Strategy-proofness in the Large*

Eduardo M. Azevedo† and Eric Budish‡

September 2, 2016

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 both by considering incentives in a large-market limit rather than finite economies and by considering incentives from an interim perspective rather than ex-post.

*First version: October 2011. For helpful discussions 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. 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) 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; Holmstrom, 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 optimal only in the limit as the market grows large. 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 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 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 the large market limit 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 (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 classification exercise suggests that the relevant distinction for practice, in markets with a

[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	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 D 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.

large number of participants, is not "SP vs. not SP", but rather "SP-L vs. not SP-L".

The last part of the paper provides conditions under which, in large markets, SP-L is no more restrictive than Bayes-Nash incentive compatibility. The result is similar in spirit to the classic revelation principle (Myerson, 1979) (for more details on the relationship see Section 5.2). Suppose we are given a mechanism that is not SP-L but that has Bayes-Nash equilibria, for any common-knowledge i.i.d. prior beliefs about the distribution of types, and that the equilibria are continuous, in a sense made precise in the text. We then construct a mechanism that is SP-L and implements approximately the same outcomes as the original mechanism. The construction uses proxy agents. Participants report their types to our mechanism, which computes the empirical distribution of types, and then plays the original mechanism on each participant's behalf using the Bayes-Nash equilibrium strategy associated with the empirical distribution of reports. This converts a mechanism with Bayes-Nash equilibria, which depend on the prior, into an SP-L mechanism that implements the same outcome in the large-market limit. We provide a detailed application of the theorem to the Boston mechanism for school assignment. The application illustrates the construction and contributes to an ongoing debate in the market design literature.

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 new MBA course allocation mechanism implemented at the Wharton School. In course allocation, impossibility theorems for SP mechanisms led the literature to conclude that random serial dictatorship is perhaps the best possible mechanism.[4] By relaxing SP to SP-L, Budish (2011) designed a new mechanism with better efficiency and fairness properties. In 2013, Wharton adopted this new mechanism in place of a previous, non SP-L mechanism. The new mechanism's improved incentives played a key role in Wharton's adoption decision, and in the mechanism's success in the first two years of use. This is evidenced by student surveys and by administrators' reports that they value the truthful preference information generated by the mechanism (Budish et al., 2015). Notably, the administrators were concerned about how easy the old mechanism was to manipulate, but were not concerned about the fact that the new mechanism is SP-L but not SP.[5]

The rest of the paper is organized as follows. Section 2 defines the environment. Section 3 defines and discusses SP-L. Section 4 presents the classification of non-SP mechanisms. Section 5 presents the result on constructing SP-L mechanisms from Bayes-Nash mechanisms. Section 6 applies the construction to the Boston mechanism. Section 7 discusses technical extensions and related literature. Section 8 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

[4] For instance, Hatfield (2009) concludes (pg. 514) "Although unfortunate, it seems that in many of these applications, the best procedure (even if it is not considered 'fair') may well be a random serial dictatorship." For related conclusions see Papai (2001), pg. 270, and Ehlers and Klaus (2003), pg. 266.

[5] In the materials Wharton uses to train its students on how to use the new mechanism Wharton spends several slides going over why it is in students' interest to report their preferences truthfully. One excerpt which highlights the role of SP-L is: "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).

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 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 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 C 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, school choice problems in which students are grouped into different priority classes, and matching markets that are divided into two sides.

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

We begin by defining our notion of the large-market limit. 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 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*

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

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.[6]

[6] The randomness in how we take the large-market limit is in contrast with early approaches to large-

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. However, we are not aware of a mechanism used in practice, or proposed for practical use, that does not have a limit. For the remainder of the paper we restrict attention to mechanisms that have limits.

Interpretation of the Limit and Relationship with Price Taking The randomness in how we take the large-market limit is economically important for the following reason: in our limit, the distribution of the empirical distribution of play is exogenous to any particular agent's own play. We state this claim formally in the Appendix as Lemma I.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 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 the limit, 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. As a second example, in Azevedo and Leshno (2011)'s matching model the empirical distribution of reports determines a set of statistics called "cutoffs" which describe the level of desirability necessary to achieve each possible match partner. In our large-market limit, each agent regards the distribution of these price-like statistics as exogenous to their own report.

---

market analysis, such as Debreu and Scarf's (1963) replicator economy and Aumann's (1964) 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.

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 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 it is optimal for each agent to report truthfully, in the large-market limit defined in Definition 2, given any full support i.i.d. distribution of opponent reports.

Definition 4. *The mechanism $\{(\Phi^n)_{\mathbb{N}}, T\}$ is* ***strategy-proof in the large (SP-L)*** *if, for any $m \in \bar{\Delta}T$ and all $t_i, t_i' \in T$*

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

Equivalently, the mechanism is 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 $t_i, t_i' \in T$

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

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 optimal only in the limit as the market grows large. In large finite markets truthful reporting is only optimal in an approximate sense. 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 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 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 notion of asymptotic strategy-proofness, which requires that reporting truthfully is approximately optimal, in a large enough market, for any realization of opponent reports.[7] 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 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.

Finally, the definition of the limit gives a useful way to think about SP-L as a generalization of price-taking. In the large-market limit the aggregate distribution of actions depends only on the distribution of one's opponents' actions, and not on one's own action. Thus, in the limit, agents take as given any statistic of the distribution of actions. In particular, in a mechanism that uses prices that are a function of the distribution of actions, agents take the distribution of prices as given. Thus, a mechanism is SP-L if reporting truthfully is optimal taking prices as given – or, more generally, taking the aggregate distribution of play as given. A mechanism is not SP-L if even an agent who takes prices as given – or, more generally, takes the aggregate distribution of play as given – wishes to misreport.

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

[7]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).

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 D.[8]

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 types t.

[8]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 D.

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 intuition for the proof 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 types, plus the gain from affecting the aggregate distribution of types (expression (9.2) in Appendix 9). Envy-freeness directly implies that the first component in this decomposition is non-positive. Lemma I.1 then implies that the second component becomes negligible in large markets. More precisely, the effect of misreporting on the distribution of the empirical distribution of reports vanishes at a rate of essentially $\sqrt{n}$, which 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.[9] Fortunately, these mechanisms satisfy a weakening of EF that we show is also sufficient. Specifically, each of these mechanisms involves a certain form of tie-breaking lottery, and after this lottery is realized no agent envies another agent with a lower lottery number. Formally,[10]

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*

[9] Both approximate CEEI and deferred acceptance include as a special case the random serial dictatorship mechanism, which Bogomolnaia and Moulin (2001) show is not envy-free.

[10] This definition is for anonymous mechanisms. The definition for semi-anonymous mechanisms, which is needed for deferred acceptance, is contained in Supplementary Appendix C. 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).

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

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

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 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 I.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 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), and the probabilistic serial mechanism (Kojima and Manea, 2010). We also obtain new results on the approximate CEEI (Budish, 2011), the Hylland and Zeckhauser (1979), and the generalized Hylland-Zeckhauser (Budish et al., 2013) pseudomarket mechanisms, 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.[11] 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.[12]

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 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)

[11]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. 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. Budish (2011) shows that approximate CEEI satisfies exact SP in a continuum economy.

[12]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).

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

In this section we will show that, in large markets, SP-L is in a well-defined sense approximately costless to impose relative to Bayes-Nash incentive compatibility. The exception is that there can be a large cost if the Bayes-Nash mechanism is very sensitive to agents' beliefs, but this itself is likely to be undesirable in practical market design settings.

5.1 Preliminaries

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

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

and given distributions $\hat{m}, m \in \Delta A$ let

$$\phi^\infty(\hat{m}, m) = \sum_{a_i \in A} \hat{m}(a_i) \cdot \phi^\infty(a_i, m)$$

Given a mechanism $\{(\Phi^n)_{\mathbb{N}}, A\}$, a **strategy** σ is defined as a map from T to ΔA. Given a strategy σ and a vector of types t, let $\sigma(t) \in \Delta(A^n)$ denote the associated distribution over vectors of actions. Given a strategy σ and a probability distribution over types $\mu \in \Delta T$, let $\sigma(\mu) \in \Delta A$ denote the distribution over actions induced by strategy σ when player types are drawn according to μ.

Definition 7. *Given a mechanism $\{(\Phi^n)_{\mathbb{N}}, A\}$ with limit $\phi^\infty(\cdot,\cdot)$, and a probability distribution over types $\mu \in \Delta T$, the strategy $\sigma^*_\mu : T \to \Delta A$ is a **limit Bayes-Nash equilibrium at prior μ** if, for all $t_i \in T$ and $a'_i \in A$:*

$$u_{t_i}[\phi^\infty(\sigma^*_\mu(t_i), \sigma^*_\mu(\mu))] \geq u_{t_i}[\phi^\infty(a'_i, \sigma^*_\mu(\mu))].$$

Typically, a mechanism's Bayes-Nash equilibria vary with the prior. For instance, in a pay-as-bid auction how much bidders shade their bid in equilibrium varies with the distribution of bidders' values, and in the Boston mechanism how students misreport their preferences in equilibrium depends on the distribution of students' preferences. We define a family of limit equilibria as a set containing a limit equilibrium for each possible prior.[13]

Definition 8. *Given a mechanism $\{(\Phi^n)_{\mathbb{N}}, A\}$ with limit $\phi^\infty(\cdot,\cdot)$, we say that $(\sigma^*_\mu)_{\mu\in\Delta T}$ is a* ***family of limit Bayes-Nash equilibria*** *if, for each $\mu \in \Delta T$, the strategy $\sigma^*_\mu(\cdot)$ is a limit BNE at prior μ.*

Our continuity condition is defined on a family of limit equilibria.

Definition 9. *Consider a mechanism $\{(\Phi^n)_{\mathbb{N}}, A\}$ with limit $\phi^\infty(\cdot,\cdot)$, and a family of limit Bayes-Nash equilibria $(\sigma^*_\mu)_{\mu\in\Delta T}$. The family of equilibria is continuous at prior $\mu_0 \in \bar{\Delta}T$ if, given $\epsilon > 0$, there exists n_0 and a neighborhood $\mathcal{N}$ of μ_0 such that, for any $n \geq n_0$, $t_i \in T$, $t_{-i} \in T^{n-1}$ where $\text{emp}[t_i, t_{-i}] \in \mathcal{N}$, and $\mu, \mu' \in \mathcal{N}$, we have:*

$$\left\|\Phi^n_i(\sigma^*_\mu(t_i), \sigma^*_\mu(t_{-i})) - \Phi^n_i(\sigma^*_{\mu'}(t_i), \sigma^*_{\mu'}(t_{-i}))\right\| < \epsilon.$$

The family of equilibria is ***continuous*** *if it is continuous at every full support prior.*

In words, a family of equilibria is continuous if a small change in the prior μ has only a small effect on agent t_i's outcome in a large enough market. We show below in Section 6 that the Boston mechanism has a continuous family of equilibria (Proposition 1). In Section 7.1 and Supplementary Appendix B we also consider a weaker notion of continuity that allows for points of discontinuity so long as they are in a certain sense knife-edge. The theorem goes through under this condition as well but in a slightly weaker form.

5.2 Construction Theorem

We now establish that, given a mechanism with a continuous family of Bayes-Nash equilibria, there exists an SP-L mechanism that implements approximately the same outcomes as the original mechanism.

Theorem 2. *Given any mechanism $\{(\Phi^n)_{\mathbb{N}}, A\}$ with a continuous family of limit Bayes-Nash equilibria $(\sigma^*_\mu)_{\mu\in\Delta T}$, there exists a direct, SP-L mechanism $\{(F^n)_{\mathbb{N}}, T\}$ such that, in the large market limit, for any prior, truthful play of the direct mechanism produces the same*

[13] In an earlier version of this paper we showed that the analysis goes through essentially unchanged if we use a family of exact Bayes-Nash equilibria in large finite markets rather than a family of limit Bayes-Nash equilibria. Please see Appendix C.2 of Azevedo and Budish (2013).

outcomes as equilibrium play of the original mechanism. Formally, letting f^∞ denote the limit of the direct mechanism, for any full-support prior μ and any type t_i we have

$$f^\infty(t_i, \mu) = \phi^\infty(\sigma^*_\mu(t_i), \sigma^*_\mu(\mu)).$$

Proof Sketch. The proof of Theorem 2 is by construction. We provide a detailed sketch as follows, with full details contained in Appendix 9.

Suppose in a market of size n the agents report types $t = (t_1, \ldots, t_n)$. Our constructed mechanism calculates the empirical distribution of reports, $\mathrm{emp}[t]$, and then plays the limit Bayes-Nash equilibrium of the original mechanism associated with this empirical distribution:

$$F^n(t) = \Phi^n(\sigma^*_{\mathrm{emp}[t]}(t)). \tag{5.2}$$

In words, F^n plays action $\sigma^*_{\mathrm{emp}[t]}(t_i)$ for agent i who reports t_i, where $\mathrm{emp}[t]$ is not the true distribution of agents' types μ_0 (which is not known to the mechanism) but rather the *empirical* distribution of agents' reported types. Intuitively, F^n acts as a proxy agent playing the original mechanism Φ^n on each agent's behalf, and uses a strategy that would be the limit Bayes-Nash equilibrium in a world in which the empirical distribution of agents' reports were in fact the true distribution of agents' preferences, and, additionally, this distribution was common knowledge.

To prove that this constructed mechanism produces outcomes under truthful play that coincide with equilibrium play of the original mechanism, suppose that the prior is μ_0 and that all agents report truthfully to the constructed mechanism. In a finite market of size n there will be sampling error, so the realized empirical will be, say, $\hat{\mu}$. Agent i who reports t_i thus receives $F^n_i(t_i, t_{-i}) = \Phi^n_i(\sigma^*_{\hat{\mu}}(t_i), \sigma^*_{\hat{\mu}}(t_{-i}))$. As the market grows large, the realized distribution of $\hat{\mu}$ converges almost surely to the true distribution μ_0, by the law of large numbers. Hence, by continuity, agent i's allocation is converging to

$$f^\infty(t_i, \mu_0) = \phi^\infty(\sigma_{\mu_0}(t_i), \sigma_{\mu_0}(\mu_0)).$$

As required, this is exactly what agent i would receive under the original mechanism, in the large-market limit, in the Bayes-Nash equilibrium corresponding to the true prior μ_0.

To prove that the constructed mechanism is SP-L, suppose that the agents other than i misreport their preferences, according to some distribution $m \neq \mu_0$. As before, in a finite market of size n, there will be sampling error, so the realized empirical will be, say, $\hat{m}$. Agent i will thus receive $F^n_i(t_i, t'_{-i}) = \Phi^n_i(\sigma^*_{\hat{m}}(t_i), \sigma^*_{\hat{m}}(t'_{-i}))$. As the market grows large, the distribution of $\hat{m}$ will converge in probability to m, so, by continuity, agent i's allocation is converging to

$$f^{\infty}(t_i, m) = \phi^{\infty}(\sigma_m(t_i), \sigma_m(m)).$$

This is what agent i would receive under the original mechanism, in the large-market limit, *in the Bayes-Nash equilibrium corresponding to prior m.* Even though the other agents are systematically misreporting their preferences, it is optimal for agent i to tell the truth, because the other agents are acting *as if* their preferences are distributed according to m, and then playing a strategy that is converging to the Bayes-Nash equilibrium corresponding to m. Thus agent i also wants to play the Bayes-Nash equilibrium strategy corresponding to m – which is exactly what happens when she reports her preferences truthfully to the constructed mechanism.[14] Hence, in the limit, it is optimal for i to report truthfully for any distribution of opponent reports, i.e., the constructed mechanism is SP-L. □

Relationship to the Revelation Principle The construction 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 $\sigma^*_{\mu_0}(\cdot)$, and plays $\sigma^*_{\mu_0}(t_i)$ 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, and our mechanism is SP-L rather than Bayes-Nash. Moreover, our mechanism is prior free and consistent with the Wilson doctrine, unlike a traditional Bayes-Nash direct revelation mechanism.

