Strategyproofness in the Large

Eduardo M. Azevedo† and Eric Budish‡

September 18, 2012

Abstract

We propose a criterion of approximate strategyproofness, \textit{strategyproofness in the large} (SP-L). A mechanism is SP-L if, for any agent, any probability distribution of the other agents’ reports, and any $\epsilon > 0$, in a large enough market the agent maximizes his expected payoff to within $\epsilon$ by reporting his preferences truthfully. In large markets, SP-L maintains many of the formal advantages of SP over Bayes-Nash or Nash equilibria, such as compliance with the Wilson doctrine and strategic simplicity. Yet, we show that SP-L is in a certain sense approximately costless to impose relative to Bayes-Nash or Nash. We interpret these results as justifying SP-L as a second-best for market design problems where SP mechanisms are unappealing. Our classification of existing non-SP mechanisms into those that are SP-L and those that are not also supports the view that SP-L is a useful second-best to SP. Specifically, all of the known mechanisms for which there is a detailed theoretical case that the mechanism has approximate incentives for truth-telling in large markets are classified as SP-L, and all of the known mechanisms for which there is empirical evidence that non-strategyproofness causes serious problems even in large markets are classified as not SP-L.

†For helpful discussions we are grateful to Nabil Al-Najjar, Susan Athey, Aaron Bodoh-Creed, Gabriel Carroll, Jeff Ely, Alex Frankel, Drew Fudenberg, Matt Gentzkow, Jason Hartline, John Hatfield, Richard Holden, Ehud Kalai, Emir Kamenica, Fuhito Kojima, Scott Kominers, Jacob Leshno, Jon Levin, Paul Milgrom, 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 are grateful to seminar audiences at Ohio State, the MFI Conference on Matching and Price Theory, UCLA, Chicago, AMMA 2011, Boston College, the NBER Conference on Market Design, Duke / UNC, Michigan, Carnegie Mellon / Pittsburgh, Montreal, Berkeley, Northwestern, Rochester, Frontiers of Market Design 2012, and ACM EC 2012.

‡Wharton, eazevedo@wharton.upenn.edu.

University of Chicago Booth School of Business, eric.budish@chicagobooth.edu.
1 Introduction

Strategyness, that playing the game truthfully is a dominant strategy, is perhaps the central notion of incentive compatibility in market design. Strategyness (SP) is frequently imposed as a design requirement in theoretical analyses, across a broad range of assignment, auction, and matching problems. And, SP played a central role in several recent real-world design reforms, including the redesign of school choice mechanisms in several cities, the redesign of the market that matches medical school graduates to residency positions, and the efforts to create mechanisms for pairwise kidney exchange (Roth, 2008).

There are several important reasons why SP is so heavily emphasized relative to other forms of incentive compatibility, such as Bayes-Nash or Nash. First, SP mechanisms are robust in the sense of Wilson (1987) and Bergemann and Morris (2005). Since reporting truthfully is a dominant strategy, equilibrium does not depend on agents’ beliefs about other agents’ preferences or information. Second, SP mechanisms are strategically simple for participants. Participants do not have to invest time and effort collecting information about others’ preferences or about what equilibrium will be played (Fudenberg and Tirole, 1991, pg. 270; Roth, 2008). 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 players (Friedman, 1991; Pathak and Sönmez, 2008; Abdulkadiroğlu et al., 2006).

However, in a wide variety of contexts, impossibility theorems indicate that strategyness severely limits what kinds of mechanisms are possible. These include Gibbard (1973) and Satterthwaite’s (1975) dictatorship theorem for general social choice problems, Hurwicz’s (1972) and Barbera and Jackson’s (1995) impossibility theorems for general equilibrium settings, the Green and Laffont (1977) VCG theorem for allocation settings with quasi-linear preferences (see also Ausubel and Milgrom (2006)), Roth’s (1982) impossibility theorem for strategyproof stable matching, Papai’s (2001) and Ehlers and Klaus’s (2003) dictatorship theorems for multi-unit demand assignment problems, Abdulkadiroğlu et al.’s (2009) impossibility theorem for strategyproof and efficient school assignment, and many others. Thus, while attractive, SP is often very costly to satisfy.

This paper proposes a criterion of approximate strategyness, and suggests that it may be a useful second-best to SP. We say that a mechanism is strategyness in the large (SP-L) if, for any agent, any full-support probability distribution of the other agents’ reports, and any $\epsilon > 0$, in a large enough market the agent maximizes his expected payoff to within $\epsilon$ by reporting his preferences truthfully. Heuristically, 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 or Walrasian mechanism, or price-like statistics in an assignment or matching mechanism (e.g., Che and Kojima, 2010; Budish, 2011; Azevedo and Leshno, 2011) – can do no better than to report her preferences truthfully. In large markets, SP-L mechanisms have many of the advantages of SP mechanisms that we described above. Specifically, SP-L mechanisms are in a certain sense compliant with the Wilson doctrine, strategically simple, and fair to unsophisticated participants. Yet our main theoretical result shows that, in large markets, SP-L is in a certain sense costless to satisfy relative to Bayes-Nash or Nash. That is, SP-L approximates the benefits of exact SP but without the costs.

Our classification of existing non-SP mechanisms into those that are SP-L and those that are not also supports the view that SP-L is a useful second-best to SP. SP-L mechanisms include several well-known mechanisms for which there is a detailed theoretical case that the mechanism has 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), Uniform-Price Auctions (Swinkels, 2001), and the Gale-Shapley Deferred Acceptance Algorithm (Immorlica and Mahdian, 2005; Kojima and Pathak, 2009). On the other hand, for each mechanism in our classification that is not SP-L, there is empirical evidence suggesting that agents strategically misreport their preferences, and that this misreporting harms design objectives such as efficiency or fairness. Examples include pay-as-bid treasury 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). Furthermore, both Milton Friedman’s critique of the pay-as-bid auction and Alvin Roth’s critique of the priority-match mechanism explicitly suggested alternative mechanisms that are not SP but that are SP-L: the uniform-price auction and deferred-acceptance mechanism, respectively. This too speaks to the intuitive appeal of the criterion.

SP-L is subtly weaker than the standard notion of approximate strategyproofness, because SP-L requires that truthful reporting is approximately optimal for any full-support probability distribution of opponent reports, as opposed to for any realization of opponent reports. At the same time, SP-L is importantly stronger than approximate Bayes-Nash incentive compatibility, which requires that truthful reporting is approximately optimal only for

---

1To the best of our knowledge, there are no empirical examples of market designs that are SP-L but which have been shown to be harmfully manipulated in large finite markets.
the equilibrium distribution of opponent reports. This positioning “in between” approximate strategyproofness and approximate Bayes-Nash is important for the results. Approximate Bayes-Nash is too weak to obtain the results on compliance with the Wilson doctrine, strategic simplicity, and fairness to unsophisticated agents. Approximate strategyproofness is too strong, both for obtaining our main theoretical result, and for obtaining our useful classification of non-SP mechanisms. For instance, the uniform-price auction is SP-L (cf. Example 1 below) but does not satisfy a stronger notion of approximate strategyproofness based on realizations of opponent reports. 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.

In addition to being necessary for the results, we view the any-probability-distribution approach as economically compelling. In a market with hundreds or thousands of participants, it seems implausible for a participant to know the exact profile of opponent play, but realistic to have an estimate of the distribution of opponent play. At the same time, if possible we do not want to design mechanisms that are sensitive to every participant having the exact same, common knowledge, estimate of this distribution.

We now describe our main theoretical result in more detail. Suppose we are given some mechanism that has Bayes-Nash (or Nash) equilibria but is not SP-L. Suppose that the mechanism is (semi-)anonymous, which is a common feature of practical market-design settings; that agents have private values, in the sense that they know their own preferences over outcomes without observing other agents’ private information; and that the mechanism and its equilibria satisfy a condition called quasi-continuity, which is roughly a continuity-almost-everywhere condition, allowing for the kinds of “knife edge” discontinuities that are common in discrete-goods allocation problems. We show that there necessarily exists another mechanism that is SP-L, and that implements approximately the same outcomes as the original mechanism, with the approximation error vanishing in the large-market limit. Thus, under suitable conditions, SP-L is approximately costless to impose relative to Bayes-Nash.

The proof is by construction of a specific SP-L mechanism, from a given mechanism that has Bayes-Nash equilibria. The construction works as follows. Agents report their types to our mechanism. Our mechanism then calculates the empirical distribution of these types, and then “activates” the Bayes-Nash equilibrium strategy of the original mechanism associated with this empirical distribution. In other words, it acts as a proxy agent playing the original mechanism on each agent’s behalf, using a strategy endogenously determined
by the empirical distribution of all reports. If agents all report their preferences truthfully, this construction will yield the same outcome as the original mechanism in the large-market limit, because the empirical distribution of reported types converges to the underlying true distribution. The subtle part of our construction is what happens if some agents misreport their preferences. Suppose the true distribution of preferences is \( \mu \), but for some reason the agents other than agent \( i \) systematically misreport their preferences, according to distribution \( m \). In a finite market, with sampling error, the empirical distribution of the other agents’ reports is say \( \hat{m} \). As the market grows large, \( \hat{m} \) is converging to \( m \), and also \( i \’s \) influence on the empirical distribution is vanishing. Thus in the limit, our construction will activate the Bayes-Nash equilibrium strategy associated with \( m \). This is the “wrong” prior – but agent \( i \) does not care. From his perspective, the other agents are reporting according to \( m \), and then playing the Bayes-Nash equilibrium strategy associated with \( m \), so \( i \) too wishes to play the Bayes-Nash equilibrium strategy associated with \( m \). This is exactly what our constructed mechanism does on \( i \’s \) behalf in the limit. Hence, no matter how the other agents play, \( i \) wishes to report his own type truthfully in the limit, i.e., the constructed mechanism is SP-L.

Our construction resembles a revelation principle construction (Myerson, 1979), in that it takes a mechanism in which agents play the game directly and produces a mechanism in which agents just report their type, and then let the center play optimally on their behalf. However, we emphasize that our construction is fundamentally distinct. In a traditional revelation mechanism, the mechanism designer knows the true prior (e.g., \( \mu \)), and then plays the Bayes-Nash equilibrium strategy associated with this true prior on agents’ behalf. It is then a Bayes-Nash equilibrium for agents to report their types truthfully. Our mechanism has two advantages relative to this benchmark. First, our mechanism is prior free: neither the agents nor the mechanism designer need know the underlying distribution of preferences a priori, because the mechanism infers the prior \( \mu \) from the empirical distribution of preferences. Second, our mechanism provides dominant-strategy incentives in the limit, whereas a traditional revelation mechanism provides just Bayes-Nash incentives even in the limit.

The most direct application of our analysis is as justification for designing an SP-L mechanism when facing a market design problem for which there are known impossibility theorems for SP. Another application of our analysis is to a recent debate in the market

---

2Imagine the proxy making the following announcement: “I do not know the distribution of preferences, and presumably neither do you. But whatever the distribution happens to be, I will play the Bayes-Nash strategy on your behalf.” Note that this construction does not provide exact incentives in finite markets. In particular, agent \( i \’s \) report affects the empirical distribution of all reports, which in turn affects what action the proxy plays on agent \( j \’s \) behalf.

3See the Conclusion for further discussion of this point.
design literature concerning the “Boston mechanism” for school assignment. The first papers on the Boston mechanisms criticized it for not being strategyproof (indeed it is not even SP-L), and suggested that a strategyproof mechanism be used instead; this advice was then acted on in practice by the Boston Public School authority (Abdulkadiroğlu and Sönmez, 2003; Abdulkadiroğlu et al., 2006; Roth, 2008). A second generation of papers argued that the Boston mechanism, while not strategyproof, has Bayes-Nash equilibria which generate greater welfare for students than do strategyproof alternatives (Abdulkadiroğlu et al., 2011; Miralles, 2009; Featherstone and Niederle, 2011). Of course, the Bayes-Nash equilibria these papers rely on to make their argument require students to have common knowledge of the preference distribution, coordinate on a specific equilibrium, make very precise strategic calculations, etc. Our analysis says that all of this complexity and non-robustness is unnecessary in a large market. Specifically, there must exist yet another mechanism that implements the same outcomes as these desirable Bayes-Nash equilibria, but that is SP-L. Moreover, our proof explicitly shows how to construct such a mechanism.

**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. An early paper in this tradition is Roberts and Postlewaite (1976) on the Walrasian mechanism, which can be seen as a response to Hurwicz’s (1972) critique that the Walrasian mechanism is not strategyproof. We discuss this literature in more detail in Section 4.2. An important thing to highlight here is that the aim of our paper is quite different from, and we think complementary to, papers in 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.

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. Another recent 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. We view this literature as providing support for the concept of SP-L, since all of the examples we are aware of in
which there is evidence of substantial harm from misreporting involves mechanisms that not only are not SP, but are not even SP-L.

Third, our main theoretical result, Theorem 1, relates to several influential theoretical ideas. We discuss its relationship to the revelation principle (Myerson, 1979) in Section 5.4.1, its relationship to Kalai (2004)’s study of the robustness of Bayes-Nash equilibria in large markets in Section 5.4.2, and its relationship to the random sampling method (e.g., Goldberg et al., 2001) in Section 5.4.3.

Fourth, our paper is related to the literature on the role of strategyproofness in practical market design. Wilson (1987) famously argued that practical market designs should aim to be robust to agents’ beliefs, and Bergemann and Morris (2005) formalized the sense in which strategyproof mechanisms are robust in the sense of Wilson. Several recent papers have argued that strategyproofness can be viewed as a design objective and not just as a constraint: papers on this theme include Abdulkadiroğlu et al. (2006), Abdulkadiroğlu et al. (2009), Pathak and Sönmez (2008), and Roth (2008). Our paper contributes to this literature by showing that our notion of SP-L approximates the formal appeal of SP, while at the same time being approximately costless to impose relative to other kinds of incentive compatibility. 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. See our discussion of one such mechanism, the Boston mechanism for school choice, in Section 5.4.4.

