of $\rho^*(m)$ are not rationed in the limit. This is so because, in this knife-edge case, average demand is exactly equal to average supply. Moreover, in both cases the price in the limit is exogenous from the perspective of each individual bidder. Even though the price is sometimes $\rho^*(m)$ and sometimes $\rho^*(m) - 1$, the probability that bidder i is pivotal in determining which of the two prices occurs converges to zero.

The argument that the uniform-price auction is SP-L is now straightforward. Choose any type t_i and any full support distribution $m \in \bar{\Delta}T$. The description of ϕ^∞ above implies that truthful reporting is a dominant strategy in the limit, hence Definition 4 is satisfied.

Last, we turn to the pay-as-bid auction. The pay-as-bid auction allocates units of the good in exactly the same way as the uniform-price auction. The difference is that winning bidders pay their bid instead of the market-clearing price $p^*(t)$. Clearly, bidders will gain from misreporting their value, even in the large-market limit. If the distribution of opponent bids is m and the limit price is $\rho^*(m)$, then a bidder of type $t_i = (v_i, q_i)$ with $v_i > \rho^*(m) + 1$ strictly prefers to misreport as $t_i' = (\rho^*(m) + 1, q_i)$: he receives the same allocation in the limit but pays a strictly lower price per unit. Hence, the pay-as-bid auction is not SP-L.□

[9] The intuition is that if a fair coin is tossed $n \to \infty$ times, the probability that at least $n/2$ of the tosses are heads converges to $1/2$, just as the probability that less than $n/2$ of the tosses are heads converges to $1/2$, with both probabilities independent of the outcome of the i^{th} toss.

Discussion: SP-L vs. Traditional Approximate SP Observe that the argument that the uniform price auction is SP-L would not go through using a stronger notion of asymptotic strategy-proofness based on realizations of opponents' reports rather than probability distributions. To see this, consider the case where there are $k = 2$ objects per bidder, and bidder i knows that all other bidders will report a demand of 2 objects for \$100. That is, that all other bidders report a type of $(2, \$100)$ for sure. Then bidder i knows that she is marginal, and can reduce the market-clearing price to 0 by asking for 1 object instead of 2. This example illustrates the importance of the interim perspective in the definition of SP-L, and why SP-L classifies mechanisms in a substantially different way than the traditional ex-post notion of approximate strategy-proofness.

Discussion: Full Support Requirement The uniform-price auction example also illustrates the importance of the full-support requirement in the SP-L definition. If a bidder believes that opponent reports equal $(2, \$100)$ for sure, then she could lower the market-clearing price from \$100 to \$0 by demanding a single unit. While this example uses a degenerate distribution with support on a single type, there are non-degenerate distributions where a bidder can manipulate the uniform-price auction.[10] For example, if bidder i believes that opponents report $(2, \$100)$ or $(2, \$200)$ with equal probability, then she can still, with high probability, drive the prices down from \$100 to \$0 by asking for one unit instead of two. Even though she is uncertain about other players' types, this uncertainty is at a part of the demand curve that is not relevant for the determination of the market-clearing price. That said, such manipulations do not seem very realistic because they require extremely detailed information about opponent play. The full support requirement in the SP-L definition is a simple way to capture the idea that agents are not likely to have that level of information.

The uniform-price auction example also suggests that, on a case by case basis, it may not always be necessary to assume full support. As long as there is uncertainty about opponents' play in a region that is relevant for price determination, bidding truthfully will be optimal in a large enough market. For example, assume that bidder i believes that her opponents report $(4, \$100)$ or $(4, \$25)$ with 50% chance each. Then, in the limit, the market-clearing price is \$100 or \$25 with 50% chance each, and bidder i cannot meaningfully affect the price. Thus, even in this adversarial case where supply intersects the expected demand curve at a discontinuity, and bidder i thinks that the distribution of opponents' reports has only two elements in its support, reporting truthfully is approximately optimal from the interim

[10] See also Swinkels (2001; Section 5) for an elegant example, with limited support, in which bidders remain pivotal with probability one even in very large markets.

perspective.

4 Classification of Non-SP Mechanisms

This section classifies a number of non-SP mechanisms into SP-L and manipulable in the large (Table 1 in the Introduction), and discusses how the classification organizes the evidence on manipulability in large markets. Specifically, all of the known mechanisms for which there is a detailed theoretical case that the mechanism has approximate incentives for truthtelling in large markets are SP-L (Section 4.2), and all of the known mechanisms for which there is empirical evidence that non-strategy-proofness causes serious problems even in large markets are manipulable in the large (Section 4.3). In particular, the classification of mechanisms based on whether or not they are SP-L predicts whether misreporting is a serious problem in practice better than the classification of mechanisms based on whether or not they are SP. These results suggest that, in large markets, SP-L versus not SP-L is a more relevant relevant dividing line than SP versus not SP.

Before proceeding, we make three brief observations regarding the classification. First, both the SP-L and the manipulable in the large columns of Table 1 include mechanisms that explicitly use prices (e.g., multi-unit auctions), as well as mechanisms that do not use prices (e.g., matching mechanisms). For the mechanisms that do use prices, the SP-L ones are exactly those where an agent who takes prices as given wishes to report truthfully, such as the uniform-price auction. Second, the table is consistent with both Milton Friedman's (1960; 1991) argument in favor of uniform-price auctions over pay-as-bid auctions, and Alvin Roth's (1990; 1991; 2002) argument in favor of deferred acceptance over priority-match algorithms. Notably, while both Friedman's criticism of pay-as-bid auctions and Roth's criticism of priority-match algorithms were made on incentives grounds, the mechanisms they suggested in their place are not SP but are SP-L. Third, with the exception of probabilistic serial, none of the SP-L mechanisms satisfy a stronger, ex-post, notion of approximate strategy-proofness. That is, the classification would not conform to the existing evidence, nor to Friedman's and Roth's arguments, without the ex-interim perspective in the definition of SP-L.

4.1 Obtaining the Classification

To show that a mechanism is not SP-L it suffices to identify an example of a distribution of play under which agents may gain by misreporting, even in the limit. For SP-L mechanisms, this section gives two easy-to-check sufficient conditions for a mechanism to be SP-L, which directly yield the classification for all of the SP-L mechanisms in Table 1. Formal definitions of each mechanism and detailed derivations are in Supplementary Appendix C.[11]

The first sufficient condition is envy-freeness, a fairness criterion which requires that no player i prefers the assignment of another player j, for any realization of the reported payoff types t.

Definition 5. *A direct mechanism $\{(\Phi^n)_{\mathbb{N}}, T\}$ is **envy-free (EF)** if, for all i, j, n, t:*

$$u_{t_i}[\Phi_i^n(t)] \geq u_{t_i}[\Phi_j^n(t)].$$

Theorem 1 below shows that EF implies SP-L. The connection between envy-freeness and incentive compatibility in large markets was first observed by Hammond (1979), who shows, in a continuum exchange economy, that EF implies SP. For related contributions see Champsaur and Laroque (1981) and Jackson and Kremer (2007).

The mathematical intuition for why EF implies SP-L is as follows. In anonymous mechanisms, the gain to player i from misreporting as player j can be decomposed as the sum of the gain from receiving j's bundle, holding fixed the aggregate distribution of reports, plus the gain from affecting the aggregate distribution of reports (expression (A.2) in Appendix A). Envy-freeness directly implies that the first component in this decomposition is non-positive. Lemma A.1 then shows that the second component becomes negligible in large markets. More precisely, even though there may exist realizations of the other players' reports where player i's gain from affecting the aggregate distribution is large (e.g., if by misreporting he affects the clearing price in the uniform-price auction), his effect on the *interim distribution of the empirical distribution* of reports vanishes with market size, at a rate of essentially $\sqrt{n}$. This yields both that EF implies SP-L and the convergence rate for EF mechanisms as stated in Theorem 1.

Most of the mechanisms in the SP-L column of Table 1 are EF, with the only exceptions being approximate CEEI and deferred acceptance.[12] To classify these mechanisms,

[11]Two of these mechanisms do not fit the framework used in the body of the paper. Deferred acceptance is a semi-anonymous mechanism, and the Walrasian mechanism has an infinite set of bundles. For details of how we accommodate these generalizations, see Supplementary Appendix C. Moreover, to define some of these mechanisms we make a selection from a correspondence. For example, in the Hylland and Zeckhauser (1979) pseudomarket mechanism, individuals report preferences for objects, and a competitive equilibrium is calculated. There exist preference profiles for which there are multiple equilibria. In the appendix, we formally define the HZ mechanism as a mechanism that picks an arbitrary selection from this correspondence. We state similar formal definitions for the approximate CEEI and Generalized HZ mechanisms (see Supplementary Appendix C).

[12]Both approximate CEEI and deferred acceptance include as a special case the random serial dictatorship

we introduce a weakening of EF that we call envy-free but for tie breaking (EF-TB). A mechanism is envy-free but for tie breaking if, after reports are submitted, the mechanism runs a tie-breaking lottery, and allocations depend on reports and on the lottery. After the lottery is realized, no participant envies another participant with a worse lottery number. The simplest example is the random serial dictatorship mechanism for allocating objects without using money. Random serial dictatorship orders participants according to a lottery, and participants then take turns picking their favorite object out of the objects that are still available. This mechanism has envy ex post, because a participant may prefer the allocation of another participant who got a better lottery number. Bogomolnaia and Moulin (2001) have shown that the mechanism can also have envy before the lottery is drawn. However, after the lottery is drawn, no participant envies another participant with a worse lottery number, which means that this mechanism is envy-free but for tie breaking. Formally, the definition is as follows.[13]

Definition 6. *A direct mechanism $\{(\Phi^n)_{\mathbb{N}}, T\}$ is **envy-free but for tie breaking (EF-TB)** if for each n there exists a function $x^n : (T \times [0,1])^N \to \Delta(X_0^n)$, symmetric over its coordinates, such that*

$$\Phi^n(t) = \int_{l \in [0,1]^n} x^n(t,l)dl$$

and, for all i, j, n, t, and l, if $l_i \geq l_j$ then

$$u_{t_i}[x_i^n(t,l)] \geq u_{t_i}[x_j^n(t,l)].$$

The following theorem shows that either condition guarantees that a mechanism is SP-L.

Theorem 1. *If a mechanism is EF-TB (and in particular if it is EF), then it is SP-L. The maximum possible gain from misreporting converges to 0 at a rate of $n^{-\frac{1}{2}+\epsilon}$ for EF mechanisms, and $n^{-\frac{1}{4}+\epsilon}$ for EF-TB mechanisms. Formally, if a mechanism is EF (EF-TB), then given $\mu \in \bar{\Delta}T$ and $\epsilon > 0$ there exists $C > 0$ such that, for all t_i, t_i' and n, the gain from deviating,*

$$u_{t_i}[\phi_i^n(t_i', \mu)] - u_{t_i}[\phi_i^n(t_i, \mu)],$$

is bounded above by

$$C \cdot n^{-\frac{1}{2}+\epsilon} \quad (C \cdot n^{-\frac{1}{4}+\epsilon}).$$

mechanism, which Bogomolnaia and Moulin (2001) show is not envy-free.

[13]This definition is for anonymous mechanisms. The definition for semi-anonymous mechanisms, which is needed for deferred acceptance, is contained in Supplementary Appendix B. The semi-anonymous version of the definition can also be used for school choice problems in which there are multiple groups of students with different priority classes (e.g., sibling priority or walk-zone priority).

The theorem shows that either condition can be used to classify new or existing mechanisms as SP-L. It also gives reasonable rates of convergence for the maximum possible gain from manipulating a mechanism.

The first claim of Theorem 1 can also be stated in the language of implementation theory, which we will introduce formally in Section 5. In this language the result is (informally): *Any social choice function F that depends only on payoff types and is EF or EF-TB is SP-L implementable.*

The proof of the theorem for the EF-TB case builds upon the argument for the EF case, by showing that EF-TB mechanisms have small amounts of envy before lotteries are drawn (Lemma A.2). This is accomplished with three basic ideas. First, how much player i envies player j prior to the lottery draw equals the average envy by all type t_i players towards type t_j players, as a consequence of anonymity. Second, it is possible to bound this average envy, after a given lottery draw l, by how evenly distributed the lottery numbers in the vector l are. Intuitively, if players of types t_i and t_j receive evenly distributed lottery numbers, average envy has to be small. The final step is an application of a probabilistic bound known as the Dvoretzky–Kiefer–Wolfowitz inequality, which guarantees that lottery numbers are typically very evenly distributed.

4.2 Relationship to the Theoretical Literature on Large Markets

The SP-L column of Table 1 organizes a large literature demonstrating the approximate incentive compatibility of specific mechanisms in large markets. Our results show that a number of mechanisms for which the literature established approximate incentive compatibility results are SP-L. This includes Walrasian mechanisms (Roberts and Postlewaite, 1976, Hammond (1979) and Jackson and Manelli, 1997), double auctions (Rustichini et al., 1994 and Cripps and Swinkels, 2006), uniform-price auctions (Swinkels, 2001), deferred acceptance mechanisms (Immorlica and Mahdian, 2005 and Kojima and Pathak, 2009),[14] and the probabilistic serial mechanism (Kojima and Manea, 2010). We also obtain new results on the Hylland and Zeckhauser (1979) pseudomarket mechanism, approximate CEEI (Budish, 2011), and the generalized Hylland-Zeckhauser mechanism (Budish et al., 2013), each of

[14]Our assumption of a finite number of types is an important limitation in matching models. The problem is that, as the market grows, we can only have a finite number of preference orderings over partners. This contrasts with other large-market matching models, such as Kojima and Pathak (2009), Kojima et al. (2013), Ashlagi et al. (2014), and Lee (2017). One interesting direction of future research is to generalize SP-L to infinite sets of types, where there is a topological notion of types that are close to one another. This could expand the applicability of the SP-L concept, especially to models like Lee (2017), where the set of types has a natural topological structure.

whose large-market incentive properties had not previously been formally studied.

The single concept of SP-L and Theorem 1 classifies all of these mechanisms. In contrast, the prior literature has employed different notions of approximate incentive compatibility and different analysis techniques, tailored for each mechanism.[15] Of course, analyses that are tailored to specific mechanisms can yield a more nuanced understanding of the exact forces pushing players away from truthful behavior in finite markets, as in the first-order condition analysis of Rustichini et al. (1994) or the rejection chain analysis of Kojima and Pathak (2009).

4.3 Relationship to Empirical Literature on Manipulability

For each of the manipulable in the large mechanisms in Table 1, there is explicit empirical evidence that participants strategically misreport their preferences in practice. Furthermore, misreporting harms design objectives such as efficiency or fairness. In this section we briefly review this evidence.[16]

Consider first multi-unit auctions for government securities. Empirical analyses have found considerable bid shading in discriminatory auctions (Hortaçsu and McAdams, 2010), but negligible bid shading in uniform-price auctions, even with as few as 13 bidders (Kastl, 2011; Hortaçsu et al. (2015)). Friedman (1991) argued that the need to play strategically in pay-as-bid auctions reduces entry of less sophisticated bidders, giving dealers a sheltered market that facilitates collusion. In uniform-price auctions, by contrast, "You do not have to be a specialist" to participate, since all bidders pay the market-clearing price. Consistent

[15]This note elaborates on the different concepts used in the literature. Roberts and Postlewaite (1976) ask that truthful reporting is ex-post approximately optimal for all opponent reports where equilibrium prices vary continuously with reports. Hammond (1979) shows that the Walrasian mechanism satisfies exact SP in a continuum economy. Rustichini et al. (1994) study the exact Bayes-Nash equilibria of double auctions in large markets, and bound the rate at which strategic misreporting vanishes with market size. Swinkels (2001) studies both exact Bayes-Nash equilibria and ϵ-Bayes-Nash equilibria of the uniform-price and pay-as-bid auctions. Kojima and Pathak (2009) study ϵ-Nash equilibria of the doctor-proposing deferred acceptance algorithm assuming complete information about preferences on the hospital side of the market and incomplete information about preferences on the doctor side of the market. In an appendix they also consider ϵ-Bayes-Nash equilibria, in which there is incomplete information about preferences on both sides of the market. Kojima and Manea (2010) show that probabilistic serial satisfies exact SP, without any modification, in a large enough finite market.

[16]We note that even for SP mechanisms preference reporting is not perfect. Rees-Jones (2015) provides survey evidence of misreporting in the US medical resident match on the doctor side of the market (for which truthful reporting is a dominant strategy), which he attributes in part to students misunderstanding the strategic environment (see also Hassidim et al., 2015). Laboratory studies have also found misreporting in SP mechanisms, though these experiments find significantly lower rates of misreporting in SP mechanisms than in easily manipulable mechanisms (Chen and Sönmez, 2006 and Featherstone and Niederle, 2011) and significantly lower rates of misreporting when it is obvious to participants why the SP mechanism is SP (Li, 2015).

with Friedman's view, Jegadeesh (1993) shows that pay-as-bid auctions depressed revenues to the US Treasury during the Salomon Squeeze in 1991, and Malvey and Archibald (1998) find that the US Treasury's adoption of uniform-price auctions in the mid-1990s broadened participation. Cross-country evidence is also consistent with Friedman's argument, as Brenner et al. (2009) find a positive relationship between a country's using uniform-price auctions and indices of ease of doing business and economic freedom, whereas pay-as-bid auctions are positively related with indices of corruption and of bank-sector concentration.

Next, consider the Boston mechanism for school choice. Abdulkadiroğlu et al. (2006) find evidence of a mix of both sophisticated strategic misreporting and unsophisticated naive truthtelling; see also recent empirical work by Agarwal and Somaini (2014) and Hwang (2014). Sophisticated parents strategically misreport their preferences by ranking a relatively unpopular school high on their submitted preference list. Unsophisticated parents, on the other hand, frequently play a dominated strategy in which they waste the highest positions on their rank-ordered list on popular schools that are unattainable for them. In extreme cases, participants who play a dominated strategy end up not receiving any of the schools they ask for.