[14] Observe that this step of the argument requires the private values assumption. It is important that i does not care per se about the other players' true types.

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 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

Denote the set of schools by $X_0 = S \cup \{\emptyset\}$. In a market of size n, there are $\lfloor q_s \cdot n \rfloor$ seats available in school 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 $\emptyset$ in excess supply. An agent of type $t_i \in T$ has a strict utility function u_{t_i} over X_0. 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. 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

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

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

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

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

6.2 Results

Let $\Sigma^*(\mu)$ denote the set of limit Bayes-Nash equilibria of the Boston mechanism given prior μ. Let $P^*(\mu)$ denote the set of vectors of probabilities of acceptance over all equilibria in $\Sigma^*(\mu)$. The next Proposition establishes existence and some regularity properties of the equilibria of the Boston mechanism.

Proposition 1 (Structure of the set of limit equilibria)**.** *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 $\bar{\Delta}T$.*

Given a prior μ, the Boston mechanism may have multiple equilibria.[15] Nevertheless, the probability of acceptance to each school is the same in any equilibrium because $P^*(\cdot)$ is single-valued. The intuition is that lowering the probability of acceptance to a school weakly reduces the set of students who would optimally point to it, and weakly increases the set of students who would 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.

Together, Proposition 1 and Theorem 2 yield the following corollary:

Corollary 1 (SP-L implementation of the Boston mechanism)**.** *The Boston mechanism has a continuous family of limit Bayes-Nash equilibria. For any such family $(\sigma_\mu^\infty)_{\mu\in\Delta T}$, 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 Bayes-Nash equilibrium play of the Boston mechanism.*

Interestingly, the SP-L mechanism that we construct according to (5.2) closely resembles the Hylland and Zeckhauser (1979) pseudo-market mechanism for single-unit assignment.[16] In our constructed mechanism, agents report their types, the mechanism computes the equilibrium market-clearing probabilities P_s^* associated with the distribution of reports, and each student 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

[15] To see why multiple equilibria are possible, consider for example an equilibrium where types t_1 and t_2 both point with probability 1/2 to each school s_1 and s_2. In such case, there are other equilibria where types t_1 and t_2 change the proportion in which they point to each school in opposite directions, keeping the probabilities of acceptance the same.

[16] 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).

mechanism computes equilibrium market-clearing prices p_s^* 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.[17] 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?

Our construction 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 our 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, 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

[17] 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.

are important issues, they are addressable. In fact the issues are similar to those dealt with in Budish and Kessler's (2015) practical implementation of an MBA 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 Relaxing Continuity

Theorem 2 assumes continuity of the given Bayes-Nash mechanism's family of equilibria. While this assumption has an intuitive appeal in that it asks that a mechanism's outcomes not be too sensitive to tiny changes in the prior, it is a strong assumption. Many 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 Supplementary Appendix B we show that a weaker version of Theorem 2 obtains under a condition that we call quasi-continuity. Quasi-continuity allows for a family of equilibria to have discontinuities, with respect to both the prior on which agents' strategies are based 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 (Theorem B.1) is as follows. If the mechanism is continuous at a given prior μ_0, then, as before, there exists an SP-L mechanism that gives agents the same outcomes as the given Bayes-Nash mechanism, in the large-market limit. If the mechanism 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 original mechanism, 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. We have counterexamples that fail quasi-continuity in which our constructed mechanism does not approximate the original Bayes-Nash mechanism, even for the weaker form of approximation described above (cf. Supplementary Appendix B.2). However, the counterexamples that we have found are far from market design applications, and also the fact that our construction leaves a gap between Bayes-Nash and SP-L only proves that our method of proof does not work, it does not prove that there is a gap. It would also be desirable to obtain results analogous to Theorems 2 and B.1 in which the continuity conditions are defined not on family of equilibria, but on mechanisms themselves.

Given these open questions, we do not see Theorems 2 and B.1 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 is 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 mechanism 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.

7.2 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 C 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 and matching markets are naturally modeled as semi-anonymous mechanisms, as are school choice mechanisms 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.

7.3 Related Literature

Our paper is related to several lines of literature. First, there is a 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. 11 for full details of the approximate incentives criteria used in this prior literature.

Second, 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 discuss 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.

Third, our paper is related to the literature on the role of strategy-proofness in practical market design. Wilson (1987) famously argued that practical market designs should aim to be robust to agents' beliefs, and Bergemann and Morris (2005) formalized the sense in which SP mechanisms are robust in the sense of Wilson. Several recent papers have argued that SP 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). 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), Carroll (2013) 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. Carroll (2013) focuses on the setting of voting rules. Like Day and Milgrom (2008), he proposes a worst-case measure of manipulability, though, like us, he considers incentives to manipulate from an ex-interim rather than ex-post perspective. He then compares voting rules based on the rate at which worst-case incentives to manipulate converge to zero. 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.

8 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 the large-market limit; 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".[18] 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.

[18]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.

_ **, Parag A. Pathak, Alvin E. Roth, and Tayfun Sönmez**, "Changing the Boston Mechanism: Strategyproofness as Equal Access," *Mimeo, Harvard University*, 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.

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

Aumann, Robert J., "Existence of Competitive Equilibria in Markets with a Continuum of Traders," *Econometrica*, 1964, pp. 1–17.

Ausubel, Lawrence M. and Paul Milgrom, "The Lovely but Lonely Vickrey Auction," *Combinatorial Auctions (Cramton et al., eds)*, 2006.

Azevedo, Eduardo and Eric Budish, "Strategyproofness in the Large," *Working Paper. First version October 2011.*, 2013.

_ **and Jacob D. Leshno**, "A Supply and Demand Framework for Two-Sided Matching Markets," *Mimeo, Harvard University*, 2011.

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 Judd Kessler**, "Changing the Course Allocation Mechanism at Wharton," *Working Paper*, 2015.

Budish, Eric B. and Estelle Cantillon, "The Multi-Unit Assignment Problem: Theory and Evidence from Course Allocation at Harvard," *American Economic Review*, 2012, *102(5)*, 2237–71.

Budish, Eric, Gerard Cachon, Judd Kessler, and Abraham Othman, "Course Match: A Large-Scale Implementation of Approximate Competitive Equilibrium from Equal Incomes for Combinatorial Allocation," *Working Paper*, 2015.

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

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.

Chen, Yan and Tayfun Sönmez, "School choice: an experimental study," *Journal of Economic theory*, 2006, *127* (1), 202–231.

Cramton, Peter, Yoav Shoham, and Richard Steinberg (eds.), *Combinatorial Auctions* 2006.

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

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

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

Ehlers, Lars and Bettina Klaus, "Coalitional Strategy-Proof and Resource-Monotonic Solutions for Multiple Assignment Problems," *Social Choice and Welfare*, 2003, *21(2)*, 265–80.

Erdil, Aytek and Paul Klemperer, "A New Payment Rule for Core-Selecting Package Auctions," *Journal of the European Economic Association*, 2010.

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

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

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

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

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

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, "Strategy-Proof, Efficient, and Nonbossy Quota Allocations," *Social Choice and Welfare*, 2009, *33(3)*, 505–15.

Hatfield, John William, Fuhito Kojima, and Scott Duke Kominers, "Strategy-Proofness, Investment Efficiency, and Marginal Returns: an Equivalence," *Mimeo, Stanford University*, 2015.

Holmstrom, Bengt, "Groves Schemes on Restricted Domains," *Econometrica*, 1979, *47*, 1137–44.

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.

_ **, 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," *Decision and Organization: A Volume in Honor of Jacob Marschak*, 1972, *12*, 297.

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

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

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

Kastl, Jakub, "Discrete Bids and Empirical Inference in Divisible Good Auctions," *Review of Economic Studies*, 2011.

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.

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.

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

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

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

Milgrom, Paul, *Putting Auction Theory to Work*, Cambridge University Press, 2004.

_ , "Critical Issues in the Practice of Market Design," *Economic Inquiry*, 2011.

Miralles, Antonio, "School Choice: The Case for the Boston Mechanism," *Auctions, Market Mechanisms and Their Applications*, 2009, pp. 58–60.

Myerson, Roger B., "Incentive Compatibility and the Bargaining Problem," *Econometrica: Journal of the Econometric Society*, 1979, pp. 61–73.

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 August 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," *American Economic Review*, 2013, *103(1)*, 80–106.

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

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

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

_ , "New Physicians: A Natural Experiment in Market Organization," *Science*, 1990, *250*, 1524–8.

_ , "A Natural Experiment in the Organization of Entry Level Labor Markets: Regional Markets for New Physicians and Surgeons in the U.K.," *American Economic Review*, 1991, *81*, 415–40.

_ , "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," *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, pp. 1041–1063.

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.

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, chapter 2, pp. 33–70.

9 Appendix: Proofs

9.1 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.[19] 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

$$\begin{aligned} & u_{t_i}[\phi_i^n(t_i', \mu)] - u_{t_i}[\phi_i^n(t_i, \mu)] \\ = & \sum_{\hat{\mu} \in \Delta T} \Pr\{\hat{\mu}|t_i', \mu, n\} \cdot u_{t_i}[\Phi_i^n(t_i'|\hat{\mu})] - \sum_{\hat{\mu} \in \Delta T} \Pr\{\hat{\mu}|t_i, \mu, n\} \cdot u_{t_i}[\Phi_i^n(t_i|\hat{\mu})]. \end{aligned} \tag{9.1}$$

We can reorder the terms on the RHS of (9.1) as

$$\begin{aligned} & \underbrace{\sum_{\hat{\mu} \in \Delta T} \Pr\{\hat{\mu}|t_i, \mu, n\} \cdot \left(u_{t_i}[\Phi_i^n(t_i'|\hat{\mu})] - u_{t_i}[\Phi_i^n(t_i|\hat{\mu})]\right)}_{\text{Envy } = \text{ Gain from reporting } t_i' \text{ holding fixed } \hat{\mu}} \\ & + \underbrace{\sum_{\hat{\mu} \in \Delta T} \left(\Pr\{\hat{\mu}|t_i', \mu, n\} - \Pr\{\hat{\mu}|t_i, \mu, n\}\right) \cdot u_{t_i}[\Phi_i^n(t_i'|\hat{\mu})]}_{\text{Gain from affecting } \hat{\mu}}. \end{aligned} \tag{9.2}$$

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

[19] 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, with the bounds converging to zero fast enough to yield the overall convergence rates stated in Theorem 1.

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 (9.2).

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

$$\Delta P(t_i, t_i', \mu, n) = \sum_{\hat{\mu} \in \Delta T} |\Pr\{\hat{\mu}|t_i', \mu, n\} - \Pr\{\hat{\mu}|t_i, \mu, n\}|. \tag{9.3}$$

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 (9.2). 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 I.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

$$E(t_i, t_i', \hat{\mu}, n) \leq C_E \cdot \min_{\tau \in T}\{\hat{\mu}(\tau) \cdot n\}^{-1/4+\epsilon}. \tag{9.4}$$

The proofs of Lemmas I.1 and I.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 I.1 and I.2 to the terms in equation (9.2), and recalling that utility is bounded above by 1, we obtain the bound

$$u_{t_i}[\phi_i^n(t_i',\mu)] - u_{t_i}[\phi_i^n(t_i,\mu)] \leq \sum_{\hat{\mu}\in\Delta T} \Pr\{\hat{\mu}|t_i,\mu,n\}\cdot E(t_i,t_i',\hat{\mu},n) + \Delta P(t_i,t_i',\mu,n). \tag{9.5}$$

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 (9.5) is nonpositive. Taking any $\epsilon > 0$, and using Lemma I.1 to bound the ΔP term in the RHS of inequality (9.5) we have that there exists $C_{\Delta P} > 0$ such that

$$u_{t_i}[\phi_i^n(t_i',\mu)] - u_{t_i}[\phi_i^n(t_i,\mu)] \leq \Pr\{\hat{\mu}(t_i') = 0|t_i,\mu,n\} + C_{\Delta P}\cdot n^{-1/2+\epsilon}. \tag{9.6}$$

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 (9.5), 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

$$\sum_{\hat{\mu}\in\Delta T} \Pr\{\hat{\mu}|t_i,\mu,n\}\cdot E(t_i,t_i',\hat{\mu},n) = \sum_{\hat{\mu}:\min_\tau \hat{\mu}(\tau)\geq\mu(\tau)-\delta} \Pr\{\hat{\mu}|t_i,\mu,n\}\cdot E(t_i,t_i',\hat{\mu},n) + \sum_{\hat{\mu}:\min_\tau \hat{\mu}(\tau)<\mu(\tau)-\delta} \Pr\{\hat{\mu}|t_i,\mu,n\}\cdot E(t_i,t_i',\hat{\mu},n). \tag{9.7}$$

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

$$\sum_{\hat{\mu}:\min_\tau \hat{\mu}(\tau)\geq\mu(\tau)-\delta} \Pr\{\hat{\mu}|t_i,\mu,n\}\cdot E(t_i,t_i',\hat{\mu},n) \leq C_E\cdot \min_{\tau\in T}\{(\mu(\tau)-\delta)n\}^{-1/4+\epsilon}. \tag{9.8}$$

To bound the second term in the RHS of 9.7, 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[20]

$$\Pr\{\hat{\mu}(\tau) \cdot n < (\mu(\tau) - \delta) \cdot n | t_i, \mu, n\} \leq C_{\delta,\mu} \cdot \exp\{-2\delta^2 n\}. \tag{9.9}$$

Take now $\delta = \min_{\tau \in T} \mu(\tau)/2$. Applying the bounds (9.8) and (9.9) to inequality (9.7), we have that

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

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

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

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 (9.5). Using the bound we just derived and Lemma I.1, we have that there exists a constant $C_{\Delta P}$ such that

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

Therefore, there exists a constant C'' such that

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

[20] 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.

□

9.1.1 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 I.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

$$\Pr\{\hat{\mu}|t_i', \mu, n\}/\Pr\{\hat{\mu}|t_i, \mu, n\} = \frac{\hat{\mu}(t_i')}{\mu(t_i')}/\frac{\hat{\mu}(t_i)}{\mu(t_i)}. \tag{9.10}$$

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 (9.10) 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

$$|\frac{\hat{\mu}(t_i')}{\mu(t_i')}/\frac{\hat{\mu}(t_i)}{\mu(t_i)} - 1| < C\delta. \tag{9.11}$$

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,[21] for any $\delta > 0$ and n,

$$\begin{aligned} \Pr\{\hat{\mu} \notin M_{\delta+\frac{1}{n}}|t_i, \mu, n\} &\leq 4 \cdot \exp(-2(n-1)\delta^2) \\ \Pr\{\hat{\mu} \notin M_{\delta+\frac{1}{n}}|t_i', \mu, n\} &\leq 4 \cdot \exp(-2(n-1)\delta^2). \end{aligned} \tag{9.12}$$

We are now ready to bound ΔP. We can decompose the sum in equation (9.3) into the terms where $\hat{\mu}$ is within or outside $M_{\delta+\frac{1}{n}}$. We then have

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

---

[21] 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 (9.12) 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

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

If we substitute equation (9.10) in the first term we obtain

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

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 (9.11) and (9.12) 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

$$\Delta P \leq C(n^{-1/2+\epsilon}+n^{-1})+8\cdot\exp(-2n^{2\epsilon}\frac{n-1}{n}), \tag{9.13}$$

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

$$\Delta P \leq C'\cdot(n^{-1/2+\epsilon}+\exp(-2n^{2\epsilon}\frac{n-1}{n})) \tag{9.14}$$

for all n.

Asymptotically, the first term in the RHS of (9.14) dominates the second term.[22] There-

[22]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 (9.13) is asymptotically much smaller than the first.