Last, our paper is conceptually related to Parkes et al. (2001), Day and Milgrom (2008), Erdil and Klemperer (2010), Pathak and Sönmez (Forthcoming), and Carroll (2011), each of which seeks to say something more useful about non-strategyproof mechanisms than simply that they are not strategyproof.4 Parkes et al. (2001), Day and Milgrom (2008) and Erdil and Klemperer, 2010 propose cardinal measures of a combinatorial auction’s manipulability, and seek to design auctions that minimize manipulability subject to other design objectives. Carroll (2011), too, proposes a cardinal measure of manipulability, and explores his measure in the context of voting problems. He derives comparisons amongst voting rules, and asymptotic lower bounds on the manipulability of any rule satisfying other desiderata. Another notable feature of (Carroll, 2011) is that he, like us, adopts an ex-interim perspective for analyzing a mechanism’s manipulability, and restricts attention to conditionally IID beliefs. (The two papers are essentially contemporaneous, with early drafts of our paper circulating just a few months before Carroll (2011)). Pathak and Sönmez (Forthcoming) propose a par-

---

4 See also Milgrom (2011) Section IV for a general discussion of these issues.
tial order by which to compare non-strategyproof mechanisms based on their vulnerability to manipulations. Mechanism $a$ is said to be more manipulable than mechanism $b$ if, for any problem instance where $b$ is manipulable by at least one agent, so too is $a$. This criterion helps to explain several recent policy decisions in which school authorities switched from one manipulable mechanism to another. We view our approach as complementary to these alternative approaches. An advantage of our approach is that it yields an explicit design desideratum, namely that mechanisms be strategyproof in the large.

**Organization of the paper** The rest of this paper is organized as follows. Section 2 describes the environment and some key assumptions. Section 3 defines strategyproofness in the large, and gives a detailed example. Section 4 presents our classification of non-SP mechanisms. Section 5 presents our main theoretical result on constructing SP-L mechanisms from Bayes-Nash mechanisms. Section 6 concludes. Proofs and other supporting materials are in the appendix.

## 2 Environment

### 2.1 Preliminaries

There is a finite set of (payoff) types $T$, and a finite set of outcomes $X_0$. 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 payment she makes. In school assignment, $X_0$ is the set of schools to which a student can be assigned. An agent’s type determines her preferences over outcomes. Specifically, for each $t_i \in T$ there is a von Neumann-Morgenstern expected utility function $u_{t_i} : X \rightarrow [0, 1]$, where $X = \Delta X_0$ denotes the set of lotteries over outcomes. Preferences are private values in the sense that an agent’s utility from her outcome depends only on her own type.

We study mechanisms that are well defined for all possible market sizes, with $X_0$ and $T$ being the same for all market sizes. For each market size $n \in \mathbb{N}$, where $n$ denotes the number of agents, there is an arbitrary set $Y_n \subseteq (X_0)^n$ of feasible allocations. For instance, in an auction or assignment setting, our 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 large.

**Definition 1.** Fix an outcome set $X_0$, a type set $T$, and a sequence of feasibility constraints $(Y_n)_\mathbb{N}$. A mechanism $\{(\Phi^n)_\mathbb{N}, A\}$ consists of a finite action space $A$ and a sequence of
allocation functions
\[ \Phi^n : A^n \to \Delta((X_0)^n), \] (2.1)

each of which satisfies feasibility: for any \( n \in \mathbb{N} \) and \( a \in A^n \), the support of \( \Phi^n(a) \) lies in the feasible set \( Y_n \).

We assume that mechanisms are **anonymous**, which requires that each agent’s outcome does not depend on her identity. Formally, a mechanism is anonymous if, for all \( n \in \mathbb{N} \), the allocation function \( \Phi^n(\cdot) \) is invariant to permutations. That is, for any \( n \), any \( a \in A^n \), and any permutation function \( \pi : \{1, \ldots, n\} \to \{1, \ldots, n\} \), we have \( \Phi^n(a) = \pi^{-1}(\Phi^n(\pi(a))) \).

Anonymity is a natural feature of many large-market settings. In Appendix C.1 we show our main result obtains if we relax anonymity to semi-anonymity as defined by Kalai (2004); semi-anonymity accommodates many additional settings in which there are asymmetries amongst classes of participants, e.g., double auctions in which there are distinct buyers and sellers, and certain kinds of two-sided matching markets.

In what follows, it will often be useful to view mechanisms from the perspective of a typical agent \( i \). Define the function
\[ \Phi^n_i : A \times A^{n-1} \to X, \] (2.2)

where \( \Phi^n_i(a_i, a_{-i}) \) denotes the marginal distribution of \( \Phi^n(a_i, a_{-i}) \) in the \( i \)-th dimension. That is, the lottery over bundles agent \( i \) receives when she plays \( a_i \) and the other agents play \( a_{-i} \). It is also useful to extend the functions \( \Phi^n \) and \( \Phi^n_i \) linearly to distributions over vectors of actions. Given a distribution \( \bar{m} \in \Delta(A^n) \) over vectors of actions, let
\[ \Phi^n(\bar{m}) = \sum_a \Phi^n(a) \cdot \bar{m}(a). \] (2.3)

Likewise, given an action \( a_i \) and a distribution over \( n - 1 \) actions \( \bar{m} \in \Delta(A^{n-1}) \), let
\[ \Phi^n_i(a_i, \bar{m}) = \sum_{a_{-i}} \Phi^n_i(a_i, a_{-i}) \cdot \bar{m}(a_{-i}). \]

### 2.2 Limit Mechanisms

We now define a key piece of notation, which is an ex- interim version of the individual allocation function defined in Equation (2.2). Consider a mechanism \( \{ (\Phi^n)_{\mathbb{N}}, A \} \), a market size \( n \), an action \( a_i \in A \), and a distribution over actions \( m \in \Delta A \). Let:
\[ \phi^n(a_i, m) = \sum_{a_{-i}} \Phi^n_i(a_i, a_{-i}) \cdot \Pr(a_{-i} | a_{-i} \sim \text{iid}(m)) \]  

(2.4)

where \( \Pr(a_{-i} | a_{-i} \sim \text{iid}(m)) \) denotes the probability that the action vector \( a_{-i} \) is realized given \( n - 1 \) independent identically distributed (iid) draws from the action distribution \( m \).

The object \( \phi^n(a_i, m) \) describes what a generic agent can expect to receive under mechanism \( \{(\Phi^n)_N, A\} \) when he plays action \( a_i \) and the other \( n - 1 \) agents play iid according to \( m \). Since each \( \Phi^n_i(a_i, a_{-i}) \) is a random outcome in \( X \equiv \Delta X_0 \), and \( X \) is convex, the object \( \phi^n(a_i, m) \) is also a random outcome in \( X \). Note that we do not use a subscript \( i \) for the function \( \phi^n(\cdot, \cdot) \), to highlight that the function does not depend on the identity of the agent, due to anonymity. We use the function \( \phi^n(\cdot) \) to define limit mechanisms.

**Definition 2.** The function \( \phi^\infty : A \times \Delta A \rightarrow X \) is the limit of mechanism \( \{(\Phi^n)_N, A\} \) if, for all \( a_i, m \):

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

where \( \phi^n \) is as defined in (2.4).

In words, \( \phi^\infty(a_i, m) \) describes what a generic agent who plays \( a_i \) receives in the large market limit of mechanism \( \{(\Phi^n)_N, A\} \), when the other agents’ play is iid according to \( m \). The randomness in how we take the large-market limit is in contrast with Debreu and Scarf’s (1963) deterministic replicator economy, and with the approach pioneered by Aumann (1964) that looks directly at a continuum economy without explicitly modeling finite economies. It is more closely inspired by the random economy method used in Immorlica and Mahdian (2005)’s and Kojima and Pathak (2009)’s studies of large matching markets.

This randomness in our definition of the limit will be useful for two reasons. First, it allows the limit \( \phi^\infty(\cdot, m) \) to be well defined for every distribution \( m \in \Delta A \), whereas a deterministic limit such as that in Debreu and Scarf (1963) is well-defined only if the coordinates of \( m \) are rational numbers. Second, it ensures that any specific empirical distribution of opponent play becomes increasingly rare as the market grows large. For instance, if there are two actions \( A = \{\text{Heads}, \text{Tails}\} \), for any iid distribution \( m \in \Delta A \), including say \( m = (.5, .5) \), the likelihood that exactly 50% of ones’ opponents play Heads goes to zero as \( n \) goes to infinity. By contrast, in a deterministic limit, each agent is assumed to know the precise empirical distribution of opponent play at each possible market size.

We note that it is very easy to construct examples of mechanisms that do not have limits. For instance, if a mechanism behaves like a uniform-price auction when \( n \) is even and like a
pay-as-bid auction when \( n \) is odd it will not have a limit. For the remainder of the analysis we impose some regularity on how the functions \( \Phi^n \) vary with market size by limiting attention to mechanisms that have limits.

## 3 Strategyproof in the Large

A mechanism is strategyproof if it is optimal for each agent to report truthfully, in any size market, given any vector of reports by her opponents.

**Definition 3.** Mechanism \( \{(\Phi^n)_{\mathbb{N}}, T\} \) is strategyproof, or \( \text{SP} \), if for all \( n \), all \( t_i, t_i' \in T \), and all \( t_{-i} \in T^{n-1} \):

\[
 u_{t_i}[\Phi^n_i(t_i, t_{-i})] \geq u_{t_i}[\Phi^n_i(t_i', t_{-i})]
\]

We say that a mechanism is strategyproof in the large if it is optimal for each agent to report truthfully, in our large-market limit, given any full-support distribution of reports by her opponents.

**Definition 4.** Mechanism \( \{(\Phi^n)_{\mathbb{N}}, T\} \) is strategyproof in the large, or \( \text{SP-L} \), if, for any full support distribution of types \( m \in \bar{\Delta} T \), and any \( t_i, t_i' \):

\[
 u_{t_i}[\phi^\infty(t_i, m)] \geq u_{t_i}[\phi^\infty(t_i', m)].
\] (3.1)

Equivalently, if for any \( \epsilon > 0 \), there exists \( n_0 \) such that if \( n > n_0 \) we have

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

Otherwise, the mechanism is manipulable in the large.

SP-L is weaker than the standard notion of approximate strategyproofness because it requires that truthful reporting is approximately optimal, in a large enough market, for any full-support probability distribution of opponent reports \( m \), as opposed to for any realization of opponent reports \( t_{-i} \).\(^5\) If a mechanism is SP-L, truthful reporting is approximately optimal for agents with quite detailed information about opponent play, as given by the ex-interim distribution of opponent play \( m \), but not necessarily for agents with exact knowledge about

\(^5\)Formally, a mechanism satisfies what we refer to as the standard notion of approximate strategyproofness if, for any \( t_i, t_i' \), and any \( \epsilon > 0 \), there exists \( n_0 \) such that if \( n > n_0 \), for any \( t_{-i} \in T^{n-1} \) we have:

\[
 u_{t_i}[\Phi^n_i(t_i, t_{-i})] \geq u_{t_i}[\Phi^n_i(t_i', t_{-i})] - \epsilon.
\] See Roberts and Postlewaite (1976) for a definition in this spirit for exchange economies, and Schummer (2004) for a quasilinear setting.
the ex-post realization of opponent play $t_{-i}$. Our notion of a large market is one where there is both a large number of players, and where players do not possess such precise information about the exact realization of opponent reports.

At the same time, SP-L is importantly stronger than approximate Bayes-Nash incentive compatibility, and SP-L mechanisms share many of the formal advantages that SP mechanisms have over Bayes-Nash incentive compatible mechanisms.\(^6\) First, in the spirit of the Wilson doctrine as formalized by Bergemann and Morris (2005), in an SP-L mechanism agents’ behavior and hence mechanism outcomes are approximately insensitive to agents’ beliefs. Specifically, reporting truthfully is approximately optimal for any beliefs $m \in \Delta T$ about the distribution of opponent play. Neither participants nor the mechanism designer need have common knowledge of the distribution of preferences, unlike in a Bayes-Nash mechanism.\(^7\) Second, SP-L mechanisms are strategically straightforward in the following sense: for any beliefs $m \in \Delta T$, a player may simply play truthfully, and she is guaranteed a nearly optimal payoff. In particular, if strategizing involves any cost $c > 0$, be it the cost of gathering information about the rules of the game, or about how opponents are likely to play, it is optimal to simply report truthfully in a large enough market. Third, SP-L mechanisms treat unsophisticated players fairly in that they are guaranteed an almost optimal payoff.\(^8\)

3.1 Example: Multi-Unit Auctions

As an example to clarify the definition of SP-L, consider multi-unit auctions for identical goods, such as government bond auctions. The two most commonly used auction formats in practice are uniform-price auctions and pay-as-bid auctions. While neither mechanism is strategyproof (Ausubel and Cramton, 2002), Milton Friedman famously argued in favor of the uniform price auction on incentive grounds (Friedman, 1960, 1991). We will show that uniform-price auctions are SP-L, whereas pay-as-bid auctions are manipulable in the large.\(^9\)

\(^6\)Formally, we say that a direct mechanism is approximately Bayes-Nash incentive compatible at a prior $\mu_0 \in \Delta T$ if, for any $t_i, t_i'$, and any $\epsilon > 0$, there exists $n_0$ such that if $n > n_0$ we have: $u_{t_i}[\phi^n(t_i, \mu_0)] \geq u_{t_i}[\phi^n(t_i', \mu_0)] - \epsilon$. See (Swinkels, 2001) for a definition along these lines for uniform-price auctions, and (Kojima and Pathak, 2009) for deferred acceptance.