Next, consider the mechanisms used in practice for the multi-unit assignment problem of course allocation. In the bidding points auction, Krishna and Ünver (2008) use both field and laboratory evidence to show that students strategically misreport their preferences, and that this harms welfare. Budish (2011) provides additional evidence that some students get very poor outcomes under this mechanism; in particular students sometimes get zero of the courses they bid for. In the Harvard Business School draft mechanism, Budish and Cantillon (2012) use data consisting of students' stated preferences and their underlying true preferences to show that students strategically misreport their preferences. They show that misreporting harms welfare relative both to a counterfactual in which students report truthfully, and relative to a counterfactual in which students misreport, but optimally. They also provide direct evidence that some students fail to play best responses, which supports the view that Bayes-Nash equilibria are less robust in practice than dominant-strategy equilibria.

For labor market clearinghouses, Roth (1990, 1991, 2002) surveys a wide variety of evidence that shows that variations on priority matching mechanisms perform poorly in practice, while variations on Gale and Shapley's deferred acceptance algorithm perform well. Roth emphasizes that the former are unstable under truthful play whereas the latter are stable under truthful play. By contrast, we emphasize that the former are not SP-L whereas the latter are SP-L.

5 SP-L is Approximately Costless in Large Markets Relative to Bayes-Nash

This section shows that, in large markets, SP-L is in a precise sense approximately costless to impose relative to Bayes-Nash incentive compatibility. Formally, we give conditions under which, if it is possible to implement a social choice function with a Bayes-Nash incentive compatible mechanism, then it is possible to approximately implement this social choice function with an SP-L mechanism. The exception is that there can be a large cost of using SP-L if the intended social choice function is discontinuous in agents' beliefs.

5.1 Social Choice Functions and Type Spaces

To formally state the result we need to introduce the notions of social choice functions and type spaces. These concepts will allow us to describe the social outcomes produced by mechanisms such as pay-as-bid auctions or the Boston mechanism, which are not SP-L but which have Bayes-Nash equilibria that depend on agents' beliefs. Our definition of type spaces is similar to that in the robust mechanism design literature as in Bergemann and Morris (2005).

A **type space** $\Omega = ((\Omega_{n,i})_{n\in\mathbb{N},i=1,\ldots,n}, \hat{t}, \hat{\pi})$ consists of a measurable set of **types** $\Omega_{n,i}$, for every market size n and agent i, and measurable maps $\hat{t}$ and $\hat{\pi}$. These maps associate, with each type ω_i in a given set $\Omega_{n,i}$, a **payoff type** $\hat{t}(\omega_i)$ in T, and **beliefs** $\hat{\pi}(\omega_i)$ over the joint distribution of opponent types $(\omega_1, \ldots, \omega_{i-1}, \omega_{i+1}, \ldots, \omega_n)$ in $\times_{j\neq i}\Omega_{n,j}$. The type space Ω_n for market size n is the product of the individual type spaces $\Omega_{n,i}$. Thus, a type encodes a participant's information about her preferences and about other participants' preferences and beliefs. For our purposes, it is sufficient to restrict attention to type spaces that are symmetric across players, so that $\Omega_{n,i}$ is the same set for all i, and with an onto function $\hat{t}$ so that all payoff types are possible.

Given a type space Ω, a **social choice function** is a sequence $F = (F^n)_{n\in\mathbb{N}}$ of functions

$$F^n : \Omega_n \longrightarrow \Delta(X_0)^n.$$

Note, importantly, that we allow social choice functions to depend on both preferences and beliefs, not only on preferences. The reason is that the outcomes of many commonly used mechanisms vary with both preferences and beliefs. For example, outcomes of the pay-as-bid auction or the Boston mechanism depend on what participants believe about other participants' preferences and information; holding an agent's preferences fixed, their beliefs affect how much they will shade their bid, or whether they will take a risk and rank a popular school first. To discuss the social choice functions implemented by these mechanisms, we need to include beliefs in the definition of a social choice function. Our definition differs from that in Bergemann and Morris (2005), where social choice functions only depend on payoff types. Our definition is similar to that in the literature on implementation with incomplete information (as in Postlewaite and Schmeidler, 1986 or Jackson, 1991), where social choice functions may depend on payoff-irrelevant information.[17] We say that a social choice function **depends only on payoff types** if, for all n, and all ω and ω' in $\Omega_{n,i}$ such that $\hat{t}(\omega_i) = \hat{t}(\omega_i')$ for all i, we have $F^n(\omega) = F^n(\omega')$.

Much applied work in mechanism design considers Bayesian equilibria where agents have a common, i.i.d. prior μ about payoff types, and know their own type.[18] To describe the outcomes of such equilibria for a range of priors μ, we need a type space that includes the union of these simple type spaces for a range of values of μ.

We now define such a type space, which we denote as Ω^*. Formally, for all n, let

$$\Omega^*_{n,i} = \{(t_i, \mu) : t_i \in T, \mu \in \bar{\Delta}T\}.$$

For any $\omega_i = (t_i, \mu) \in \Omega^*_{n,i}$, let $\hat{t}(\omega_i) = t_i$. The beliefs of type ω_i are

$$\hat{\pi}(\omega_i)(\omega_{-i}) = 0$$

if, for any $j \neq i$, the first element of ω_j is not μ, and the beliefs are

$$\hat{\pi}(\omega_i)(\omega_{-i}) = \Pi_{j \neq i} \mu(\hat{t}(\omega_j))$$

[17] Restricting attention to social choice functions that only depend on payoff types is reasonable if the social choice function is interpreted as a map from preference profiles to socially optimal alternatives, as in the interpretation in Maskin (1999) p. 24, which goes back to Arrow (1951). Following Bergemann and Morris (2005), the robust mechanism design literature often restricts attention to social choice functions that only depend on payoff types. However, we use social choice functions to describe the allocations produced by equilibria of Bayesian mechanisms, which in many applications depend on beliefs. For that reason, we follow the literature on implementation with incomplete information, which typically allows for social choice functions to depend on payoff-irrelevant information. For example, Jackson (1991) p. 463 defines a social choice function as a function from a set of states to allocations. In particular, the social choice function may produce different outcomes in states where agents have the same preferences, but different information about the preferences of other agents. Maskin and Sjöström's (2002) (pp. 276-277) discussion of the Bayesian implementation literature uses a similar definition. Likewise, Postlewaite and Schmeidler's (1986) social welfare correspondences and Palfrey and Srivastava's (1989) social choice sets may depend on payoff-irrelevant information.

[18] This type space is a particular case of what Chung and Ely (2007) and Bergemann and Morris (2005) call naive type spaces.

otherwise. We will refer to the type space $\Omega^* = ((\Omega^*_{n,i})_{n\in\mathbb{N},i=1,\ldots,n}, \hat{t}, \hat{\pi})$, as the **union of all common prior, i.i.d., full-support type spaces**.

If a social choice function F is defined on Ω^*, denote by $f^n(t_i,\mu)$ the bundle that a type (μ, t_i) agent expects to receive in a market of size n. Formally,

$$f^n(t_i,\mu) = \sum_{t_{-i}\in T^{n-1}} \Pr\{t_{-i}|t_{-i} \sim iid(\mu)\} \cdot F^n_i((t_1,\mu),\ldots,(t_n,\mu)).$$

5.2 Limit Bayes-Nash and SP-L Implementability

This subsection defines the implementability notions that we need to state the theorem, and a regularity condition on social choice functions.

We begin by defining implementability in limit Bayes-Nash equilibria. It will be useful to extend the function Φ^n linearly to distributions over vectors of actions. Given a distribution $\bar{m} \in \Delta(A^n)$ over vectors of actions, let

$$\Phi^n(\bar{m}) = \sum_{a\in A^n} \bar{m}(a)\cdot\Phi^n(a). \tag{5.1}$$

Likewise, given an action a_i and a distribution $\bar{m} \in \Delta(A^{n-1})$ over $n-1$ actions, let

$$\Phi^n_i(a_i,\bar{m}) = \sum_{a_{-i}\in A^{n-1}} \bar{m}(a_{-i})\cdot\Phi^n_i(a_i,a_{-i}).$$

Given a mechanism $\{(\Phi^n)_\mathbb{N}, A\}$ and type space Ω, a **strategy** σ is defined as a map from $\Omega_{n,i}$ to ΔA. Given a strategy σ, market size n and a vector of belief types $\omega \in \Omega_n$, let $\sigma(\omega) \in \Delta(A^n)$ denote the associated distribution over vectors of actions.

Definition 7. *Given a mechanism $\{(\Phi^n)_\mathbb{N}, A\}$ and type space Ω, the strategy σ^* is a* ***limit Bayes-Nash equilibrium*** *if each participant's strategy becomes arbitrarily close to optimal as the market grows large. Formally,*[19] *for all $\epsilon > 0$ there exists n_0 such that, for all $n \geq n_0$, $\omega_i \in \Omega_{n,i}$, a_i in the support of $\sigma^*(\omega_i)$, and $a'_i \in A$:*

$$\int_{\omega_{-i}} u_{t_i}[\Phi^n_i(a'_i,\sigma^*(\omega_{-i}))] - u_{t_i}[\Phi^n_i(a_i,\sigma^*(\omega_{-i}))]\, d\hat{\pi}(\omega_i)(\omega_{-i}) \leq \epsilon.$$

A social choice function F is **limit Bayes-Nash implementable** if there exists a mech-

[19]The integral equals the expected gain for participant i with type ω_i to change her strategy. It is an expectation over all possible types ω_{-i} of other players, taken with respect to the measure $\hat{\pi}(\omega_i)$, which represents participant i's belief about the types of other participants.

anism $((\Phi^n)_{n\in\mathbb{N}}, A)$ with a limit Bayes-Nash equilibrium σ^* such that

$$F^n(\omega) = \Phi^n(\sigma^*(\omega))$$

for all ω in the type space Ω for which F is defined.

The theorem also requires the following regularity condition.

Definition 8. *A social choice function $(F^n)_{n\in\mathbb{N}}$ defined over Ω^* is* ***continuous*** *at a prior μ_0 if, given $\epsilon > 0$, there exists n_0 and a neighborhood $\mathcal{N}$ of μ_0 such that the following holds. Consider any $n \geq n_0$, and any two vectors of types ω and ω', with $\omega_i = (t_i, \mu)$ and $\omega_i' = (t_i, \mu')$ for all i, where μ, μ', and* emp$[t]$ *belong to $\mathcal{N}$. Then, for any such n, ω, and ω',*

$$\|F^n(\omega) - F^n(\omega')\| < \epsilon.$$

The social choice function is ***continuous*** *if it is continuous at every full support prior.*

That is, a social choice function defined over Ω^* is continuous if, in large enough markets, social outcomes vary continuously with beliefs. This is a substantial restriction, even for social choice functions that are Bayes-Nash implementable. In the working paper version of this article we established a version of Theorem 2 that used a weaker condition, called quasi-continuity. See Section 7.2 for further discussion.

A direct revelation mechanism $((\Phi^n)_{n\in\mathbb{N}}, T)$ **approximately implements** a social choice function $F = (F^n)_{n\in\mathbb{N}}$ defined over Ω^* if, for every $\epsilon > 0$ and prior μ in $\bar{\Delta}T$, there exists n_0 such that, for all $n \geq n_0$ and $t_i \in T$,

$$\|f^n(t_i, \mu) - \phi^n(t_i, \mu)\| < \epsilon.$$

We say that F is **approximately SP-L implementable** if there exists an SP-L mechanism that approximately implements F.

5.3 Construction Theorem

We now state the main result of this section.

Theorem 2. *Consider a social choice function $F = (F^n)_{n\in\mathbb{N}}$ defined over Ω^*, the union of all common prior, i.i.d., full support type spaces. If F is continuous and limit Bayes-Nash implementable, then F is approximately SP-L implementable.*

Proof Sketch. The proof of Theorem 2 is by construction. We provide a detailed sketch as follows, with full details contained in Appendix A.

$(F_n)_{n\in\mathbb{N}}$ is limit Bayes-Nash implementable. Therefore, there exists a mechanism $((\Phi^n)_{n\in\mathbb{N}}, A)$ with a limit Bayes-Nash equilibrium σ^* such that

$$F^n(\omega) = \Phi^n(\sigma^*(\omega))$$

for all $n \geq 0$ and vectors of types ω in Ω_n^*. Construct the direct mechanism $((\Psi^n)_{n\in\mathbb{N}}, T)$ as follows. Given a vector of payoff types t, let $\text{emp}[t] \in \Delta T$ be the empirical distribution of payoff types in t. Given a market size n, let

$$\Psi^n(t) = \Phi^n(\sigma^*((t_1, \text{emp}[t]), \ldots, (t_n, \text{emp}[t]))). \tag{5.2}$$

In words, Ψ^n plays action $\sigma^*(t_i, \text{emp}[t])$ for agent i who reports t_i, where $\text{emp}[t]$ is the *empirical* distribution of reported payoff types. The constructed mechanism Ψ^n can be interpreted as a proxy mechanism. Ψ^n plays the original mechanism Φ^n on each agent's behalf, using the limit Bayes-Nash equilibrium strategy σ^*, and assuming that players believe that payoff types are i.i.d according to the empirical distribution of payoff types $\text{emp}[t]$.

We need to establish two facts. First, the constructed mechanism yields approximately the same outcome as the social choice function. This follows from continuity and from the law of large numbers. Specifically, assume that participant i reports t_i, and that other participants' reports are i.i.d. according to a distribution μ in $\bar{\Delta}T$. The law of large numbers implies that $\text{emp}[t]$ converges to m in probability as the market grows large. Continuity then implies that the expected bundle received by agent i is close to $f^n(t_i, \mu)$.

Second, we need to show that the constructed mechanism is SP-L. Suppose that agent i's payoff type is t_i but that she reports t_i', and that other participants report i.i.d. according to a distribution μ in $\bar{\Delta}T$. We have already established that agent i receives a bundle that is close to $f^n(t_i', \mu)$. Because the social choice function is limit Bayes-Nash implementable, agent i's utility for $f^n(t_i', \mu)$ cannot be much higher than her utility for $f^n(t_i, \mu)$. As the market grows, these approximations improve, and the maximum possible gain from misreporting converges to 0. This shows that the constructed mechanism is SP-L.

Relationship to the Revelation Principle The construction used in the proof of Theorem 2 is related to the traditional Bayes-Nash direct revelation mechanism construction (Myerson, 1979). In a traditional Bayes-Nash direct revelation mechanism, the mechanism designer and participants have a common knowledge prior about payoff types, say μ_0. The mechanism announces a Bayes-Nash equilibrium strategy $\sigma^*(\cdot, \mu_0)$, and plays $\sigma^*(t_i, \mu_0)$ on behalf of an agent who reports t_i. Truthful reporting is a Bayes-Nash equilibrium.

In contrast, our constructed mechanism does not depend on a prior. Instead, the mechanism *infers* a prior from the empirical distribution of agents' play (cf. Segal (2003); Baliga and Vohra (2003)). If agents indeed play truthfully, this inference is correct in the limit. But if the agents misreport, so that the empirical $\hat{m}$ is very different from the prior μ_0, our mechanism adjusts each agent's play to be the Bayes-Nash equilibrium play in a world where the prior was in fact $\hat{m}$. As a result, an agent who reports her preferences truthfully remains happy to have done so even if the other agents misreport, unlike in a traditional Bayes-Nash direct revelation mechanism, so the constructed mechanism is SP-L rather than Bayes-Nash. Moreover, the constructed mechanism is prior free and consistent with the Wilson doctrine, unlike a traditional Bayes-Nash direct revelation mechanism. The mechanism designer need not know the prior to run the mechanism, and the participants need not know the prior to play optimally.

6 Application: The Boston Mechanism

The school choice literature has debated the desirability of the commonly used Boston mechanism for student assignment. While the mechanism has good efficiency properties, it has been criticized because it gives students strong incentives to misreport preferences. This section applies Theorem 2 to show that there exists a mechanism that produces approximately the same outcomes as the Boston mechanism, but is SP-L. We begin by giving a formal definition of the Boston mechanism and our results, and then discuss how this contributes to the debate in the literature.

6.1 Definition of the Boston Mechanism

The set of bundles is a set of schools $X_0 = S \cup \{\emptyset\}$. In a market of size n, there are $\lfloor q_s \cdot n \rfloor$ seats available in school s in S, where $q_s \in (0, 1)$ denotes the proportion of the market that s can accommodate and $\lfloor \cdot \rfloor$ is the floor function. It is assumed that X_0 includes a null school $\emptyset$ that is in excess supply. An agent of payoff type $t_i \in T$ has a utility function u_{t_i} over X_0, with no indifferences. The utility of the null school is normalized to 0. In particular, all agents strictly prefer any of the proper schools to the null school.

We consider a simplified version of the Boston mechanism with a single round.[20] The action space is the set of proper schools $A = S$, so that each student points to a school. If the number of students pointing to school s is lower than the number of seats, then all of those students are allocated to school s. If there are more students who point to s than its capacity, then students are randomly rationed, and those who do not obtain a seat in s are allocated to the null school. Formally, given a vector of reports a, the allocation $\Phi_i^n(a)$ assigns i to school a_i with probability

$$\min\left\{\frac{\lfloor q_{a_i} \cdot n \rfloor}{\mathrm{emp}_{a_i}[a] \cdot n}, 1\right\},$$

and to the null school with the remaining probability. Consequently, the limit mechanism is

$$\phi^\infty(s, m) = \min\left\{\frac{q_s}{m_s}, 1\right\} \cdot s,$$

which denotes receiving school s with the probability $\min\{q_s/m_s, 1\}$, which we term the probability of acceptance to school s, and school $\emptyset$ with the remaining probability.

6.2 Results

In the appendix, we show that the Boston mechanism has equilibria σ^* where $\sigma^*(t_i, \mu)$ depends continuously on beliefs μ for $\mu \in \bar{\Delta}T$. Theorem 2 then yields the following corollary:

Corollary 1 (SP-L implementation of the Boston mechanism)**.** *The Boston mechanism has limit Bayes-Nash equilibria that depend continuously on beliefs. For any such equilibrium σ^*, the direct mechanism constructed according to equation (5.2) is SP-L, and, in the large market limit, for any prior, truthful play of the direct mechanism produces the same outcomes as equilibrium play of σ^*.*

Interestingly, the SP-L mechanism constructed by (5.2) closely resembles the Hylland and Zeckhauser (1979) pseudo-market mechanism for single-unit assignment.[21] In the constructed mechanism, agents report their types, the mechanism computes the equilibrium market-clearing probabilities associated with the distribution of reports, and each student

[20]This simplified version of the Boston mechanism streamlines the exposition. However, this simplification means that the result in this section is stylized. An interesting question for future research is to extend the result to the standard version of the Boston mechanism, and to variations such as the adaptive Boston mechanism (Harless, 2014; Dur, 2015; Mennle and Seuken, 2015).