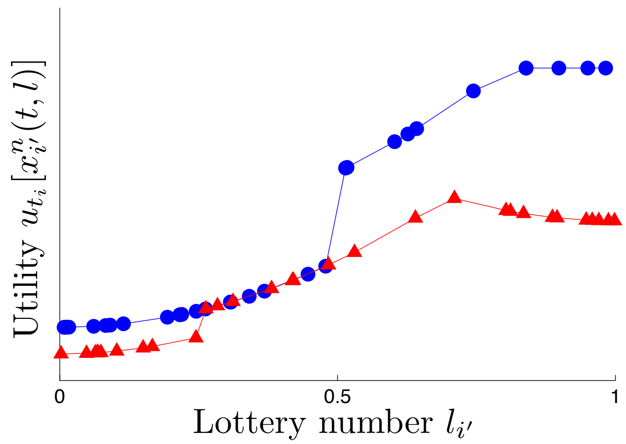

Figure I.1: A scatter plot of the lottery numbers $l_{i'}$ of different agents i' on the horizontal axis, and the utility $u_{t_i}[x^n_{i'}(t,l)]$ of type t_i agents from the bundles i' receives in the vertical axis. Balls represent agents with $t_{i'} = t_i$, and triangles agents with $t_{i'} = t_j$. 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 I.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 I.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 I.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 I.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 I.2). We now formalize these ideas.

Proof of Lemma I.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

$$\left|\frac{|I_k(i'|t,l)|}{|I(i'|t)|} - \frac{1}{K}\right| < \frac{1}{K^2}. \tag{9.15}$$

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 I.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

$$\sum_{j' \in I(j)} \frac{u_{t_i}[x^n_{j'}(t,l)]}{|I(j|t)|} - \sum_{i' \in I(i)} \frac{u_{t_i}[x^n_{i'}(t,l)]}{|I(i|t)|} \leq \frac{3}{K(l,t,i,j)}. \tag{9.16}$$

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 $j' \in I_k(j)$,

$$u_{t_i}[x^n_{j'}(t,l)] \leq v_{k+1}. \tag{9.17}$$

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

$$v_{k+1} \leq u_{t_i}[x^n_{i'}(t,l)]. \tag{9.18}$$

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.

$$\begin{aligned} & \sum_{j' \in I(j)} \frac{u_{t_i}[x^n_{j'}(t,l)]}{|I(j)|} \\ = \; & \sum_{k=1}^{K} \sum_{j' \in I_k(j)} \frac{|I_k(j)|}{|I(j)|} \cdot \frac{u_{t_i}[x^n_{j'}(t,l)]}{|I_k(j)|} \\ \leq \; & \sum_{k=1}^{K} \frac{|I_k(j)|}{|I(j)|} \cdot v_{k+1}. \end{aligned} \tag{9.19}$$

The second line follows from breaking the sum over the K sets $I_k(j)$, and the third line follows from inequality (9.17). 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 (9.15) we can bound the expression above as

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

The equation in the first line follows from rearranging the sum. The second line follows from $v_k \leq 1$, 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 (9.15). 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 (9.18) we have

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

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 (9.19), the bound (9.16) 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 I.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

$$u_{t_i}[\Phi_j^n(t)] - u_{t_i}[\Phi_i^n(t)] \leq C_E \cdot \min_{i'=i,j}\{|I(i|t)|\}^{-1/4+\epsilon} \tag{9.20}$$

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

$$\Phi_{i'}^n(t) = \int_{l\in[0,1]^n} \sum_{i''\in I(i')} \frac{x_{i''}^n(t,l)}{|I(i')|}\, dl. \tag{9.21}$$

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

$$u_{t_i}[\Phi_j^n(t)] - u_{t_i}[\Phi_i^n(t)] = \int_{l\in[0,1]^n} \sum_{j'\in I(j)} \frac{u_{t_i}[x_{j'}^n(t,l)]}{|I(j|t)|} - \sum_{i'\in I(i)} \frac{u_{t_i}[x_{i'}^n(t,l)]}{|I(i|t)|}\, dl.$$

Claim I.1 then implies that envy is bounded by

$$u_{t_i}[\Phi_j^n(t)] - u_{t_i}[\Phi_i^n(t)] \leq \int_{l\in[0,1]^n} \frac{3}{K(l,t,i,j)}\, dl. \tag{9.22}$$

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 $\delta > 0$,

$$\Pr\{\sup_x |\hat{F}_{i'}(x|l) - x| > \delta\} \leq 2\exp(-2|I(i')|\delta^2). \tag{9.23}$$

Fixing a partition size K, the conditions in (9.15) 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 (9.23), 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 (9.15), 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 (9.22). 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 (9.22) 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.[23] Therefore, there exists a constant C_E such that equation (9.20)

[23]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 I.2. Take $\epsilon > 0$, and consider a constant C_E as in the statement of Claim I.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 I.2. The final inequality follows because $\min_{i'=i,j}\{|I(i|t)|\}$ is weakly greater than $\min_{\tau \in T}\{\hat{\mu}(\tau) \cdot n\}$.

□

9.1.2 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 I.1 and I.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 D.1.4.

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

9.2 Proof of Theorem 2

This proof makes extensive use of the notation defined in Section 5. We construct the direct mechanism $\{(F^n)_{n\in\mathbb{N}}, T\}$ as in equation (5.2). We establish each part of the theorem statement in turn.

Part 1: For any type t_i and $\mu \in \bar{\Delta}T$, we have $f^\infty(t_i, \mu) = \phi^\infty(\sigma^*_\mu(t_i), \sigma^*_\mu(\mu))$.

We demonstrate the result by showing that, given any $\epsilon > 0$, there exists n_0 such that, for all $n \geq n_0$,

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

By the triangle inequality this expression is bounded by[24]

$$\sum_{t_{-i}\in T^{n-1}} \Pr\{t_{-i}|t_{-i} \sim iid(\mu)\} \cdot \|F^n_i(t_i, t_{-i}) - \Phi^n_i(\sigma^*_\mu(t_i), \sigma^*_\mu(t_{-i}))\|. \tag{9.24}$$

By the definition of continuity of a family of equilibria there exists n_0 and a neighborhood $\mathcal{N}$ of μ such that, for all $n \geq n_0$ and vector of n types t with emp$[t]$ in $\mathcal{N}$,

$$\|\Phi^n_i(\sigma^*_{\text{emp}[t]}(t)) - \Phi^n_i(\sigma^*_\mu(t))\| < \epsilon/2.$$

The left term inside the norm equals $F^n_i(t)$. Hence,

$$\|F^n_i(t) - \Phi^n_i(\sigma^*_\mu(t))\| < \epsilon/2. \tag{9.25}$$

The sum (9.24) can be broken down into

$$\sum_{\text{emp}[t_i,t_{-i}]\in\mathcal{N}} \Pr\{t_{-i}|t_{-i} \sim iid(\mu)\} \cdot \|F^n_i(t_i, t_{-i}) - \Phi^n_i(\sigma^*_\mu(t_i), \sigma^*_\mu(t_{-i}))\|$$
$$+ \sum_{\text{emp}[t_i,t_{-i}]\notin\mathcal{N}} \Pr\{t_{-i}|t_{-i} \sim iid(\mu)\} \cdot \|F^n_i(t_i, t_{-i}) - \Phi^n_i(\sigma^*_\mu(t_i), \sigma^*_\mu(t_{-i}))\|.$$

Consider $n \geq n_0$. The first sum is smaller than $\epsilon/2$ by inequality (9.25). Moreover, by the law of large numbers, we may take n_0 such that the second sum is smaller than $\epsilon/2$. Hence,

$$\|f^n(t_i, \mu) - \phi^n(\sigma^*_\mu(t_i), \sigma^*_\mu(\mu))\| < \epsilon/2 + \epsilon/2 = \epsilon,$$

as desired.

---

[24]This sum refers to the values of t_{-i} that are drawn with positive probability, as defined earlier.

Part 2: The constructed mechanism is SP-L.

For any t_i and t_i' in T and $\mu \in \bar{\Delta}T$, we have

$$u_{t_i}[f^{\infty}(t_i,\mu)] = u_{t_i}[\phi^{\infty}(\sigma_{\mu}(t_i),\sigma_{\mu}(\mu))] \geq u_{t_i}[\phi^{\infty}(\sigma_{\mu}(t_i'),\sigma_{\mu}(\mu))] = u_{t_i}[f^{\infty}(t_i',\mu)].$$

The equalities follow from part 1 and the inequality follows from the definition of a family of limit equilibria. This implies that the direct mechanism is SP-L.

9.3 Proof of Proposition 1

We establish the Proposition in a series of claims.

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

Proof. Payoffs

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

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

Claim I.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 σ. Let the mass of students pointing to school s in this equilibrium be

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

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 σ', 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 σ, 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 σ'. Note that the probability of obtaining entry to any school in S^+ is strictly lower at σ' than at σ from how we constructed S^+. Similarly, the probability of obtaining entry to any school not in S^+ is weakly higher. Therefore, in the equilibrium σ', 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. $\square$

Claim I.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. $\square$

Claim I.6. Σ^* is convex-valued.

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

Claim I.7. Consider a prior $\mu_0 \in \bar{\Delta}T$, and associated equilibrium σ_0 such that, for some t_i and s_0, we have $\sigma_0(t_i)(s_0) > 0$. 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,

$$P^*_{s_0}(\mu_k) \cdot u_{t'_i}(s_0) < P^*_{s_1}(\mu_k) \cdot u_{t'_i}(s_1). \tag{9.26}$$

Denote the mass of t'_i types originally pointing to school s_0 as the strictly positive constant

$$C = \sigma_0(t'_i)(s_0) \cdot \mu_0(t'_i).$$

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

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

Let σ_k be an equilibrium associated with μ_k. The mass of students pointing to schools in S^+ under σ_k minus the mass of students pointing to schools in S^+ under σ_0 equals

$$\sum_{s\in S^+, t_i\in T} \sigma_k(t_i)(s) \cdot \mu_k(t_i) - \sum_{s\in S^+, t_i\in T} \sigma_0(t_i)(s) \cdot \mu_0(t_i).$$

This sum can be decomposed as

$$\begin{aligned} &\sum_{s\in S^+, t_i\in T} (\sigma_k(t_i)(s) - \sigma_0(t_i)(s)) \cdot \mu_0(t_i) \\ &+ \sum_{s\in S^+, t_i\in T} \sigma_k(t_i)(s) \cdot (\mu_k(t_i) - \mu_0(t_i)). \end{aligned} \tag{9.27}$$

Students who point to schools in S^+ under σ_0 continue to do so under σ_k. And, because equation (9.26) holds, the mass of students who point to schools in $S \backslash S^+$ under σ_0 but who point to schools in S^+ under σ_k is at least C. Hence, the first term in expression (9.27) 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 σ_k is strictly larger than the mass of students pointing to schools in S^+ under σ_0.

This implies that there exists a school $s^+ \in S^+$ such that the mass of students pointing to s^+ is strictly greater under σ_k than under σ_0. And there exists a school $s^- \in S \backslash S^+$ such that the mass of students pointing to s^- is strictly smaller under σ_k than under σ_0. However, from the way we constructed S^+ we have that $\rho_{s^+}(\mu_k) > \rho_{s^-}(\mu_k)$, which is a contradiction. □

Claim I.8. Consider a prior μ_0, and associated equilibrium σ_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:

$$\sum_{t_i \in T} \sigma_0(t_i)(s_0) \cdot \mu_0(t_i) < q_{s_0}.$$

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

$$C = q_{s_0} - \sum_{t_i \in T} \sigma_0(t_i)(s_0) \cdot \mu_0(t_i).$$

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 σ_k be an equilibrium given μ_k. The fact that the probability of acceptance at s_0 is lower than 1 under σ_k implies that the difference between the mass of students pointing to s_0 under σ_k and σ_0 is bounded below by C. That is,

$$\sum_{t_i \in T} \sigma_k(t_i)(s_0) \cdot \mu_k(t_i) - \sum_{t_i \in T} \sigma_0(t_i)(s_0) \cdot \mu_0(t_i) > C.$$

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

$$\sum_{t_i \in T} (\sigma_k(t_i)(s_0) - \sigma_0(t_i)(s_0)) \cdot \mu_0(t_i) > C/2. \tag{9.28}$$

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 $\rho_s(\mu_k) = P^*_s(\mu_k)/P^*_s(\mu_0)$. We can assume, passing to a subsequence if necessary, that the ordering of schools according to $\rho_s(\mu_k)$ 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

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

All students who point to a school in $S^- \cup \{s_0\}$ under σ_k point to schools in $S^- \cup \{s_0\}$ under σ_0. Thus,

$$\sum_{s \in S^- \cup \{s_0\}, t_i \in T} (\sigma_k(t_i)(s) - \sigma_0(t_i)(s)) \cdot \mu_0(t_i) \leq 0.$$

Substituting inequality (9.28) we have that, for large enough k,

$$\sum_{s \in S^-, t_i \in T} (\sigma_k(t_i)(s) - \sigma_0(t_i)(s)) \cdot \mu_0(t_i) < -C/2. \tag{9.29}$$

The mass of students pointing to schools in S^- under σ_k minus the mass of students pointing to schools in S^- under σ_0 equals

$$\sum_{s \in S^-, t_i \in T} \sigma_k(t_i)(s) \cdot \mu_k(t_i) - \sum_{s \in S^-, t_i \in T} \sigma_0(t_i)(s) \cdot \mu_0(t_i).$$

This sum can be decomposed into

$$\begin{aligned} &\sum_{s \in S^-, t_i \in T} (\sigma_k(t_i)(s) - \sigma_0(t_i)(s)) \cdot \mu_0(t_i) \\ &+ \sum_{s \in S^-, t_i \in T} \sigma_k(t_i)(s) \cdot (\mu_k(t_i) - \mu_0(t_i)). \end{aligned}$$

By inequality (9.29), 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 σ_k is strictly lower than the mass of students pointing to schools in S^- under σ_0. Hence, for at least one school s^- in S^-, we have $\rho_{s^-}(\mu_k) \geq 1$. But this contradicts $\rho_{s^-}(\mu_k) \leq \rho_{s_0}(\mu_k) < 1$. □

Claim I.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 σ_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 $(\sigma_k)_{k \geq 1}$, associated with the μ_k, which converges to a strategy profile with distance lower than ϵ to σ_0.

Part 1: Define the candidate sequence of equilibria.

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

$$\sigma_k(t_i) = \sigma'_k(t_i) + (1-\epsilon)\cdot[\sigma_0(t_i) - \sigma'_0(t_i)]\cdot\frac{\mu_0(t_i)}{\mu_k(t_i)}.$$

Note that this sequence converges to $\epsilon\cdot\sigma'_0 + (1-\epsilon)\cdot\sigma_0$. Hence, it converges to a point within ϵ distance from σ_0.

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

Because the sum $\sum_s \sigma_k(t_i)(s) = 1$, we only have to demonstrate that every $\sigma_k(t_i)(s)$ is nonnegative. To see this, note that σ_k converges to $\epsilon\cdot\sigma'_0 + (1-\epsilon)\cdot\sigma_0$. Hence, if either $\sigma_0(t_i)(s) > 0$ or $\sigma'_0(t_i)(s) > 0$, then $\sigma_k(t_i)(s) > 0$ for sufficiently large k. The remaining case is when $\sigma_0(t_i)(s) = \sigma'_0(t_i)(s) = 0$. In this case we have that $\sigma_k(t_i)(s) = \sigma'_k(t_i)(s) \geq 0$.