\(^7\)Bergemann and Morris (2005) formalize the Wilson doctrine as requiring that equilibrium outcomes of a mechanism do not depend on participants’ beliefs (more precisely, equilibrium outcomes depend on agents’ payoff types but not their belief types). SP mechanisms satisfy the Wilson doctrine as defined this way because truthful reporting is optimal for any beliefs. In an SP-L mechanism, truthful reporting is approximately optimal for a wide range of beliefs.

\(^8\)By unsophisticated players we mean players who are able to express their own preferences but who do not have the information or strategic sophistication to misreport their preferences optimally. An example is the parents that choose dominated strategies in school choice mechanisms, as in Pathak and Sönmez (2008).

\(^9\)Pathak and Sönmez (Forthcoming) provide a complementary perspective on the incentive comparison between the uniform-price and pay-as-bid auctions. Pathak and Sönmez (Forthcoming) show that any
In addition to illustrating the basic distinction between SP-L and manipulable in the large, the example highlights the role of several key aspects of the definition of SP-L, including the ex-interim as opposed to ex-post perspective, and the importance of the assumption of full support.

**Example 1. (Multi-Unit Auctions).** There are \( kn \) units of a homogeneous good. To simplify notation, we assume that agents assign a constant per-unit value to the good, up to a capacity limit. Specifically, each agent \( i \)'s type \( t_i \) 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 \( X_0 = (V \times Q) \cup \{0\} \), 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 \( p \).

Formally, \( D(p; t) = \sum_{i=1}^{n} q_i \cdot 1\{v_i \geq p\} \), where the notation \( 1\{\text{statement}\} \) denotes the indicator function that returns 1 if the statement is true and 0 if the statement is false.

That is, \( p^\ast(t) \) is the highest price at which demand weakly exceeds supply. The uniform-price auction allocates each bidder \( i \) her demanded quantity at \( p^\ast(t) \), with the exception that bids with \( v_i = p^\ast(t) \) are rationed with equal probability. Formally, \( \Phi^\ast_n(t) \) allots each bidder the following number of units of the good, at a price per unit of \( p^\ast(t) \). The probability of a bid being rationed \( \bar{r} \) is set to clear the market.

<table>
<thead>
<tr>
<th>Reported Value</th>
<th>Expected Number of Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>( v_i &lt; p^\ast(t) )</td>
<td>0</td>
</tr>
<tr>
<td>( v_i = p^\ast(t) )</td>
<td>( \bar{r} \cdot q_i )</td>
</tr>
<tr>
<td>( v_i &gt; p^\ast(t) )</td>
<td>( q_i )</td>
</tr>
</tbody>
</table>

agent who can profitably manipulate the uniform-price auction in a given finite economy can also profitably manipulate the pay-as-bid auction in that same finite economy. Moreover, the latter manipulation is always larger in utility terms. Thus, Pathak and Sönmez (Forthcoming) suggests that the pay-as-bid auction is more manipulable than the uniform-price auction in any given finite economy, whereas our analysis highlights that the pay-as-bid auction’s manipulability persists even in the large-market limit, whereas the uniform-price auction is strategyproof in the large.

\(^{10}\)Formally, \( D(p; t) = \sum_{i=1}^{n} q_i \cdot 1\{v_i \geq p\} \), where the notation \( 1\{\text{statement}\} \) denotes the indicator function that returns 1 if the statement is true and 0 if the statement is false.

\(^{11}\)Since preferences are linear up to the capacity limit, the exact form of the rationing is immaterial in the
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 \( m \):

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

(3.3)

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 case, as the market grows large, the realized price as defined in (3.2) will be equal to \( \rho^*(m) \) with probability converging to one. Therefore, the limit mechanism allocates each bidder their demand at \( \rho^*(m) \), again with the exception that bidders with value exactly equal to \( \rho^*(m) \) are rationed, and with all winning bidders paying \( \rho^*(m) \) per unit. Formally, \( \phi^\infty(t_i, m) \) gives player \( i \)

<table>
<thead>
<tr>
<th>Reported Value</th>
<th>Expected Number of Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>( v_i &lt; \rho^*(m) )</td>
<td>0</td>
</tr>
<tr>
<td>( v_i = \rho^*(m) )</td>
<td>( \bar{r} \cdot q_i )</td>
</tr>
<tr>
<td>( v_i &gt; \rho^*(m) )</td>
<td>( q_i )</td>
</tr>
</tbody>
</table>

at a per unit price of \( \rho^*(m) \), and the rationing probability \( \bar{r} \) is set so that the market clears on average.\(^\text{12}\) 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 also 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 \( \frac{1}{2} \).\(^\text{13}\) Therefore, the price in our limit will be \( \rho^*(m) \) with probability

\[
\bar{r} = \frac{kn - D(p + 1; t)}{D(p; t) - D(p + 1; t)}.
\]

\(^\text{12}\) That is, \( \bar{r} \) satisfies

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

\(^\text{13}\) The intuition is that if a fair coin is tossed \( n \to \infty \) times, the probability that \( \geq \frac{n}{2} \) of the tosses are
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,

<table>
<thead>
<tr>
<th>Reported Value</th>
<th>Expected Number of Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>$v_i &lt; \rho^*(m)$</td>
<td>0</td>
</tr>
<tr>
<td>$v_i \geq \rho^*(m)$</td>
<td>$q_i$</td>
</tr>
</tbody>
</table>

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

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

Note that our argument would not go through if we used a stronger notion of approximate strategyproofness 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 (3.1) to obtain for all probability distributions $m \in \Delta T$, rather than for all full support probability distributions $m \in \Delta T$. Full support ensures that the probability that any particular bidder is pivotal goes to zero as the market grows large.\footnote{In the uniform-price auction, it is possible to replace the full support assumption with a weaker assumption that still delivers that demand is not deterministic at the market-clearing price. The reason we use full support in the definition of SP-L is that it is a simple assumption on the support of $m$ that delivers approximate incentive compatibility for a wide variety of mechanisms.} 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 $v_i > \rho^*(m) + 1$ heads converges to $\frac{1}{2}$, just as the probability that $< \frac{1}{2}$ of the tosses are heads converges to $\frac{1}{2}$, with both probabilities independent of the outcome of the $i^{th}$ toss.
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. Notice that $t_i$’s optimal misreport depends on $m$, and that an unsophisticated bidder who bids truthfully can suffer a large loss relative to optimal behavior. Hence, in contrast to the uniform-price auction, the pay-as-bid auction is neither strategically simple nor fair to unsophisticated bidders. □

The fact that uniform price auctions are SP-L while pay-as-bid auctions are not is consistent with Friedman’s critique of pay-as-bid auctions, and with intuition. In the next Section we apply the SP-L versus Manipulable in the Large classification to a much broader set of market design problems.

4 Classification of Non-Strategyproof Mechanisms

This Section classifies a wide variety of non-strategyproof mechanisms into SP-L and Manipulable in the Large (Table 1), including the multi-unit auctions analyzed in Example 1 as well as other auction, single- and multi-unit assignment, and matching mechanisms, and discusses how the classification organizes the extant theoretical and empirical 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 truth-telling in large markets are classified as SP-L (Section 4.2), and all of the known mechanisms for which there is empirical evidence that non-strategyproofness causes serious problems even in large markets are classified as Manipulable in the Large (Section 4.3). We regard this evidence as further support for the appeal of SP-L as a second-best to 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., the multi-unit auctions), as well as mechanisms that do not use prices (e.g., the 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 (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 strategyproofness. 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 produce an example of a profitable manipulation, as we did for pay-as-bid auctions. For SP-L mechanisms, this Section gives two simple sufficient conditions for a mechanism to be SP-L, which directly imply the classification for all of the SP-L mechanisms in Table 1. Formal definitions of each mechanism and detailed derivations are in Appendix B.

The first sufficient condition is envy-freeness, that no player \( i \) prefers the assignment of another player \( j \), for any realization of the reported types \( t \).

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

\[
 u_{ti}[\Phi^n_i(t)] \geq u_{ti}[\Phi^n_j(t)].
\]

The connection between envy-freeness, which is usually thought of as a fairness criterion, and the incentives criterion of SP-L is the following. In anonymous mechanisms, the gain

\[\text{footnote 5 for a formal definition of approximate strategyproofness. See footnote 19 for a detailed discussion of the approximate incentive compatibility concepts used in previous studies of the mechanisms we classify as SP-L.}\]
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 (Equation (A.3) in Appendix A). Envy-freeness directly implies that the first component in this decomposition is non-positive. A simple probabilistic argument then establishes that the second component becomes negligible in large markets, which establishes that EF implies SP-L. Most of the mechanisms in the SP-L column of Table 1 are envy-free, with the only exceptions being Approximate CEEI and Deferred Acceptance. Fortunately, these mechanisms satisfy a weakening of envy-freeness that we show is also sufficient. Specifically, each of these mechanisms involves a certain form of uniform random tie breaking, and after the lottery is realized no agent envies another agent with a lower lottery number. Formally,

**Definition 6.** A direct mechanism $\{(\Phi^n)_n, T\}$ is envy-free but for tie breaking (EF-TB) if for each $n$ there exists a function $x^n : (T \times [0, 1])^N \to \Delta(X^n_0)$, 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

$$u_{t_i}[x^n_i(t, l)] \geq u_{t_j}[x^n_j(t, l)].$$

In words, Definition 6 asks that a mechanism can be written in terms of an allocation function $x^n(\cdot, \cdot)$ that depends on both agents’ reports and the outcome of $n$ uniform random lottery draws, such that the allocation function obeys the following weakening of envy-freeness: if agent $i$ has a better lottery draw than agent $j$, then $i$ does not envy $j$. The following Proposition shows that either condition guarantees that a mechanism is SP-L.

**Proposition 1.** If a mechanism is EF-TB (in particular if it is envy-free), then it is SP-L.

The proof of the Proposition is in Appendix A.1. The appendix also shows that incentives to misreport vanish at a rate of $n^{1/2-\epsilon}$ for envy-free mechanisms, and $n^{1/4-\epsilon}$ for EF-TB

---

16 Both Approximate CEEI and Deferred Acceptance include as a special case the random serial dictatorship mechanism, which Bogomolnaia and Moulin (2001) showed is not envy-free.

17 The definition for semi-anonymous mechanisms, which is needed for Deferred Acceptance, is contained in Appendix C.1.

18 It might seem contrary to intuition that a mechanism can be EF-TB, but not envy-free. To see why this is indeed possible, consider the following example. There are $n = 2$ players, and $2$ possible types $R, D$. We denote the player with the higher (lower) lottery number the winner (loser). If the winner is a type $R$, then the winner (loser) gets a large (small) rock. If the winner is a type $D$, then the winner (loser) gets a large (small) diamond. If all players prefer larger objects, but attribute greater marginal utility to a larger diamond than to a larger rock, then a type $R$ envies a type $D$. 

mechanisms.

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. We obtain results for Walrasian mechanisms as in Roberts and Postlewaite (1976) and (Jackson and Manelli, 1997), Double Auctions as in Rustichini et al. (1994) and (Cripps and Swinkels, 2006), Uniform-Price Auctions as in Swinkels (2001), Deferred Acceptance mechanisms as in Immorlica and Mahdian (2005) and Kojima and Pathak (2009), the Probabilistic Serial mechanism as in Kojima and Manea (2010), Approximate CEEI as in Budish (2011), and the Hylland and Zeckhauser (1979) and Budish et al. (Forthcoming) Pseudomarket Mechanisms, whose large market incentive properties had not previously been studied.

One sense in which SP-L organizes these findings is that the single, well-defined concept of SP-L properly classifies all the mechanisms above. In contrast, the literature has employed different notions of approximate incentive compatibility, tailored for each mechanism.

The notion of SP-L, along with our results described in Section 4.1, also delivers reasonable rates of convergence for these mechanisms. Proposition A1 in the Appendix implies that incentives to misreport vanish at a rate of $n^{\frac{3}{4}-\epsilon}$ for all SP-L mechanisms in the table, save for Deferred Acceptance and Approximate CEEI, which have a rate of $n^{\frac{1}{2}-\epsilon}$.

The breadth and simplicity of these results comes at a cost relative to analyses that are tailored to specific mechanisms. One cost is that SP-L is weaker than several previous notions of incentive compatibility in large markets, in the sense that it is ex-interim and not ex-post and assumes full support. A second cost is that we require finite type and action spaces in our definition of each mechanism. Third, analyses of specific mechanisms 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

\[19\] 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, Satterthwaite and Williams (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 $\epsilon$-Bayes-Nash equilibria of the uniform-price and pay-as-bid auctions. Kojima and Pathak (2009) study $\epsilon$-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 $\epsilon$-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.
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 empirical evidence that participants strategically misreport their preferences in practice. For many of these mechanisms there is also evidence that the manipulations undermine efficiency and fairness.

Consider first multi-unit auctions for government securities, as in Example 1. 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)). Moreover, there is some evidence that manipulability has real effects. Friedman (1991) criticized pay-as-bid auctions, arguing that the need to play strategically reduces entry of less sophisticated bidders, giving dealers a “sheltered market” that facilitates collusion. In uniform-price auctions, however, “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) 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. This evidence that manipulability has real costs is perhaps surprising, as treasury auctions are an ideal setting for Bayesian mechanisms to work well, with sophisticated bidders who can devote substantial resources to computing optimal behavior, and who play repeatedly and hence can learn equilibrium. In all of the other mechanisms we review below, the case for participants’ sophistication and a stable environment is not as compelling, and the evidence that manipulability has real costs is stronger and more direct.

Consider now the Boston mechanism for school choice, used to allocate seats in public schools in many cities across the United States. Unlike treasury auctions, school choice involves a mixture of sophisticated and unsophisticated players, many of whom will play the game only once, having no opportunity to learn. Abdulkadiroğlu et al. (2006) indeed find evidence of both sophisticated and unsophisticated play. Sophisticated parents, such as those in the West Zone Parents Group for which extensive email archives were discovered, 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. The strategic complexity of the mechanism runs counter to policymakers’ goals of a “level playing field” (Pathak and Sönmez, 2008).