[21]See also Miralles (2009), which contains a very nice description of the connection between the Boston mechanism's Bayes-Nash equilibria and Hylland and Zeckhauser (1979).

points to their most-preferred school given their reported types and the computed probabilities. In Hylland and Zeckhauser (1979)'s mechanism, agents report their types, the mechanism computes equilibrium market-clearing prices given the distribution of reports, and each student purchases the lottery they like best given their reported types and the computed prices.

6.3 Discussion: the Debate over the Boston Mechanism

Our analysis offers a new perspective to an ongoing market design debate concerning the Boston mechanism. The earliest papers on the Boston mechanism, Abdulkadiroğlu and Sönmez (2003) and Abdulkadiroğlu et al. (2006), criticized the mechanism on the grounds that it is not SP, and proposed that the Gale-Shapley deferred acceptance algorithm be used instead.[22] These papers had a major policy impact as they led to the Gale-Shapley algorithm's eventual adoption for use in practice (cf. Roth, 2008).

A second generation of papers on the Boston mechanism, Abdulkadiroğlu et al. (2011); Miralles (2009); Featherstone and Niederle (2011), made a more positive case for the mechanism. They argued that while the Boston mechanism is not SP, it has Bayes-Nash equilibria that are attractive. In particular, it has Bayes-Nash equilibria that yield greater student welfare than do the dominant strategy equilibria of the Gale-Shapley procedure. Perhaps, these papers argue, the earlier papers were too quick to dismiss the Boston mechanism.

However, these second-generation papers rely on students being able to reach the attractive Bayes-Nash equilibria. This raises several potential questions: is common knowledge a reasonable assumption? Will students be able to calculate the desired equilibrium? Will unsophisticated students be badly harmed?

Corollary 1 shows that, in a large market, it is possible to obtain the attractive welfare properties of the Bayes-Nash equilibria identified by these second-generation papers on the Boston mechanism, but without the robustness problems associated with Bayes-Nash mechanisms.

We make three caveats regarding whether the constructed mechanism is appropriate for practical use. First, participants may find that a proxy mechanism like ours, or similar mechanisms like the Hylland and Zeckhauser (1979) pseudomarket mechanism, are too difficult to understand (i.e., opaque). Second, reporting von Neumann-Morgenstern preferences accurately may be difficult for participants. Therefore, with respect to these first two caveats,

[22] In two-sided matching, the Gale-Shapley algorithm is strategy proof for the proposing side of the market and SP-L for the non-proposing side of the market. In school choice only the student side of the market is strategic, with schools being non-strategic players whose preferences are determined by public policy.

to take the proxy mechanism seriously for practice one needs to explain it in a transparent way, and to design and validate a user interface for accurately reporting preferences. While these are important issues, they are addressable; the issues are similar to those dealt with in Budish and Kessler's (2016) practical implementation of an SP-L course allocation mechanism. The third caveat is that there is an ongoing empirical debate on the magnitude of the welfare gains at stake. That is, on the difference in welfare between Bayes-Nash equilibrium play of the Boston mechanism and truthful play of the Gale-Shapley mechanism (Agarwal and Somaini, 2014; Casalmiglia et al., 2014; Hwang, 2014). If these gains are small, then the simpler Gale-Shapley mechanism is likely more desirable.

7 Extensions and Discussion

7.1 Semi-Anonymity

Our analysis focuses on mechanisms that are anonymous, meaning that each agent's outcome is a symmetric function of her own action and the distribution of all actions. In Supplementary Appendix B we generalize key definitions and results to the case of semi-anonymous mechanisms, as defined in Kalai (2004). A mechanism is semi-anonymous if each agent belongs to one of a finite number of groups, and her outcome is a symmetric function of her own action, her group, and the distribution of actions within each group. This generalization is useful for two reasons. First, it allows our analysis to cover more mechanisms. For instance, double auctions are semi-anonymous if buyers and sellers belong to distinct groups; two-sided matching markets are semi-anonymous under the assumption that the number of possible types of match partners is finite (cf. footnote 14); and school choice mechanisms are semi-anonymous if there are multiple priority classes. Second, it allows results and concepts stated for i.i.d. distributions to be extended to more general distributions. See also the discussion below in Section 7.5 about generalizing i.i.d. to distributions with aggregate uncertainty.

7.2 Relaxing Continuity

Theorem 2 assumes continuity of the given social choice function. While this assumption has an intuitive appeal, it is a substantial assumption. Some well-known mechanisms violate it. For example, in pay-as-bid and uniform-price auctions, even though a small change in the prior typically has only a small effect on agents' bids, this small change in bids can have a large (i.e., discontinuous) effect on the number of units some bidder wins or the market-clearing price.

In the working paper version of this article (Azevedo and Budish, 2016), we show that a weaker version of Theorem 2 obtains under a condition that we call quasi-continuity. Quasi-continuity allows for the desired social choice function to have discontinuities, with respect to both the prior and the empirical distribution of reports, but requires that the discontinuities are in a certain sense knife-edge. Roughly, any discontinuity is surrounded by regions in which outcomes are continuous. Under this condition, the conclusion of the theorem is as follows. If the social choice function is continuous at a given prior μ_0, then, as before, there exists an SP-L mechanism that gives agents the same outcomes in the large-market limit. If the social choice function is not continuous at μ_0, then there exists an SP-L mechanism that gives agents a convex combination of the outcomes they would obtain under the desired social choice function, for a set of priors in an arbitrarily small neighborhood of μ_0.

A question that remains open for future research is to fully characterize the conditions under which there is no gap between Bayes-Nash and SP-L in large markets. The working paper lists counterexamples that fail quasi-continuity, and in which the construction does not approximate the desired social choice function, even for the weaker form of approximation described above. However, the counterexamples are far from market design applications. Moreover, the fact that the construction leaves a gap between Bayes-Nash and SP-L proves that our method of proof does not work, but does not prove that there is a gap.

Given these open questions, we do not see Theorem 2 as providing definitive proof that there is never an advantage to using Bayes-Nash over SP-L in large markets. Rather, we see the results as suggesting that, for the purposes of practical market design, a researcher may be justified searching in the space of SP-L mechanisms rather than broadening her search to include Bayes-Nash. For there to be a meaningful gain to using Bayes-Nash over SP-L in large markets, the Bayes-Nash social choice function must fail quasi-continuity, which means that its outcomes are extremely sensitive to agents' beliefs and reports. In addition, the researcher must believe the usual conditions required for Bayes-Nash equilibrium, such as common knowledge and strategic sophistication, which seems unrealistic in the context of a highly discontinuous mechanism.

7.3 Discussion: Strategic Simplicity and Fairness

In the introduction we emphasized strategic simplicity and fairness as reasons why the market design literature has found SP so compelling relative to Bayes-Nash, alongside the traditional robustness arguments associated with Wilson (1987) and Bergemann and Morris (2005). In this sub-section we briefly formalize the sense in which SP-L mechanisms approximate these appeals of SP.

SP-L mechanisms are strategically straightforward in the following sense: for any full-support beliefs $m \in \bar{\Delta}T$ about opponent play, and any cost $c > 0$ of calculating an optimal response, in a large enough market it is optimal to simply report truthfully and avoid the cost c. This cost c might be the cost of gathering information about the rules of the game, or about opponents' preferences and beliefs, or about which equilibrium will be played, etc. It could also be a real cost of lying about one's type, as in Kartik (2009). In the terminology of Roth (2008), truthful reporting is a safe strategy.

SP-L mechanisms are fair to unsophisticated players in the following sense: for any distribution of play $m \in \bar{\Delta}T$, and any $c > 0$, in a large enough market the cost of being unsophisticated and just reporting truthfully is less than c. By unsophisticated players we mean players who are able to express their own preferences, but who do not have the information or strategic sophistication to misreport their preferences optimally. Examples include parents who choose dominated strategies in school choice mechanisms, as in Pathak and Sönmez (2008), and individual investors participating in government bond auctions, as discussed by Friedman (1991).

7.4 Discussion: Voting, Public Goods and Strict SP-L

Our motivation and analysis has focused on canonical problems in market design. A natural question is whether SP-L is a useful concept in mechanism design problems outside of market design, such as voting and public goods provision. We suggest that the answer is no, and that thinking about the difference between such problems and market design problems suggests a modest strengthening of SP-L that may be useful.

The key difference between voting and public goods problems, on the one hand, and the market design problems we have emphasized, is that in voting and public goods provision every agent gets the same outcome.[23] Technically, this can be accommodated in our frame-

[23]For expositional simplicity our discussion of public goods provision focuses on the case where agents' payments vary with the societal decision (e.g., whether or not to build a park, or whether to build an expensive park, a cheap park, or no park) but not, conditional on the societal decision, with their specific report. For public goods mechanisms in which agents' payments do vary with their specific report (conditioning on the societal decision), and preferences are quasi-linear, the conclusions are as follows. Such mechanisms are SP-L if and only if, in the large market limit, all types make the same expected payment from the interim perspective, for any full-support i.i.d. distribution of others' reports. Such mechanisms are never strictly SP-L. If they satisfy the my play matters condition, it is because agents' reports can affect their payments in the large market limit, not the social decision; in such a case the mechanism is manipulable in the large.

work by letting X_0 denote the set of social alternatives and defining the set Y_n of feasible allocations in a market of size n to be the set $(x, x, \ldots, x)_{x \in X_0}$, i.e., the set of allocations in which all n agents get the same outcome. This environment allows for a wide range of voting mechanisms. For example, agents report their preferences and the mechanism chooses the alternative that is the majority winner, or the Borda count winner, and so on. In this environment, *any* anonymous mechanism is SP-L. This follows from the observation that any mechanism is trivially envy-free, because all agents get the same outcome, and Theorem 1. However, such mechanisms are SP-L for the trivial reason that in the large market limit each agent has zero impact on their outcome. By contrast, in market design problems, while each agent has zero effect on aggregate statistics such as prices, *each agent has a large effect on what they themselves receive* given the aggregates.

To clarify the distinction between the market design problems of interest in the present paper and problems such as voting and public goods provision in which all agents get the same outcome, we introduce the following mild strengthening of SP-L:

Definition 9. *The direct mechanism $\{(\Phi^n)_{\mathbb{N}}, T\}$ is **strictly strategy-proof in the large (strictly SP-L)** if it is SP-L, and, in addition, for any $t_i \in T$ there exists a $t_i' \in T$ and $m \in \bar{\Delta}T$ such that:*

$$\liminf_{n\to\infty} u_{t_i}[\phi^n(t_i, m)] - u_{t_i}[\phi^n(t_i', m)] > 0.$$

If the mechanism has a limit, this additional requirement is equivalent to, for any $t_i \in T$, there exists a $t_i' \in T$ and $m \in \bar{\Delta}T$ such that

$$u_{t_i}[\phi^\infty(t_i, m)] > u_{t_i}[\phi^\infty(t_i', m)].$$

In words, strict SP-L requires that truthful reporting is at least approximately as good as all other reports (as in Definition 4), and, in addition, is strictly preferred to at least some other report for at least some distribution. This is a modest additional requirement and it is easy to see that all of the SP-L mechanisms in Table 1 satisfy it. At the same time, voting and public goods mechanisms, in which all agents get the same outcome, are easily seen to fail this condition.[24]

(To avoid potential confusion, note that in the Vickrey-Clarke-Groves mechanism, while agents' payments do vary with their reports in finite markets – e.g., if they are pivotal in the decision to provide the expensive park – in the large market limit all types' payments are zero because no agent is pivotal.)

[24] Carroll (2013) introduces an interesting approach to studying approximate incentive compatibility for voting rules. His criterion, like SP-L, evaluates incentives to misreport from the interim perspective with respect to i.i.d. distributions of others' play. But, since in voting all mechanisms are trivially SP-L, he studies the *rate* at which incentives to misreport vanish as the market grows large. Also of interest are d'Aspremont

Theorems 1 and 2 are stated and proved for SP-L, not strict SP-L, but fortunately they can be applied as is to reach conclusions for strict SP-L under a mild additional condition.

Definition 10. *The direct mechanism* $\{(\Phi^n)_{\mathbb{N}}, T\}$ *satisfies the* ***my play matters*** *condition if, for any* $t_i \in T$*, there exists a* $t_i' \in T$ *and* $m \in \bar{\Delta}T$ *such that the limit at distribution of play* m *exists and*

$$u_{t_i}[\phi^\infty(t_i, m)] \neq u_{t_i}[\phi^\infty(t_i', m)].$$

In words, the my play matters condition requires that reporting truthfully affects one's utility as compared to at least some potential misreport and distribution. It is easy to see that all of the market design mechanisms we mention throughout the paper, both those that are SP-L and those that are manipulable in the large, satisfy the condition, whereas, again, voting and public goods mechanisms fail it. The following is immediate:

Remark 1. *If a mechanism is SP-L and satisfies the my play matters condition, it is strictly SP-L.*

Remark 1 allows results for SP-L to be translated into results for strict SP-L by checking the my play matters condition. Theorem 1 shows that if a mechanism is EF or EF-TB, it is SP-L. Therefore, by Remark 1, if the mechanism also satisfies the my play matters condition, it is strictly SP-L. Similarly, in Theorem 2, if the mechanism constructed in the proof satisfies the my play matters condition, then it is SP-L.[25]

One could imagine a stronger notion of strict SP-L in which, instead of requiring that for each true type t_i there exists some misreport t_i' and distribution m for which the preference to report truthfully is strict, we instead require that for every true type t_i and misreport t_i', there exists such a distribution m. This condition would require a correspondingly stronger my play matters condition for Remark 1 to follow. The advantage of our version of strict

and Peleg (1988) and Majumdar and Sen (2004), who study ordinal Bayesian incentive compatibility (OBIC). OBIC strengthens BIC by requiring that reporting one's ordinal preferences truthfully is optimal for any cardinal representation of an agent's true ordinal preferences, but at the same time OBIC is weaker than SP because it evaluates incentives to misreport with respect to a common-knowledge prior about the distribution of others' reports. While mathematically unrelated to SP-L, OBIC is in a similar spirit in that it identifies a compelling criterion between BIC and SP.

[25]This is a condition on the social choice function F considered, in the sense that the mechanism is a well-defined construction that depends on the social choice function. The condition can alternatively be stated directly in terms of the social choice function. The condition of my play matters on the constructed mechanism is equivalent to the following condition on the social choice function. For every t_i in T, there exists m in $\bar{\Delta}T$ and t_i' in T such that the limits of $f^n(t_i, m)$ and $f^n(t_i', m)$ as n goes to infinity exist, and

$$u_{t_i}[\lim_{n\to\infty} f^n(t_i, m)] \neq u_{t_i}[\lim_{n\to\infty} f^n(t_i', m)].$$

SP-L is that both it and the corresponding my play matters condition are easier to check, and it is strong enough to usefully distinguish canonical market design mechanisms (construed broadly to include Walrasian allocation as well as auctions, assignment, matching, etc.) from voting and public goods mechanisms. That said, some researchers may find the stronger notion to be more compelling for their particular application.

7.5 Discussion: Aggregate Uncertainty

SP-L requires that agents find it approximately optimal to report truthfully for any full-support i.i.d. distribution of opponent play. We motivated this assumption by arguing that it is more realistic to assume that agents have beliefs about the distribution of opponents' play, rather than about the precise realization of opponents' play, as in SP, or in the traditional notion of approximate SP. Still, this is a lot of information. A natural question is to what extent SP-L is robust to participants having even less information. More specifically, what happens in the realistic case where participants have aggregate uncertainty?

We can formalize this question as follows. Consider a participant whose aggregate uncertainty about the distribution of others' play can be modeled by a finite set of states of the world, and beliefs about opponent play that are i.i.d. conditional on the state. Formally, she believes that with probability p_k her opponents play i.i.d. with probability distribution $m_k \in \bar{\Delta}T$, out of a finite set of possible full-support distributions $m_1, \ldots, m_K$. Observe that her belief about her opponents' overall distribution of play is not i.i.d., but is i.i.d. conditional on the resolution of aggregate uncertainty as indexed by k. If a mechanism is SP-L, will a participant who faces aggregate uncertainty wish to report optimally in a large market?

It is straightforward to see that the answer to this question is affirmative. SP-L requires that reporting truthfully is approximately optimal, in a large enough market, for any full support i.i.d. distribution of play. This implies that reporting truthfully is approximately optimal, in a large enough market, for any finite mixture of full support i.i.d. distributions of play. Formally, for any $\epsilon > 0$, there exists n_k such that the gain from misreporting conditional on opponent play being i.i.d. according to m_k is less than ϵ. Therefore, if n is greater than the maximum of the n_k, the participant cannot gain more than ϵ by misreporting.

Since a mechanism being SP-L implies that a mechanism is incentive compatible in a large market under aggregate uncertainty, the conclusions of Theorems 1 and 2 generalize to accommodate aggregate uncertainty, and all of the mechanisms in the SP-L column of Table 1 are incentive compatible in large markets under aggregate uncertainty. The key thing to emphasize is that, while SP-L evaluates the incentives to misreport from the perspective of an agent who perceives others' play as full-support i.i.d., it requires that the agent finds truthful reporting to be approximately optimal for *any* such distribution. Thus, while it may seem that SP-L only gives incentives for truth-telling if a participant has these very specific i.i.d. beliefs, in fact if a mechanism is SP-L truthful reporting is approximately optimal for any convex combintation of i.i.d. full-support priors. This means that reporting truthfully is approximately optimal in an SP-L mechanism under a much more general set of circumstances, including when there is aggregate uncertainty.

8 Related Literature

Our paper is related to three broad lines of literature: the literature on how large markets ease incentive constraints for specific mechanisms; the literature on implementation theory; and the literature on the role of strategy-proofness in market design. We discuss each in turn.