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

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

$$\sum_{t_i}\sigma_k(t_i)(s)\cdot\mu_k(t_i) = \sum_{t_i}\sigma'_k(t_i)(s)\cdot\mu_k(t_i) + (1-\epsilon)\cdot\sum_{t_i}[\sigma_0(t_i)(s) - \sigma'_0(t_i)(s)]\cdot\mu_0(t_i). \quad (9.30)$$

There are two cases. The first case is when the mass of students pointing to s is strictly lower than q_s under either σ_0 or σ'_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 σ'_0 and σ_0. The mass of students pointing to school s under σ_k converges to

$$\epsilon\cdot(\sum_{t_i\in T}\sigma'_0(t_i)(s)) + (1-\epsilon)\cdot(\sum_{t_i\in T}\sigma_0(t_i)(s)).$$

That is, to an average of the mass of students pointing to s under σ'_0 and σ_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 σ_k is 1. This is equal to the probability of acceptance under σ'_k, by Claim I.8.

The second case is when the mass of students pointing to school s is at least equal to q_s both under σ_0 and under σ'_0. If this is the case, then the mass of students pointing to school s is the same under σ_0 and under σ'_0, because probabilities of acceptance are the same in any equilibrium under μ_0. Therefore, the sum

$$\sum_{t_i}[\sigma_0(t_i)(s) - \sigma_0'(t_i)(s)] \cdot \mu_0(t_i) = 0.$$

Substituting this in Equation (9.30), we have that the probabilities of acceptance under σ_k and σ_k' are equal, as desired.

To complete the proof we show that, for large enough k, the strategies σ_k are optimal given $P^*(\mu_k)$. Consider a school s with $\sigma_k(t_i)(s) > 0$. Therefore, either $\sigma_k'(t_i)(s) > 0$ or $\sigma_0(t_i)(s) > 0$. If $\sigma_k'(t_i)(s) > 0$, then it must be optimal for type t_i to point to s under $P^*(\mu_k)$, because σ_k' is an equilibrium. Likewise, if $\sigma_0(t_i)(s) > 0$, then Claim I.7 implies that, for large enough k, it is optimal for type t_i to report s under $P^*(\mu_k)$. $\square$

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

9.4 Proof of Corollary 1

By Proposition 1, Σ^* is non-empty, lower hemi-continuous, and convex-valued. The Michael Selection Theorem implies that Σ^* has a continuous selection. Because outcomes of the Boston mechanism vary continuously with the empirical distribution of types, the selection is a continuous family of limit Bayes-Nash equilibria. The corollary then follows from Theorem 1.

Supplementary Appendix to "Strategy-proofness in the Large" (for Online Publication)

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

September 2, 2016

B Relaxing Continuity to Quasi-Continuity

In this section we relax the continuity requirement in Theorem 2 to a condition we call quasi-continuity. Roughly, quasi-continuity relaxes continuity to allow for discontinuities so long as they are knife-edge. This relaxation is useful because some mechanisms are discontinuous at some knife-edge configurations. For example, the families of equilibria of pay-as-bid and uniform-price auctions described in Appendix D are not continuous, but are quasi-continuous.

Definition B.1. *Consider a mechanism $\{(\Phi^n)_{\mathbb{N}}, A\}$ with limit $\phi^\infty(\cdot,\cdot)$, and a family of limit Bayes-Nash equilibria $(\sigma^*_\mu)_{\mu\in\Delta T}$. The family of equilibria is **quasi-continuous** if, for every $\mu_0 \in \bar{\Delta}T$ and every $\epsilon > 0$, there exists a neighborhood $\mathcal{N}$ of μ_0 that can be decomposed as $\mathcal{N} = \cup_{1\le k\le K}\mathcal{A}_k \cup \mathcal{B}$ with each $\mathcal{A}_k$ open, such that:*

1. *If types are drawn i.i.d. according to μ_0, then the probability that the realized empirical distribution of types is within distance $1/n$ of $\mathcal{B}$ goes to zero as n grows large. Formally,*

$$\lim_{n\to\infty} \Pr\{\text{distance}(\text{emp}[t], \mathcal{B}) \le 1/n | t \in T^n, t \sim iid(\mu_0)\} = 0.$$

2. *Within each set $\mathcal{A}_k$, in a large enough market, agents' outcomes are continuous with respect to changes in the empirical distribution of opponents' types and the strategy that agents use. Formally, there exists n_0 such that for each $\mathcal{A}_k$, for any $n \ge n_0$, and*

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

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

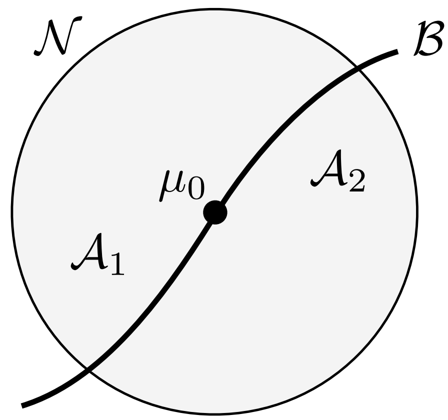

Figure B.1: The Figure illustrates the quasi-continuity definition, with $K = 2$. Around the prior $\mu_0 \in \bar{\Delta}T$, there exists a neighborhood $\mathcal{N}$ that can be decomposed into the sets $\mathcal{A}_1$ and $\mathcal{A}_2$, where equilibrium outcomes vary continuously, and a small "knife edge" set $\mathcal{B}$ where equilibrium outcomes may be discontinuous.

any $\mu, \mu', \text{emp}[t_i, t_{-i}], \text{emp}[t_i, t'_{-i}] \in \mathcal{A}_k$, *we have:*

$$\|\Phi_i^n(\sigma_\mu^*(t_i), \sigma_\mu^*(t_{-i})) - \Phi_i^n(\sigma_{\mu'}^*(t_i), \sigma_{\mu'}^*(t'_{-i}))\| < \epsilon.$$

In words, quasi-continuity allows the family of equilibria to be discontinuous at prior μ_0, but it requires that the discontinuity is knife-edge in the following sense: a small enough neighborhood $\mathcal{N}$ of μ_0 can be decomposed as a finite number of subsets $\mathcal{A}_k$ where the outcomes vary continuously, and a set $\mathcal{B}$ where the empirical distribution of a randomly drawn type profile lands with vanishingly small probability. This decomposition is illustrated in Figure B.1. Heuristically, $\mathcal{B}$ is a small discontinuity set, and is surrounded by sets $\mathcal{A}_k$ where outcomes vary continuously. Note that quasi-continuity requires that, within each region $\mathcal{A}_k$, outcomes vary continuously with both types and strategies. In contrast, continuity as defined in Definition 9 only requires that outcomes vary continuously with strategies, which is less restrictive.

Theorem B.1. *Given any mechanism* $\{(\Phi^n)_{\mathbb{N}}, A\}$ *with a quasi-continuous family of limit Bayes-Nash equilibria* $(\sigma_\mu^*)_{\mu \in \Delta T}$, *there exists a direct, SP-L mechanism* $\{(F^n)_{\mathbb{N}}, T\}$ *with the following properties.*

1. *If the original mechanism is continuous at a prior* $\mu_0 \in \bar{\Delta}T$ *then, in the limit, truthful play of the direct mechanism produces the same outcomes as equilibrium play of the*

original mechanism. Formally, for any $t_i \in T$, we have

$$f^\infty(t_i, \mu_0) = \phi^\infty(\sigma^*_{\mu_0}(t_i), \sigma^*_{\mu_0}(\mu_0)),$$

where f^∞ is the limit of the direct mechanism.

2. *For any prior $\mu_0 \in \bar{\Delta}T$, in the large market limit, truthful play of the direct mechanism produces the same outcomes as a convex combination of equilibrium play of the original mechanism under priors that are close to μ_0. Formally, for any $\epsilon > 0$, there exists n_0, an integer K, numbers π^n_k with $\sum_{k=1,\cdots,K} \pi^n_k = 1$, and priors μ_k with $\|\mu_k - \mu_0\| < \epsilon$ such that, for all $n \geq n_0$ and $t_i \in T$, we have*

$$\|f^n(t_i, \mu_0) - \sum_{k=1,\cdots,K} \pi^n_k \cdot \phi^n(\sigma^*_{\mu_k}(t_i), \sigma^*_{\mu_k}(\mu_k))\| < \epsilon,$$

where f^n is the function representing the direct mechanism from an interim perspective, as defined in equation (3.1).

Theorem B.1 says that, if the original mechanism is quasi-continuous rather than continuous, there exists an SP-L mechanism that approximates the Bayes-Nash mechanism, but in a weaker sense than in the continuous case: the SP-L mechanism approximates a convex combination of outcomes of the original Bayes-Nash mechanism, for a set of priors arbitrarily close to the prior μ_0.

B.1 Proof of Theorem B.1

Let F be defined as in equation (5.2). The proof of Theorem B.1 is based on the following approximation lemma.

Lemma B.1. *Fix a prior $\mu_0 \in \bar{\Delta}T$ and $\epsilon > 0$. Then there exists a neighborhood $\mathcal{N} = \mathcal{A}_1 \cup \cdots \cup \mathcal{A}_K \cup \mathcal{B}$, with each $\mathcal{A}_k$ open, and n_0, with the following property. For each $k = 1, \cdots, K$ we can take a prior μ_k in $\mathcal{A}_k$ with $\|\mu_k - \mu_0\| < \epsilon$ such that, for all $n \geq n_0$, there exist positive weights π^n_k with $\sum_{1\leq k\leq K} \pi^n_k = 1$, such that, for all t_i,*

$$\|f^n(t_i, \mu_0) - \sum_{k=1}^{K} \pi^n_k \cdot z_k(t_i)\| < \epsilon, \tag{B.1}$$

where

$$z_k(t_i) = \phi^\infty(\sigma^*_{\mu_k}(t_i), \sigma^*_{\mu_k}(\mu_k)). \tag{B.2}$$

The lemma states that the bundle received by an agent playing t_i in the direct mechanism can be approximated by a convex combination of the bundles type t_i receives in the limit Bayes-Nash equilibria of the original mechanism, with the elements in the convex combination corresponding to priors in a small neighborhood of μ_0, with one prior for each region $\mathcal{A}_k$.

The lemma deals with one of the key difficulties in the proof. Namely, the lemma implies that each agent can only have a small effect on the aggregate outcome of the constructed mechanism, in the sense that the weights π_k^n do not depend on agent i's report t_i. Note also that the approximation formula is not generally true without the quasi-continuity condition, as shown with an example in Appendix B.2. Before proving the lemma, we use it to prove Theorem B.1.

Proof of Theorem B.1. If the family of Bayes-Nash equilibria is continuous at a prior $\mu_0 \in \bar{\Delta}T$, then the proof of Theorem 2 implies that truthful play of f under μ_0 produces the same outcomes as equilibrium play of the original mechanism. That is, Part 1 of of the theorem statement follows from the proof of Theorem 2.

Consider now a prior $\mu_0 \in \bar{\Delta}T$, and $\epsilon > 0$. Recall that, by assumption, the family of limit Bayes-Nash equilibria is quasi-continuous, but may not be continuous. We will show that there exists n_0 such that, for all $n \geq n_0$, the gain from misreporting is lower than ϵ for any type (which proves that the direct mechanism is SP-L) and that Part 2 of the theorem statement holds.

The proof is based on the following approximation. By Lemma B.1 (using $\frac{\epsilon}{2|X_0|}$ as the constant), there exists a neighborhood $\mathcal{N} = \mathcal{A}_1 \cup \cdots \cup \mathcal{A}_K \cup \mathcal{B}$ of μ_0, priors $\mu_k \in \mathcal{N}_k$, weights π_k^n, and n_0 with the following properties. For all t_i' in T, and $n \geq n_0$,

$$\begin{aligned} \sum_{k=1}^{K} \pi_k^n &= 1, \qquad\qquad (B.3) \\ \|\mu_k - \mu_0\| &< \epsilon, \text{ and} \\ \|f^n(t_i', \mu_0) - \sum_{k=1}^{K} \pi_k^n \cdot z_k(t_i')\| &< \frac{\epsilon}{2|X_0|} \leq \frac{\epsilon}{2}, \end{aligned}$$

where z_k is given by equation (B.2).

Step 1: the gain from misreporting is no greater than ϵ.

Consider, for any pair of types t_i and t_i', the gain of a type t_i player from deviating to t_i', when opponents play i.i.d. according to μ_0. Using the approximation formula, we can bound the gain from deviating, for $n \geq n_0$, by

$$\begin{aligned} u_{t_i}[f^n(t_i', \mu_0)] - u_{t_i}[f^n(t_i, \mu_0)] &\leq \\ \sum_{k=1}^{K} \pi_k^n \cdot \{u_{t_i}[z_k(t_i')] - u_{t_i}[z_k(t_i)]\} & \\ +|u_{t_i}[f^n({t'}_i, \mu_0)] - \sum_{k=1}^{K} \pi_k^n \cdot u_{t_i}[z_k({t'}_i)]| & \\ +|u_{t_i}[f^n(t_i, \mu_0)] - \sum_{k=1}^{K} \pi_k^n \cdot u_{t_i}[z_k(t_i)]| &< \\ 0 + \epsilon/2 + \epsilon/2 &= \epsilon. \end{aligned}$$

The first inequality follows from rearranging the LHS, and then taking absolute values of the two last terms in the RHS. As for the second inequality, the first term is weakly negative by equation (B.2) and the fact that $\sigma^*_{\mu_k}$ is a limit equilibrium. The second and third terms are smaller than $\epsilon/2$ by the bounds (B.3), the fact that utility is always between 0 and 1, and that the set of random bundles has $|X_0|$ dimensions. Since $\mu_0 \in \bar{\Delta} T$ and $\epsilon > 0$ are arbitrary, we have that the constructed mechanism is SP-L.

Step 2: outcomes of the constructed mechanism $\{(F^n)_{\mathbb{N}}, T\}$ approximate a convex combination of equilibrium outcomes of $\{(\Phi^n)_{\mathbb{N}}, A\}$ under $(\sigma^*_\mu)_{\mu \in \Delta T}$ at μ_0.

By the triangle inequality we have

$$\begin{aligned} & \|f^n(t_i, \mu_0) - \sum_{k=1}^{K} \pi_k^n \cdot \phi^n(\sigma^*_{\mu_k}(t_i), \sigma^*_{\mu_k}(\mu_k))\| \\ \leq \; & \|f^n(t_i, \mu_0) - \sum_{k=1}^{K} \pi_k^n \cdot \phi^\infty(\sigma^*_{\mu_k}(t_i), \sigma^*_{\mu_k}(\mu_k))\| \\ & + \sum_{k=1}^{K} \pi_k^n \cdot \|\phi^\infty(\sigma^*_{\mu_k}(t_i), \sigma^*_{\mu_k}(\mu_k)) - \phi^n(\sigma^*_{\mu_k}(t_i), \sigma^*_{\mu_k}(\mu_k))\|. \end{aligned}$$

The first term on the RHS of the inequality is bounded by $\epsilon/2$, by the bound (B.3). By the definition of the limit, and the fact that the π_k^n sum to 1, we may take n_0 to be large enough such that the second term is also bounded by $\epsilon/2$. Moreover, this bound can be taken uniform for all $t_i \in T$. Therefore, we have that

$$\|f^n(t_i, \mu_0) - \sum_{k=1}^{K} \pi_k^n \cdot \phi^n(\sigma^*_{\mu_k}(t_i), \sigma^*_{\mu_k}(\mu_k))\| < \epsilon/2 + \epsilon/2 = \epsilon, \tag{B.4}$$

as desired.

□

B.1.1 Proof of Lemma B.1

In the proof we will use the following notation. If t' is a vector of types, and $\mathcal{S} \subseteq \Delta T$, we will say that $t' \in \mathcal{S}$ iff $\text{emp}[t'] \in \mathcal{S}$. Throughout the proof, we use the shorthand $\hat{\mu} = \text{emp}[t]$. The expression

$$\Pr(t'_{-i}|t'_{-i} \sim \mu)$$

denotes the probability that the vector of types t'_{-i} is realized if each player's type is drawn i.i.d. according to the distribution μ.