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 is consistent with 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, which is closely related to the property that we emphasize, which is that the former are not SP-L whereas the latter are SP-L.20

Thus, 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, this misreporting harms design objectives such as revenue or welfare. By contrast, to the best of our knowledge, there are no empirical examples where an SP-L market design is shown to be harmfully manipulated in a large market (though we caution that there is not yet empirical evidence one way or the other for several of the SP-L mechanisms in

---

20Roth (2002) describes the incentives problems created by unstable mechanisms as follows: “Even in a large market, it is not hard to ascertain if an outcome is unstable, because the market can do a lot of parallel processing. Consider a worker, for example, who has received an offer from her third choice firm. She only needs to make two phone calls to find out if she is part of a blocking pair.”
the table). The evidence to date thus suggests that the relevant distinction for practice, in contexts with a large number of participants, is not “SP vs. not SP”, but rather “SP-L vs. not SP-L”. Or, more conservatively, “SP vs. SP-L vs. not SP-L”.

5 Constructing SP-L Mechanisms from Bayes-Nash Mechanisms

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 (or Nash) incentive compatibility. Together with the analysis of Sections 3 and 4, the result completes our argument that SP-L is a useful second-best to SP.

The result is proved by construction of a specific SP-L mechanism, from a given mechanism that has Bayes-Nash equilibria. We describe the construction formally in Section 5.1. Section 5.2 provides a regularity condition, quasi-continuity, that is sufficient for the constructed mechanism to be SP-L. Section 5.3 states the main theorem of this section, sketches the proof, and describes several extensions that are contained in Appendix C. Section 5.4 describes the relationship of our result to several related ideas, including the revelation principle, Kalai’s (2004) study of large Bayesian games, random sampling mechanisms, and the debate concerning the Boston mechanism.

5.1 The Construction

We begin by defining a limit Bayes-Nash equilibrium, which is the standard notion of Bayes-Nash equilibrium but applied to the limit mechanism \( \phi^\infty(\cdot, \cdot) \) as opposed to a finite economy mechanism. In the text we will state the results for limit equilibria, which tend to be more analytically tractable in applications. The Appendix proves a version of the construction Theorem 1 for exact Bayes-Nash equilibria. Therefore, the basic fact that SP-L mechanisms are not considerably more restrictive than Bayesian mechanisms holds regardless of whether a particular analyst is interested in limit equilibria or exact Bayes-Nash equilibria.

Given a mechanism \( \{(\Phi^n)_{n}, A\} \), a strategy \( \sigma \) is defined as a map from \( T \) to \( \Delta A \). Given a prior \( \mu \) and strategy \( \sigma \), we denote by \( \sigma(\mu) \) the distribution over actions induced by strategy \( \sigma \) when player types are drawn according to \( \mu \).

**Definition 7.** Given a mechanism \( \{(\Phi^n)_{n}, A\} \) with limit \( \phi^\infty(\cdot, \cdot) \), and a probability distribution over types \( \mu \in \Delta T \), the strategy \( \sigma^\mu_\mu : T \rightarrow \Delta A \) is a limit \( \mu \)-BNE if, for all \( t_i \in T \) and
where \( \sigma^*_{\mu}(\mu) \) denotes the probability distribution over actions generated by drawing types i.i.d. according to \( \mu \) and then playing, for each drawn type, according to \( \sigma^*_{\mu}(\cdot) \).

Often, a mechanism’s Bayes-Nash equilibria will 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.

**Definition 8.** Given a mechanism \( \{(\Phi^n)_N, A\} \) 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 \( \mu \)-BNE.

The construction takes as given a family of limit BNE because, when researchers claim that a particular mechanism has certain properties in Bayes-Nash equilibrium, they typically make a claim about a family of equilibria. For instance, Abdulkadiroğlu et al. (2011) show that, for each possible distribution of student preferences, the Boston mechanism has a Bayes-Nash equilibrium that leads to an ex-ante Pareto efficient allocation.

We are now ready to formally define our construction. Take as input a mechanism \( \{(\Phi^n)_N, A\} \) and a family of limit equilibria \( (\sigma^*_{\mu})_{\mu \in \Delta T} \). Given a vector of types \( t \) and strategy \( \sigma \), let \( \sigma(t) \in \Delta(A^n) \) denote the associated distribution over vectors of actions (cf. (2.3)), and let \( \text{emp}[t] \in \Delta T \) denote the empirical distribution of \( t \) on \( T \). Construct a new mechanism \( \{(F^n)_N, T\} \) from the original mechanism and the family of equilibria according to:

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

(5.1)

Intuitively, \( F^n(\cdot) \) acts as a proxy agent playing the original mechanism \( \Phi^n(\cdot) \) on each agent’s behalf. The key feature is the strategy that the proxy uses: rather than use the Bayes-Nash equilibrium strategy associated with the “true prior”, which need not be known by the mechanism designer, it uses the strategy, \( \sigma_{\text{emp}[t]}(\cdot) \), associated with the empirical distribution of reports \( \text{emp}[t] \), playing \( \sigma_{\text{emp}[t]}(t_i) \) on behalf of an agent who reports \( t_i \). That is, it uses the strategy that would be a limit Bayes-Nash equilibrium in a world in which the empirical distribution of agents’ reports were in fact the true, common-knowledge, prior.
Notice that each agent’s report affects the empirical distribution, and hence which strategy gets activated. If agent \(t_i\) misreports as \(t'_i\), then our constructed mechanism will use strategy \(\sigma_{\text{emp}}[t'_i,t_{-i}]()\) on each agent’s behalf, instead of \(\sigma_{\text{emp}}[t_i,t_{-i}]()\).

Notice as well that the construction allows the researcher to specify the family of limit equilibria. In particular, if there are multiple equilibria of varying desirability, the researcher can specify a family consisting of the equilibrium that is most desirable for each prior.

Our main result in this section shows that the mechanism \(\{(F^n)_N,T\}\) constructed according to (5.1) gives agents the same outcomes in the large-market limit as the Bayes-Nash equilibria of the original mechanism \(\{(\Phi^n)_N,A\}\), but in a manner that is SP-L. The result requires a mild continuity condition on the family of equilibria, which we turn to next.

5.2 Quasi-Continuity

We now define our continuity requirement, called quasi-continuity.

**Definition 9.** Consider a mechanism \(\{(\Phi^n)_N,A\}\) 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** at full support prior \(\mu_0 \in \Delta T\) if, for every \(\epsilon > 0\), there exists a neighborhood \(\mathcal{N}\) of \(\mu_0\) that can be decomposed as \(\mathcal{N} = \bigcup_{1 \leq k \leq K} A_k \cup B\) with each \(A_k\) open, such that:

1. If types are drawn iid according to \(\mu_0\), then the probability that the empirical distribution of types lands within distance \(1/n\) of \(B\) goes to zero as \(n\) grows large. Formally,

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

2. Within each set \(A_k\), 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, for each \(A_k\), there exists \(n_0\) such that for any \(n \geq n_0\), and any \(\mu, \mu', \text{emp}[t_i,t_{-i}], \text{emp}[t'_i,t'_{-i}] \in A_k\), we have:

   \[
   |\Phi^n_1(\sigma^*_\mu(t_i),\sigma^*_\mu(t_{-i}))-\Phi^n_1(\sigma^*_\mu'(t_i),\sigma^*_\mu'(t_{-i}))| < \epsilon.
   \]

The family of equilibria is **continuous at \(\mu_0\)** if, for the prior \(\mu_0\), Condition 2 holds with respect to the entire neighborhood \(\mathcal{N}\), that is, \(B = \emptyset\) and \(K = 1\). A family of limit equilibria is (quasi-)continuous if it is (quasi-)continuous for every full support prior \(\mu_0 \in \Delta T\).
Note that continuity and quasi-continuity are defined for families of limit equilibria of mechanisms, not for mechanisms per se. Continuity asks that around any prior $\mu_0$, as per Condition 2, the allocation that an agent of type $t_i$ receives varies continuously, as long as opponents have types with an empirical distribution close to $\mu_0$, and play is according to a strategy $\sigma^*_\mu$ with $\mu$ close to $\mu_0$. Quasi-continuity, while considerably weaker, has a more elaborate definition. Quasi-continuity allows the family of equilibria to be discontinuous at prior $\mu_0$. However, it imposes some regularity in how outcomes of the family of equilibria vary close to $\mu_0$. Quasi-continuity requires that a small enough neighborhood $\mathcal{N}$ of $\mu_0$ 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. Heuristically, $\mathcal{B}$ is a “knife-edge” discontinuity set, and is surrounded by sets $\mathcal{A}_k$ of local continuity.

To further clarify this definition, we provide the following example. Consider a uniform price multi-unit auction, as defined in Example 1. Take $\mu_0 \in \Delta T$, and a family of limit equilibria $(\sigma^*_\mu)_{\mu \in \Delta T}$ such that agents report truthfully for all $\mu \in \Delta T$. Assume that, in expectation, the price $1 < \rho^*(\mu_0) < \bar{v}$ clears the market exactly. That is, $\mu_0$ is a “knife edge” case as described in Example 1 where expected per-capita demand is exactly equal to per-capita supply at the price $\rho^*(\mu_0)$ that clears the market in expectation:

$$E[D(\rho^*; t_i) | t_i \sim \mu_0] = k.$$ 

Take a neighborhood $\mathcal{N}$ of $\mu_0$ small enough such that, for all $\mu$ in $\mathcal{N}$, there is excess expected demand at price $\rho^*(\mu_0) - 1$ and excess supply at price $\rho^*(\mu_0) + 1$. If agents report their types truthfully, and the vector of reported types satisfies $\text{emp}[t] \in \mathcal{N}$, then the realized market clearing price $p^*(t)$ will either be $\rho^*(\mu_0)$ or $\rho^*(\mu_0) - 1$. Within this set $\mathcal{N}$, price is not a continuous function of the empirical distribution of reports. However, $\mathcal{N}$ can be divided into a set of priors where expected demand at $\rho^*(\mu_0)$ is strictly higher than supply $k$,

$$\mathcal{A}_1 = \{\mu \in \mathcal{N} : E[D(\rho^*(\mu_0); t_i) | t_i \sim \mu] > k\},$$

a set of priors where expected demand at $\rho^*$ is strictly lower than supply $k$,

$$\mathcal{A}_2 = \{\mu \in \mathcal{N} : E[D(\rho^*(\mu_0); t_i) | t_i \sim \mu] < k\},$$

and a residual set $\mathcal{B} = \mathcal{N} \setminus (\mathcal{A}_1 \cup \mathcal{A}_2)$. $\mathcal{B}$ is then the subset of priors $\mu \in \mathcal{N}$ where the
market clears exactly in expectation. Intuitively, the \( B \) set is where price is discontinuous with respect to small changes in demand. Dividing \( N \) in this way, the market clearing price is always \( \rho^*(\mu_0) \) when agents have a vector of types \( t \) with empirical distribution in \( A_1 \), and \( \rho^*(\mu_0) - 1 \) when agents have a vector of types with empirical distribution in \( A_2 \). Moreover, as \( n \) grows large, the probability that a vector of types drawn randomly according to \( \mu_0 \) has an empirical distribution closer than \( 1/n \) to \( B \) is converges to zero. Therefore, the family of equilibria \( (\sigma^*_\mu)_{\mu \in \Delta T} \) is quasi-continuous. Appendix D describes a family of equilibria of the pay-as-bid auction in detail and shows that it too satisfies quasi-continuity. Appendix C.6 gives an example where quasi-continuity is not satisfied, and shows that the conclusions of the construction Theorem do not hold in that case.

Quasi-continuity is related to the equicontinuity condition proposed by Kalai (2004), and similarly to the continuity required by Roberts and Postlewaite (1976), but it is stronger in an important respect and weaker in an important respect. Quasi-continuity is stronger in that it imposes continuity conditions with respect to both the empirical distribution of agents’ play and the equilibrium strategies they use. Kalai equicontinuity asks for continuity with respect to just a single Bayes-Nash equilibrium.\(^{21}\) That is, quasi-continuity is a property of a family of equilibria, whereas Kalai equicontinuity is a property of a game. Quasi-continuity is weaker, however, in that it allows for certain kinds of discontinuities, with respect to both the distribution of agents’ play and the prior on which they base equilibrium behavior. This added generality is important since such discontinuities are common in discrete-goods allocation problems.

### 5.3 The Construction Theorem

The following result states formally that the mechanism constructed according to (5.1) is SP-L, and gives agents approximately the same outcomes as the Bayes-Nash equilibria of the original mechanism.

We first define the sense in which the original mechanism approximates the constructed mechanism. Consider a mechanism \( \{\Phi^n_{\mu}, A\} \) with a family of limit equilibria \( (\sigma^*_\mu)_{\mu \in \Delta T} \). The outcomes of the direct mechanism **approximate** equilibrium outcomes of the original mechanism at \( \mu_0 \) if, for any \( \epsilon > 0 \), there exists \( n_0 \) such that, for all \( n \geq n_0 \) and \( t_i \in T \),

\[
|f^n(t_i|\mu_0) - \phi^n(\sigma^*_\mu_0(t_i)|\sigma^*_\mu_0(\mu_0))| < \epsilon,
\]

\(^{21}\)Formally, Kalai (2004) equicontinuity requires that, in a sequence of games, the payoff of any given player varies uniformly continuously with the aggregate distribution of play.
where \( f \) is defined from \( F \) as in Equation (2.4).

The definition asks that, for a prior \( \mu_0 \), truthful play of the constructed mechanism (the \( f^n \) term) approaches the ideal theoretical prediction of play in the original mechanism (the \( \phi^n \) term). This is convenient for it means that, if an analyst believes that the original mechanism yields desirable outcomes in Bayesian equilibrium, then the new mechanism leads to the same distribution of outcomes.