Large Markets Our paper is most closely related to the large theory literature that has studied how market size can ease incentive constraints for specific mechanisms. We discussed this literature in detail in Section 4.2. It is important to highlight that the aim of our paper is quite different from, and complementary to, this literature. Whereas papers such as Roberts and Postlewaite (1976) provide a defense of a *specific pre-existing mechanism* based on its approximate incentives properties in large markets, our paper aims to justify SP-L as a *general desideratum* for market design. In particular, our paper can be seen as providing justification for focusing on SP-L when designing *new* mechanisms. Another point of difference versus this literature is that our criterion itself is new; see fn. 15 for full details of the approximate incentives criteria used in this prior literature.

Implementation Theory Our paper is closely related to the implementation theory literature, both in our high-level goals and in specific ideas. The goal of implementation theory is to determine what social choice rules can be implemented by some mechanism, under different solution concepts, and to find applicable necessary and/or sufficient conditions for implementation (Maskin and Sjöström, 2002). In the same fashion, our paper proposes an incentive compatibility concept, SP-L. Theorem 1 gives applicable sufficient conditions for a mechanism to be SP-L. Theorem 2 shows that the set of social choice functions that can be implemented is not much more restrictive than under Bayes-Nash equilibria (with the caveats above). More specifically, our paper is related to the literature on partial implementation, which considers whether a social choice rule is implemented by at least one equilibrium of a mechanism, as in Hurwicz (1972) and Bergemann and Morris (2005). This is in contrast to full implementation, which considers whether a social choice rule is implemented by every equilibrium of a mechanism, as in much of the implementation theory literature.[26] Specific ideas in our paper relate to three strands of the literature.

First, our paper is related to the literature on implementation in dominant strategies. One of the key findings of the early implementation theory literature is that implementation in dominant strategies is extremely restrictive. Hurwicz (1972) showed that it is impossible to implement the Walrasian correspondence with a strategy-proof mechanism. This impossibility is due to incentive compatibility conditions, so that dominant strategy implementation is restrictive even if we ignore the multiple equilibrium issue of full implementation. Subsequent studies showed that dominant strategy implementation is restrictive in other settings, such as quasilinear preferences, social choice, and matching (Green and Laffont, 1977; Gibbard, 1973; Satterthwaite, 1975; Roth, 1982).

Second, our paper is related to the robust implementation literature. Robust implementation considers mechanisms where truthful reporting is robust to a broad set of beliefs that the participants may have. For example, Bergemann and Morris (2005) consider direct mechanisms where it is ex interim optimal to report truthfully for arbitrary beliefs. They give conditions under which this notion of implementability is equivalent to ex post implementability, which is equivalent to dominant strategy implementability in the private values case. SP-L requires truthful reporting to be optimal under a meaningfully broader set of beliefs than in Bayesian implementation, but narrower than the set of beliefs allowed in the robust mechanism design literature. Thus, the goal of SP-L is to retain some of the benefits of robustness, while being less restrictive than robust and dominant strategies implementation.

Third, our paper uses approximations, which is an old idea in implementation theory. In implementation theory, approximations are used in the literature on large markets (e.g., Hammond, 1979 as a response to Hurwicz, 1972), which we discussed above, and in the literature on virtual implementation (Matsushima, 1988; Abreu and Sen, 1991; Abreu and Matsushima, 1992). The literature on virtual implementation revisited full implementation under classic solution concepts, but allowing for an approximate notion of implementation. Namely, the social choice function F is virtually implementable if there exists a social choice

[26] The issue of guaranteeing implementation under every equilibrium is central in the implementation theory literature. In fact, Jackson (2001) p. 660 classifies the study of full implementation as "implementation theory", and the study of partial implementation and incentive compatibility as "mechanism design".

function that is exactly implementable and whose outcomes are arbitrarily close to the outcomes of F, in a probabilistic sense. Virtual implementation differs from our approximation because the virtual implementation approximation does not depend on market size, and is for allocations and not for incentives. The main finding in this literature is that virtual implementation under many solution concepts, such as full Nash implementation, is extremely permissive, with most social choice functions being implementable (Matsushima, 1988; Abreu and Sen, 1991). In our setting, the approximation in the definition of SP-L also makes the set of mechanisms that are SP-L considerably larger than the set of mechanism that are SP. However, the notion of SP-L retains considerable bite, as many existing mechanisms are not SP-L.

Strategy-proofness in Market Design Our paper is related to three strands of literature on the role of strategy-proofness in market design. First, there is an empirical literature that studies how participants behave in real-world non-SP market designs. One example is Abdulkadiroğlu et al. (2006), who show, in the context of the school choice system in Boston, that sophisticated students strategically misreport their preferences, while unsophisticated students frequently play dominated strategies; see Hwang (2014) and Agarwal and Somaini (2014) for related studies. Another example is Budish and Cantillon (2012), who show that students at Harvard Business School strategically misreport their preferences for courses, often sub-optimally, and that this misreporting harms welfare relative to both truthful play and optimal equilibrium behavior. We discussed this literature in more detail in Section 4.3. This literature supports the SP-L concept, because all of the examples in which there is evidence of harm from misreporting involve mechanisms that not only are not SP, but are not even SP-L.

Second, several recent papers in the market design literature have argued that strategy-proofness can be viewed as a design objective and not just as a constraint. Papers on this theme include Abdulkadiroğlu et al. (2006), Abdulkadiroğlu et al. (2009), Pathak and Sönmez (2008), Roth (2008), Milgrom (2011) Section IV, Pathak and Sönmez (2013) and Li (2015). The overall argument for SP market design traces to Wilson (1987). Our paper contributes to this literature by showing that our notion of SP-L approximates the appeal of SP, while at the same time being considerably less restrictive. Also, the distinction we draw between mechanisms that are SP-L and mechanisms that are manipulable even in large markets highlights that many mechanisms in practice are manipulable in a preventable way.

Last, our paper is closely conceptually related to Parkes et al. (2001), Day and Milgrom (2008), Erdil and Klemperer (2010), and especially Pathak and Sönmez (2013). Each of these papers – motivated, like us, by the restrictiveness of SP – proposes a method to compare the manipulability of non-SP mechanisms based on the magnitude of their violation of SP. Parkes et al. (2001), Day and Milgrom (2008) and Erdil and Klemperer (2010) focus on the setting of combinatorial auctions. They propose cardinal measures of a combinatorial auction's manipulability based, respectively, on Euclidean distance from Vickrey prices, the worst-case incentive to misreport, and marginal incentives to misreport. Each of these papers then seeks to design a combinatorial auction that minimizes manipulability subject to other design objectives. Pathak and Sönmez (2013), most similarly to us, use a general mechanism design environment that encompasses a wide range of market design problems. They propose the following partial order over non-SP mechanisms: mechanism ψ is said to be more manipulable than mechanism φ if, for any problem instance where φ is manipulable by at least one agent, so too is ψ. This concept helps to explain several recent policy decisions in which school authorities in Chicago and England switched from one non-SP mechanism to another. This concept also yields an alternative formalization of Milton Friedman's argument for uniform-price auctions over pay-as-bid auctions: whereas we show that uniform-price auctions are SP-L and pay-as-bid auctions are not, Pathak and Sönmez (2013) show that the pay-as-bid auction is more manipulable than the uniform-price auction according to their partial order. We view our approach as complementary to these alternative approaches. Two important advantages of our approach are that it yields the classification of non-SP mechanisms as displayed in Table 1, and yields an explicit second-best criterion for designing new mechanisms, namely that they be SP-L.

9 Conclusion

A potential interpretation of our results is that they suggest that SP-L be viewed as a necessary condition for good design in large anonymous and semi-anonymous settings. Our criterion provides a common language for criticism of mechanisms ranging from Friedman's (1960) criticism of pay-as-bid auctions, to Roth's (1990; 1991) criticism of priority-matching mechanisms, to Abdulkadiroğlu and Sönmez's (2003) criticism of the Boston mechanism for school choice. The issue is not simply that these mechanisms are manipulable, but that they are manipulable even in large markets; even the kinds of agents we think of as "price takers" will want to misreport their preferences. The evidence we review in Section 4 suggests that manipulability in the large is a costly problem in practice, whereas the record for SP-L mechanisms, though incomplete, is positive. Our result in Section 5 then indicates that manipulability in the large can be avoided at approximately zero cost. Together, these results suggest that using a mechanism that is manipulable in the large is a preventable design mistake.

Whether SP-L can also be viewed as sufficient depends upon the extent to which the large-market abstraction is compelling in the problem of interest. Unfortunately, even with convergence rates such as those stated in Theorem 1, there rarely is a simple bright-line answer to the question of "how large is large".[27] But – just as economists in other fields instinctively understand that there are some contexts where it is necessary to explicitly model strategic interactions, and other contexts where it may be reasonable to assume price-taking behavior – we hope that market designers will pause to consider whether it is necessary to restrict attention to SP mechanisms, or whether SP-L may be sufficient for the problem at hand.

[27]Even in theoretical analyses of the convergence properties of specific mechanisms, rarely is the analysis sufficient to answer the question of, e.g., "is 1000 participants large?" Convergence is often slow or includes a large constant term. A notable exception is double auctions. For instance, Rustichini et al. (1994) are able to show, in a double auction with unit demand and uniformly distributed values, that 6 buyers and sellers is large enough to approximate efficiency to within one percent. Of course, in any specific context, the analyst's case that the market is large can be strengthened with empirical or computational evidence; see, for instance, Roth and Peranson (1999).

References

Abdulkadiroğlu, Atila and Tayfun Sönmez, "School Choice: A Mechanism Design Approach," *The American Economic Review*, 2003, *93* (3), 729–747.

_ , Nikhil Agarwal, and Parag A. Pathak, "The Welfare Effects of Coordinated Assignment: Evidence from the NYC HS Match," *NBER Working Paper No 21046*, 2016.

_ , Parag A. Pathak, Alvin E. Roth, and Tayfun Sönmez, "Changing the Boston School Choice Mechanism," *NBER Working Paper No 11965*, 2006.

_ , _ , and _ , "Strategy-Proofness versus Efficiency in Matching with Indifferences: Redesigning the NYC High School Match," *The American Economic Review*, 2009, *99* (5), 1954–1978.

_ , Yeon-Koo Che, and Yosuke Yasuda, "Resolving Conflicting Preferences in School Choice: The 'Boston Mechanism' Reconsidered," *The American Economic Review*, 2011, *101* (1), 399–410.

Abreu, Dilip and Arunava Sen, "Virtual Implementation in Nash Equilibrium," *Econometrica: Journal of the Econometric Society*, 1991, pp. 997–1021.

_ and Hitoshi Matsushima, "Virtual Implementation in Iteratively Undominated Strategies: Complete Information," *Econometrica*, 1992, pp. 993–1008.

Agarwal, Nikhil and Paulo Somaini, "Demand Analysis using Strategic Reports: An Application to a School Choice Mechanism," *NBER Working Paper No 20775*, 2014.

Arrow, Kenneth J, *Social choice and individual values*, Wiley, 1951.

Ashlagi, Itai, Mark Braverman, and Avinatan Hassidim, "Stability in Large Matching Markets with Complementarities," *Operations Research*, 2014, *62* (4), 713–732.

Aumann, Robert J., "Existence of Competitive Equilibria in Markets with a Continuum of Traders," *Econometrica*, 1966, *34* (1), 1–17.

Ausubel, Lawrence M. and Paul Milgrom, "The Lovely but Lonely Vickrey Auction," in Peter Cramton, Yoav Shoham, and Richard Steinberg, eds., *Combinatorial Auctions*, Cambridge, MA: The MIT Press, 2006, pp. 57–95.

_ and Peter Cramton, "Demand Reduction and Inefficiency in Multi-Unit Auctions," 2002. Mimeo, University of Maryland.

Azevedo, Eduardo M. and Eric Budish, "Strategy-proofness in the Large," *Mimeo, University of Chicago*, 2016.

Baliga, Sandeep and Rakesh Vohra, "Market Research and Market Design," *Advances in Theoretical Economics*, 2003, *3* (1).

Bergemann, Dirk and Stephen Morris, "Robust Mechanism Design," *Econometrica*, 2005, *73* (6), 1771–1813.

Bogomolnaia, Anna and Herve Moulin, "A New Solution to the Random Assignment Problem," *Journal of Economic Theory*, 2001, *100* (2), 295–328.

Brenner, Menachem, Dan Galai, and Orly Sade, "Sovereign Debt Auctions: Uniform or Discriminatory?," *Journal of Monetary Economics*, 2009, *56* (2), 267–74.

Budish, Eric, "The Combinatorial Assignment Problem: Approximate Competitive Equilibrium from Equal Incomes," *Journal of Political Economy*, 2011, *119* (6), 1061–1103.

_ **and Estelle Cantillon**, "The Multi-Unit Assignment Problem: Theory and Evidence from Course Allocation at Harvard," *The American Economic Review*, 2012, *102* (5), 2237–2271.

_ **and Judd Kessler**, "Bringing Real Market Participants' Real Preferences into the Lab: An Experiment that Changed the Course Allocation Mechanism at Wharton," *Working Paper*, 2016.

_ **, Gerard Cachon, Judd Kessler, and Abraham Othman**, "Course Match: A Large-Scale Implementation of Approximate Competitive Equilibrium from Equal Incomes for Combinatorial Allocation," *Operations Research, forthcoming*, 2015.

_ **, Yeon-Koo Che, Fuhito Kojima, and Paul Milgrom**, "Designing Random Allocation Mechanisms: Theory and Applications," *The American Economic Review*, 2013, *103* (2), 585–623.

Carroll, Gabriel, "A Quantitative Approach to Incentives: Application to Voting Rules," *Mimeo, Stanford*, 2013.

Casalmiglia, Caterina, Chao Fu, and Maia Güell, "Structural Estimation of a Model of School Choices: The Boston Mechanism vs. Its Alternatives," *Mimeo, CEMFI*, 2014.

Champsaur, Paul and Guy Laroque, "Fair Allocations in Large Economies," *Journal of Economic Theory*, 1981, *25*, 269–282.

Chen, Yan and Tayfun Sönmez, "School Choice: An Experimental Study," *Journal of Economic theory*, 2006, *127* (1), 202–231.

Chung, Kim-Sau and Jeffrey C Ely, "Foundations of Dominant-Strategy Mechanisms," *The Review of Economic Studies*, 2007, *74* (2), 447–476.

Cramton, Peter, Yoav Shoham, and Richard Steinberg (eds.), *Combinatorial Auctions*, Cambridge, MA: The MIT Press, 2006.

Cripps, Martin W. and Jeroen M. Swinkels, "Efficiency of Large Double Auctions," *Econometrica*, 2006, *74* (1), 47–92.

d'Aspremont, Claude and Bezalel Peleg, "Ordinal Bayesian Incentives Compatible Representations of Committees," *Social Choice and Welfare*, 1988, *5* (4), 261–279.

Day, Robert and Paul Milgrom, "Core-Selecting Package Auctions," *International Journal of Game Theory*, 2008, *36* (3-4), 393–407.

Debreu, Gerard and Herbert Scarf, "A Limit Theorem on the Core of an Economy," *International Economic Review*, 1963, *4* (3), 235–246.

Dur, Umut, "The modified Boston mechanism," Technical Report, Mimeo, North Carolina State University 2015.

Ehlers, Lars and Bettina Klaus, "Coalitional Strategy-Proof and Resource-Monotonic Solutions for Multiple Assignment Problems," *Social Choice and Welfare*, 2003, *21* (2), 265–280.

Erdil, Aytek and Paul Klemperer, "A New Payment Rule for Core-Selecting Package Auctions," *Journal of the European Economic Association*, 2010, *8* (2-3), 537–547.

Featherstone, Clayton and Muriel Niederle, "School Choice Mechanisms Under Incomplete Information: An Experimental Investigation," *Mimeo, Harvard Business School*, 2011.

Friedman, Milton, *A Program for Monetary Stability*, Vol. 541 of *The Miller Lectures*, New York, NY: Fordham University Press, 1960.

_ , "How to Sell Government Securities," *Wall Street Journal*, 1991, p. A8.

Fudenberg, Drew and Jean Tirole, *Game Theory*, Cambridge, MA: MIT Press, 1991.

Gershkov, Alex, Benny Moldovanu, and Xianwen Shi, "Bayesian and Dominant Strategy Implementation Revisited," 2011. Mimeo, University of Bonn.

Gibbard, Allan, "Manipulation of Voting Schemes: A General Result," *Econometrica*, 1973, pp. 587–601.

Green, Jerry and Jean-Jacques Laffont, "Characterization of Satisfactory Mechanisms for the Revelation of Preferences for Public Goods," *Econometrica*, 1977, *45* (2), 427–438.

Hammond, Peter J., "Straightforward Individual Incentive Compatibility in Large Economies," *Review of Economic Studies*, 1979, *46* (2), 263–282.

Harless, Patrick, "A School Choice Compromise: Between Immediate and Deferred Acceptance," *Mimeo, University of Rochester*, 2014.

Hassidim, Avinatan, Deborah Marciano-Romm, Assaf Romm, and Ran I. Shorrer, "'Strategic' Behavior in a Strategy-Proof Environment," *Mimeo, Havard Kennedy School*, 2015.

Hatfield, John William, "Strategy-Proof, Efficient, and Nonbossy Quota Allocations," *Social Choice and Welfare*, 2009, *33* (3), 505–515.

_ **, Fuhito Kojima, and Scott Duke Kominers**, "Strategy-Proofness, Investment Efficiency, and Marginal Returns: An Equivalence," *Mimeo, Stanford University*, 2015.

Holmström, Bengt, "Groves' Schemes on Restricted Domains," *Econometrica*, 1979, *47* (5), 1137–1144.

Hortaçsu, Ali and David McAdams, "Mechanism Choice and Strategic Bidding in Divisible Good Auctions: An Empirical Analysis of the Turkish Treasury Auction Market," *Journal of Political Economy*, 2010, *118* (5), 833–865.

_ **, Jakub Kastl, and Allen Zhang**, "Bid Shading and Bidder Surplus in U.S. Treasury Auction System," *Working Paper*, 2015.

Hurwicz, Leonid, "On Informationally Decentralized Systems," in C.B. McGuire and Roy Radner, eds., *Decision and Organization: A Volume in Honor of Jacob Marschak*, Vol. 12 of *Studies in Mathematical and Managerial Economics*, Amsterdam: North-Holland, 1972, pp. 297–336.