Proof of Lemma B.1. We begin by constructing the neighborhood in the statement of the lemma. By the quasi-continuity condition, we take a neighborhood $\mathcal{N} = \cup_{k=1}^{K} \mathcal{A}_k \cup \mathcal{B}$ and n_0 such that Conditions 1 and 2 of Definition B.1 are satisfied with $\epsilon/5$ in place of ϵ. We take $\mathcal{N}$ to be convex, which is without loss of generality.[1] We let the μ_k be any priors in $\mathcal{A}_k$ such that $\|\mu_k - \mu_0\| < \epsilon$. With that, to prove the lemma we have to show that there exist weights π_k^n such that the approximation formula (B.1) holds. The proof involves three steps.

Step 1: approximation of $F^n(t)$ for vectors of types with empirical distribution of types in each region $\mathcal{A}_k$.

Claim B.1. The integer n_0 can be taken such that, for all $n \geq n_0$, $k = 1, \cdots, K$, and $t \in \mathcal{A}_k$ we have

$$\|F_i^n(t) - z_k(t_i)\| < 3\epsilon/5.$$

Proof. We begin with the term $F_i^n(t)$, and derive two inequalities that together yield the desired bound. The first inequality bounds the distance between $F_i^n(t)$ and $\phi^n(\sigma^*_{\mu_k}(t_i), \sigma^*_{\mu_k}(\mu_k))$. By definition, we have that

$$\phi^n(\sigma^*_{\mu_k}(t_i), \sigma^*_{\mu_k}(\mu_k)) = \sum_{t'_{-i}} \Pr(t'_{-i}|t'_{-i} \sim \mu_k) \cdot \Phi_i^n(\sigma^*_{\mu_k}(t_i), \sigma^*_{\mu_k}(t'_{-i})). \tag{B.5}$$

[1] To see this, note that, if $\mathcal{N}$ is a non-convex neighborhood satisfying the requirements, we can take a ball $\mathcal{N}' \ni \mu_0$ contained in $\mathcal{N}$, and define the sets $\mathcal{A}'_k = \mathcal{A}_k \cap \mathcal{N}'$ and $\mathcal{B}' = \mathcal{B} \cap \mathcal{N}'$. It follows that Conditions 1 and 2 hold for the new sets, as each $\mathcal{A}'_k \subseteq \mathcal{A}$ and $\mathcal{B}' \subseteq \mathcal{B}$.

Therefore, we have that

$$
\begin{aligned}
& \|F_i^n(t) - \phi^n(\sigma^*_{\mu_k}(t_i), \sigma^*_{\mu_k}(\mu_k))\| \qquad \text{(B.6)} \\
= \; & \|F_i^n(t) - \sum_{t_{-i}} \Pr(t'_{-i}|t'_{-i} \sim \mu_k) \cdot \Phi_i^n(\sigma^*_{\mu_k}(t_i), \sigma^*_{\mu_k}(t'_{-i}))\| \\
\leq \; & \sum_{t'_{-i} \in T^{n-1}} \Pr(t'_{-i}|t'_{-i} \sim \mu_k) \cdot \|F_i^n(t) - \Phi_i^n(\sigma^*_{\mu_k}(t_i), \sigma^*_{\mu_k}(t'_{-i}))\| \\
= \; & \sum_{t'_{-i}:\mathrm{emp}[t_i, t'_{-i}] \in \mathcal{A}_k} \Pr(t'_{-i}|t'_{-i} \sim \mu_k) \cdot \|F_i^n(t) - \Phi_i^n(\sigma^*_{\mu_k}(t_i), \sigma^*_{\mu_k}(t'_{-i}))\| + \\
& \sum_{t'_{-i}:\mathrm{emp}[t_i, t'_{-i}] \notin \mathcal{A}_k} \Pr(t'_{-i}|t'_{-i} \sim \mu_k) \cdot \|F_i^n(t) - \Phi_i^n(\sigma^*_{\mu_k}(t_i), \sigma^*_{\mu_k}(t'_{-i}))\|.
\end{aligned}
$$

The first equality follows by substituting the definition of ϕ^n from equation (B.5). The inequality follows from the triangle inequality and the fact that the probabilities must sum to 1. The last equality simply breaks the sum into two parts, the t'_{-i} for which $\mathrm{emp}[t_i, t'_{-i}]$ is in $\mathcal{A}_k$, and the ones for which it is not.

Consider now the expression in the RHS of inequality (B.6). From the way we construct $F^n(\cdot)$ and using the convention that $\hat{\mu} = \mathrm{emp}[t]$, the first term equals

$$
\begin{aligned}
& \sum_{t'_{-i}:\mathrm{emp}[t_i, t'_{-i}] \in \mathcal{A}_k} \Pr(t'_{-i}|t'_{-i} \sim \mu_k) \cdot \|F_i^n(t) - \Phi_i^n(\sigma^*_{\mu_k}(t_i), \sigma^*_{\mu_k}(t'_{-i}))\| \\
= \; & \sum_{t'_{-i}:\mathrm{emp}[t_i, t'_{-i}] \in \mathcal{A}_k} \Pr(t'_{-i}|t'_{-i} \sim \mu_k) \cdot \|\Phi_i^n(\sigma^*_{\hat{\mu}}(t_i), \sigma^*_{\hat{\mu}}(t_{-i})) - \Phi_i^n(\sigma^*_{\mu_k}(t_i), \sigma^*_{\mu_k}(t'_{-i}))\|.
\end{aligned}
$$

Condition 2 in Definition B.1 implies that, for all $n \geq n_0$, this expression is bounded above by $\epsilon/5$. As for the second term in the RHS of inequality (B.6), by the weak law of large numbers, we may take n_0 large enough such that the total probability that $\mathrm{emp}[t_i, t'_{-i}] \notin \mathcal{A}_k$ is lower than $\epsilon/5$ for all $n \geq n_0$. This bounds the second term by $\epsilon/5$. Substituting these bounds in inequality (B.6) then yields

$$
\|F_i^n(t) - \phi^n(\sigma^*_{\mu_k}(t_i), \sigma^*_{\mu_k}(\mu_k))\| < \epsilon/5 + \epsilon/5 = 2\epsilon/5. \qquad \text{(B.7)}
$$

Finally, by the definition of the limit, we may take n_0 such that, for all $n \geq n_0$,

$$
\begin{aligned}
& \|\phi^n(\sigma^*_{\mu_k}(t_i), \sigma^*_{\mu_k}(\mu_k)) - z_k(t_i)\| \qquad \text{(B.8)} \\
= \; & \|\phi^n(\sigma^*_{\mu_k}(t_i), \sigma^*_{\mu_k}(\mu_k)) - \phi^\infty(\sigma^*_{\mu_k}(t_i), \sigma^*_{\mu_k}(\mu_k))\| < \epsilon/5.
\end{aligned}
$$

Note that these bounds are uniform for all $t \in \mathcal{A}_k$. Moreover, since K is finite, n_0 can be taken such that the bounds hold for all $k = 1, \cdots, K$. The claim then follows from inequalities (B.7) and (B.8), as, for $n \geq n_0$,

$$\begin{aligned} & \|F_i^n(t) - z_k(t)\| \\ \leq \; & \|F_i^n(t) - \phi^n(\sigma^*_{\mu_k}(t_i), \sigma^*_{\mu_k}(\mu_k))\| + \|\phi^n(\sigma^*_{\mu_k}(t_i), \sigma^*_{\mu_k}(\mu_k)) - z_k(t_i)\| \\ < \; & 2\epsilon/5 + \epsilon/5 = 3\epsilon/5, \end{aligned}$$

completing the proof. □

The next step shows that the probability that a vector (t_i, t_{-i}) falls within region $\mathcal{A}_k$, when t_{-i} is distributed randomly, does not vary too much with t_i in large markets. This is a key step in our argument, as it implies that an individual agent cannot appreciably change the probability that t falls within each $\mathcal{A}_k$, and therefore cannot have a large effect on the aggregate allocation.

Step 2: approximation of the probability that the empirical distribution of types is in region $\mathcal{A}_k$.

Claim B.2. The integer n_0 can be taken such that, for all $n \geq n_0$ there exist weights $\pi_1^n, \ldots, \pi_K^n$ such that $\sum_{k=1}^K \pi_k^n = 1$ and

$$|\Pr\{(t_i, t_{-i}) \in \mathcal{A}_k | t_{-i} \in T^{n-1}, t_{-i} \sim \mu_0\} - \pi_k^n| < \epsilon/5K$$

for all k and all t_i.

Proof. Let $\epsilon' = \epsilon/5K$. We begin by constructing numbers $\bar{\pi}_k^n$ that are approximately equal to the weights π_k^n in the statement of the claim. Let

$$\bar{\pi}_k^n = \Pr\{t' \in \mathcal{A}_k | t' \in T^n, t' \sim \mu_0\}$$

be the probability that a vector of n types drawn independently according to μ_0 is in $\mathcal{A}_k$. We will show that, for large n, for any type t_i, the $\bar{\pi}_k^n$ are very close to the probability

$$\Pr\{(t_i, t_{-i}) \in \mathcal{A}_k | t_{-i} \in T^{n-1}, t_{-i} \sim \mu_0\}$$

that a vector with agent i's type fixed at t_i and the other types drawn i.i.d. is in $\mathcal{A}_k$. To see this, consider the difference between the probability of a vector of types falling within region

$\mathcal{A}_k$ when i's type is fixed as t_i, minus the probability of falling within $\mathcal{A}_k$ when i's type is drawn randomly. This difference equals

$$\begin{aligned} & \Pr\{(t_i, t'_{-i}) \in \mathcal{A}_k | t'_{-i} \in T^{n-1}, t'_{-i} \sim \mu_0\} - \bar{\pi}^n_k \\ = \; & \Pr\{(t_i, t'_{-i}) \in \mathcal{A}_k \text{ and } t' \notin \mathcal{A}_k | t' \in T^n, t' \sim \mu_0\} \\ - \; & \Pr\{(t_i, t'_{-i}) \notin \mathcal{A}_k \text{ and } t' \in \mathcal{A}_k | t' \in T^n, t' \sim \mu_0\}. \end{aligned} \tag{B.9}$$

This expression equals the probability of drawing a vector $t' \in T^n$ where changing agent i's type from t'_i to t_i moves the vector of types from outside $\mathcal{A}_k$ to inside $\mathcal{A}_k$, minus the probability of choosing a vector where changing i's type from t'_i to t_i moves the vector from inside $\mathcal{A}_k$ to outside $\mathcal{A}_k$. We now show that the probability of such vectors being drawn is very small in a sufficiently large market.

Consider the case where $(t_i, t'_{-i}) \notin \mathcal{A}_k$, but $(t'_i, t'_{-i}) \in \mathcal{A}_k$. One possibility is that $(t_i, t'_{-i}) \notin \mathcal{N}$. By the law of large numbers, we may take n_0 large enough such that for $n \geq n_0$ the probability of this happening is less than $\epsilon'/8$. The other possibility is that $(t_i, t'_{-i}) \in \mathcal{N}$, but $(t_i, t'_{-i}) \notin \mathcal{A}_k$. In this case, the line segment connecting $\text{emp}[t_i, t'_{-i}]$ and $\text{emp}[t']$ contains a point in $\mathcal{B}$, because $\mathcal{N}$ is convex and each $\mathcal{A}_k$ is open. This means that the distance between $\text{emp}[t']$ and $\mathcal{B}$ is at most $1/n$. By Condition 1 of Definition B.1, we may take n_0 such that this probability is less than $\epsilon'/8$. This argument implies that we may take n_0 such that, for all $n \geq n_0$,

$$\Pr\{(t_i, t'_{-i}) \notin \mathcal{A}_k, t' \in \mathcal{A}_k | t' \in T^n, t' \sim \mu_0\} < \epsilon'/8 + \epsilon'/8 = \epsilon'/4.$$

An analogous argument proves that n_0 can be chosen such that, for $n \geq n_0$,

$$\Pr\{(t_i, t'_{-i}) \in \mathcal{A}_k, t' \notin \mathcal{A}_k | t' \in T^n, t' \sim \mu_0\} < \epsilon'/4.$$

Substituting these two inequalities into the RHS of equation (B.9) yields that

$$|\Pr\{(t_i, t'_{-i}) \in \mathcal{A}_k | t'_{-i} \in T^{n-1}, t'_{-i} \sim \mu_0\} - \bar{\pi}^n_k| < \epsilon'/4 + \epsilon'/4 = \epsilon'/2. \tag{B.10}$$

Note, however, that the $\bar{\pi}^n_k$ do not necessarily sum to 1, as it may be the case that $t' \notin \cup_k \mathcal{A}_k$. To complete the proof, define

$$\pi^n_k = \bar{\pi}^n_k / \sum_{k'=1}^{K} \bar{\pi}^n_{k'}. \tag{B.11}$$

We have that the probability that $t' \notin \cup_k \mathcal{A}_k$ converges to 0. Therefore, we may take n_0 such that for $n \geq n_0$

$$|1 - 1/\sum_{k'=1}^{K} \bar{\pi}_{k'}^n| < \epsilon'/2. \tag{B.12}$$

We now apply these bounds to prove the claim. We have

$$\begin{aligned}
& |\Pr\{(t_i, t'_{-i}) \in \mathcal{A}_k | t'_{-i} \in T^{n-1}, t'_{-i} \sim \mu_0\} - \pi_k^n| \\
\leq \;& |\Pr\{(t_i, t'_{-i}) \in \mathcal{A}_k | t'_{-i} \in T^{n-1}, t'_{-i} \sim \mu_0\} - \bar{\pi}_k^n| + |\pi_k^n - \bar{\pi}_k^n| \\
< \;& \epsilon'/2 + |\pi_k^n - \bar{\pi}_k^n| \\
= \;& \epsilon'/2 + |\bar{\pi}_k^n / (\sum_{k'=1}^{K} \bar{\pi}_{k'}^n) - \bar{\pi}_k^n| \\
= \;& \epsilon'/2 + |1 - 1/(\sum_{k'=1}^{K} \bar{\pi}_{k'}^n)| \cdot \bar{\pi}_k^n \\
< \;& \epsilon'/2 + \epsilon'/2 = \epsilon/5K.
\end{aligned}$$

The series of steps in the above derivation were as follows. The second line follows from the triangle inequality. The third line uses the bound from inequality (B.10). The fourth line uses the definition of π_k^n from equation (B.11). Finally, the fifth line follows from multiplying $\bar{\pi}_k^n$ out of the right term, and the sixth line comes from inequality (B.12) and $\bar{\pi}_k^n \leq 1$. □

Step 3: completing the proof.