If for some prior the original equilibrium is not continuous, the outcomes of the constructed mechanism will match those of the original mechanism in a weaker sense. We say that the outcomes of the constructed mechanism approximate a convex combination of equilibrium outcomes of the original mechanism at \( \mu_0 \) if, for any \( \epsilon > 0 \), there exists \( n_0 \), an integer \( K \), numbers \( \pi_k^n \) with \( \sum_{k=1}^{K} \pi_k^n = 1 \), and priors \( \mu_k \) with \( |\mu_k - \mu_0| < \epsilon \) such that, for all \( n \geq n_0 \) and \( t_i \in T \),

\[
|f^n(t_i|\mu_0) - \sum_{k=1}^{K} \pi_k^n \cdot \phi^n(\sigma^*_{\mu_k}(t_i)|\sigma^*_{\mu_k}(\mu_k))| < \epsilon.
\]

With these definitions, we have the following Theorem.

**Theorem 1.** Consider a mechanism \( \{(\Phi^n)_\mathbb{N}, A\} \) with a quasi-continuous family of limit equilibria \( (\sigma^*_{\mu})_{\mu \in \Delta T} \). Define the direct mechanism \( \{(F^n)_\mathbb{N}, T\} \) as in equation (5.1). Fix an arbitrary full-support prior \( \mu_0 \in \Delta T \). Then,

1. The direct mechanism is SP-L.

2. If the family \( (\sigma^*_{\mu})_{\mu \in \Delta T} \) is continuous at \( \mu_0 \), then outcomes of the direct mechanism \( \{(F^n)_\mathbb{N}, T\} \) approximate equilibrium outcomes of \( \{(\Phi^n)_\mathbb{N}, A\} \) under \( (\sigma^*_{\mu})_{\mu \in \Delta T} \) at \( \mu_0 \).

3. If the family \( (\sigma^*_{\mu})_{\mu \in \Delta T} \) is quasi-continuous at \( \mu_0 \), then outcomes of the direct 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 \( \mu_0 \).

\( ^{22} \)Note that what we refer to as types are payoff types, a player’s actual preferences over outcomes, as opposed to the Harsanyi (1967-68) and Mertens and Zamir (1985) notion of types, which include payoff types and beliefs about opponents’ types. While the terminology we adopt is standard in the applied mechanism design literature, since the epistemic terminology has also been influential in mechanism design, many readers might find it helpful to recast our definition of approximation in terms of the epistemic definition of types. In these terms, the outcomes of the original mechanism \( \Phi^n(\sigma^*_{\mu}(t)) \) depend both on a vector of payoff types, and an assumed common iid prior belief \( \mu \) that players have over the payoff types of others. We can write these outcomes simply as a function \( G^n(t, \mu) \equiv \Phi^n(\sigma^*_{\mu}(t)) \). In contrast, outcomes of the direct mechanism \( F^n(t) \) only depend on the vector of payoff types. Our notion of approximation is that, for a fixed \( \mu_0 \), when types in the vector \( t_{-i} \) are drawn according to \( \mu_0 \), the distribution of outcomes \( F^n(t, t_{-i}) \) is close to the distribution of \( G^n(t, t_{-i}, \mu_0) \).
The full proof of Theorem 1 is contained in Appendix A. Here we provide a detailed sketch.

Proof Sketch. First, suppose that agents report their preferences truthfully, according to the true prior \( \mu_0 \), and that the limit equilibria of \( \{(\Phi^n_n, A)\} \) are continuous at \( \mu_0 \). 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 \( t_i \) 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 \( \mu_0 \), by the law of large numbers. Hence, by continuity, agent \( i \)'s allocation \( \Phi^n_i(\sigma^*_\hat{\mu}(t_i), \sigma^*_\hat{\mu}(t_{-i})) \) is converging as \( n \to \infty \) to \( \phi^\infty(\sigma^*_{\mu_0}(t_i), \sigma^*_{\mu_0}(\mu_0)) \), exactly what he receives under Bayes-Nash equilibrium of the original mechanism. This is what is required by Part 2 of the theorem statement.

Now, suppose that the agents other than \( i \) misreport their preferences, according to some distribution \( m \in \Delta T \). Assume for now that the limit equilibria of \( \{(\Phi^n_n, A)\} \) are continuous at \( m \). 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 \( \phi^\infty(\sigma^*_m(t_i), \sigma^*_m(m)) \). This is what agent \( i \) would receive under the original mechanism in the Bayes-Nash equilibrium corresponding to prior \( m \). Even though the other agents are systematically misreporting their preferences, our agent \( i \) remains happy 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 he reports his preferences truthfully to \( \{(F^n^n, T)\} \). Hence, in the limit, we get dominant-strategy incentives, i.e., our constructed mechanism is SP-L. This is what is required by Part 1 of the theorem statement.

The last step of the proof sketch is to describe what happens in the event that the equilibrium of the original mechanism is not continuous at some prior; for instance, in the uniform-price auction example described in Section 5.2, consider the priors that clear the market exactly in expectation. This requires a technical lemma (Lemma A3 in the appendix) which says that, for any full-support prior \( m \in \Delta T \), the allocation an agent receives under \( \{(F^n^n, T)\} \) can be approximated by a convex combination of the allocations he would receive in the limit Bayes-Nash equilibria of \( \{(\Phi^n_n, A)\} \), for priors close to \( m \).

---

23 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.
The key to the proof of the lemma is that, in a large enough market, a single agent cannot appreciably change the probability that the aggregate profile lands in each region $A_k$, as defined in Definition 9. This allows us to exploit the continuity within each region $A_k$, and the vanishing likelihood that the aggregate profile lands near the discontinuity region $B$, to obtain the desired approximation.

Extensions and Robustness Appendix C discusses several extensions of Theorem 1 and robustness to alternative formulations. We show that extensions of Theorem 1 still hold if we relax anonymity to semi-anonymity (C.1), if the given mechanism has a family of complete information Nash equilibria instead of Bayes-Nash equilibria (C.2), and if we consider a family of finite-economy equilibria instead of a family of limit equilibria (C.3). Appendix C.4 discusses aggregate uncertainty. Appendix C.5 provides an additional result showing that SP-L mechanisms have a strong ex-post robustness property when they are equicontinuous as defined by Kalai (2004).

5.4 Discussion

5.4.1 Relation to the Revelation Principle

It is important to emphasize how our constructed mechanism $\{(F^n)_N, T\}$ differs from a traditional Bayes-Nash direct revelation mechanism (Fudenberg and Tirole, 1991; Section 7.2). In a traditional Bayes-Nash DRM, the mechanism designer and participants have common knowledge of the prior $\mu_0$. The mechanism then announces a BNE strategy $\sigma^{\ast}_{\mu_0}(\cdot)$, and plays $\sigma^{\ast}_{\mu_0}(t_i)$ on behalf of an agent who reports $t_i$.

Our approach does not assume common knowledge of the prior. Instead, our constructed mechanism infers a prior from the empirical distribution of agents’ play. If agents indeed play truthfully, this inference is exactly correct in the limit, and our detail-free mechanism coincides with the traditional Bayes-Nash DRM. But if the agents other than $i$ misreport, so that the empirical $\text{emp}[t]$ is very different from the prior $\mu_0$, then our mechanism automatically adjusts each agent’s play to be the Bayes-Nash equilibrium play in a world where the prior was in fact $\text{emp}[t]$. As a result, an agent who reports his preferences truthfully remains happy to have done so even if the other agents misreport, which is not the case in a traditional Bayes-Nash DRM.

To summarize, the two key differences between our mechanism and a traditional Bayes-Nash DRM are: (1) our mechanism is prior free, whereas a traditional DRM assumes common knowledge of the prior; (2) our mechanism gives each agent incentive to report truthfully
for any distribution of opponent play, whereas a traditional DRM gives agents incentive to report truthfully only if all of his opponents do as well.

5.4.2 Relation to Kalai (2004)

We briefly remark on the relationship between our Theorem 1 and Theorem 1 of Kalai (2004). Our environment and assumptions are substantially similar to (and inspired by) Kalai’s. The main difference in the two papers’ setups is the nature of the continuity assumptions. Kalai’s continuity requires that a given player’s outcome is a (uniformly equi)continuous function of the empirical distribution of other players’ reports. Our notion of continuity requires that a given player’s outcome is continuous both with respect to the empirical distribution of other players’ reports, and with respect to changes in the equilibrium strategy they use. Put differently, Kalai assumes continuity with respect to a single Bayes-Nash equilibrium, for a single prior, whereas we impose continuity with respect to a family of equilibria, for all possible priors. Our condition is much stronger in this regard. Our condition is weaker in another aspect, namely that we allow for certain kinds of discontinuities as discussed above.

Kalai’s Theorem 1 shows that Bayes-Nash equilibria are approximately ex-post Nash. In words, if a large number of agents with private information about their types play some Bayes-Nash equilibrium, then ex-post – i.e., after seeing each agent’s chosen action – agents have vanishingly little incentive to revise their play. This is an important robustness result for Bayes-Nash equilibria. However, the theorem’s usefulness for market design depends on the analyst’s confidence that agents can reach the Bayes-Nash equilibrium in the first place. Robustness in the Kalai sense is not the same as robustness in the Wilson (1987) and Bergemann and Morris (2005) sense, and the latter kind of robustness may also be important in practice. For instance, the Bayes-Nash equilibria of the Boston mechanism for school choice studied by Abdulkadiroğlu et al. (2011) could be ex-post robust in the Kalai sense without implying that Bayes-Nash equilibrium is a good prediction of play in that market (cf. Section 5.4.4).

In environments where the analyst is less confident that the Bayes-Nash equilibrium outcome can be reached – failure of common knowledge, unsophisticated players, coordination problems stemming from multiple equilibria, etc. – our Theorem 1 may be more useful. Our Theorem 1 shows that it is approximately costless to use a mechanism that is not just Bayes-Nash, but SP-L. Moreover, if we assume continuity and not just quasi-continuity, SP-L mechanisms satisfy a notion of ex-post robustness that is stronger than the ex-post robustness obtained by Kalai for Bayes-Nash equilibria. See Appendix C.5 for details.
5.4.3 Relation to VCG and Random-Sampling Mechanisms

Our proof technique is related to, but distinct from, the ideas behind Vickrey-Clarke-Groves (VCG) and random-sampling mechanisms. In VCG, each agent $i$ faces prices that depend on the empirical distribution of the $n-1$ agents other than himself. In random-sampling mechanisms (Cordoba and Hammond, 1998; Goldberg et al., 2001; Segal, 2003; Baliga and Vohra, 2003), each agent faces prices that depend on a random sample of the agents other than himself; e.g., a typical approach is to randomly divide the population of $n$ agents into two distinct groups of $n/2$ agents, and have each group face prices that depend on the empirical distribution of the agents in the other group. See Hartline and Karlin (2007) for a recent survey.

The advantage of these two approaches is that, since the prices each agent faces depend only on agents other than himself, the mechanisms are able to provide exact dominant-strategy incentives. By contrast, in our approach, since the empirical distribution of all $n$ agents’ reports is used to activate a single Bayes-Nash equilibrium, each agent affects which strategy is activated (our analogue of affecting prices), and we are able to provide only approximate incentives.

The disadvantage of the VCG and random-sampling approaches is closely related to the advantage: since the prices each agent faces depend only on agents other than himself, different agents necessarily face different prices. In some settings, such as combinatorial auctions and monopoly pricing (e.g., Segal (2003)), this is not problematic, but in general this approach can cause violations of feasibility. Suppose we applied the VCG idea to our basic construction above. If in a market of size $n$ the agents report $t = (t_1, \ldots, t_n)$, we might try to construct $\bar{F}_n^{\pi}(\cdot)$ from $\Phi_n^{\pi}(\cdot)$ as:

$$
\bar{F}_n^{\pi}(t_i, t_{-i}) = \Phi_n^{\pi}(\sigma_{\text{emp}[t_{-i}]}(t_i), \sigma_{\text{emp}[t_{-i}]}(t_{-i})),
$$

(5.2)

that is, for agent $i$’s allocation, we activate the Bayes-Nash equilibrium of $\Phi_n^{\pi}(\cdot)$ that corresponds to the empirical distribution $\text{emp}[t_{-i}]$ of the agents other than $i$. However, there is no reason to expect that the allocation that results from $n$ applications of (5.2), once for each agent $i$, will be feasible. We know that $\Phi_n^{\pi}(\cdot)$ itself produces a feasible outcome for any profile of $n$ actions (cf. Definition 1), but in the outcome constructed according to (5.2) there are different action profiles used for different agents, because a different strategy

\[\text{Page 30}\]

\[\text{AZEVEDO AND BUDISH}\]

\[\text{5.4.3 Relation to VCG and Random-Sampling Mechanisms}\]

Our proof technique is related to, but distinct from, the ideas behind Vickrey-Clarke-Groves (VCG) and random-sampling mechanisms. In VCG, each agent $i$ faces prices that depend on the empirical distribution of the $n-1$ agents other than himself. In random-sampling mechanisms (Cordoba and Hammond, 1998; Goldberg et al., 2001; Segal, 2003; Baliga and Vohra, 2003), each agent faces prices that depend on a random sample of the agents other than himself; e.g., a typical approach is to randomly divide the population of $n$ agents into two distinct groups of $n/2$ agents, and have each group face prices that depend on the empirical distribution of the agents in the other group. See Hartline and Karlin (2007) for a recent survey.

The advantage of these two approaches is that, since the prices each agent faces depend only on agents other than himself, the mechanisms are able to provide exact dominant-strategy incentives. By contrast, in our approach, since the empirical distribution of all $n$ agents’ reports is used to activate a single Bayes-Nash equilibrium, each agent affects which strategy is activated (our analogue of affecting prices), and we are able to provide only approximate incentives.