Hwang, Sam, "A Robust Redesign of High School Match," *Working Paper*, 2014.

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

Immorlica, Nicole and Mohammad Mahdian, "Marriage, Honesty, and Stability," in "Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms" Society for Industrial and Applied Mathematics 2005, pp. 53–62.

Jackson, Mathew O, "Bayesian Implementation," *Econometrica*, 1991, pp. 461–477.

Jackson, Matthew O, "A Crash Course in Implementation Theory," *Social choice and welfare*, 2001, *18* (4), 655–708.

Jackson, Matthew O. and Alejandro M. Manelli, "Approximately Competitive Equilibria in Large Finite Economies," *Journal of Economic Theory*, 1997, *77* (2), 354–376.

_ **and Ilan Kremer**, "Envy-freeness and Implementation in Large Economies," *Review of Economic Design*, 2007, *11* (3), 185–198.

Jegadeesh, Narasimhan, "Treasury Auction Bids and the Salomon Squeeze," *Journal of Finance*, 1993, *48* (4), 1403–1419.

Kalai, Ehud, "Large Robust Games," *Econometrica*, 2004, *72* (6), 1631–1665.

Kartik, Navin, "Strategic Communication with Lying Costs," *Review of Economic Studies*, 2009, *76(4)*, 1359–1395.

Kastl, Jakub, "Discrete Bids and Empirical Inference in Divisible Good Auctions," *Review of Economic Studies*, 2011, *78* (3), 974–1014.

Kojima, Fuhito and Mihai Manea, "Incentives in the Probabilistic Serial Mechanism," *Journal of Economic Theory*, 2010, *145* (1), 106–123.

_ **and Parag A. Pathak**, "Incentives and Stability in Large Two-Sided Matching Markets," *The American Economic Review*, 2009, *99* (3), 608–627.

_ **, Parag A Pathak, and Alvin E Roth**, "Matching with Couples: Stability and Incentives in Large Markets," *Quarterly Journal of Economics*, 2013, *128* (4), 1585–1632.

Krishna, Aradhna and Utku Ünver, "Improving the Efficiency of Course Bidding at Business Schools: Field and Laboratory Studies," *Marketing Science*, 2008, *27* (2), 262–282.

Lee, Sang Mok, "Incentive Compatibility of Large Centralized Matching Markets," *Review of Economic Studies*, January 2017, *84* (1).

Li, Shengwu, "Obviously Strategy-Proof Mechanisms," *Working Paper*, 2015.

Liu, Qingmin and Marek Pycia, "Ordinal Efficiency, Fairness, and Incentives in Large Markets," *Mimeo, UCLA*, 2011.

Majumdar, Dipjyoti and Arunava Sen, "Ordinally Bayesian Incentive Compatible Voting Rules," *Econometrica*, 2004, *72* (2), 523–540.

Malvey, Paul F. and Christine M. Archibald, "Uniform-Price Auctions: Update of the Treasury Experience," *Department of the Treasury, Office of Market Finance*, 1998.

Manelli, Alejandro M. and Daniel R. Vincent, "Bayesian and Dominant-Strategy Implementation in the Independent Private-Values Model," *Econometrica*, 2010, *78* (6), 1905–1938.

Maskin, Eric, "Nash Equilibrium and Welfare Optimality," *The Review of Economic Studies*, 1999, *66* (1), 23–38.

_ **and Tomas Sjöström**, "Implementation Theory," in A.K. Sen K.J Arrow and K. Suzumura, eds., *Handbook of social Choice and Welfare*, Elsevier, 2002, pp. 237–288.

Matsushima, Hitoshi, "A New Approach to the Implementation Problem," *Journal of Economic Theory*, 1988, *45* (1), 128–144.

Mennle, Timo and Sven Seuken, "Trade-offs in School Choice: Comparing Deferred Acceptance, the Naıve and the Adaptive Boston Mechanism," Technical Report, Mimeo, University of Zurich 2015.

Milgrom, Paul, *Putting Auction Theory to Work*, Cambridge; New York: Cambridge University Press, 2004.

_ , "Critical Issues in the Practice of Market Design," *Economic Inquiry*, 2011, *49* (2), 311–320.

Miralles, Antonio, "School Choice: The Case for the Boston Mechanism," in "Auctions, Market Mechanisms and Their Applications," Vol. 14 of *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering* AMMA Springer Berlin 2009, pp. 58–60.

Myerson, Roger B., "Incentive Compatibility and the Bargaining Problem," *Econometrica*, 1979, *47* (1), 61–73.

Palfrey, Thomas R and Sanjay Srivastava, "Implementation with Incomplete Information in Exchange Economies," *Econometrica*, 1989, pp. 115–134.

Papai, Szilvia, "Strategyproof and Nonbossy Multiple Assignments," *Journal of Public Economic Theory*, 2001, *3* (3), 257–271.

Parkes, David C., Jayant R. Kalagnanam, and Marta Eso, "Achieving Budget-Balance with Vickrey-Based Payment Schemes in Exchanges," in "Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence," Vol. 17 AAAI Press 2001, pp. 1161–1168.

Pathak, Parag A. and Tayfun Sönmez, "Leveling the Playing Field: Sincere and Sophisticated Players in the Boston Mechanism," *The American Economic Review*, 2008, *98* (4), 1636–1652.

_ **and** _ , "School Admissions Reform in Chicago and England: Comparing Mechanisms by their Vulnerability to Manipulation," *The American Economic Review*, 2013, *103* (1), 80–106.

Postlewaite, Andrew and David Schmeidler, "Equilibrium, Decentralization, and Differential Information: Implementation in Differential Information Economies," *Journal of Economic Theory*, 1986, *39* (1), 14–33.

Rees-Jones, Alex, "Suboptimal Behavior in Strategy-Proof Mechanisms: Evidence from the Residency Match," *Working Paper*, 2015.

Roberts, Donald J. and Andrew Postlewaite, "The Incentives for Price-Taking Behavior in Large Exchange Economies," *Econometrica*, 1976, *44* (1), 115–127.

Roth, Alvin E., "The Economics of Matching: Stability and Incentives," *Mathematics of Operations Research*, 1982, *7* (4), 617–628.

_ , "New Physicians: A Natural Experiment in Market Organization," *Science*, 1990, *250* (4987), 1524–1528.

_ , "A Natural Experiment in the Organization of Entry Level Labor Markets: Regional Markets for New Physicians and Surgeons in the United Kingdom," *The American Economic Review*, 1991, *81* (3), 415–440.

_ , "The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics," *Econometrica*, 2002, *70* (4), 1341–1378.

_ , "What Have We Learned from Market Design?," *The Economic Journal*, 2008, *118* (527), 285–310.

_ **and Elliott Peranson**, "The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design," *The American Economic Review*, 1999, *89* (4), 748–780.

Rustichini, Aldo, Mark A. Satterthwaite, and Steven R. Williams, "Convergence to Efficiency in a Simple Market with Incomplete Information," *Econometrica*, 1994, *62* (5), 1041–1063.

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

Segal, Ilya, "Optimal Pricing Mechanisms with Unknown Demand," *The American Economic Review*, 2003, *93* (3), 509–529.

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

Swinkels, Jeroen M., "Efficiency of Large Private Value Auctions," *Econometrica*, 2001, *69* (1), 37–68.

Wharton, "Course Match – Motivation and Description.," *Powerpoint Presentation*, 2013.

Wilson, Robert, "Game-Theoretic Analyses of Trading Processes," in Truman F. Bewley, ed., *Advances in Economic Theory: Fifth World Congress*, Cambridge, UK: Cambridge University Press, 1987, pp. 33–70.

A Appendix: Proofs

A.1 Proof of Theorem 1

We first define notation that will be used in the proof of Theorem 1. Given $\hat{\mu} \in \Delta T$, let $\Phi_i^n(t_i|\hat{\mu})$ denote the bundle $\Phi_i^n(t_i, t_{-i})$, where t_{-i} is an arbitrary vector of $n-1$ types such that $\text{emp}[t_i, t_{-i}] = \hat{\mu}$, if such t_{-i} exists.[28] If there is no such t_{-i}, which is the case for example if $\hat{\mu}(t_i) = 0$, then $\Phi_i^n(t_i|\hat{\mu})$ is defined as the random bundle placing equal weight on all outcomes in X_0. Note that bundles $\Phi_i^n(t_i|\hat{\mu})$ which do not correspond to any t_{-i} do not play any role in the results. They are defined only to simplify the notation in the proof below. Let $\Pr\{\hat{\mu}|t_i', \mu, n\}$ be the probability that the empirical distribution of (t_i', t_{-i}) is $\hat{\mu}$, given a fixed t_i' and that the vector t_{-i} of $n-1$ types is drawn i.i.d. according to μ. Throughout the proof we consider sums over infinite sets, but where only a finite number of the summands are nonzero. We adopt the convention that these are finite sums of only the positive terms.

Fix a prior $\mu \in \bar{\Delta}T$, market size n, and consider the utility a type t_i agent expects to obtain if she reports t_i'. This equals

$$u_{t_i}[\phi_i^n(t_i', \mu)] = \sum_{\hat{\mu} \in \Delta T} \Pr\{\hat{\mu}|t_i', \mu, n\} \cdot u_{t_i}[\Phi_i^n(t_i'|\hat{\mu})].$$

The interim gain from misreporting as type t_i' instead of type t_i equals

$$\begin{aligned} & u_{t_i}[\phi_i^n(t_i', \mu)] - u_{t_i}[\phi_i^n(t_i, \mu)] \\ = & \sum_{\hat{\mu} \in \Delta T} \Pr\{\hat{\mu}|t_i', \mu, n\} \cdot u_{t_i}[\Phi_i^n(t_i'|\hat{\mu})] - \sum_{\hat{\mu} \in \Delta T} \Pr\{\hat{\mu}|t_i, \mu, n\} \cdot u_{t_i}[\Phi_i^n(t_i|\hat{\mu})]. \end{aligned} \tag{A.1}$$

We can reorder the terms on the RHS of (A.1) as

$$\begin{aligned} & \underbrace{\sum_{\hat{\mu} \in \Delta T} \Pr\{\hat{\mu}|t_i, \mu, n\} \cdot \left(u_{t_i}[\Phi_i^n(t_i'|\hat{\mu})] - u_{t_i}[\Phi_i^n(t_i|\hat{\mu})]\right)}_{\text{Envy} = \text{Gain from reporting } t_i' \text{ holding fixed } \hat{\mu}} \\ & + \underbrace{\sum_{\hat{\mu} \in \Delta T} \left(\Pr\{\hat{\mu}|t_i', \mu, n\} - \Pr\{\hat{\mu}|t_i, \mu, n\}\right) \cdot u_{t_i}[\Phi_i^n(t_i'|\hat{\mu})]}_{\text{Gain from affecting } \hat{\mu}}. \end{aligned} \tag{A.2}$$

That is, the gain from misreporting can be decomposed into two terms. The first term

[28] Recall that anonymity implies that, if t_{-i} and t_{-i}' have the same empirical distribution, then $\Phi_i^n(t_i, t_{-i}) = \Phi_i^n(t_i, t_{-i}')$.

is the expected gain, over all possible empirical distributions $\hat{\mu}$, of reporting t_i' instead of t_i, holding fixed the empirical distribution of types. This quantity equals how much type t_i players envy type t_i' players, in expectation. The second term is the sum, over all possible empirical distributions $\hat{\mu}$, of how much changing the report from t_i to t_i' increases the likelihood of $\hat{\mu}$, times the utility of receiving the bundle given to a type t_i' agent. That is, how much player i gains by manipulating the expected empirical distribution of reports $\hat{\mu}$. Our goal is to show that, if a mechanism is EF or EF-TB, then both of these terms are bounded above in large markets.

The proof is based on two lemmas. The first lemma bounds the effect that a single player can have on the probability distribution of the realized empirical distribution of types. This will allow us to bound the second term in expression (A.2).

Lemma A.1. *Define, given types t_i and t_i', distribution of types $\mu \in \Delta T$, and market size n, the function*

$$\Delta P(t_i, t_i', \mu, n) = \sum_{\hat{\mu} \in \Delta T} |\Pr\{\hat{\mu}|t_i', \mu, n\} - \Pr\{\hat{\mu}|t_i, \mu, n\}|. \tag{A.3}$$

Then, for any $\mu \in \bar{\Delta} T$, and $\epsilon > 0$, there exists a constant $C_{\Delta P} > 0$ such that, for any t_i, t_i' and n we have

$$\Delta P(t_i, t_i', \mu, n) \leq C_{\Delta P} \cdot n^{-1/2+\epsilon}.$$

The second lemma will help us bound the first term in expression (A.2). Note that this term is always weakly negative for EF mechanisms, by definition, but that it can be positive for EF-TB mechanisms. The lemma provides a bound on the maximum amount of envy in an EF-TB mechanism, based on the minimum number of agents of a given type.

Lemma A.2. *Fix an EF-TB mechanism $\{(\Phi^n)_{\mathbb{N}}, T\}$. Define, given types t_i and t_i', empirical distribution of types $\hat{\mu} \in \Delta T$, and market size n, the function*

$$E(t_i, t_i', \hat{\mu}, n) = u_{t_i}[\Phi_i^n(t_i'|\hat{\mu})] - u_{t_i}[\Phi_i^n(t_i|\hat{\mu})],$$

which measures the envy of t_i for t_i'. Then, for any $\epsilon > 0$, there exists C_E such that, for all $t_i, t_i' \in T$, n, and $\hat{\mu} \in \bar{\Delta} T$ such that $\hat{\mu}$ corresponds to the empirical distribution of types for some vector in T^n, we have

$$E(t_i, t_i', \hat{\mu}, n) \leq C_E \cdot \min_{\tau \in T} \{\hat{\mu}(\tau) \cdot n\}^{-1/4+\epsilon}. \tag{A.4}$$

The proofs of Lemmas A.1 and A.2 are given below. We now use the two lemmas to prove Theorem 1

Proof of Theorem 1, Case 1: EF mechanisms. Applying the notation of Lemmas A.1 and A.2 to the terms in equation (A.2), and recalling that utility is bounded above by 1, we obtain the bound

$$u_{t_i}[\phi_i^n(t_i',\mu)] - u_{t_i}[\phi_i^n(t_i,\mu)] \leq \sum_{\hat{\mu}\in\Delta T} \Pr\{\hat{\mu}|t_i,\mu,n\}\cdot E(t_i,t_i',\hat{\mu},n) + \Delta P(t_i,t_i',\mu,n). \tag{A.5}$$

If a mechanism is EF and $\hat{\mu}(t_i') > 0$, i.e., the empirical $\hat{\mu}$ has at least one report of t_i', then the first term in the RHS of inequality (A.5) is nonpositive. Taking any $\epsilon > 0$, and using Lemma A.1 to bound the ΔP term in the RHS of inequality (A.5) we have that there exists $C_{\Delta P} > 0$ such that

$$u_{t_i}[\phi_i^n(t_i',\mu)] - u_{t_i}[\phi_i^n(t_i,\mu)] \leq \Pr\{\hat{\mu}(t_i') = 0|t_i,\mu,n\} + C_{\Delta P}\cdot n^{-1/2+\epsilon}. \tag{A.6}$$

Since the probability that $\hat{\mu}(t_i') = 0$ goes to 0 exponentially with n, we have the desired result.

□

Proof of Theorem 1, Case 2: EF-TB mechanisms. We begin by bounding the envy term in inequality (A.5), which is weakly negative for EF mechanisms but can be strictly positive in EF-TB mechanisms. We can, for any $\delta \geq 0$, decompose the envy term as

$$\sum_{\hat{\mu}\in\Delta T} \Pr\{\hat{\mu}|t_i,\mu,n\}\cdot E(t_i,t_i',\hat{\mu},n) = \sum_{\hat{\mu}:\min_\tau \hat{\mu}(\tau)\geq\mu(\tau)-\delta} \Pr\{\hat{\mu}|t_i,\mu,n\}\cdot E(t_i,t_i',\hat{\mu},n) + \sum_{\hat{\mu}:\min_\tau \hat{\mu}(\tau)<\mu(\tau)-\delta} \Pr\{\hat{\mu}|t_i,\mu,n\}\cdot E(t_i,t_i',\hat{\mu},n). \tag{A.7}$$

By Lemma A.2, for any $\epsilon > 0$ there exists a constant C_E such that

$$\sum_{\hat{\mu}:\min_\tau \hat{\mu}(\tau)\geq\mu(\tau)-\delta} \Pr\{\hat{\mu}|t_i,\mu,n\}\cdot E(t_i,t_i',\hat{\mu},n) \leq C_E\cdot \min_{\tau\in T}\{(\mu(\tau)-\delta)n\}^{-1/4+\epsilon}. \tag{A.8}$$

To bound the second term in the RHS of A.7, begin by noting that $\hat{\mu}(\tau)\cdot n$ equals the number of agents who draw type τ. This number is the outcome of $n-1$ i.i.d. draws of agents different than i, plus 1 if $t_i = \tau$. Using Hoeffding's inequality, for any τ, we can bound the probability that the realized value of $\hat{\mu}(\tau) \cdot n$ is much smaller than $\mu(\tau) \cdot n$. We have that, for any $\delta > 0$, there exists a constant $C_{\delta,\mu} > 0$ such that[29]

$$\Pr\{\hat{\mu}(\tau) \cdot n < (\mu(\tau) - \delta) \cdot n | t_i, \mu, n\} \leq C_{\delta,\mu} \cdot \exp\{-2\delta^2 n\}. \tag{A.9}$$

Take now $\delta = \min_{\tau \in T} \mu(\tau)/2$. Applying the bounds (A.8) and (A.9) to inequality (A.7), we have that

$$\begin{aligned} \sum_{\hat{\mu} \in \Delta T} \Pr\{\hat{\mu} | t_i, \mu, n\} \cdot E(t_i, t_i', \hat{\mu}, n) \quad \leq \quad & C_E \cdot \min_{\tau \in T} \{(\mu(\tau) - \delta) n\}^{-1/4+\epsilon} \\ & + |T| \cdot C_{\delta,\mu} \cdot \exp\{-2\delta^2 n\}. \end{aligned}$$

Multiplying n out of the first term in the RHS then yields

$$\begin{aligned} \sum_{\hat{\mu} \in \Delta T} \Pr\{\hat{\mu} | t_i, \mu, n\} \cdot E(t_i, t_i', \hat{\mu}, n) \quad \leq \quad & C_E \cdot \min_{\tau \in T} \{\mu(\tau) - \delta\}^{-1/4+\epsilon} \cdot n^{-1/4+\epsilon} \\ & + |T| \cdot C_{\delta,\mu} \cdot \exp\{-2\delta^2 n\}. \end{aligned}$$

Therefore, there exists a constant C' such that for all n, t_i', and t_i,

$$\sum_{\hat{\mu} \in \Delta T} \Pr\{\hat{\mu} | t_i, \mu, n\} \cdot E(t_i, t_i', \hat{\mu}, n) \leq C' \cdot n^{-1/4+\epsilon}.$$

Return now to inequality (A.5). Using the bound we just derived and Lemma A.1, we have that there exists a constant $C_{\Delta P}$ such that

$$\begin{aligned} u_{t_i}[\phi_i^n(t_i', \mu)] - u_{t_i}[\phi_i^n(t_i, \mu)] \quad \leq \quad & C' \cdot n^{-1/4+\epsilon} \\ & + C_{\Delta P} \cdot n^{-1/2+\epsilon}. \end{aligned}$$

Therefore, there exists a constant C'' such that

$$u_{t_i}[\phi_i^n(t_i', \mu)] - u_{t_i}[\phi_i^n(t_i, \mu)] \leq C'' \cdot n^{-1/4+\epsilon},$$

[29] Hoeffding's inequality states that, given n i.i.d. binomial random variables with probability of success p, and $z > 0$, the probability of having fewer than $(p - z)n$ successes is bounded above by $\exp\{-2z^2 n\}$. Note that, in the bound below, t_i is fixed, while the $n - 1$ coordinates of t_{-i} are drawn i.i.d. according to μ. For that reason, the Hoeffding bound must be modified to include a constant that depends on δ and μ, which we denote $C_{\delta,\mu}$. The reason why a constant suffices is that, conditional on δ and μ, the bound taking into account the $n - 1$ draws converges to 0 at the same rate as the bound considering n draws.

as desired.