Finally, we apply the results from Steps 1 and 2 to prove the lemma. We have

$$f^n(t_i, \mu_0) - \sum_{k=1}^{K} \pi_k^n \cdot z_k(t_i) = \sum_{t_{-i} \in T^{n-1}} \Pr(t_{-i} | t_{-i} \sim \mu_0) \cdot F_i^n(t) - \sum_{k=1}^{K} \pi_k^n \cdot z_k(t_i).$$

This sum can be decomposed depending on whether $\hat{\mu} = \text{emp}[t]$ is in each of the $\mathcal{A}_k$ sets or outside the union of the $\mathcal{A}_k$ sets. We have

$$\begin{aligned}
& f^n(t_i, \mu_0) - \sum_{k=1}^{K} \pi_k^n \cdot z_k(t_i) \qquad \text{(B.13)} \\
= \;& \sum_{k=1}^{K} (\{\sum_{t_{-i} : \hat{\mu} \in \mathcal{A}_k} \Pr(t_{-i} | t_{-i} \sim \mu_0) \cdot F_i^n(t)\} - \pi_k^n \cdot z_k^n(t_i)) + \sum_{t_{-i} : \hat{\mu} \notin \cup_k \mathcal{A}_k} \Pr(t_{-i} | t_{-i} \sim \mu_0) \cdot F_i^n(t).
\end{aligned}$$

We begin by looking at the terms where $\hat{\mu}$ is in one of the $\mathcal{A}_k$. We will show that for each k these terms are small. We have that, for each k,

$$\begin{aligned}
&\|(\sum_{t_{-i}:\hat{\mu}\in\mathcal{A}_k} \Pr(t_{-i}|t_{-i}\sim\mu_0)\cdot F_i^n(t)) - \pi_k^n \cdot z_k(t_i)\| \\
= \| &\sum_{t_{-i}:\hat{\mu}\in\mathcal{A}_k} \Pr(t_{-i}|t_{-i}\sim\mu_0)\cdot (F_i^n(t) - z_k(t_i)) \\
&+(\{\sum_{t_{-i}:\hat{\mu}\in\mathcal{A}_k} \Pr(t_{-i}|t_{-i}\sim\mu_0)\} - \pi_k^n)\cdot z_k(t_i)\| \\
\leq \quad &\sum_{t_{-i}:\hat{\mu}\in\mathcal{A}_k} \Pr(t_{-i}|t_{-i}\sim\mu_0)\cdot \|F_i^n(t) - z_k(t_i)\| \\
&+|\{\sum_{t_{-i}:\hat{\mu}\in\mathcal{A}_k} \Pr(t_{-i}|t_{-i}\sim\mu_0)\} - \pi_k^n|.
\end{aligned} \tag{B.14}$$

The equality in the second line follows from rearranging the expression. The inequality in the third line follows from the triangle inequality, and the fact that the norm of the vector $z_k(t_i) \in X$ is weakly less than 1.

Consider now the right hand side of inequality (B.14). By Claim B.1, we may take n_0 such that for all $n \geq n_0$ and t_{-i} such that $\hat{\mu} \in \mathcal{A}_k$,

$$\|F_i^n(t) - z_k(t_i)\| < 3\epsilon/5.$$

By Claim B.2, n_0 may be taken such that, for $n \geq n_0$, the second term is bounded by

$$|\{\sum_{t_{-i}:\hat{\mu}\in\mathcal{A}_k} \Pr(t_{-i}|t_{-i}\sim\mu_0)\} - \pi_k^n| < \frac{1}{5K}\epsilon.$$

Substituting these two bounds in inequality (B.14) we have that, for all $n \geq n_0$,

$$\begin{aligned}
&\|\{\sum_{t_{-i}:\hat{\mu}\in\mathcal{A}_k} \Pr(t_{-i}|t_{-i}\sim\mu_0)\cdot F_i^n(t)\} - \pi_k^n \cdot z_k^n(t_i)\| \\
< \quad &\frac{3}{5}\epsilon \cdot \sum_{t_{-i}:\hat{\mu}\in\mathcal{A}_k} \Pr(t_{-i}|t_{-i}\sim\mu_0) + \frac{1}{5K}\epsilon.
\end{aligned}$$

Summing over all k we get

$$\sum_{k=1}^{K} \|\{\sum_{t_{-i}:\hat{\mu}\in\mathcal{A}_k} \Pr(t_{-i}|t_{-i}\sim\mu_0)\cdot F_i^n(t)\} - \pi_k^n \cdot z_k^n(t_i)\| < \frac{3}{5}\epsilon + K\frac{1}{5K}\epsilon = 4\epsilon/5.$$

Using the triangle inequality, the sum operator can be brought into the norm, yielding the inequality

$$\| \sum_{k=1}^{K} (\{ \sum_{t_{-i}:\hat{\mu}\in\mathcal{A}_k} \Pr(t_{-i}|t_{-i}\sim\mu_0)\cdot F_i^n(t)\} - \pi_k^n \cdot z_k^n(t_i))\| < 4\epsilon/5. \tag{B.15}$$

The argument above bounds the terms in the RHS of equation (B.13) that correspond to t within the sets $\mathcal{A}_k$. To bound the last term, note that we may take n_0 to be large enough so that, for all $n \geq n_0$, the probability that $t \notin \cup_k \mathcal{A}_k$ is strictly less than $\epsilon/5$. That is,

$$\sum_{t_{-i}:\hat{\mu}\notin\cup_k\mathcal{A}_k} \Pr(t_{-i}|t_{-i}\sim\mu_0) < \epsilon/5. \tag{B.16}$$

Plugging equations (B.15) and (B.16) into equation (B.13) we obtain

$$\|f^n(t_i,\mu_0) - \sum_{k=1}^{K} \pi_k^n \cdot z_k(t_i)\| < \epsilon,$$

completing the proof of Step 3, and hence the lemma.

□

B.2 An Example without Quasi-Continuity

This section shows, by example, that it is necessary to impose regularity conditions on a family of limit equilibria to obtain the results in Theorem B.1. We consider an example of a non quasi-continuous family of equilibria, and show that the mechanism constructed in the proof of Theorem B.1 does not satisfy any of the implications of the theorem. Namely, the constructed mechanism is not SP-L, and outcomes of the constructed mechanism do not approximate the outcomes of the original mechanism, nor a convex combination of the outcomes. We consider the case of BNE of the finite mechanism to highlight that, even for finite economy BNE, the construction used to prove our Theorems 2 and B.1 depends on the quasi-continuity condition. It is a simple matter to extend the logic of the example to limit BNE.

Consider a set of two objects $O = \{o_1, o_2\}$. The set of bundles is $X_0 = O \times \{0, -10\}$, so that a bundle x_0 specifies an object $x_0(1) = o_1$ or o_2, and a transfer $x_0(2) = 0$ or -10 of a numeraire. Therefore, agents either receive no transfer, or pay a fine of 10. The set of types is $T = O = \{o_1, o_2\}$, with an agent's type denoting her favorite object. Utility is given by

$$u_{t_i}(x_0) = \mathbf{1}\{x_0(1) = t_i\} + x_0(2).$$

That is, an agent has utility 1 for receiving an object matching her type, and quasilinear utility on the transfer. Consider the set of actions

$$A = O \times \{f, nf\}.$$

An action a_i specifies an object preference $a_i(1) = o_1$ or o_2, and a message $a_i(2) = f$ (standing for fine) or nf (standing for no fine). We define the mechanism $\{\Phi^n, A\}$ as follows.

- If all $a_j(2) = nf$, $j = 1, \cdots, n$, then $\Phi_i^n(a) = (a_i(1), 0)$. That is, if all agents choose the no fine option, then each agent receives her favorite object and no one is fined.

- If some $a_j(2) = f$, $j = 1, \cdots, n$, then some agents will be fined, depending on whether the number of agents asking for object o_1 is odd or even.

  - If $\#\{j : a_j(1) = o_1\}$ is odd, then agents asking for object o_1 are fined:

$$\Phi_i^n(a) = (a_i(1), -10 \cdot \mathbf{1}\{a_i(1) = o_1\}).$$

  - If $\#\{j : a_j(1) = o_1\}$ is even, then agents asking for object o_2 are fined:

$$\Phi_i^n(a) = (a_i(1), -10 \cdot \mathbf{1}\{a_i(1) = o_2\}).$$

We now define a sequence of families of BNE. Let n_0 be a sufficiently large number, and $\delta > 0$ a small positive constant. Let μ_0 be the distribution putting equal weight on o_1 and o_2. Define now the following subset of $\mathbb{N} \times \bar{\Delta}T$,

$$S = \{(n, \mu) \in \mathbb{N} \times \bar{\Delta}T : n \cdot \mu(o_1) \text{ is an odd integer}, n \geq n_0, \|\mu - \mu_0\| < \delta\}.$$

That is, S is the set of all pairs of a number of players and a distribution over types such that, in a type profile with n types and empirical distribution of types μ, the number of players with $t_i = o_1$ is odd. Moreover, n has to be larger than n_0 and μ sufficiently close to μ_0.

Consider now the following sequence of families $(\sigma_\mu^n)_{\mu \in \Delta T, n \in \mathbb{N}}$ of BNE of this mechanism.

- If $(n,\mu) \in S$, then σ^n_μ specifies that agents play actions that match their types $a_i(1) = t_i$, and send the fine message $a_i(2) = f$.

- Otherwise agents play actions that match their types $a_i(1) = t_i$, but send the no fine message $a_i(2) = nf$.

Note that, for suitably chosen n_0 and δ, this is a family of equilibria. If $(n,\mu) \notin S$, then it is optimal for agents to request their favorite object, as no agents are fined. If $(n,\mu) \in S$, then sending the f or nf message is immaterial, as fines are always activated since all other players send the f message in equilibrium. Moreover, it is optimal to request one's favorite object ($a_i(1) = t_i$), as the probability that agents requesting objects o_1 or o_2 are fined are both approximately equal to $1/2$.

Note also that this family of BNE is not quasi-continuous at μ_0. To see this formally, take a neighborhood $\mathcal{N}$ of μ_0 small enough such that the set

$$\{\mu \in \mathcal{N} : \exists n \text{ with } (n,\mu) \in S\} \tag{B.17}$$

is relatively dense with respect to $\mathcal{N}$. Any open subset $\mathcal{A}_k$ of this neighborhood therefore contains infinite points in this set (B.17), and infinite points outside this set (B.17). In particular, for any n_0, there exist $n \geq n_0$ and μ and $\text{emp}[t]$ in $\mathcal{A}_k$ where type o_1 agents are not fined, i.e., $\Phi^n_i(o_1, \sigma^n_\mu(t_{-i})) = (o_1, 0)$, and similarly there exist $n \geq n_0$ and μ and $\text{emp}[t]$ in $\mathcal{A}_k$ where type o_1 agents are fined, i.e., $\Phi^n_i(o_1, \sigma^n_\mu(t_{-i})) = (o_1, -10)$. This violates Condition 2 of the definition of quasi-continuity.

Define now the direct mechanism $((F^n)_{n\in\mathbf{N}}), T)$ such that

$$F^n(t) = \Phi^n(\sigma^n_{\text{emp}[t]}(t)).$$

This is the construction used in the proof of Theorem 2. We now show that this mechanism is neither SP-L, nor does it approximate outcomes of the σ^n_μ equilibria. Consider a type profile t such that $(n, \text{emp}[t]) \notin S$. Then the no fine equilibrium is played and

$$F^n_i(t) = (t_i, 0).$$

That is, the mechanism simply assigns the requested object to each agent, and there are no fines. However, if $(n, \text{emp}[t]) \in S$, we have

$$F^n_i(t) = (t_i, -10 \cdot \{t_i = o_1\}).$$

That is, the mechanism assigns the requested object to each agent, but only fines the $t_i = o_1$ types. This happens because the equilibria σ_μ^n where agents send the fine message are played exactly at the profiles where the number of o_1 reports is odd, and therefore where agents reporting o_1 are fined. In contrast, agents reporting o_2 are never fined. Note that, if types are distributed according to μ_0, the probability that $(n,t) \in S$ converges to $1/2$ as the number of players grows. We have that the constructed mechanism has a limit

$$\begin{aligned} f_i^\infty(t_i, \mu_0) &= \frac{1}{2}(o_1, 0) + \frac{1}{2}(o_1, -10) \text{ if } t_i = o_1 \\ & (o_2, 0) \text{ if } t_i = o_2. \end{aligned}$$

In particular, the constructed mechanism is not SP-L, as a type $t_i = o_1$ agent would prefer to report being a type o_2. Moreover, the above allocation does not approximate a convex combination of allocations received in the sequence of families of equilibria.

C Semi-Anonymity

Our main analysis considers anonymous mechanisms, where agents' outcomes depend on their own report and the distribution of all reports. The analysis generalizes straightforwardly, though at some notational burden, to the case of semi-anonymous mechanisms, as defined by Kalai (2004). In this setting, agents are divided into a number of groups, and agents within each group can be treated differently by the mechanism.

In this section, agents belong to **groups** g in a finite set G. The set of types is partitioned into subsets

$$T = T_{g_1} \cup T_{g_2} \cup \cdots \cup T_{g_G}.$$

A **semi-anonymous mechanism** is defined as $\{(\Phi^n)_{n\in\mathbb{N}}, (A_g)_{g\in G}\}$, where the A_g are the sets of actions available to each group g, and

$$A = A_{g_1} \cup \cdots \cup A_{g_G}$$

is the set of actions. As in the anonymous case, the Φ^n are functions

$$\Phi^n : A^n \to \Delta({X_0}^n).$$

The difference with respect to anonymous mechanisms is that agents in group g are restricted to play strategies in A_g. That is, if $t_i \in T_g$ then the support of any strategy $\sigma(t_i)$

is contained in A_g. In a matching setting, for example, the groups may specify whether an agent is a man or a woman, and the agent's traits. Agents are then permitted to misreport their preferences over other match partners, but they cannot misrepresent their gender or their traits. **Limit mechanisms** are defined as in Section 3.1. In particular, we define limit mechanisms with respect to a single distribution $\mu \in \Delta T$, and not distributions of types within groups. Alternatively, one could assume that the number of agents in each group grows in a specific way, and that types are drawn i.i.d. within each group. We now formally define a two-sided matching mechanism, to clarify the definition.

Example C.1. (Two-Sided Matching) This example shows that semi-anonymous mechanisms include matching mechanisms in two-sided markets (Gale and Shapley, 1962). Agents are men and women, who differ on a set of traits. Groups g index both sex and the traits, so that the set of groups is

$$G = \{m_1, m_2, \cdots, m_M\} \cup \{w_1, w_2, \cdots, w_W\}.$$

That is, there are M groups of men and W groups of women. Men and women within each group have the same traits, and are equally good marriage partners. However, within each group, agents may differ in their preferences over the other groups. The way in which the semi-anonymous framework differs from the anonymous setting is that men and women may misreport their preferences, but cannot misreport their sex nor traits.

Formally, agent i's type is

$$t_i = (g_{t_i}, u_{t_i}),$$

where $g_{t_i} \in G$ is the agent's group, and u_{t_i} is a strictly positive utility function over the groups of the opposite sex. The set of outcomes $X_0 = G \cup \emptyset$. That is, each agent only cares about which type of man (woman) she (he) is matched to, or whether she (he) is unmatched. Utilities of each type t_i are given by $u_{t_i}(g)$ if she is matched to someone of the opposite sex. We extend u_{t_i} so that it is 0 if the agent is unmatched or matched to a group of the same sex.

Consider now a stable matching mechanism, using a tie-breaking lottery, as in school choice mechanisms (Abdulkadiroğlu et al., 2009). The mechanism is direct, so that $A_g = T_g$ for each $g \in G$. Men and women report a vector of types t, and therefore traits. This implies a weak preference ordering of each man over each woman and vice versa. The mechanism assigns a lottery number l_i to each agent, uniformly and independently distributed between 0 and 1. Lottery numbers are used to break ties between preferences. That is, preferences are refined to strict preferences, by using the lottery numbers to break ties. Conditional on a vector of lotteries l and a vector of reported types t, the mechanism implements a stable matching $x^n(t,l)$. The function $x^n(t,l)$ is taken to be symmetric, to conform to the semi-anonymity assumption. The mechanism is then defined as

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

□

Define a semi-anonymous mechanism as SP-L if no agent wants to misreport as a different type within the same group.