The disadvantage of the VCG and random-sampling approaches is closely related to the advantage: since the prices each agent faces depend only on agents other than himself, different agents necessarily face different prices. In some settings, such as combinatorial auctions and monopoly pricing (e.g., Segal (2003)), this is not problematic, but in general this approach can cause violations of feasibility. Suppose we applied the VCG idea to our basic construction above. If in a market of size $n$ the agents report $t = (t_1, \ldots, t_n)$, we might try to construct $\bar{F}_n^{\pi}(\cdot)$ from $\Phi_n^{\pi}(\cdot)$ as:

$$
\bar{F}_n^{\pi}(t_i, t_{-i}) = \Phi_n^{\pi}(\sigma_{\text{emp}[t_{-i}]}(t_i), \sigma_{\text{emp}[t_{-i}]}(t_{-i})),
$$

(5.2)

that is, for agent $i$’s allocation, we activate the Bayes-Nash equilibrium of $\Phi_n^{\pi}(\cdot)$ that corresponds to the empirical distribution $\text{emp}[t_{-i}]$ of the agents other than $i$. However, there is no reason to expect that the allocation that results from $n$ applications of (5.2), once for each agent $i$, will be feasible. We know that $\Phi_n^{\pi}(\cdot)$ itself produces a feasible outcome for any profile of $n$ actions (cf. Definition 1), but in the outcome constructed according to (5.2) there are different action profiles used for different agents, because a different strategy

\[\text{See, for instance, Parkes et al. (2001) on the feasibility issues that arise in the context of combinatorial exchange (two-sided combinatorial auctions), and Kovalenkov (2002) on the feasibility issues that arise in the context of a Walrasian exchange economy.}\]
σ^{*}_{emp[t-i]} gets activated for each \( i \). By contrast, our mechanism constructed according to (5.1) is always feasible, because it inputs a single action profile into \( \Phi^{n}(\cdot) \).

### 5.4.4 The Debate on the Boston Mechanism

Our analysis contributes to an ongoing market design debate concerning the Boston mechanism for student assignment (see Appendix B for a formal description). Abdulkadiroğlu and Sönmez (2003) and Abdulkadiroğlu et al. (2006) criticized the Boston mechanism on the grounds that it is not SP, and proposed that the strategyproof Gale-Shapley deferred acceptance algorithm be used instead.\(^{25}\) The Gale-Shapley algorithm was eventually adopted for use in practice (cf. Roth (2008)).

A second generation of papers on the Boston mechanism then argued that the Boston mechanism has Bayes-Nash equilibria that yield greater student welfare than do the dominant strategy equilibria of the Gale-Shapley procedure (Abdulkadiroğlu et al. (2011); Miralles (2009); Featherstone and Niederle (2011)). Perhaps, these papers argue, the earlier papers were too quick to dismiss the Boston mechanism in favor of strategyproof deferred acceptance. These papers are a reminder that SP often comes with a cost.

However, Bayes-Nash equilibria have their own costs. If advocating the Boston mechanism for use in practice, based on its Bayes-Nash equilibria, a market designer must confront several questions: is common knowledge a reasonable assumption? will students be able to calculate the desired equilibrium? will students be able to coordinate in the event of multiple equilibria? will unsophisticated students be badly harmed?

Our Theorem 1 says that all of this complexity and non-robustness is unnecessary in a large market. Specifically, there must exist yet another mechanism that implements the same outcomes as these desirable Bayes-Nash equilibria of the Boston mechanism, but that is SP-L.\(^{26}\)

### 6 Conclusion

This paper proposes strategyproofness in the large (SP-L) as a useful second-best alternative to strategyproofness (SP). First, we show that many of the arguments favoring SP design

---

\(^{25}\)In two-sided matching, the Gale-Shapley algorithm is strategyproof 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.

\(^{26}\)We do not take a position on which is preferable between the SP-L implementation of the Boston mechanism and the SP deferred acceptance procedure.
over Bayes-Nash or Nash support SP-L design as well: compliance with the Wilson doctrine, strategic simplicity, and fairness for unsophisticated agents. Then, our main result shows that, in large anonymous markets, SP-L is in a certain sense no more restrictive than Bayes-Nash and Nash implementation. That is, in large markets, SP-L has many of the benefits of SP design but without the costs.

We also show that SP-L usefully organizes the prior theoretical and empirical evidence on manipulability in large markets. On the theory side, all of the known mechanisms for which there is a detailed theoretical case that the mechanism has approximate incentives for truth-telling in large markets are correctly classified as SP-L. On the empirical side, all of the known instances of mechanisms that have been empirically documented as having costly incentive problems in real-world large markets not only are not SP, but are not even SP-L. That is, the empirical record suggests that the relevant question for practice is not whether or not the mechanism is SP, but instead whether or not the mechanism is SP-L. More conservatively, we suggest that mechanisms be classified based on whether they are SP, SP-L, or not even SP-L.

We hope that our analysis will have two effects. First, we hope to embolden market design researchers, when facing a problem for which it is known that there are no good strategyproof mechanisms, to instead consider designing a mechanism that is SP-L. Examples of papers that have taken this approach are Milgrom (2009) and Budish (2011), for the problems of combinatorial exchange and combinatorial assignment, respectively. Our analysis justifies the decision to focus on SP-L mechanisms, because it shows that the researcher would not gain much by broadening their search to include Bayes-Nash mechanisms, which typically are much less tractable to work with, and which require the researcher to model the information environment.

Second, we hope that researchers who derive negative results for SP mechanisms, or who design SP mechanisms that sacrifice large amounts of welfare in the service of exact dominant strategy incentives, will pause and consider whether SP is too strong a requirement. Just as economists in other fields instinctively understand that there are some contexts where it is necessary to explicitly model strategic interactions, and some contexts where it is reasonable to assume price-taking behavior, we hope that market design researchers will pause to ask whether SP or SP-L is the more appropriate incentives requirement for the problem at hand.
A Proofs

A.1 Proof of Proposition 1

Proposition 1 in the text follows from the following more general Proposition, which gives rates of convergence for the incentives to misreport in EF and EF-TB mechanisms.

Proposition A1. Consider a mechanism \( \{ (\Phi^n)_{i=1}^{N}, T \} \), a distribution \( \mu \in \Delta T \), and \( \epsilon > 0 \).

- If the mechanism is EF, then there exists a constant \( C \) such that, for all \( t'_i \in T \) and all \( n \),
  \[
  u_{t'_i}[\phi^n_i(t'_i, \mu)] - u_{t_i}[\phi^n_i(t_i, \mu)] \leq C \cdot n^{-1/2+\epsilon}.
  \]

- If the mechanism is EF-TB, then there exists a constant \( C' \) such that, for all \( t'_i \in T \) and all \( n \),
  \[
  u_{t'_i}[\phi^n_i(t'_i, \mu)] - u_{t_i}[\phi^n_i(t_i, \mu)] \leq C' \cdot n^{-1/4+\epsilon}.
  \]

The proof of Proposition A1 will use some additional notation. We use \( \Phi^n_i(t_i|\hat{\mu}) \) to denote \( \Phi^n_i(t_i, t_{-i}) \) where \( t_{-i} \) is any vector of \( n - 1 \) types such that \( \text{emp}[t_i, t_{-i}] = \hat{\mu} \). If \( \hat{\mu}(t_i) = 0 \), \( \Phi^n_i(t_i|\hat{\mu}) \) is defined as an arbitrary bundle. 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 drawing \( t_{-i} \) iid according to \( \mu \); let \( \Pr\{t|t'_i, \mu, n\} \) be the probability that \( (t_i, t'_{-i}) = t \). Given a vector of types \( t \), denote the set of indexes of players with type equal to \( i \) as

\[
I(i|t) = \{i' : t_{i'} = t_i\}.
\]

When there is no risk of confusion, this set will be denoted \( I(i) \). The number of elements in a set \( I \) is denoted by \( |I| \). Throughout the proof we will consider sums over infinite sets, but where only a finite number of the summands are positive. We use the convention that these are finite sums of only the positive terms.

The proof is based on two Lemmas. The first Lemma shows that a single player cannot appreciably change the probability distribution over empirical distributions of types in the population.

Lemma A1. Define the function

\[
\Delta P(t'_i, t_i, \mu, n) = \sum_{\hat{\mu} \in \Delta T} |\Pr\{\hat{\mu}|t'_i, \mu, n\} - \Pr\{\hat{\mu}|t_i, \mu, n\}|.
\]

(A.1)
Fix \( \mu \in \bar{\Delta} T \), and \( \epsilon > 0 \). Then there exists a constant \( C_{\Delta P} > 0 \) such that, for any \( t'_i, t_i \in T \) and \( n \in \mathbb{N} \),

\[
\Delta P(t'_i, t_i, \mu, n) \leq C_{\Delta P} \cdot n^{-1/2+\epsilon}.
\]

The second Lemma is only necessary for the EF-TB case. The Lemma shows that anonymous EF-TB mechanisms are approximately EF in markets with many agents with types \( t_i \) and \( t'_i \).

**Lemma A2.** Fix an EF-TB mechanism \( \{(\Phi^n)_n, T\} \) and \( \epsilon > 0 \). Define, for a vector of \( n \) types \( t \), and \( t'_i \in T \), the function

\[
E(t'_i, t, n) = u_{t_i}[\Phi^n_j(t)] - u_{t_i}[\Phi^n_i(t)],
\]

where \( j \) is any player such that \( t_j = t'_i \), if one exists, and \( E(t'_i, t, n) = 1 \) otherwise. Then there exists a constant \( C_E \) such that, for all \( t'_i, t, n \), envy is bounded by

\[
E(t'_i, t, n) \leq C_E \cdot \min_{i'=1,2,\ldots,n} |I(i'|t)|^{-1/4+\epsilon}. \quad (A.2)
\]

With the Lemmas in hand, the argument is straightforward. The gain from misreporting can be decomposed in the gain of reporting another type, holding fixed the empirical distribution of types in the population, and the gain from manipulating the empirical distribution of types. The Lemmas guarantee that both effects vanish in large markets.

**Proof of Proposition A1.** Fix a prior \( \mu \in \bar{\Delta} T \), fix \( n \), and consider the utility a type \( t_i \) agent expects to obtain if she reports \( t'_i \). We may write this expectation as

\[
\begin{align*}
 u_{t_i}[\phi^n_i(t'_i, \mu)] &= \sum_{\hat{\mu} \in \Delta T} \Pr\{\hat{\mu}|t'_i, \mu, n\} \cdot \Phi^n_i(t'_i|\hat{\mu}).
\end{align*}
\]

The gain from misreporting from type \( t_i \) to \( t'_i \) equals

\[
\begin{align*}
 &u_{t_i}[\phi^n_i(t'_i, \mu)] - u_{t_i}[\phi^n_i(t_i, \mu)] \\
 &= \sum_{\hat{\mu}} \Pr\{\hat{\mu}|t'_i, \mu, n\} \cdot u_{t_i}[\Phi^n_i(t'_i|\hat{\mu})] - \sum_{\hat{\mu}} \Pr\{\hat{\mu}|t_i, \mu, n\} \cdot u_{t_i}[\Phi^n_i(t_i|\hat{\mu})].
\end{align*}
\]
We can reorder the terms as\(^\text{27}\)

\[
\sum_{\hat{\mu}} \Pr\{\hat{\mu}|t_i, \mu, n\} \cdot (u_t[\Phi_n^n(t_i|\hat{\mu})] - u_t[\Phi_n^n(t_i|\mu)])
\]

\[
\text{Envy } = \text{Gain from reporting } t_i' \text{ holding fixed } \hat{\mu}
\]

\[
+ \sum_{\hat{\mu}} (\Pr\{\hat{\mu}|t_i', \mu, n\} - \Pr\{\hat{\mu}|t_i, \mu, n\}) \cdot u_t[\Phi_n^n(t_i'|\hat{\mu})].
\]

\[\text{(A.3)}\]

That is, the gain from misreporting can always be decomposed in two terms. The first term is the average 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 \(t_j\) players. The second term is the sum over all possible empirical distributions \(\hat{\mu}\) of how much changing the report 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 \(\hat{\mu}\).

With the notation of Lemmas A1 and A2, the envy term and the change in probability term can be bounded by

\[
u_t[\phi_i^n(t_i', \mu)] - u_t[\phi_i^n(t_i, \mu)] \leq \sum_i \Pr\{t|t_i, \mu, n\} \cdot E(t_i', t, n)
\]

\[+ \Delta P(t_i', t_i, \mu, n). \]

\[\text{(A.4)}\]

**Case 1, envy-free mechanisms.** Note that, if a mechanism is envy free, the first term in the RHS of inequality (A.4) is nonpositive whenever \(\text{emp}[t](t_i') \neq 0\). Taking any \(\epsilon > 0\), and using Lemma A1 to bound the second term in the RHS of inequality (A.4) we have

\[
u_t[\phi_i^n(t_i', \mu)] - u_t[\phi_i^n(t_i, \mu)] \leq \Pr\{\text{emp}[t](t_i') = 0|t_i, \mu, n\}
\]

\[+ C\Delta P \cdot n^{-1/2+\epsilon}. \]

\[\text{(A.5)}\]

Since the probability that \(\text{emp}[t](t_i') = 0\) goes to 0 exponentially with \(n\), we have the desired result.