□

A.1.1 Proof of the Lemmas

We now prove the lemmas. Throughout the proofs, we consider the case $\epsilon < 1/4$, which implies the results for $\epsilon \geq 1/4$.

Proof of Lemma A.1. To show that a single player cannot appreciably affect the distribution of $\hat{\mu}$, we start by calculating the effect of changing i's report on the probability of an individual value of $\hat{\mu}$ being drawn. Consider any $\hat{\mu}$ that is the empirical distribution of some vector of types with n agents.

Enumerate the elements of T as

$$T = \{\tau_1, \tau_2, \cdots \tau_{|T|}\}.$$

Since $\hat{\mu}$ follows a multinomial distribution, for any $t_i \in T$, the probability $\Pr\{\hat{\mu}|t_i, \mu, n\}$ equals

$$\binom{n-1}{n\hat{\mu}(\tau_1), \cdots, n\hat{\mu}(t_i) - 1, \cdots, n\hat{\mu}(\tau_{|T|})} \cdot \mu(\tau_1)^{n\hat{\mu}(\tau_1)} \cdots \mu(t_i)^{n\hat{\mu}(t_i)-1} \cdots \mu(\tau_{|T|})^{n\hat{\mu}(\tau_{|T|})},$$

where the term in parentheses is a multinomial coefficient. Note that the $n\hat{\mu}(\tau)$ terms in this expression are integers, since this is the number of agents with a given type in a realization $\hat{\mu}$ of the distribution of types. Moreover, t_i only enters the formula in one factorial term in the denominator, and a power term in the numerator. With this observation, we have that

$$\Pr\{\hat{\mu}|t_i', \mu, n\} / \Pr\{\hat{\mu}|t_i, \mu, n\} = \frac{\hat{\mu}(t_i')}{\mu(t_i')} / \frac{\hat{\mu}(t_i)}{\mu(t_i)}. \tag{A.10}$$

For the rest of the proof, we will consider separately values of $\hat{\mu}$ which are close to μ, and those that are very different from μ. We will show that player i can only have a small effect on the probability of the former, while the latter occur with very small probability.

We derive bounds as functions of a variable δ. Initially, we derive bounds valid for any $\delta > 0$, and, later in the proof, we consider the case where δ is a particular function of n. Define, for any $\delta > 0$, the set M_δ of empirical distributions $\hat{\mu}$ that are sufficiently close to the true distribution μ as

$$M_\delta = \{\hat{\mu} \in \Delta T : |\hat{\mu}(t_i) - \mu(t_i)| < \delta \text{ and } |\hat{\mu}(t_i') - \mu(t_i')| < \delta\}.$$

Note that, when $\hat{\mu}(t_i) = \mu(t_i)$ and $\hat{\mu}(t_i') = \mu(t_i')$, the ratio on the right of equation (A.10) equals 1 and is continuously differentiable in $\hat{\mu}(t_i)$ and $\hat{\mu}(t_i')$. Consequently, there exists a constant $C > 0$, and $\bar{\delta} > 0$ such that, for all $\delta \leq \bar{\delta}$, if $\hat{\mu} \in M_\delta$ then

$$|\frac{\hat{\mu}(t_i')}{\mu(t_i')} / \frac{\hat{\mu}(t_i)}{\mu(t_i)} - 1| < C\delta. \tag{A.11}$$

Moreover, we can bound the probability that the empirical distribution of types $\hat{\mu}$ is not in $M_{\delta+\frac{1}{n}}$. By Hoeffding's inequality,[30] for any $\delta > 0$ and n,

$$\begin{aligned} \Pr\{\hat{\mu} \notin M_{\delta+\frac{1}{n}} | t_i, \mu, n\} &\leq 4 \cdot \exp(-2(n-1)\delta^2) \\ \Pr\{\hat{\mu} \notin M_{\delta+\frac{1}{n}} | t_i', \mu, n\} &\leq 4 \cdot \exp(-2(n-1)\delta^2). \end{aligned} \tag{A.12}$$

We are now ready to bound ΔP. We can decompose the sum in equation (A.3) into the terms where $\hat{\mu}$ is within or outside $M_{\delta+\frac{1}{n}}$. We then have

$$\begin{aligned} \Delta P &= \sum_{\hat{\mu} \in M_{\delta+\frac{1}{n}}} |\Pr\{\hat{\mu}|t_i', \mu, n\} - \Pr\{\hat{\mu}|t_i, \mu, n\}| \\ &\quad + \sum_{\hat{\mu} \notin M_{\delta+\frac{1}{n}}} |\Pr\{\hat{\mu}|t_i', \mu, n\} - \Pr\{\hat{\mu}|t_i, \mu, n\}|. \end{aligned}$$

[30] Hoeffding's inequality yields

$$\Pr\{|\hat{\mu}(t_i) - \frac{n-1}{n}\mu(t_i) - \frac{1}{n}| > \delta | t_i, \mu, n\} < 2\exp\{-2(n-1)\delta^2\}.$$

Moreover,

$$\begin{aligned} |\hat{\mu}(t_i) - \mu(t_i)| &= |\hat{\mu}(t_i) - \frac{n-1}{n}\mu(t_i) - \frac{1}{n} + \frac{1}{n}(1-\mu(t_i))| \\ &\leq |\hat{\mu}(t_i) - \frac{n-1}{n}\mu(t_i) - \frac{1}{n}| + \frac{1}{n}|1-\mu(t_i)|. \end{aligned}$$

Hence,

$$\Pr\{|\hat{\mu}(t_i) - \mu(t_i)| > \delta + \frac{1}{n} | t_i, \mu, n\} < 2\exp\{-2(n-1)\delta^2\}.$$

By a similar argument,

$$\Pr\{|\hat{\mu}(t_i') - \mu(t_i')| > \delta + \frac{1}{n} | t_i, \mu, n\} < 2\exp\{-2(n-1)\delta^2\}.$$

Adding these two bounds implies the bound (A.12) when player i plays t_i, and the case where player i plays t_i' is analogous.

Rearranging the first term, and using the triangle inequality in the second term we have

$$\begin{aligned}\Delta P \;\leq\; & \sum_{\hat{\mu}\in M_{\delta+\frac{1}{n}}} |\Pr\{\hat{\mu}|t_i',\mu,n\}/\Pr\{\hat{\mu}|t_i,\mu,n\}-1|\cdot\Pr\{\hat{\mu}|t_i,\mu,n\} \\ & + \sum_{\hat{\mu}\notin M_{\delta+\frac{1}{n}}} (\Pr\{\hat{\mu}|t_i',\mu,n\}+\Pr\{\hat{\mu}|t_i,\mu,n\}).\end{aligned}$$

If we substitute equation (A.10) in the first term we obtain

$$\begin{aligned}\Delta P \;\leq\; & \sum_{\hat{\mu}\in M_{\delta+\frac{1}{n}}} |\frac{\hat{\mu}(t_i')}{\mu(t_i')}/\frac{\hat{\mu}(t_i)}{\mu(t_i)}-1|\cdot\Pr\{\hat{\mu}|t_i,\mu,n\} \\ & + \sum_{\hat{\mu}\notin M_{\delta+\frac{1}{n}}} (\Pr\{\hat{\mu}|t_i',\mu,n\}+\Pr\{\hat{\mu}|t_i,\mu,n\}).\end{aligned}$$

We can bound the first sum using the fact that the ratio being summed is small for $\hat{\mu}\in M_{\delta+\frac{1}{n}}$, and bound the second sum since the total probability that $\hat{\mu}\notin M_{\delta+\frac{1}{n}}$ is small. Formally, using equations (A.11) and (A.12) we have that, for all n and δ with $\delta+\frac{1}{n}\leq\bar{\delta}$,

$$\Delta P \leq C(\delta+\frac{1}{n}) + 8\cdot\exp(-2(n-1)\delta^2).$$

To complete the proof we will substitute δ by an appropriate function of n. Note that the first term is increasing in δ, while the second term is decreasing in δ. In particular, for the second term to converge to 0, asymptotically δ has to be greater than $n^{-1/2}$. If we take $\delta = n^{-1/2+\epsilon}$, we obtain the bound

$$\Delta P \leq C(n^{-1/2+\epsilon}+n^{-1}) + 8\cdot\exp(-2n^{2\epsilon}\frac{n-1}{n}), \tag{A.13}$$

for all n large enough such that $\delta+\frac{1}{n} = n^{-1/2+\epsilon}+n^{-1}\leq\bar{\delta}$. Therefore, we can take a constant C' such that

$$\Delta P \leq C'\cdot(n^{-1/2+\epsilon}+\exp(-2n^{2\epsilon}\frac{n-1}{n})) \tag{A.14}$$

for all n.

Asymptotically, the first term in the RHS of (A.14) dominates the second term.[31] There-

[31] To see this, note that the logarithm of $n^{-1/2+\epsilon}$ is $-(1/2+\epsilon)\log n$, while the logarithm of $\exp(-2n^{2\epsilon}\frac{n-1}{n})$ equals $-2n^{2\epsilon}\frac{n-1}{n}$. Since $n^{2\epsilon}\frac{n-1}{n}$ is asymptotically much larger than $\log n$, we have that the second term in equation (A.13) is asymptotically much smaller than the first.

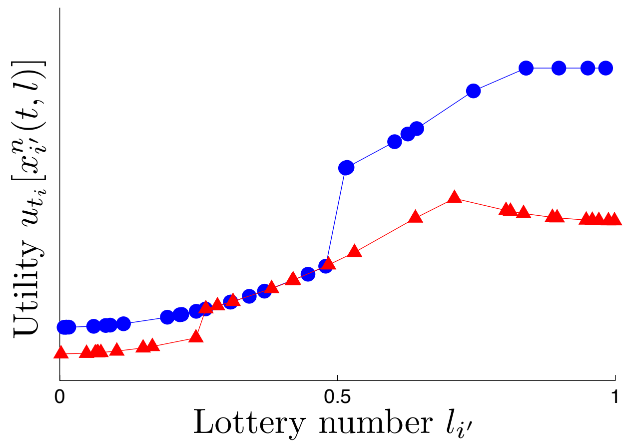

Figure A.1: A scatter plot of the lottery numbers $l_{i'}$ of different agents i' on the horizontal axis, and the utility $u_{t_i}[x^n_{i'}(t,l)]$ of type t_i agents from the bundles i' receives in the vertical axis. Balls represent agents with $t_{i'} = t_i$, and triangles agents with $t_{i'} = t_j$. The values are consistent with EF-TB, as the utilities of type t_i agents are always above the utilities from bundles of any agent with lower lottery number.

fore, we can find a constant $C_{\Delta P}$ such that

$$\Delta P \leq C_{\Delta P} \cdot n^{-1/2+\epsilon},$$

completing the proof.

□

We now prove Lemma A.2. The result would follow immediately if we restricted attention to mechanisms that are EF. The difficulty in establishing the result is that mechanisms that are EF-TB but not EF can have large amounts of envy ex-post, i.e., $u_{t_i}[\Phi^n_j(t)] - u_{t_i}[\Phi^n_i(t)]$ can be large. To see why this can be the case, fix two players i and j and consider Figure A.1. The figure plots, for several players i' whose types are either $t_{i'} = t_i$ or $t_{i'} = t_j$, lottery numbers $l_{i'}$ in the horizontal axis and the utility of a type t_i for the bundle i' receives in the vertical axis. Players with $t_{i'} = t_i$ are plotted as balls, and players with $t_{i'} = t_j$ as triangles. Note that the figure is consistent with EF-TB. In particular, if $l_j \leq l_i$, then player i prefers his own bundle to player j's bundle. However, if player j received a higher lottery number, $l_j > l_i$, it is perfectly consistent with EF-TB that player i prefers player j's bundle. That is, a player corresponding to a ball may envy a player corresponding to a triangle in the picture, as long as the triangle player has a higher lottery number. In fact, player i can envy player j by a large amount, so EF-TB mechanisms can have a lot of envy ex-post.

Figure A.1 also suggests a way to prove the lemma, despite this difficulty. The proof exploits two basic insights. First, note that the curve formed by the balls – the utility player i derives from the bundles assigned to the type t_i players – is always above the curve formed by the triangles – the utility player i derives from the bundles assigned to the type t_j players. Hence, for type t_i agents to, on average, have a large amount of ex-post envy of type t_j agents, the lottery outcome must be very uneven, favoring type t_j players over type t_i players. We can bound this average ex-post envy as a function of how well distributed lottery numbers are (see Claim A.1). Second, due to symmetry, how much player i envies player j ex-ante (i.e., before the lottery) equals how much player i prefers the bundles received by type t_j players over the bundles received by type t_i players, averaging over all type t_i and t_j players, and all possible lottery draws. Since lottery draws are likely to be very evenly distributed in a large market, it follows that player i's envy with respect to player j, before the lottery draw, is small (see Claim A.2). We now formalize these ideas.

Proof of Lemma A.2. The proof of the lemma has three steps. The first step bounds how much players of a given type envy players of another type, on average, conditional on a vector of reports t and lottery draw l, as a function of how evenly distributed the lottery numbers are. The second step bounds envy between two players, conditional on a vector of reports t, but before the lottery is drawn. Finally, the third step uses these bounds to prove the result.

Step 1. Bounding average envy after a lottery draw.

We begin by defining a measure of how evenly distributed a vector of lottery numbers is. Fix a market size n, vector of types $t \in T^n$, vector of lottery draws l and players i and j. Partition the set of players in groups according to where their lottery number falls among K uniformly-spaced intervals $L_1 = [0, 1/K)$, $L_2 = [1/K, 2/K)$, $\cdots, L_K = [(K-1)/K, 1]$. Denote the set of all type $t_{i'}$ players by

$$I(i'|t) = \{i'' : t_{i''} = t_{i'}\},$$

and denote the set of type $t_{i'}$ players with lottery numbers in L_k by

$$I_k(i'|t, l) = \{i'' \in I(i'|t) : l_{i''} \in L_k\}.$$

When there is no risk of confusion, these sets will be denoted by $I(i')$ and $I_k(i')$, respectively. The number of elements in a set of players $I(i')$ is denoted by $|I(i')|$.

Given the lottery draw l, we choose the number of partitions $K(l, t, i, j)$ such that the type t_i and type t_j players' lottery numbers are not too unevenly distributed over the L_k sets.

Specifically, let $K(l,t,i,j)$ be the largest integer K such that, for $i' = i, j$, and $k = 1, \cdots, K$, we have

$$|\frac{|I_k(i'|t,l)|}{|I(i'|t)|} - \frac{1}{K}| < \frac{1}{K^2}. \tag{A.15}$$

Such an integer necessarily exists, as $K = 1$ satisfies this condition. Intuitively, the larger is $K(l,t,i,j)$, the more evenly distributed the lottery numbers l are. When there is no risk of confusion, we write $K(l)$ or K for $K(l,t,i,j)$.

The following claim bounds the average envy of type t_i players towards type t_j players, after a lottery draw, as a function of $K(l,t,i,j)$.