Definition C.1. *The semi-anonymous mechanism* $\{(\Phi^n)_{\mathbb{N}}, (T_g)_{g\in G}\}$ *is **strategy-proof in the large (SP-L)** if, for any* $m \in \bar{\Delta}T$, $g \in G$, *and* $t_i, t_i' \in T_g$,

$$u_{t_i}[\phi^\infty(t_i, m)] \geq u_{t_i}[\phi^\infty(t_i', m)]. \tag{C.1}$$

Equivalently, the mechanism is SP-L if, for any $m \in \bar{\Delta}T$ *and* $\epsilon > 0$, *there exists* n_0 *such that, for all* $n \geq n_0$, $g \in G$, *and* $t_i, t_i' \in T_g$,

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

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

The sufficient conditions for a mechanism to be SP-L also have straightforward extensions. The extension of the EF-TB condition is that no agent envies another agent in the same group, and with lower lottery number.

Definition C.2. *A direct semi-anonymous mechanism* $\{(\Phi^n)_{\mathbb{N}}, (T_g)_{g\in G}\}$ *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* $l_i \geq l_j$, *and if* t_i *and* t_j *belong to the same group, then*

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

With this definition, an extension of Theorem 1 to semi-anonymous mechanisms follows from essentially the same proof.[2] This implies that the stable matching procedure in example C.1 is SP-L, because an agent envying another agent with a lower lottery number would violate the stability condition.

We now extend the definition of limit BNE to this setting, and state and prove an extension of Theorem B.1. The conclusions of the theorem are unchanged, and the only difference is that it considers a family of limit equilibria of a semi-anonymous mechanism, and not an anonymous mechanism. The proof uses a construction identical to that in Theorem B.1. The proof follows from noting that the argument in the anonymous case implies that the approximation formulas in Theorem B.1 hold, and then showing that this implies that the constructed semi-anonymous mechanism is SP-L.

We must first extend the concept of a limit BNE. The difference with respect to the anonymous case is that in the semi-anonymous case it is only necessary to rule out deviations where agents of group g play other actions in A_g, or, in a direct mechanism, report being a different type in T_g. A **strategy** is defined as a map $\sigma : T \to \Delta A$ such that if $t_i \in T_g$ then the support of $\sigma(t_i)$ is contained in A_g.

Definition C.3. *Given a semi-anonymous mechanism $\{(\Phi^n)_{\mathbb{N}}, (A_g)_{g\in G}\}$ with limit $\phi^\infty(\cdot,\cdot)$, and a probability distribution over types $\mu \in \Delta T$, the strategy $\sigma^*_\mu : T \to \Delta A$ is a* ***limit Bayes-Nash Equilibrium at prior*** *μ if, for all $g \in G$, $t_i \in T_g$ and $a'_i \in A_g$:*

$$u_{t_i}[\phi^\infty(\sigma^*_\mu(t_i), \sigma^*_\mu(\mu))] \geq u_{t_i}[\phi^\infty(a'_i, \sigma^*_\mu(\mu))].$$

The definition of **(quasi-) continuous** families of limit equilibria is identical to the anonymous case. With these definitions, the statement of the semi-anonymous version of Theorem B.1 is similar to the anonymous case, with the key difference being the larger class of mechanisms considered.

Theorem C.1 (Extension of Theorem B.1 to semi-anonymous mechanisms)**.** *Given any semi-anonymous mechanism $\{(\Phi^n)_{\mathbb{N}}, (A_g)_{g\in G}\}$ with a quasi-continuous family of limit Bayes-Nash equilibria $(\sigma^*_\mu)_{\mu\in\Delta T}$, there exists a direct, SP-L, semi-anonymous mechanism $\{(F^n)_{\mathbb{N}}, (T_g)_{g\in G}\}$ with the following properties.*

[2] Lemma A.1 holds as is, since it is a statement about the empirical distribution of randomly drawn vectors of types, and therefore does not rely on the definition of a mechanism. Lemma A.2 holds for any two types t_i and t'_i in the same group, using the same proof, as for any such pairs of types the EF-TB condition in the semi-anonymous case implies the same properties as in the anonymous case. Given the two lemmas, the argument in the proof of Theorem 1 in Appendix 1.1 holds as is, as long as we take t'_i to be in the same group as t_i, which is all that is needed for the definition of SP-L for semi-anonymous mechanisms.

1. *If the original mechanism is continuous at a prior $\mu_0 \in \bar{\Delta}T$ then, in the limit, truthful play of the direct mechanism produces the same outcomes as equilibrium play of the original mechanism. Formally, for any $t_i \in T$, we have*

$$f^{\infty}(t_i, \mu_0) = \phi^{\infty}(\sigma^*_{\mu_0}(t_i), \sigma^*_{\mu_0}(\mu_0)),$$

where f^{∞} is the limit of the direct mechanism.

2. *For any prior $\mu_0 \in \bar{\Delta}T$, in the large market limit, truthful play of the direct mechanism produces the same outcomes as a convex combination of equilibrium play of the original mechanism under priors that are close to μ_0. Formally, for any $\epsilon > 0$, there exists n_0, an integer K, numbers π^n_k with $\sum_{k=1,\cdots,K} \pi^n_k = 1$, and priors μ_k with $\|\mu_k - \mu_0\| < \epsilon$ such that, for all $n \geq n_0$ and $t_i \in T$, we have*

$$\|f^n(t_i, \mu_0) - \sum_{k=1,\cdots,K} \pi^n_k \cdot \phi^n(\sigma^*_{\mu_k}(t_i), \sigma^*_{\mu_k}(\mu_k))\| < \epsilon,$$

where f^n is the function representing the direct mechanism from an interim perspective, as defined in equation (3.1).

Proof. Let F be defined as in equation (5.2). Parts 1 and 2 of the theorem statement follow from the same argument as in the proof of Theorem B.1. This is the case because $\{(\Phi^n)_{\mathbb{N}}, \cup_{g \in G} A_g\}$ is an anonymous mechanism. Moreover, parts 1 and 2 do not rely on agents playing optimally under σ^*_μ, and the definitions of (quasi-) continuity of a family of limit equilibria are the same in the anonymous and semi-anonymous case. Likewise, Lemma B.1 holds in the semi-anonymous setting, with essentially the same proof.

It only remains to be proven that the direct mechanism is SP-L. To establish this, we employ a small modification of the original argument, as now the σ^*_μ are limit equilibria of a semi-anonymous mechanism. As in the proof of Theorem B.1, take $\mu_0 \in \bar{\Delta}T$, and $\epsilon > 0$. By Lemma B.1 (with $\frac{\epsilon}{2|X_0|}$ as the constant), there exist a neighborhood $\mathcal{N}$ of μ_0, priors μ_k and weights π^n_k for $k = 1, \cdots, K$, and n_0 such that, for all $n \geq n_0$ and t'_i in T we have

$$\begin{aligned} \sum_{k=1}^{K} \pi^n_k &= 1, \\ \|\mu_k - \mu_0\| &< \epsilon, \text{ and} \\ \|f^n(t'_i, \mu_0) - \sum_{k=1}^{K} \pi^n_k \cdot z_k(t'_i)\| &< \frac{\epsilon}{2|X_0|} \leq \frac{\epsilon}{2}, \end{aligned} \tag{C.2}$$

where

$$z_k(t_i') = \phi^\infty(\sigma^*_{\mu_k}(t_i'), \sigma^*_{\mu_k}(\mu_k)).$$

Take now any two types t_i and t_i' in the same group. We have that

$$\begin{aligned}
& u_{t_i}[f^n(t_i', m)] - u_{t_i}[f^n(t_i, m)] \\
\leq\ & \sum_{k=1,\dots,K} \pi_k^n \cdot \{u_{t_i}[z_k(t_i')] - u_{t_i}[z_k(t_i)]\} \\
& + |u_{t_i}[f^n(t_i', \mu_0)] - u_{t_i}[\sum_{k=1}^K \pi_k^n \cdot z_k(t_i')]| + |u_{t_i}[f^n(t_i, \mu_0)] - u_{t_i}[\sum_{k=1}^K \pi_k^n \cdot z_k(t_i)]|.
\end{aligned}$$

Consider now the RHS of this inequality. From the definition of z_k and of a limit equilibrium, we have that the first sum is nonpositive. Moreover, by the bound (C.2), the fact that utility is in $[0, 1]$, and that the set of random bundles has X_0 dimensions, the second and third terms are each bounded by $\epsilon/2$. Therefore, we have that

$$u_{t_i}[f^n(t_i', m)] - u_{t_i}[f^n(t_i, m)] < 0 + \epsilon/2 + \epsilon/2 = \epsilon.$$

Since this holds for all t_i and t_i' in the same group, we have that the constructed semi-anonymous mechanism is SP-L. □

D Details for Table 1

This Section provides supporting details for the classification of non-SP mechanisms presented as Table 1. For each mechanism we provide a formal definition of the mechanism in our setting, a formal proof of the classification, and relevant references. The analysis of multi-unit auctions in Section D.1.1 is especially detailed and illustrates why the interim approach to taking the large-market limit is crucial for obtaining the classification.

D.1 Anonymous Mechanisms.

D.1.1 Multi-Unit Auctions

We consider multi-unit auctions for identical goods, such as government bond auctions. The two most common 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 show that uniform-price auctions are SP-L, whereas pay-as-bid auctions are manipulable in the large.

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 consists of a per-unit value v_i and a maximum capacity q_i. The set of possible values is $V = \{1, \ldots, \bar{v}\}$, the set of possible capacity limits is $Q = \{0, 1, \ldots, \bar{q}\}$ 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, let $D(p; t)$ denote the demand for the object at price p.[3] The market-clearing price is

$$p^*(t) = \max\{p \in V : \frac{D(p; t)}{n} \geq k\}. \tag{D.1}$$

That is, $p^*(t)$ is the highest price at which demand weakly exceeds supply. 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 $p^*(t)$. The probability of a bid being rationed $\bar{r}$ is set to clear the market.[4]

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:

$$\rho^*(m) = \max\{p \in V : E[D(p; t_i)|t_i \sim m] \geq k\}. \tag{D.2}$$

[3] Formally, $D(p; t) = \sum_{i=1}^{n} q_i \cdot 1\{v_i \geq p\}$, where the notation $1\{\cdot\}$ denotes the indicator function.

[4] Because preferences are linear up to the capacity limit, the exact form of the rationing is immaterial. The rationing probability is

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

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

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

In this generic case, as the market grows large, the realized price as defined in (D.1) equal $\rho^*(m)$ with probability converging to one. Therefore, the limit mechanism allocates each bidder their demand at $\rho^*(m)$, 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.[5] Note that, in this generic case, the price in the limit is deterministic and is 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, $E[D(\rho^*(m);t_i)|t_i \sim m] = k$. 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 $\frac{1}{2}$.[6] 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

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

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

[6] 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.

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.

Note that this argument would not go through had we used a stronger notion of approximate strategy-proofness based on realizations of opponents' reports rather than probability distributions. In any size market, it is always possible to construct a profile of opponent bids t_{-i} where, ex-post, bidder t_i can profitably lower the market-clearing price by shading his quantity demanded. Similarly, our argument would not go through if SP-L required equation (3.2) to obtain for all probability distributions $m \in \Delta T$, rather than for all full support probability distributions $m \in \bar{\Delta}T$. Full support ensures that the probability that any particular bidder is pivotal goes to zero as the market grows large. See Swinkels (2001; Section 5) for an elegant example, with limited support, in which bidders remain pivotal with probability one even in very large markets.

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.

D.1.2 Single-Unit Assignment

In single-unit assignment problems, each agent is to be assigned at most one indivisible object, and there are no transfers. We refer the reader to Kojima and Manea (2010) and references therein for a detailed description of the environment and applications.

Formally, we define single-unit assignment as follows. Denote the set of object types by X_0. In a market of size n there are $\{q_{x_0} \cdot n\}$ units of object type x_0 available.[7] An agent of type $t_i \in T$ has a strict utility function u_{t_i} over X_0. It is assumed that X_0 includes a

[7] A bracketed expression denotes the nearest integer to the real number within brackets.

null object $\emptyset$, in supply $n - \sum_{x_0' \neq \emptyset}\{q_{x_0'} \cdot n\} \geq 0$, so that the total quantity of objects equals n. The utility of the null object is normalized to 0. Therefore, we assume that all agents strictly prefer any other object (termed a proper object) to the null object.

Boston Mechanism

The Boston mechanism is a mechanism used in many cities to allocate seats in public schools. Abdulkadiroğlu and Sönmez (2003) show that the Boston mechanism is not SP, and Abdulkadiroğlu et al. (2006) document that it was extensively manipulated in practice. We now formally define the Boston mechanism and show that it is not SP-L. This complements an example given by Kojima and Pathak (2009), in a formally different environment, where the Boston mechanism can be manipulated in a large market.

We now define the Boston mechanism. Fix a vector of reports t. To be consistent with the literature we will use the terminology of schools (the objects) and students (the agents). The mechanism first assigns to each student a lottery number l_i, uniformly and independently distributed in $[0, 1]$. The mechanism then proceeds in rounds, following the algorithm below.

1. The mechanism begins in $\texttt{round} = 1$. All students are initially unassigned.

2. Students that are still present in the mechanism take turns, in the order of their lottery number, with higher lottery numbers going first. In her turn student i is permanently assigned to her $\texttt{round}^{\text{th}}$ choice, as given by u_{t_i}, if there are still seats in that school, or remains unassigned otherwise.

3. If all students have been assigned, finish, otherwise increase `round` by 1 and go to Step 2.

Note that the algorithm must finish, as eventually all students are assigned either to a proper school or to the null school $x_0 = \emptyset$. Therefore, conditional on a vector of types t and lottery numbers l the mechanism produces a well-defined outcome $x^n(t, l)$. Before lottery draws, the mechanism is defined as

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

We now show that the Boston mechanism is not SP-L. Consider an economy with two proper schools, $x_0 = A, B$, and the null school $x_0 = \emptyset$, corresponding to being unmatched. That is, $X_0 = \{A, B, \emptyset\}$. Let $q_A = q_B = 1/6$. Consider a distribution $m \in \bar{\Delta}T$ such that 2/3 of the agents prefer school A, while only 1/3 prefer school B. Then, in a large market, the proper schools are filled in the first round with probability close to 1. Therefore, an agent has a negligible chance of getting her second choice. The chance of getting her first choice is $(1/6)/(2/3) = 1/4$ for school A and $(1/6)/(1/3) = 1/2$ for school B. That is, the limit mechanism is

$$\begin{aligned} \phi^{\infty}(t_i, m) &= \frac{1}{4} \cdot A + \frac{3}{4} \cdot \emptyset \text{ if } u_{t_i}[A] > u_{t_i}[B] \\ & \quad \frac{1}{2} \cdot B + \frac{1}{2} \cdot \emptyset \text{ otherwise.} \end{aligned} \tag{D.3}$$

Note in particular that an agent who prefers school A faces a tradeoff when reporting her preferences. If she announces that she prefers school A, she will be assigned to it with 1/2 the chance she has of receiving school B. Therefore, it is not optimal for an agent with $u_{t_i}[A] > u_{t_i}[B] > u_{t_i}[A]/2$ to report truthfully.

Probabilistic Serial Mechanism

The probabilistic serial mechanism has been proposed as a solution to the assignment problem by Bogomolnaia and Moulin (2001). The mechanism works as follows. With time running continuously, agents "eat" probability shares of their favorite object, out of all objects still available. After probability shares of all objects are assigned, the objects are randomly assigned to agents according to these probabilities. We refer the reader to Kojima and Manea (2010) page 110 for a formal definition of the mechanism, as their analysis includes ours as a particular case.

Bogomolnaia and Moulin (2001) show that the mechanism is EF. Consequently, Theorem 1 guarantees that it is SP-L. Note that the fact that this mechanism is SP-L is a particular case of Kojima and Manea's Theorem 1.