**Case 2, EF-TB mechanisms.** Fix \(\epsilon > 0\). We being by bounding the envy term in

\[\text{Note that this expression uses } \Phi_n^n(t_i|\hat{\mu}) \text{ terms where } \hat{\mu}(t_i) = 0. \text{ As discussed above, we define } \Phi_n^n(t_i|\hat{\mu}) \text{ to be an arbitrary bundle in that case. Note that, since all such arbitrarily defined terms cancel each other, the expression still equals the gain from misreporting. Moreover, the bounds derived below are valid regardless of the arbitrary values chosen for these terms.}\]
Inequality (A.4). By Lemma A2 and the fact that envy is bounded by 1 we have

\[ \sum_t \Pr\{t|t_i, \mu, n\} \cdot E(t'_i, t, n) \leq \sum_t \Pr\{t|t_i, \mu, n\} \cdot \min\{1, C_E \cdot \min_{i'=1,2,\ldots,n} |I(i'|t)|^{-1/4+\epsilon}\}. \quad (A.6) \]

Using Hoeffding’s inequality,\(^{28}\) we can bound for any \(i'\) the probability that there are too few agents with the same type. We have that, for any \(\delta > 0\), there exists a constant \(C_\delta > 0\) such that

\[ \Pr\{I(i'|t) \geq (\mu(t_i) - \delta) \cdot n|t_i, \mu, n\} \geq 1 - C_\delta \cdot \exp\{-2\delta^2 n\}. \]

Note that the constant \(C_\delta\) is necessary in this application of Hoeffding’s inequality, since the value of \(t_i\) is deterministic, and \(I(i', t)\) is determined with only \(n-1\) draws. Moreover, since there are only \(|T|\) types, the probability that this inequality is satisfied for all \(i'\) is at least \(1 - |T| \cdot C_\delta \cdot \exp\{-2\delta^2 n\}\). Applying this to inequality (A.6) we have that

\[ \sum_t \Pr\{t|t_i, \mu, n\} \cdot E(t'_i, t, n) \leq C_E \cdot \min_{i'=1,2,\ldots,n} ((\mu(t_i) - \delta) \cdot n)^{-1/4+\epsilon} + C_\delta \cdot |T| \cdot \exp\{-2\delta^2 n\}. \]

Now take \(\delta\) small enough such that

\[ \min_{\tau \in T}((\mu(\tau) - \delta) > 0. \]

We then have that

\[ \sum_t \Pr\{t|t_i, \mu, n\} \cdot E(t'_i, t, n) \leq C_E \cdot n^{-1/4+\epsilon} \cdot \min_{i'=1,2,\ldots,n} ((\mu(t_i) - \delta)^{-1/4+\epsilon} + C_\delta \cdot |T| \cdot \exp\{-2\delta^2 n\}. \]

Therefore, there exists a constant \(C'\) such, that for all \(n\), \(t'_i\), and \(t_i\),

\[ \sum_t \Pr\{t|t_i, \mu, n\} \cdot E(t'_i, t, n) \leq C' \cdot n^{-1/4+\epsilon}. \]

Return now to inequality (A.4). Using the bound we just derived and Lemma A1, we

\(^{28}\)Hoeffding’s inequality states that given \(n\) iid 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 \(\exp\{-2nz^2\}\).
have

\[ u_{t_i} [\phi^n_i(t', \mu)] - u_{t_i} [\phi^n_i(t, \mu)] \leq C' \cdot n^{-1/4 + \epsilon} \\
+ C_{\Delta P} \cdot n^{-1/2 + \epsilon}. \]

Therefore, there exists a constant \( C'' \) such that

\[ u_{t_i} [\phi^n_i(t', \mu)] - u_{t_i} [\phi^n_i(t, \mu)] \leq C'' \cdot n^{-1/4 + \epsilon}, \]

as desired. \( \Box \)

### A.1.1 Proof of the Lemmas

We now prove the Lemmas.

**Proof of Lemma A1.** 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.

Enumerate the elements of \( T \) as

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

As \( \hat{\mu} \) follows a multinomial distribution, we can write, for any \( t_i \in T \),

\[ \Pr\{\hat{\mu}|t_{i'}, \mu, n\} = \left( \binom{n - 1}{n\hat{\mu}(\tau_1), \cdots, n\hat{\mu}(t_i) - 1, \cdots, n\hat{\mu}(\tau_{|T|})} \cdot \mu(\tau_1)^{n\hat{\mu}(\tau_1)} \cdots \mu(t_i)^{n\hat{\mu}(t_i) - 1} \cdots \mu(\tau_{|T|})^{n\hat{\mu}(\tau_{|T|})} \right). \]

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, note that \( t_i \) only enters the formula in one factorial term in the denominator, and a power term in the numerator. With this observation, we have that

\[ \Pr\{\hat{\mu}|t_{i'}, \mu, n\} / \Pr\{\hat{\mu}|t_i, \mu, n\} = \frac{\hat{\mu}(t_{i'})}{\mu(t_{i'})} \cdot \frac{\hat{\mu}(t_i)}{\mu(t_i)}. \quad (A.7) \]

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

Define, for any \( \delta > 0 \), the set \( M_\delta \) of empirical distributions \( \hat{\mu} \) that are sufficiently close to the true distribution \( \mu \) as

\[
M_\delta = \{ \hat{\mu} \in \Delta T : |\hat{\mu}(t_i) - \mu(t_i)| < \delta \}.
\]

Since the expression on the right of Equation A.7 is differentiable, and equals 1 when the empirical distributions equal the expected prior distributions, there exists a constant \( C > 0 \) such that if \( \hat{\mu} \in M_\delta \) then

\[
\left| \frac{\hat{\mu}(t'_i)}{\mu(t'_i)} \frac{\hat{\mu}(t_i)}{\mu(t_i)} - 1 \right| < C\delta. \tag{A.8}
\]

Moreover, we can bound the probability that the empirical distribution of types \( \hat{\mu} \) is not in \( M_\delta \). By Hoeffding’s inequality, there exists \( C' > 0 \) such that

\[
\Pr\{ \hat{\mu} / \notin M_\delta | t_i, \mu, n \} \leq C' \cdot \exp(-2n\delta^2). \tag{A.9}
\]

We are now ready to bound \( \Delta P \). We can decompose the sum in Equation (A.1) among the terms where \( \hat{\mu} \) is within or outside \( M_\delta \). We then have

\[
\Delta P = \sum_{\hat{\mu} \in M_\delta} |\Pr\{ \hat{\mu}|t'_i, \mu, n \} - \Pr\{ \hat{\mu}|t_i, \mu, n \}| + \sum_{\hat{\mu} / \notin M_\delta} |\Pr\{ \hat{\mu}|t'_i, \mu, n \} - \Pr\{ \hat{\mu}|t_i, \mu, n \}|.
\]

Using equation (A.7) to substitute for the first summand, and the triangle inequality in the second summand,

\[
\Delta P \leq \sum_{\hat{\mu} \in M_\delta} |\Pr\{ \hat{\mu}|t'_i, \mu, n \} / \Pr\{ \hat{\mu}|t_i, \mu, n \} - 1| \cdot \Pr\{ \hat{\mu}|t_i, \mu, n \} + \sum_{\hat{\mu} / \notin M_\delta} (\Pr\{ \hat{\mu}|t'_i, \mu, n \} + \Pr\{ \hat{\mu}|t_i, \mu, n \}).
\]

We can bound the first sum using the fact that the ratio being summed is small for \( \hat{\mu} \in M_\delta \), and bound the second sum since the total probability that \( \hat{\mu} / \notin M_\delta \) is small. Formally, using equations (A.8) through (A.10) we have

\[
\Delta P \leq C\delta + 2C' \cdot \exp(-2n\delta^2).
\]
If we take $\delta = n^{-1/2+\epsilon}$, we obtain the bound

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

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

for a suitably chosen constant $C_{\Delta P} > 0$.

We now prove Lemma A2. The difficulty with establishing the result is that some mechanisms are EF-TB but not EF. To see why this can be the case, consider Figure A.1. The figure plots, for several players $i'$, their lottery numbers $l_{i'}$ in the horizontal axis, and the utility of a type $t_i$ for the bundle $i'$ receives. 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 $j$’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.

The key idea in the proof is that how much $i$ envies $j$ in expectation, before the lottery is realized, is equal to how much an average type $t_i$ player envies an average type $t_j$ player. Therefore, we only have to show is that, in expectation, the average vertical coordinates of the balls are higher than those of the triangles. The proof then uses the fact that lottery numbers are very evenly distributed in the interval $[0, 1]$ with high probability. Whenever this is the case, the EF-TB condition implies that average envy has to be small.

**Proof of Lemma A2.** Fix $n$, a profile of types $t$, player $i$ with type $t_i$ and alternative report $t'_{i'}$, and player $j$ such that $t_j = t'_{i'}$. To prove the Lemma we will show that the inequality (A.2) must hold.

Note we can write the expected bundle $\Phi^n_{i'}(t)$ received by player $i'$ as the expected bundle received by all players with the same type, over all realizations of $l$. That is,

$$\Phi^n_{i'}(t) = \int_{t \in [0, 1]^n} \sum_{i'' \in I(i')} \frac{x^n_{i''}(t, l)}{|I(i')|}. \quad (A.11)$$

This will play a key role in the proof, as we will use this formula to evaluate $\Phi^n_{i'}(t)$, so that we can take advantage of the largeness of the market to bound the envy of player $i$.

**Part 1.** Bounding average envy after an arbitrary lottery draw.
Figure A.1: A scatter plot of the lottery numbers \( l' \) of different agents \( i' \) on the horizontal axis, and the utility \( u_{t_i}[x^n_n(t, l)] \) 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.

Fix a realization \( l \) of the lottery. We will partition the set of players in groups according to where their lottery number falls among \( K \) uniformly-spaced intervals \( L_1 = [0, 1/K) \), \( L_2 = [1/K, 2/K) \), \( L_K = [k-1/K, 1] \). Denote the number of type \( t_{i'} \) players with lottery number in \( L_k \) by

\[
I_k(i') = \{ i'' \in I(i') : l_{i''} \in L_k \}.
\]

Given the lottery draw \( l \), we will choose the number of partitions \( K(l) \) such that the players with types \( t_i \) and \( t_j \) are not too unevenly distributed over the \( I_k(i') \) sets. Let \( K(l) \) be the largest integer \( K \) such that for \( i' = i, j \) and all \( k \)

\[
\left| \frac{|I_k(i')|}{|I(i')|} - \frac{1}{K} \right| < \frac{1}{K^2}.
\]

(A.12)

Such an integer necessarily exists, as \( K = 1 \) satisfies this condition. Intuitively, \( K(l) \) is a measure of how evenly distributed the lottery numbers \( l \) are.

We will use these sets to bound how much players of type \( t_i \) envy players of type \( t_j \), on average. Denote the minimum utility received by a player with type \( t_i \) and lottery number in \( L_k \) as

\[
\bar{u}_k(l) = \min\{ u_{t_i}[x^n_n(t, l)] : i' \in I_k(i) \}.
\]

Define \( \bar{u}_{K(l)+1}(l) = 1 \). Although \( \bar{u}(l) \) and \( K(l) \) depend on \( l \), we will omit this dependence.
when there is no risk of confusion. The EF-TB condition gives that, for all \( j' \in I_k(j) \),

\[
  u_{t_i} [x^n_{j'}(t, l)] \leq \bar{u}_{k+1}(l),
\]

and by definition we have that, for all \( i' \in I_{k+1}(i) \),

\[
  \bar{u}_{k+1}(l) \leq u_{t_i} [x^n_{i'}(t, l)].
\]

We can now bound the average utility for a type \( t_i \) of the bundles received by all players with type \( t_j \) as follows.

\[
  \sum_{j' \in I(j)} \frac{u_{t_i} [x^n_{j'}(t, l)]}{|I(j)|} \leq \sum_{k=1, \ldots, K} \left[ \frac{|I_k(j)|}{|I(j)|} \cdot \sum_{j' \in I_k(j)} \frac{|I_k(j)|}{|I(j)|} \cdot u_{t_i} [x^n_{j'}(t, l)] \right] \leq \sum_{k=1, \ldots, K} \frac{|I_k(j)|}{|I(j)|} \cdot \bar{u}_{k+1}.
\]

The second line follows from breaking the sum over the \( K \) sets \( I_k(j) \), and the third line follows from the bound (A.13). We will now use the fact that the sets \( I_k(i) \) and \( I_k(j) \) have approximately the same size, due to the way we chose \( K \). Using condition (A.12) we can bound the expression above as

\[
  \sum_{k=1, \ldots, K} \frac{|I_k(j)|}{|I(j)|} \cdot \bar{u}_{k+1} \leq \sum_{k=2, \ldots, K} \frac{|I_k(i)|}{|I(j)|} \cdot \bar{u}_k + \frac{3}{K}.
\]

This follows because the first \( K - 1 \) terms in the sum in the LHS are bounded by the terms in the sum in the RHS, plus \( 2(K - 1) \) error terms of at most \( 1/K^2 \), due to lottery numbers not being exactly evenly distributed in the \( L_k \) intervals, as per inequality (A.12). The final term \( |I_K(j)|/|I(j)| \cdot \bar{u}_{K+1} \) has to be bounded by the maximum possible utility, \( \bar{u}_{K+1} = 1 \), times the proportion of lottery numbers in the last interval, which cannot exceed \( 1/K + 1/K^2 \).