Claim A.1. Fix a market size n, vector of types $t \in T^n$, lottery draws $l \in [0,1]^n$, and players i and j. Then the average envy of type t_i players towards type t_j players is bounded by

$$\sum_{j' \in I(j)} \frac{u_{t_i}[x^n_{j'}(t,l)]}{|I(j|t)|} - \sum_{i' \in I(i)} \frac{u_{t_i}[x^n_{i'}(t,l)]}{|I(i|t)|} \leq \frac{3}{K(l,t,i,j)}. \tag{A.16}$$

Proof. Denote the minimum utility received by a player with type t_i and lottery number in L_k as

$$v_k(l) = \min\{u_{t_i}[x^n_{i'}(t,l)] : i' \in I_k(i)\}.$$

Define $v_{K(l)+1}(l) = 1$. Although $v_k(l)$ and $K(l)$ depend on l, we will omit this dependence when there is no risk of confusion. Note that, by the EF-TB condition, for all $j' \in I_k(j)$,

$$u_{t_i}[x^n_{j'}(t,l)] \leq v_{k+1}. \tag{A.17}$$

Moreover, for all $i' \in I_{k+1}(i)$,

$$v_{k+1} \leq u_{t_i}[x^n_{i'}(t,l)]. \tag{A.18}$$

We now bound the average utility a type t_i agent derives from the bundles received by all players with type t_j as follows.

$$\begin{aligned} &\sum_{j' \in I(j)} \frac{u_{t_i}[x^n_{j'}(t,l)]}{|I(j)|} \\ = \; &\sum_{k=1}^{K} \sum_{j' \in I_k(j)} \frac{|I_k(j)|}{|I(j)|} \cdot \frac{u_{t_i}[x^n_{j'}(t,l)]}{|I_k(j)|} \\ \leq \; &\sum_{k=1}^{K} \frac{|I_k(j)|}{|I(j)|} \cdot v_{k+1}. \end{aligned} \tag{A.19}$$

The second line follows from breaking the sum over the K sets $I_k(j)$, and the third line follows from inequality (A.17). We now use the fact that K was chosen such that both $|I_k(i)|/|I(i)|$ and $|I_k(j)|/|I(j)|$ are approximately equal to $1/K$. Using condition (A.15) we can bound the expression above as

$$\begin{aligned}\sum_{k=1}^{K}\frac{|I_k(j)|}{|I(j)|}\cdot v_{k+1} &= \sum_{k=2}^{K}\frac{|I_k(i)|}{|I(i)|}\cdot v_k+\sum_{k=2}^{K}[\frac{|I_{k-1}(j)|}{|I(j)|}-\frac{|I_k(i)|}{|I(i)|}]\cdot v_k+\frac{|I_K(j)|}{|I(j)|}\cdot v_{K+1}\\ &\le \sum_{k=2}^{K}\frac{|I_k(i)|}{|I(i)|}\cdot v_k+(K-1)\frac{2}{K^2}+(\frac{1}{K}+\frac{1}{K^2})\\ &\le \sum_{k=2}^{K}\frac{|I_k(i)|}{|I(i)|}\cdot v_k+\frac{3}{K}.\end{aligned}$$

The equation in the first line follows from rearranging the sum. The second line follows from $v_k \le 1$, and from the fact that the fractions $I_k(i)/I(i)$ and $I_k(j)/I(j)$ are in the interval $[\frac{1}{K}-\frac{1}{K^2},\frac{1}{K}+\frac{1}{K^2}]$ as per inequality (A.15). The inequality in the third line follows from summing the second and third terms of the RHS of the second line.

We now bound the RHS of this expression using the fact that type t_i agents in the interval $I_k(i)$ receive utility of at least v_k. Using inequality (A.18) we have

$$\begin{aligned}&\sum_{k=2}^{K}\frac{|I_k(i)|}{|I(i)|}\cdot v_k+\frac{3}{K}\\ \le\ &\sum_{k=2}^{K}\sum_{i'\in I_k(i)}\frac{|I_k(i)|}{|I(i)|}\cdot\frac{u_{t_i}[x^n_{i'}(t,l)]}{|I_k(i)|}+\frac{3}{K}\\ \le\ &\sum_{k=1}^{K}\sum_{i'\in I_k(i)}\frac{|I_k(i)|}{|I(i)|}\cdot\frac{u_{t_i}[x^n_{i'}(t,l)]}{|I_k(i)|}+\frac{3}{K}.\end{aligned}$$

The first inequality follows from v_k being lower than the utility of any player in $I_k(i)$, and the second inequality follows because the latter sum equals the first plus the $k=1$ term. Since we started from inequality (A.19), the bound (A.16) follows, completing the proof.

□

Step 2: Bounding envy before the lottery draw.

We now bound the envy between two players i and j given a profile of types t, before the lottery is drawn.

Claim A.2. Given $\epsilon > 0$, there exists a constant $C_E > 0$ such that, for any $t \in T^n$ and

$i, j \leq n$, player i's envy with respect to player j is bounded by

$$u_{t_i}[\Phi^n_j(t)] - u_{t_i}[\Phi^n_i(t)] \leq C_E \cdot \min_{i'=i,j}\{|I(i|t)|\}^{-1/4+\epsilon} \tag{A.20}$$

Proof. Given a vector of types t and a player i', using anonymity, we can write the expected bundle $\Phi^n_{i'}(t)$ received by player i' as the expected bundle received by all players with the same type, over all realizations of l:

$$\Phi^n_{i'}(t) = \int_{l\in[0,1]^n} \sum_{i''\in I(i')} \frac{x^n_{i''}(t,l)}{|I(i')|}\, dl. \tag{A.21}$$

Hence, player i's envy of player j can be written as:

$$u_{t_i}[\Phi^n_j(t)] - u_{t_i}[\Phi^n_i(t)] = \int_{l\in[0,1]^n} \sum_{j'\in I(j)} \frac{u_{t_i}[x^n_{j'}(t,l)]}{|I(j|t)|} - \sum_{i'\in I(i)} \frac{u_{t_i}[x^n_{i'}(t,l)]}{|I(i|t)|}\, dl.$$

Claim A.1 then implies that envy is bounded by

$$u_{t_i}[\Phi^n_j(t)] - u_{t_i}[\Phi^n_i(t)] \leq \int_{l\in[0,1]^n} \frac{3}{K(l,t,i,j)}\, dl. \tag{A.22}$$

We need to show that, on average over all lottery realizations, $K(l)$ is large enough such that the integral above is small. Given a lottery draw l denote by $\hat{F}_{i'}(x|l)$ the fraction of agents in $I(i')$ with lottery number no greater than x. Formally,

$$\hat{F}_{i'}(x|l) = |\{i'' \in I(i') : l_{i''} \leq x\}|/|I(i')|.$$

That is, $\hat{F}_{i'}$ is the empirical distribution function of the lottery draws of type $t_{i'}$ agents. Since the lottery numbers are i.i.d., we know that the $\hat{F}_{i'}(x|l)$ functions are very likely to be close to the actual distribution of lottery draws $F(x) = x$. By the Dvoretzky–Kiefer–Wolfowitz inequality, for any $\delta > 0$,

$$\Pr\{\sup_x |\hat{F}_{i'}(x|l) - x| > \delta\} \leq 2\exp(-2|I(i')|\delta^2). \tag{A.23}$$

Fixing a partition size K, the conditions in (A.15) for the number of agents in each interval to be close to $1/K$ can be written as

$$|[\hat{F}_{i'}(\frac{k}{K}|l) - \hat{F}_{i'}(\frac{k-1}{K}|l)] - \frac{1}{K}| \leq \frac{1}{K^2},$$

for $k = 1, \ldots, K$ and $i' = i, j$. Applying the inequality (A.23), using $\delta = 1/2K^2$, we have that the probability that each such condition is violated is bounded by

$$\Pr\{|\frac{|I_k(i')|}{|I(i')|} - \frac{1}{K}| > \frac{1}{K^2}\} \leq 2 \cdot \exp(-|I(i')|/2K^4).$$

Consider now an arbitrary integer $\bar{K} > 0$. Note that the probability that $K(l) \geq \bar{K}$ is at least as large as the probability that $K = \bar{K}$ satisfies all of the conditions (A.15), since $K(l)$ by construction is the largest integer that satisfies these conditions. Therefore,

$$\begin{aligned}\Pr\{K(l) < \bar{K}\} &\leq 2\bar{K}[\exp(-|I(i)|/2\bar{K}^4) + \exp(-|I(j)|/2\bar{K}^4)] \\ &\leq 4\bar{K}\exp(-\min_{i'=i,j}\{|I(i')|\}/2\bar{K}^4).\end{aligned}$$

Using this, we can bound the integral in the right side of equation (A.22). Note that the integrand $3/K(l)$ is decreasing in $K(l)$, and attains its maximum value of 3 when $K(l) = 1$. Therefore, the integral in equation (A.22) can be bounded by

$$\begin{aligned}\int_{l\in[0,1]^n} \frac{3}{K(l,t,i,j)}\,dl &\leq \frac{3}{\bar{K}} + 3\Pr\{K(l) < \bar{K}\} \\ &\leq \frac{3}{\bar{K}} + 12\bar{K}\exp(-\min_{i'=i,j}\{|I(i')|\}/2\bar{K}^4),\end{aligned}$$

Note that the first term on the RHS is decreasing in $\bar{K}$, while the second term is increasing in $\bar{K}$. Taking $\bar{K} = \lfloor \min_{i'=i,j} |I(i')|^{1/4-\epsilon} \rfloor$, we have that this last expression is bounded by

$$\begin{aligned}&3/\min_{i'=i,j}\lfloor\{|I(i')|\}^{1/4-\epsilon}\rfloor \\ &+12\min_{i'=i,j}\{|I(i')|\}^{1/4-\epsilon}\exp\{-\min_{i'=i,j}\{|I(i')|\}^{4\epsilon}/2\}.\end{aligned}$$

Note that, as $\min_{i'=i,j}\{|I(i')|\}$ grows, the second term is asymptotically negligible compared to the first term.[32] Therefore, there exists a constant C_E such that equation (A.20)

[32]This can be shown formally by taking logs of both terms. The log of the first term equals approximately

$$\log 3 - (\frac{1}{4} - \epsilon)\log\min_{i'=i,j}\{|I(i')|\},$$

while the log of the second term equals

$$\log 12 + (\frac{1}{4} - \epsilon)\log\min_{i'=i,j}\{|I(i')|\} - \min_{i'=i,j}\{|I(i')|\}^{4\epsilon}/2.$$

As $\min_{i'=i,j}\{|I(i')|\}$ grows, the difference between the second term and the first term goes to $-\infty$, because holds, proving the claim.

□

Step 3: Completing the proof.

The lemma now follows from Claim A.2. Take $\epsilon > 0$, and consider a constant C_E as in the statement of Claim A.2. Consider t_i, t_i', $\hat{\mu}$, and n as in the statement of the lemma. Recall that, since $\hat{\mu} \in \bar{\Delta}T$, we have $\hat{\mu}(\tau) > 0$ for all $\tau \in T$. Additionally, since $\hat{\mu}$ equals the empirical distribution of some vector of types, there exists t_{-i} and j such that $\hat{\mu} = \text{emp}[t]$ and $t_j = t_i'$. Therefore, we have

$$\begin{aligned}
E(t_i, t_i', \hat{\mu}, n) &= u_{t_i}[\Phi_i^n(t_i'|\hat{\mu})] - u_{t_i}[\Phi_i^n(t_i|\hat{\mu})] \\
&= u_{t_i}[\Phi_j^n(t)] - u_{t_i}[\Phi_i^n(t)] \\
&\leq C_E \cdot \min_{i'=i,j}\{|I(i|t)|\}^{-1/4+\epsilon} \\
&\leq C_E \cdot \min_{\tau \in T}\{\hat{\mu}(\tau) \cdot n\}^{-1/4+\epsilon}.
\end{aligned}$$

The first equation is the definition of $E(t_i, t_i', \hat{\mu}, n)$. The equation in the second line follows from the way we defined t. The inequality in the third line follows from Claim A.2. The final inequality follows because $\min_{i'=i,j}\{|I(i|t)|\}$ is weakly greater than $\min_{\tau \in T}\{\hat{\mu}(\tau) \cdot n\}$.

□

A.1.2 Infinite Set of Bundles

We close this Section by highlighting that the assumption of a finite set of bundles X_0 is not necessary for Theorem 1.

Remark 1. For the proof of Theorem 1 and Lemmas A.1 and A.2, we do not have to assume X_0 finite. The proofs follow verbatim with the following assumptions. X_0 is a measurable subset of Euclidean space. Agents' utility functions over X_0 are measurable and have range $[-\infty, 1]$. The utility of reporting truthfully is at least 0. That is, for all n and $t \in T^n$,

$$u_{t_i}[\Phi_i^n(t)] \geq 0.$$

The theorem holds with otherwise arbitrary X_0 satisfying these assumptions. The added generality is important for classifying the Walrasian mechanism in Appendix C.1.4.

$\min_{i'=i,j}\{|I(i')|\}^{4\epsilon}$ grows much more quickly than $\log \min_{i'=i,j}\{|I(i')|\}$.

A.2 Proof of Theorem 2

Because $(F^n)_{n\in\mathbb{N}}$ is limit Bayes-Nash implementable, there exists a mechanism $((\Phi^n)_{n\in\mathbb{N}}, A)$ with a limit Bayes Nash equilibrium σ^* such that

$$F^n(\omega) = \Phi^n(\sigma^*(\omega))$$

for all n and vectors of n types ω in Ω^*_n. Define the direct mechanism $((\Psi^n)_{n\in\mathbb{N}}, T)$ by

$$\Psi^n(t) = \Phi^n(\sigma^*((t_1, \mathrm{emp}[t]), \ldots, (t_n, \mathrm{emp}[t]))).$$

Denote by $\psi^n(t_i, \mu)$ the bundle a participant who reports t_i expects to receive from Ψ^n if the other participants report i.i.d. according to μ.

Part 1: $((\Psi^n)_{n\in\mathbb{N}}, T)$ approximately implements $(F^n)_{n\in\mathbb{N}}$.

We must prove that, given t_i in T, μ in $\bar{\Delta}T$, and $\epsilon > 0$, there exists n_0 such that, for all $n \geq n_0$

$$\|f^n(t_i, \mu) - \psi^n(t_i, \mu)\| < \epsilon. \tag{A.24}$$

By the definition of $f^n(t_i, \mu)$ we have

$$f^n(t_i, \mu) = \sum_{t_{-i}\in T^{n-1}} \Pr\{t_{-i}|t_{-i} \sim iid(\mu)\} \cdot F^n_i((t_1, \mu), \ldots, (t_n, \mu)).$$

Likewise, by the definition of $\psi^n(t_i, \mu)$ we have

$$\begin{aligned}\psi^n(t_i, \mu) &= \sum_{t_{-i}\in T^{n-1}} \Pr\{t_{-i}|t_{-i} \sim iid(\mu)\} \cdot \Phi^n(\sigma^*((t_1, \mathrm{emp}[t]), \ldots, (t_n, \mathrm{emp}[t])))\\ &= \sum_{t_{-i}\in T^{n-1}} \Pr\{t_{-i}|t_{-i} \sim iid(\mu)\} \cdot F^n((t_1, \mathrm{emp}[t]), \ldots, (t_n, \mathrm{emp}[t])).\end{aligned}$$

Therefore, by the triangle inequality,

$$\|f^n(t_i, \mu) - \psi^n(t_i, \mu)\| \leq \sum_{t_{-i}\in T^{n-1}} \Pr\{t_{-i}|t_{-i} \sim iid(\mu)\} \cdot \Delta(t_{-i}), \tag{A.25}$$

where

$$\Delta(t_{-i}) = \|F^n_i((t_1, \mu), \ldots, (t_n, \mu)) - F^n_i((t_1, \mathrm{emp}[t]), \ldots, (t_n, \mathrm{emp}[t]))\|.$$

Moreover, because the social choice function $(F^n)_{n\in\mathbb{N}}$ is continuous, there exists a neighborhood $\mathcal{N}$ of μ and n_0 in $\mathbb{N}$ such that, for any t_{-i} with $\text{emp}[t] \in \mathcal{N}$ and $n \geq n_0$,

$$\Delta(t_{-i}) < \epsilon/2.$$

By the law of large numbers, we can take n_0 to be large enough so that the probability that $\text{emp}[t] \notin \mathcal{N}$ is lower than $\epsilon/2$.

We can decompose the difference in inequality (A.25) as

$$\begin{aligned} \|f^n(t_i,\mu) - \psi^n(t_i,\mu)\| \leq & \sum_{t_{-i}:\text{emp}[t]\in\mathcal{N}} \Pr\{t_{-i}|t_{-i} \sim iid(\mu)\} \cdot \Delta(t_{-i}) \\ & + \sum_{t_{-i}:\text{emp}[t]\notin\mathcal{N}} \Pr\{t_{-i}|t_{-i} \sim iid(\mu)\} \cdot \Delta(t_{-i}). \end{aligned}$$

Each of the terms on the right hand side is bounded above by $\epsilon/2$, which establishes inequality (A.24).

Part 2: $((\Phi^n)_{n\in\mathbb{N}}, T)$ **is SP-L.**

We must show that, for any t_i and t_i' in T, μ in $\bar{\Delta}T$, and $\epsilon > 0$, there exists n_0 such that, for all $n \geq n_0$,

$$u_{t_i}[\psi^n(t_i',\mu)] - u_{t_i}[\psi^n(t_i,\mu)] \leq \epsilon. \tag{A.26}$$

From the triangle inequality we have that

$$\begin{aligned} u_{t_i}[\psi^n(t_i',\mu)] - u_{t_i}[\psi^n(t_i,\mu)] \leq & u_{t_i}[f^n(t_i',\mu)] - u_{t_i}[f^n(t_i,\mu)] \\ & + \|f^n(t_i',\mu) - \psi^n(t_i',\mu)\| \\ & + \|f^n(t_i,\mu) - \psi^n(t_i,\mu)\|. \end{aligned}$$

By the definition of f^n and the fact that σ^* is a limit Bayes-Nash equilibrium, there exists n_0 such that, for $n \geq n_0$, the first term in the right-hand side is bounded above by $\epsilon/3$. Moreover, by step 1 of this proof, we can take n_0 such that the second and third terms are bounded above by $\epsilon/3$. This implies inequality (A.26).