Hylland and Zeckhauser Pseudo-Market Mechanism

Hylland and Zeckhauser (1979) proposed a pseudo-market mechanism for single-unit assignment, in which agents are endowed with equal budgets of an imaginary currency which they use to purchase probability shares of the objects. The mechanism works as follows. First, agents report their types, t. Second, the mechanism allocates each agent an equal budget $B > 0$ of an artificial currency. Third, the mechanism computes a competitive equilibrium price vector $p^* \in \mathbb{R}_+^{|X_0|}$. That is, a vector of prices, one for each object type, such that when each agent is allocated her most-preferred affordable bundle of probability shares, based on her reported preferences, the market clears. Last, each agent is allocated her most-preferred affordable bundle at p^*, given her reported preferences. We refer the reader to the original paper for full details.

Hylland and Zeckhauser (1979) prove existence of competitive equilibrium prices in a setting that is strictly more general than ours (in particular, they allow for indifferences).

For each market size n and each possible reported vector of types $t \in T^n$, choose one such price vector in an anonymous manner, and use this price vector to define the resulting allocation $\Phi^n(t)$. As Hylland and Zeckhauser (1979) observe on page 307, since each agent has the same budget and faces the same prices, such a mechanism is EF. Consequently, Theorem 1 guarantees that it is SP-L.

D.1.3 Multi-Unit Assignment

In multi-unit assignment problems, each agent is to be assigned a finite number of indivisible objects. Transfers of a numeraire are not allowed. A prototypical application is the allocation of courses to students at business schools. For further details we refer the reader to Budish (2011).

Denote the finite set of object types by J. Each object j is available in supply $\{q_j \cdot n\}$. A bundle $x_0 \in X_0 = \mathcal{P}(J)$ specifies a subset of the object types.[8] A type t_i specifies a utility function u_{t_i} over bundles. We will adopt the terminology of course allocation, denoting object types by courses, and agents by students.

HBS Draft Mechanism

The mechanism used by Harvard Business School to allocate MBA courses was studied empirically by Budish and Cantillon (2012). Using survey data, they showed that students often misreport their preferences. Here we formally define the mechanism and show that it is not SP-L.

The HBS draft mechanism does not allow students to express preferences over bundles of courses. Instead, students submit a preference ordering over single courses. To examine the possibility of truthful reporting, we restrict our attention to preferences over bundles that are responsive to preferences over individual courses, with preferences over individual courses strict. We will say that a student of type t_i prefers course j_A to course j_B if she prefers a bundle consisting only of course j_A to a bundle consisting only of course j_B, that is, $u_{t_i}(\{j_A\}) > u_{t_i}(\{j_B\})$.

The HBS draft mechanism works as follows. First, each student is assigned a lottery number l_i, uniformly distributed in $[0, 1]$. In the first round, students take turns ordered by their lottery number, with higher lottery numbers going first. At her turn, student i chooses her favorite course out of the ones that are still available. In round two, the same procedure is repeated, but with students with lower lottery numbers going first. The procedure is repeated in the following rounds, with higher lottery numbers going first in the odd rounds

[8] $\mathcal{P}(J)$ denotes the power set of J.

and last in the even rounds. The mechanism ends after k rounds, where k is the number of courses required per student.

To see that this mechanism is not SP-L, consider the following example based closely on Example 1 of Budish and Cantillon (2012). There are 4 proper courses, $J = \{j_A, j_B, j_C, j_D\}$, of which students require $k = 2$ courses each. Each course has capacity for $\frac{2}{3}$ of the population, that is $q_j = \frac{2}{3}$ for each $j \in J$. Consider a probability distribution over students' reports where $\frac{1}{3}$ of the population lists courses in the order j_A, j_B, j_C, j_D, $\frac{1}{3}$ lists courses in the order j_B, j_A, j_C, j_D, and $\frac{1}{3}$ lists courses in the order j_A, j_C, j_D, j_B. Given this distribution of reports, the probability that course j_A reaches capacity either in the end of the first round, or early in the second round converges to 1, as the market grows large. Therefore, a student that ranks course j_A as her first choice has probability close to 1 of receiving it, while a student who ranks j_A second has probability close to 0 of receiving it. In contrast, course j_B is very likely to reach capacity either late in the second round, or early in the third round, in a large market. Consequently, a student who ranks course j_B either first or second is very likely to receive it. For this reason, a student whose true preference order is j_B, j_A, j_C, j_D profits by misreporting as j_A, j_B, j_C, j_D. By doing so, the student receives both j_A and j_B, her two favorite courses, rather than courses j_B and j_C if she reports truthfully.[9]

The Bidding Points Auction Mechanism

The bidding points auction mechanism is used by several business schools to allocate MBA courses. It has been described by Sönmez and Ünver (2010) and Krishna and Ünver (2008), who demonstrated that the mechanism is flawed in several important ways, despite its widespread use. We now define the bidding points auction mechanism and show that it is not SP-L.

The mechanism works as follows. Students report vectors of bids, with one bid per course. Students can only spend up to a budget of B points, so that the set of actions is the set of all vectors of bids that sum to at most B. We restrict the bids to be integers, so that

$$A = \{a_i \in \{0, 1, \cdots, B\}^J_+ : \sum_j a_{i,j} \leq B\}.$$

Given a vector of bids, the mechanism starts with the highest bid and allocates the course to the student, as long as the course still has capacity. Ties are broken randomly.

To examine the possibility of truthful reporting, we assume that students' preferences are

[9]This particular profitable misreport is valid for any cardinal preferences consistent with the ordinal preferences j_B, j_A, j_C, j_D. In other examples the profitability of a particular misreport might depend on cardinal preference information.

additive, meaning that their utility for a bundle of courses is the sum of their utilities from the component courses in that bundle. This allows us to interpret a student's bid vector as an expression of their individual course preferences, and allows us to interpret the bidding points auction as a direct mechanism with $T = A$.

Consider the case where there are three courses, $J = \{j_A, j_B, j_C\}$. Consider an agent who likes the three courses j_A, j_B, j_C equally, and derives no utility of being unmatched. That is,

$$\begin{aligned} u_{t_i}(j_A) &= u_{t_i}(j_B) = u_{t_i}(j_C) = B/3, \\ u_{t_i}(\emptyset) &= 0. \end{aligned} \tag{D.4}$$

Consider a distribution of play m, such that, in the large-market limit, the last accepted bid for the courses j_A, j_B, j_C is $2B/3$ with very high probability. In that case, the agent should not report her true preferences, with bids equal to her utility. If bids are given by equation (D.4), then with very high probability the agent does not receive any course. If instead she bids B for one of the courses she likes, and 0 for the others, she receives at least one of the courses. Therefore, the mechanism is not SP-L.

Approximate Competitive Equilibrium from Equal Incomes (A-CEEI)

Budish (2011) proposed a pseudo-market mechanism for multi-unit assignment problems. Budish's setting is a strict generalization of ours. For that reason, we do not repeat all formal definitions, and refer the reader to the original paper for further details. In our setting, the A-CEEI mechanism can be defined as follows. First, assign each student a lottery number l_i uniformly and identically distributed in $[0, 1]$. Then give each student a budget in an imaginary currency of $1 + l_i \cdot \beta_{(n)}$, where $\beta_{(n)}$ is a strictly positive constant that is weakly decreasing in n, as defined in Budish (2011) page 1081. Budish's Theorem 1 guarantees that given these budgets there exists an approximate competitive equilibrium of the economy where agents purchase courses using the imaginary currency. The A-CEEI mechanism selects one such equilibrium, anonymously, and gives each agent his equilibrium allocation. This defines a function $x^n(\cdot, \cdot)$ giving an assignment of bundles $x^n(t, l) \in X_0^n$, for each vector of types t and lottery draws l. The A-CEEI mechanism is defined as

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

To show that this mechanism is SP-L, we use Theorem 1. By the definition of approximate competitive equilibrium (Budish's Definition 1), after lotteries are drawn, no agent envies another agent with a lower lottery number. Therefore, the CEEI mechanism is EF-TB, and therefore SP-L.

The Generalized Hylland and Zeckhauser Pseudo-Market

Budish et al. (2013) have proposed an extension of the Hylland and Zeckhauser pseudo-market mechanism that can be used for multi-unit assignment problems. In the simplest setting they consider, students have additive preferences over courses. We therefore assume that T only includes additive preferences. With this assumption, their setting is a strict generalization of ours. Budish et al. (2013) then formally define the mechanism. It works similarly to the Hylland and Zeckhauser mechanism, with students purchasing probability shares of courses using a fake currency. The mechanism then calculates a competitive equilibrium allocation of probability shares. Finally, the mechanism implements a lottery over allocations that gives each agent her equilibrium probability share. Budish et al.'s Theorem 6 and Corollary 3 guarantee that the mechanism is well-defined, as both an equilibrium exists and can be implemented by a lottery over feasible assignments. Budish et al.'s Theorem 8 shows that the mechanism is envy-free. Along with our Theorem 1, this implies that the mechanism is SP-L.

D.1.4 Exchange Economies

Walrasian Mechanism

A Walrasian mechanism implements competitive equilibrium allocations in an exchange economy. Several contributions in the literature have considered approximate incentive compatibility of Walrasian mechanisms in large markets, including the classic paper by Roberts and Postlewaite (1976). We refer the reader to Jackson and Manelli (1997) for an overview and references. We note that this example has an infinite set of bundles X_0, which does not fit the framework in the body of the paper. However, the mechanism fits the more general framework considered in Appendix 1.1.2, which allows us to use Theorem 1 to classify it as SP-L.

We consider an exchange economy with J goods. A type $t_i = (e_{t_i}, v_{t_i})$ specifies

- An endowment vector $e_{t_i} \in \mathbb{R}_+^J$.
- A continuous utility function v_{t_i} over bundles of goods in $\mathbb{R}_+^J$, taking values in $[0, 1]$.

Assume that the finite set of types T is such that, for any finite n and type vector $t \in T^n$, there always exists at least one competitive equilibrium where all agents of the same type receive the same bundle. This is guaranteed under standard assumptions on the set of utility functions and endowment vectors.

Given a type t_i, we define the utility function u_{t_i} over net trades $x_0 \in \mathbb{R}^J$ as

$$\begin{aligned} u_{t_i} &= v_{t_i}(e_{t_i} + x_0) \text{ if } e_{t_i} + x_0 \in \mathbb{R}^J_+ \\ & -\infty \text{ if } e_{t_i} + x_0 \notin \mathbb{R}^J_+. \end{aligned}$$

We let X_0 be $\mathbb{R}^J$, the set of all possible vectors of net trades.

Having defined X_0 and T, we now define the mechanism. For all n, t, $\Phi^n(t)$ anonymously selects a competitive equilibrium allocation of an economy with the n agents of types in the vector t, such that agents of the same type receive the same bundle, and assigns each agent i her vector of net trades in that equilibrium.

Note that the Walrasian mechanism is EF, as each agent receives her preferred vector of net trades given prices. Furthermore, while X_0 is not finite, it does satisfy the more general assumptions in Remark 1. Namely, X_0 is a measurable subset of Euclidean space, utility is measurable and bounded above by 1, and the utility of telling reporting truthfully is at least 0. Therefore, by Theorem 1, the Walrasian mechanism is SP-L.

D.2 Semi-Anonymous Mechanisms

Semi-anonymity generalizes anonymity to allow a mechanism to treat agents differently if they belong to identifiably distinct groups. Examples include treating men and women differently in a matching mechanism, and treating buyers and sellers differently in a double auction. While the body of the paper deals with the notationally simpler case of anonymous mechanisms, semi-anonymous mechanisms are analyzed in Appendix C. This subsection classifies some of these mechanisms.

D.2.1 Double Auctions

Double auctions have been extensively studied as a simplified model of price formation. We consider auctions where buyers and sellers submit bids, and prices are given as the average of marginal winning and losing bids. See for example Rustichini et al. (1994) for further details and references.

Types t_i specify whether an agent is a potential buyer or seller, and a value. That is, types specify the agent's group, which is $g_{t_i} = b(\text{uyer})$ or $s(\text{eller})$, and her value for the object, which is v_{t_i}. Sellers are endowed with a unit of the object, while buyers are not. The set of types is $T = G \times V$, with $G = \{b, s\}$ and $V = \{1, \cdots, \bar{v}\}$. A bundle x_0 specifies whether the agent trades or not, with a dummy $d_{x_0} = 0$ or 1, and the price of the transaction

$$p_{x_0} \in P = \{(p' + p'')/2 : p', p'' \in V\}.$$

We have $X_0 = \{0, 1\} \times P$. Buyers and sellers have quasilinear utility. The utility of a bundle is 0 if the agent does not trade. If the bundle prescribes a trade, utility is $v_{t_i} - p_{x_0}$ for a buyer, and $p_{x_0} - v_{t_i}$for a seller.

The mechanism works as follows. Given t, let $n_s(t)$ be the number of sellers, and therefore the number of objects. The market clearing price is the average of the $n_s(t)^{\text{st}}$ and $n_s(t)+1^{\text{st}}$ highest valuations. The mechanism assigns bundles x_0 with this price to all agents. The objects are assigned to the agents with the $n_s(t)$ highest valuations, with uniform tie-breaking for agents tied with the lowest winning valuation. Formally, the mechanism $\Phi^n(t)$ assigns bundles x_0 specifying trade to all buyers with valuations higher than the price, all sellers with valuations lower than the price, and randomly rations agents with valuations equal to the price.

Note that the mechanism is envy-free. This is so because all agents pay the same price, and therefore do not envy the price paid by other agents. Moreover, at this price, agents who trade with probability 1 would rather trade than not trade, and likewise agents that trade with probability 0 would rather not trade. Agents that are rationed are indifferent between trading or not trading, and therefore the mechanism is envy-free.[10] Therefore, double auctions are SP-L.

D.2.2 Matching

This setting is defined formally in Section C, Example C.1. That section also defines stable matching mechanisms, which are shown to be SP-L using a semi-anonymous version of the EF-TB condition.

Priority Match

Priority match mechanisms are described by Roth (1991), who proved that these mechanisms can produce unstable outcomes. Roth also documented that labor market clearinghouses using priority matching mechanisms were very likely to fail, and hypothesized that the reason why they failed is that they produce unstable outcomes.

The priority match works as follows. Given a man i (woman) and a woman (man) j define the rank of i on j's preferences as 1 plus the number of men (women) who are strictly

[10]Note that agents are only rationed in the case of a tie between the marginal winning and losing bids, and therefore both of these bids equal the price.

preferred to i. Assign to the pair i, j the priority $p_{i,j}$ equal to the rank of the man in the woman's preferences, times the rank of the woman in the man's preferences. The mechanism then proceeds by matching pairs with the lowest priorities first, breaking ties randomly.

To see that the priority match mechanism is not SP-L, consider the case where there is a single trait for men. Then women are indifferent over all men. In this case, the priority match mechanism coincides with the Boston mechanism, which is not SP-L.

It is interesting to note that Roth (1991) conjectured that the reason why stable matching mechanisms seem to succeed in practice, while priority matching mechanisms lead to unravelling and market failures, is stability. Our analysis, however, shows that stable matching mechanisms are SP-L, while priority matching mechanisms are not. Therefore, Roth's empirical finding can be phrased equivalently as saying that SP-L mechanisms succeed while non SP-L mechanisms fail.

References

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

_ , Parag A. Pathak, Alvin E. Roth, and Tayfun Sönmez, "Changing the Boston Mechanism: Strategyproofness as Equal Access," *Mimeo, Harvard University*, 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.

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

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

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

Budish, Eric B. and Estelle Cantillon, "The Multi-Unit Assignment Problem: Theory and Evidence from Course Allocation at Harvard," *American Economic Review*, 2012, *102(5)*, 2237–71.

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

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

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

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

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

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

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

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.

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.

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

Roth, Alvin E., "A Natural Experiment in the Organization of Entry Level Labor Markets: Regional Markets for New Physicians and Surgeons in the U.K.," *American Economic Review*, 1991, *81*, 415–40.

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

Sönmez, Tayfun and Utku Ünver, "Course Bidding at Business Schools," *International Economic Review*, 2010.

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