A.3 Proof of Corollary 1

In this section, we denote the Boston mechanism by $((\Phi)_{n\in\mathbb{N}}, S)$. The corollary uses some facts about limit equilibria of the boston mechanism given a common identically independently distributed prior over payoff types. Let $\Sigma^*(\mu)$ denote the set of limit equilibria of the Boston mechanism given a prior μ in ΔT . Formally, denote by Σ^{**} be the set of limit

Bayes-Nash equilibria of the Boston mechanism in the type space Ω^*. Then we define

$$\Sigma^*(\mu) = \{\rho \in \mathbb{R}_+^{T\times S} : \exists \sigma^* \in \Sigma^{**} \text{ such that } \rho(s, t_i) = \sigma^*(s, (t_i, \mu)) \text{ for all } s \in S, t_i \in T\}.$$

That is, each element ρ of $\Sigma^*(\mu)$ specifies the probability $\rho(s, t_i)$ with which type t_i agents play action s in a limit equilibrium of the game with a common identically idenpendently distributed prior μ over payoff types. In other words, ρ is an equilibrium strategy profile of the Boston mechanism with set of types T and a common iid prior μ. Let $P^*(\mu)$ be the set of vectors of probability of acceptance to each school in equilibrium. We then have the following result:

Proposition 1. *The correspondence $\Sigma^*(\mu)$ is non- empty, convex-valued and continuous in $\bar{\Delta}T$. The correspondence $P^*(\mu)$ is non-empty, single-valued, and continuous in $\bar{\Delta}T$.*

The Proposition shows that, given a prior μ, the Boston mechanism may have multiple equilibria. Nevertheless, the probability of acceptance to each school is the same in any equilibrium. The intuition is that lowering the probability of acceptance to a school weakly reduces the set of students who want to point to it, and weakly increases the set of students who want to point to other schools. Therefore, an argument similar to uniqueness arguments in competitive markets with gross substitutes shows that equilibrium probabilities of acceptance are unique. Moreover, equilibrium delivers well-behaved outcomes because probabilities of acceptance vary continuously.

Before proving the Proposition, we use it to establish Corollary 1.

Proposition 1 implies that Σ^* is non-empty, lower hemi-continuous, and convex-valued. The Michael Selection Theorem implies that Σ^* has a continuous selection. Thus, there exists a limit Bayes Nash equilibrium σ^* of the Boston mechanism defined over the type space Ω^*, and moreover this equilibrium $\sigma^*(t_i, \mu)$ varies continuously with μ in $\bar{\Delta}T$. Because outcomes of the Boston mechanism vary continuously with the empirical distribution of types, the social choice function $(F^n)_{n\in\mathbb{N}}$ defined by

$$F^n(\omega) = \Phi^n(\sigma^*(\omega))$$

is continuous and limit Bayes-Nash implementable. Corollary 1 then follows from Theorem 2.

A.3.1 Proof of Proposition 1

The Boston mechanism has a limit

$$\phi^{\infty}(s,m) = \min\{\frac{q_s}{m_s}, 1\}.$$

Therefore, a strategy profile ρ^* is in $\Sigma^*(\mu)$ if and only if, for all t_i and t_i' in T,

$$u_{t_i}[\phi^{\infty}(\rho^*(t_i'), \rho^*(\mu))] \leq u_{t_i}[\phi^{\infty}(\rho^*(t_i), \rho^*(\mu))].$$

In that case, we say that ρ^* is a limit Bayes-Nash equilibrium of the Boston mechanism given μ. Given a prior μ and strategy profile ρ, denote by $\rho(\mu)$ the induced distribution over actions.

We establish the Proposition in a series of claims.

Claim A.3. The correspondence Σ^* is non-empty and upper hemi-continuous.

Proof. Payoffs

$$u_{t_i}[\phi^{\infty}(\rho(t_i), \rho(\mu))]$$

vary continuously with σ and μ. Therefore, Σ^* is non-empty and upper hemi-continuous (see Fudenberg and Tirole (1991) p. 30). □

Claim A.4. For a fixed $\mu \in \Delta T$, the probabilities of acceptance to each school are the same in any limit Bayes Nash equilibrium.

Proof. Consider an equilibrium ρ. Let the mass of students pointing to school s in this equilibrium be

$$m_s = \sum_{t_i} \rho(t_i)(s) \cdot \mu(t_i)$$

and let the probability of acceptance at school s be p_s. Let the vectors $p = (p_s)_{s\in S}$ and $m = (m_s)_{s\in S}$. To establish the result, consider another equilibrium ρ', with associated vectors of the mass of students pointing to each school m' and probabilities of acceptance p'. Define the set of schools for which $p_s > p_s'$ as S^+ and the set of schools for which $p_s < p_s'$ as S^-.

Consider now the types who, in the equilibrium ρ, choose a school in S^+ with positive probability. All agents with types in

$$T^+ = \{t_i \in T : \max_{s\in S^+} u_{t_i} \cdot p_s > \max_{s\notin S^+} u_{t_i} \cdot p_s\}$$

must choose a school in S^+. That is, all agents who strictly prefer some school in S^+ to any school not in S^+ must point to one of the S^+schools in equilibrium. Therefore,

$$\sum_{t_i \in T^+} \mu_{t_i} \leq \sum_{s \in S^+} m_s.$$

Consider the types who choose a school in S^+ in the equilibrium ρ'. Note that the probability of obtaining entry to any school in S^+ is strictly lower at ρ' than at ρ from how we constructed S^+. Similarly, the probability of obtaining entry to any school not in S^+ is weakly higher. Therefore, in the equilibrium ρ', only agents in T^+ possibly choose a school in S^+ with positive probability. That is,

$$\sum_{s \in S^+} m'_s \leq \sum_{t_i \in T^+} \mu_{t_i}.$$

These two inequalities then imply that

$$\sum_{s \in S^+} m'_s \leq \sum_{s \in S^+} m_s.$$

However, for any $s \in S^+$ we have

$$m_s < m'_s,$$

because $p_s > p'_s$, and because probabilities of acceptance are determined by the mass of students pointing to each school. Taken together, these equations imply that $S^+ = \emptyset$. Analogously, we can prove that $S^- = \emptyset$, so $p = p'$ as desired. □

Claim A.5. P^* is non-empty, single-valued, and continuous.

Proof. The previous claims show that P^* is non-empty and single-valued. Moreover, P^* is upper hemi-continuous, because Σ^* is upper hemi-continuous and probabilities of acceptance depend continuously on equilibrium strategies and the distribution of types. Finally, P^* is continuous because continuity is equivalent to upper hemi-continuity for single-valued and non-empty correspondences. □

Claim A.6. Σ^* is convex-valued.

Proof. Fix μ, and consider two equilibria ρ and ρ', and let $\bar{\rho}$ be a convex combination of ρ and ρ'. We must show that the strategy profile $\bar{\rho}$ is an equilibrium. By Claim A.4, the probability of acceptance to each school is the same under ρ and ρ'. Therefore, the probability of acceptance is the same under $\bar{\rho}$. Because the support of $\bar{\rho}$ is contained in the union of the supports of ρ and ρ', all types play optimally under $\bar{\rho}$. □

Claim A.7. Consider a prior $\mu_0 \in \bar{\Delta}T$, and associated equilibrium ρ_0 such that, for some t_i and s_0, we have $\rho_0(t_i)(s_0) > 0$. Then there exists a neighborhood of μ_0 such that, for all μ in this neighborhood, school s_0 is optimal for t_i given $P^*(\mu)$. That is, for any $s \in S$,

$$P^*_{s_0}(\mu) \cdot u_{t_i}(s_0) \geq P^*_s(\mu) \cdot u_{t_i}(s).$$

Proof. To reach a contradiction, assume that this is not the case for some type t_i' and school s_0. Then there exists a school s_1 and sequence of priors $(\mu_k)_{k\in\mathbb{N}}$ converging to μ_0 such that, for all k,

$$P^*_{s_0}(\mu_k) \cdot u_{t_i'}(s_0) < P^*_{s_1}(\mu_k) \cdot u_{t_i'}(s_1). \tag{A.27}$$

Denote the mass of t_i' types originally pointing to school s_0 as the strictly positive constant

$$C = \rho_0(t_i')(s_0) \cdot \mu_0(t_i').$$

Denote the relative increase in probability of acceptance at school s from prior μ_0 to prior μ_k by $r_s(\mu_k) = P^*_s(\mu_k)/P^*_s(\mu_0)$. We can assume, passing to a subsequence if necessary, that the ordering of schools according to $r_s(\mu_k)$ is the same for all k. Denote the schools where the probability of acceptance increases relatively more than at school s_0 as

$$S^+ = \{s : r_s(\mu_k) > r_{s_0}(\mu_k)\}.$$

Let ρ_k be an equilibrium associated with μ_k. The mass of students pointing to schools in S^+ under ρ_k minus the mass of students pointing to schools in S^+ under ρ_0 equals

$$\sum_{s\in S^+, t_i\in T} \rho_k(t_i)(s) \cdot \mu_k(t_i) - \sum_{s\in S^+, t_i\in T} \rho_0(t_i)(s) \cdot \mu_0(t_i).$$

This sum can be decomposed as

$$\begin{aligned} &\sum_{s\in S^+, t_i\in T} (\rho_k(t_i)(s) - \rho_0(t_i)(s)) \cdot \mu_0(t_i) \\ &+ \sum_{s\in S^+, t_i\in T} \rho_k(t_i)(s) \cdot (\mu_k(t_i) - \mu_0(t_i)). \end{aligned} \tag{A.28}$$

Students who point to schools in S^+ under ρ_0 continue to do so under ρ_k. And, because equation (A.27) holds, the mass of students who point to schools in $S \backslash S^+$ under ρ_0 but who point to schools in S^+ under ρ_k is at least C. Hence, the first term in expression (A.28) is bounded below by C. Moreover, the second term converges to 0, because μ_k converges to μ_0. Therefore, for large enough k, the mass of students pointing to schools in S^+ under ρ_k is strictly larger than the mass of students pointing to schools in S^+ under ρ_0.

This implies that there exists a school $s^+ \in S^+$ such that the mass of students pointing to s^+ is strictly greater under ρ_k than under ρ_0. And there exists a school $s^- \in S \backslash S^+$ such that the mass of students pointing to s^- is strictly smaller under ρ_k than under ρ_0. However, from the way we constructed S^+ we have that $r_{s^+}(\mu_k) > r_{s^-}(\mu_k)$, which is a contradiction. $\square$

Claim A.8. Consider a prior μ_0, and associated equilibrium ρ_0 such that, for some t_i and school s_0, the mass of students pointing to s_0 is strictly lower than its capacity:

$$\sum_{t_i \in T} \rho_0(t_i)(s_0) \cdot \mu_0(t_i) < q_{s_0}.$$

Then there exists a neighborhood of μ_0 such that, for all μ in this neighborhood, $P^*_{s_0}(\mu) = 1$.

Proof. Denote the excess supply of school s_0 as the strictly positive constant

$$C = q_{s_0} - \sum_{t_i \in T} \rho_0(t_i)(s_0) \cdot \mu_0(t_i).$$

To reach a contradiction, assume that the claim's conclusion does not hold. Then there exists a sequence of priors $(\mu_k)_{k \in \mathbb{N}}$ converging to μ_0 such that, for all k, $P^*_{s_0}(\mu_k) < 1$. Let ρ_k be an equilibrium given μ_k. The fact that the probability of acceptance at s_0 is lower than 1 under ρ_k implies that the difference between the mass of students pointing to s_0 under ρ_k and ρ_0 is bounded below by C. That is,

$$\sum_{t_i \in T} \rho_k(t_i)(s_0) \cdot \mu_k(t_i) - \sum_{t_i \in T} \rho_0(t_i)(s_0) \cdot \mu_0(t_i) > C.$$

Because μ_k converges to μ_0, this implies that, for large enough k,

$$\sum_{t_i \in T} (\rho_k(t_i)(s_0) - \rho_0(t_i)(s_0)) \cdot \mu_0(t_i) > C/2. \tag{A.29}$$

As in the previous claim's proof, denote the relative increase in the probability of acceptance at school s from prior μ_0 to prior μ_k by $r_s(\mu_k) = P^*_s(\mu_k)/P^*_s(\mu_0)$. We can assume, passing to a subsequence if necessary, that the ordering of schools according to $r_s(\mu_k)$ is the same for all k. Denote the set of schools where the relative probability of acceptance does not increase more than in s_0 by

$$S^- = \{s : r_s(\mu_k) \leq r_{s_0}(\mu_0)\} \backslash \{s_0\}.$$

All students who point to a school in $S^- \cup \{s_0\}$ under ρ_k point to schools in $S^- \cup \{s_0\}$ under ρ_0. Thus,

$$\sum_{s \in S^- \cup \{s_0\}, t_i \in T} (\rho_k(t_i)(s) - \rho_0(t_i)(s)) \cdot \mu_0(t_i) \leq 0.$$

Substituting inequality (A.29) we have that, for large enough k,

$$\sum_{s \in S^-, t_i \in T} (\rho_k(t_i)(s) - \rho_0(t_i)(s)) \cdot \mu_0(t_i) < -C/2. \tag{A.30}$$

The mass of students pointing to schools in S^- under ρ_k minus the mass of students pointing to schools in S^- under ρ_0 equals

$$\sum_{s \in S^-, t_i \in T} \rho_k(t_i)(s) \cdot \mu_k(t_i) - \sum_{s \in S^-, t_i \in T} \rho_0(t_i)(s) \cdot \mu_0(t_i).$$

This sum can be decomposed into

$$\begin{aligned} &\sum_{s \in S^-, t_i \in T} (\rho_k(t_i)(s) - \rho_0(t_i)(s)) \cdot \mu_0(t_i) \\ &+ \sum_{s \in S^-, t_i \in T} \rho_k(t_i)(s) \cdot (\mu_k(t_i) - \mu_0(t_i)). \end{aligned}$$

By inequality (A.30), for large enough k, the first term in the expression above is smaller than $-C/2$. Because the second term converges to 0, we have that, for sufficiently large k, the mass of students pointing to schools in S^- under ρ_k is strictly lower than the mass of students pointing to schools in S^- under ρ_0. Hence, for at least one school s^- in S^-, we have $r_{s^-}(\mu_k) \geq 1$. But this contradicts $r_{s^-}(\mu_k) \leq r_{s_0}(\mu_k) < 1$. □

Claim A.9. The correspondence Σ^* is lower hemi-continuous in $\bar{\Delta} T$.

Proof. To prove lower hemi-continuity, fix μ_0, an associated limit equilibrium ρ_0, and consider a sequence $(\mu_k)_{k \geq 1}$ converging to μ_0. Fix $\epsilon > 0$. We will show that there exists a sequence of equilibria $(\rho_k)_{k \geq 1}$, associated with the μ_k, which converges to a strategy profile with distance lower than ϵ to ρ_0.

Part 1: Define the candidate sequence of equilibria.

Let ρ'_k be an equilibrium associated with μ_k. Passing to a subsequence, we can assume that $(\rho'_k)_{k\geq 1}$ converges to an equilibrium ρ'_0 associated with μ_0. Define

$$\rho_k(t_i) = \rho'_k(t_i) + (1-\epsilon)\cdot[\rho_0(t_i) - \rho'_0(t_i)]\cdot\frac{\mu_0(t_i)}{\mu_k(t_i)}.$$

Note that this sequence converges to $\epsilon\cdot\rho'_0 + (1-\epsilon)\cdot\rho_0$. Hence, it converges to a point within ϵ distance from ρ_0.

Part 2: For large enough k, ρ_k is a strategy profile.

Because the sum $\sum_s \rho_k(t_i)(s) = 1$, we only have to demonstrate that every $\rho_k(t_i)(s)$ is nonnegative. To see this, note that ρ_k converges to $\epsilon\cdot\rho'_0 + (1-\epsilon)\cdot\rho_0$. Hence, if either $\rho_0(t_i)(s) > 0$ or $\rho'_0(t_i)(s) > 0$, then $\rho_k(t_i)(s) > 0$ for sufficiently large k. The remaining case is when $\rho_0(t_i)(s) = \rho'_0(t_i)(s) = 0$. In this case we have that $\rho_k(t_i)(s) = \rho'_k(t_i)(s) \geq 0$.

Part 3: For sufficiently large k, the ρ_k are equilibria.

We will begin by proving that, for sufficiently large k, the probabilities of acceptance under ρ_k equal those under ρ'_k. That is, the probabilities of acceptance under ρ_k equal $P^*(\mu_k)$. To see this, note that the mass of agents pointing to school s under ρ_k equals

$$\sum_{t_i}\rho_k(t_i)(s)\cdot\mu_k(t_i) = \sum_{t_i}\rho'_k(t_i)(s)\cdot\mu_k(t_i) + (1-\epsilon)\cdot\sum_{t_i}[\rho_0(t_i)(s) - \rho'_0(t_i)(s)]\cdot\mu_0(t_i). \quad \text{(A.31)}$$

There are two cases. The first case is when the mass of students pointing to s is strictly lower than q_s under either ρ_0 or ρ'_0. In this case, we have $P^*_s(\mu_0) = 1$, so that, in the mass of students pointing to s is at most equal to q_s under both ρ'_0 and ρ_0. The mass of students pointing to school s under ρ_k converges to

$$\epsilon\cdot(\sum_{t_i\in T}\rho'_0(t_i)(s)) + (1-\epsilon)\cdot(\sum_{t_i\in T}\rho_0(t_i)(s)).$$

That is, to an average of the mass of students pointing to s under ρ'_0 and ρ_0. Because both quantities are weakly smaller than q_s, and at least one of them is strictly lower than q_s, this average is strictly lower than q_s. Thus, for large enough k, the probability of acceptance to s under ρ_k is 1. This is equal to the probability of acceptance under ρ'_k, by Claim A.8.

The second case is when the mass of students pointing to school s is at least equal to q_s both under ρ_0 and under ρ'_0. If this is the case, then the mass of students pointing to school s is the same under ρ_0 and under ρ'_0, because probabilities of acceptance are the same in any equilibrium under μ_0. Therefore, the sum

$$\sum_{t_i} [\rho_0(t_i)(s) - \rho_0'(t_i)(s)] \cdot \mu_0(t_i) = 0.$$

Substituting this in Equation (A.31), we have that the probabilities of acceptance under ρ_k and ρ_k' are equal, as desired.

To complete the proof we show that, for large enough k, the strategies ρ_k are optimal given $P^*(\mu_k)$. Consider a school s with $\rho_k(t_i)(s) > 0$. Therefore, either $\rho_k'(t_i)(s) > 0$ or $\rho_0(t_i)(s) > 0$. If $\rho_k'(t_i)(s) > 0$, then it is optimal for type t_i to point to s under $P^*(\mu_k)$, because ρ_k' is an equilibrium. Likewise, if $\rho_0(t_i)(s) > 0$, then Claim A.7 implies that, for large enough k, it is optimal for type t_i to report s under $P^*(\mu_k)$. □

The proposition then follows from Claims A.3, A.5, and A.9.