Now we will 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 $\bar{u}_k$. Using inequality (A.14) we have

$$\sum_{k=2,\ldots,K} \frac{|I_k(i)|}{|I(i)|} \cdot \bar{u}_k + \frac{3}{K} \leq \sum_{k=2,\ldots,K} \sum_{i'\in I_k(i)} \frac{|I_k(i)|}{|I(i)|} \cdot \frac{u_{i_k} [x_{i'k}(t,l)]}{|I_k(i)|} + \frac{3}{K}.$$ 

The first term is lower than the average utility received by type $t_i$ agents.\footnote{The first term is lower than average utility because the sum does not include the term for $k = 1$.} Therefore, remembering that we started with the average utility a type $t_i$ agent derives from the bundle of a type $t_j$ agent in inequality (A.15), we have found that this average utility is bounded by

$$\sum_{i'\in I(j)} \frac{u_{i_k} [x_{i'k}(t,l)]}{|I(j)|} \leq \sum_{i\in I(j)} \frac{u_{i_k} [x_{i'k}(t,l)]}{|I(i)|} + \frac{3}{K(l)}.$$ \hfill (A.16)

**Part 2: Bounding envy before the lottery draw.**

We can now use Equation (A.11) to bound how much player $i$ envies player $j$. Integrating the bound (A.16) for all lottery draws we have

$$E(t', t, n) \equiv u_{t_i} [\Phi_j^n(t)] - u_{i_k} [\Phi^n_i(t)] \leq \int_{l\in[0,1]^n} \frac{3}{K(l)} dl.$$ \hfill (A.17)

The last step of the proof is showing that, averaging 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 iid, 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 $\delta > 0$,

$$\Pr \{ \sup_x |\hat{F}_{i'}(x|l) - x| > \delta \} \leq 2e^{-2|I(i')|\delta^2}.$$ \hfill (A.18)

Fixing a partition size $K$, the conditions in (A.12) for the number of agents in each
interval to be close to $1/K$ can be written as
\[
|\hat{F}_v(k + 1/K|l) - \hat{F}_v(k|l) - 1/K| \leq 1/K^2.
\]
for $k = 1, \ldots, K - 1$ and $i' = i, j$. Applying the inequality (A.18), using $\delta = 1/2K^2$, we have that the probability that each such condition is violated is bounded by
\[
\Pr\{|I_k(i')| - 1/K > 1/K^2\} \leq 2 \cdot \exp(-|I(i')|/2K^4).
\]
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 the conditions above. Therefore,
\[
\Pr\{K(l) \geq \bar{K}\} \geq 1 - 2\bar{K} \exp(-|I(i')|/2\bar{K}^4) + \exp(-|I(j)|/2\bar{K}^4).
\]
Using this, we can bound the integral in the right side of Equation (A.17). 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 is at most equal to
\[
\bar{E}_{i,j}(t) \leq \frac{3}{\bar{K}} + 3 \Pr\{K(l) < \bar{K}\} \\
\leq \frac{3}{\bar{K}} + 6\bar{K} \exp(-|I(i)|/2\bar{K}^4) + \exp(-|I(j)|/2\bar{K}^4).
\]
Taking $\bar{K} = \lceil\min_{i' = i,j}|I(i')|^{1/4-\epsilon}\rceil$, and noting that $\bar{K} \geq 1$ we have that there exists $C_E$, uniform over all $i, j, t, n$, such that
\[
E(t_i, t, n) \leq C_E \cdot \min_{i' = i,j}|I(i')|^{-1/4+\epsilon}.
\]
In particular, this implies the result in the statement of the Lemma.

\[\square\]

A.2 Proof of Theorem 1

The core of the proof is the following approximation Lemma.

**Lemma A3.** Fix a prior $\mu_0$ and $\epsilon > 0$. Let $\mathcal{N}$ be a neighborhood as in Definition 9. For each $k = 1, \cdots, K$ let $\mu_k$ be a prior in $\mathcal{A}_k$, with $|\mu_k - \mu_0| < \epsilon$. Then there exists $n_0$ such
that, for all \( n \geq n_0 \), there exist positive weights \( \pi^n_k \) with \( \sum_{1 \leq k \leq K} \pi^n_k = 1 \), such that for all \( t_i \)

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

where

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

The Lemma states that the bundle received by an agent playing \( t_i \) in the constructed mechanism can be approximated by a convex combination of the bundles received when playing the original equilibrium within each region \( A_k \). Each \( z_k(t_i) \) is defined as the bundle an agent receives when playing \( t_i \) when the probability distribution of opponents’ types and the prior on which equilibrium is selected are each within \( A_k \). The key assertion that the approximation Lemma makes is that the \( \pi^n_k \) do not depend on \( t_i \). That is, irrespective of the type an agent reports, the approximation weights can be taken to be the same. This reflects the fact that a single agent has a very small effect on the probability of the distribution of types falling within each region \( A_k \).

With the Lemma in hand, Theorem 1 follows from an intuitive argument.

**Proof of Theorem 1. Part 1.**

To see that the constructed mechanism is SP-L, consider the gain for type \( t_i \) from deviating to \( \hat{t}_i \) when opponents play \( \mu_0 \). That is,

\[
u_{t_i}[f^n(\hat{t}_i, \mu_0)] - u_{t_i}[f^n(t_i, \mu_0)].
\]

By the approximation Lemma, and the boundedness of \( u \), given \( \epsilon > 0 \), there exists \( n_0, \pi^n_k, \mu_k \), and \( z_k \), for \( k = 1, \ldots, K \), as in the statement of the Lemma such that, for all \( n > n_0 \):

\[
|u_{t_i}[f^n(\hat{t}_i, \mu_0)] - \sum_{k=1, \ldots, K} \pi^n_k \cdot u_{t_i}[z_k(\hat{t}_i)]| < \epsilon/2 \quad (A.19)
\]

\[
|u_{t_i}[f^n(t_i, \mu_0)] - \sum_{k=1, \ldots, K} \pi^n_k \cdot u_{t_i}[z_k(t_i)]| < \epsilon/2. \quad (A.20)
\]

Also, by the definition of \( z_k(\cdot) \), since \( \sigma^*_{\mu_k}(\cdot) \) is a limit Bayes-Nash equilibrium, we have that

\[
u_{t_i}[z_k(t_i)] \geq u_{t_i}[z_k(\hat{t}_i)]. \quad (A.21)
\]
Therefore, we may bound the gain from deviating for $n > n_0$ by

$$u_t_i[f^n(t_i, \mu_0)] - u_t_i[f(t_i, \mu_0)] \leq 
\sum_k \pi_k^n \cdot \{u_t_i[z_k(t_i)] - u_t_i[z_k(t_i)]\} 
+ |u_t_i[f^n(t_i, \mu_0)] - \sum_k \pi_k^n \cdot u_t_i[z_k(t_i)]| 
\leq 0 + \epsilon/2 + \epsilon/2 = \epsilon.$$

The first inequality follows from the triangle inequality, and the second inequality from the bounds in inequalities (A.19), (A.20), (A.21).

**Part 3.**

Part 3 follows from the approximation Lemma. Given $\mu_0 \in \Delta T, \epsilon > 0$, by the Lemma we may take $n_0, \mu_k$, for each $k = 1 \cdots K$, such that $|\mu_k - \mu_0| < \epsilon$, and for all $n \geq n_0$

$$|f^n(t_i, \mu_0) - \sum_k \pi_k^n \cdot \phi^{\infty}(\sigma^{*}_{\mu_k}(t_i), \sigma^{*}_{\mu_k}(\mu_k))| < \epsilon/2.$$

By the definition of the limit, we may take $n_0$ such that for all $k, t_i, n \geq n_0$

$$|\phi^{\infty}(\sigma^{*}_{\mu_k}(t_i), \sigma^{*}_{\mu_k}(\mu_k)) - \phi^n(\sigma^{*}_{\mu_k}(t_i), \sigma^{*}_{\mu_k}(\mu_k))| < \epsilon/2.$$

By the triangle inequality and inequalities (A.22) and (A.23) we have that

$$|f^n(t_i, \mu_0) - \sum_k \pi_k^n \cdot \phi^{n}(\sigma^{*}_{\mu_k}(t_i), \sigma^{*}_{\mu_k}(\mu_k))| < \epsilon/2 + \epsilon/2 = \epsilon.$$

**Part 3.**

Finally for Part 2(a), note that we may take $\mathcal{N} = \mathcal{A}_1$ and $\mu_1 = \mu_0$ in the continuous case. Therefore $\pi_k^n = 1$, and equation (A.24) becomes

$$|f^n(t_i, \mu_0) - \phi^n(\sigma^{*}_{\mu_0}(t_i), \sigma^{*}_{\mu_0}(\mu_0)| < \epsilon.$$
A.2.1 Proof of Lemma A3

In the proof we will use the following notation. If \( t' \) is a vector of types, and \( S \in \Delta T \), we will say that \( t' \in S \) if \( \text{emp}[t'] \in S \). 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 type is iid according to the distribution \( \mu \).

The proof of the Lemma involves three steps. The first step shows that the approximation formula holds within each region \( A_k \).

Step 1.

There exists \( n_0 \) such that, for all \( n > n_0 \) and \( t \in A_k \) we have
\[
|\Phi^n_i(\sigma^*_\mu(t_i), \sigma^*_\mu(t_{-i})) - \phi^\infty(\sigma^*_\mu(t_i), \sigma^*_\mu_k(\mu_k))| < 4\epsilon.
\]

Using the \( z_k(t_i) \) notation, this is
\[
|\Phi^n_i(\sigma^*_\mu(t_i), \sigma^*_\mu(t_{-i})) - z_k(t_i)| < 4\epsilon.
\]

Proof. First note that, by Condition 2 of Definition 9 we may take \( n_1 \) such that for \( n \geq n_1 \)
\[
|\Phi^n_i(\sigma^*_\mu(t_i), \sigma^*_\mu(t_{-i})) - \Phi^n_i(\sigma^*_\mu_k(t_i), \sigma^*_\mu_k(t_{-i}))| < \epsilon. \tag{A.25}
\]

Note that the left term \( \Phi^n_i(\sigma^*_\mu(t_i), \sigma^*_\mu(t_{-i})) \) is the term whose distance to \( z_k(t_i) \) we wish to bound. We will do so by showing that \( \Phi^n_i(\sigma^*_\mu_k(t_i), \sigma^*_\mu_k(t_{-i})) \) is close to \( \phi^n(\sigma^*_\mu_k(t_i), \sigma^*_\mu_k(\mu_k)) \), and then showing that \( \phi^n(\sigma^*_\mu_k(t_i), \sigma^*_\mu_k(\mu_k)) \) is close to \( z_k(t_i) \).

By definition we have that
\[
\phi^n(\sigma^*_\mu_k(t_i), \sigma^*_\mu_k(\mu_k)) = \sum_{t'_{-i}} \Pr(t'_{-i}|t'_{-i} \sim \mu_k) \cdot \Phi^n_i(\sigma^*_\mu_k(t_i), \sigma^*_\mu_k(t'_{-i})). \tag{A.26}
\]

We will now bound the distance between \( \phi^n(\sigma^*_\mu_k(t_i), \sigma^*_\mu_k(\mu_k)) \) and \( \Phi^n_i(\sigma^*_\mu_k(t_i), \sigma^*_\mu_k(t_{-i})) \). For
all \( t \in \mathcal{A}_k \) we have

\[
\left| \Phi^n_i(\sigma^*_k(t_i), \sigma^*_h(t_{-i})) - \phi^n(\sigma^*_k(t_i), \sigma^*_h(\mu_k)) \right| \\
= \left| \Phi^n_i(\sigma^*_k(t_i), \sigma^*_h(t_{-i})) - \sum_{t'_{-i}} \Pr(t'_{-i} | t'_{-i} \sim \mu_k) \cdot \Phi^n_i(\sigma^*_k(t_i), \sigma^*_h(t'_{-i})) \right| \\
\leq \sum_{t'_{-i}} \Pr(t'_{-i} | t'_{-i} \sim \mu_k) \cdot \left| \Phi^n_i(\sigma^*_k(t_i), \sigma^*_h(t_{-i})) - \Phi^n_i(\sigma^*_k(t_i), \sigma^*_h(t'_{-i})) \right| \\
= \sum_{t'_{-i} : \text{emp}[t_i, t'_{-i}] \in \mathcal{A}_k} \Pr(t'_{-i} | t'_{-i} \sim \mu_k) \cdot \left| \Phi^n_i(\sigma^*_k(t_i), \sigma^*_h(t_{-i})) - \Phi^n_i(\sigma^*_k(t_i), \sigma^*_h(t'_{-i})) \right| + \\
\sum_{t'_{-i} : \text{emp}[t_i, t'_{-i}] \notin \mathcal{A}_k} \Pr(t'_{-i} | t'_{-i} \sim \mu_k) \cdot \left| \Phi^n_i(\sigma^*_k(t_i), \sigma^*_h(t_{-i})) - \Phi^n_i(\sigma^*_k(t_i), \sigma^*_h(t'_{-i})) \right| \\
(A.27)
\]

The first equality follows by substituting the definition of \( \phi^n \) from Equation (A.26). 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 \( \text{emp}[t_i, t'_{-i}] \) is in \( \mathcal{A}_k \), and the ones for which it is not. Consider now the expression on the right side of the last equality. Note that we may take \( n_1 \) such that the first term is bounded by

\[
\sum_{t'_{-i} : \text{emp}[t_i, t'_{-i}] \in \mathcal{A}_k} \Pr(t'_{-i} | t'_{-i} \sim \mu_k) \cdot \left| \Phi^n_i(\sigma^*_k(t_i), \sigma^*_h(t_{-i})) - \Phi^n_i(\sigma^*_k(t_i), \sigma^*_h(t'_{-i})) \right| < \epsilon,
\]

which follows from Condition 2 in Definition 9. As for the second term, by the law of large numbers, we may take \( n_2 \) large enough such that the total probability that \( \text{emp}[t_i, t'_{-i}] \in \mathcal{A}_k \) is greater than \( 1 - \epsilon \). This bounds the second term by

\[
\sum_{t'_{-i} : \text{emp}[t_i, t'_{-i}] \notin \mathcal{A}_k} \Pr(t'_{-i} | t'_{-i} \sim \mu_k) \cdot \left| \Phi^n_i(\sigma^*_k(t_i), \sigma^*_h(t_{-i})) - \Phi^n_i(\sigma^*_k(t_i), \sigma^*_h(t'_{-i})) \right| < \epsilon.
\]

Substituting these bounds in inequality (A.27) then yields

\[
\left| \Phi^n_i(\sigma^*_k(t_i), \sigma^*_h(t_{-i})) - \phi^n(\sigma^*_k(t_i), \sigma^*_h(\mu_k)) \right| < \epsilon + \epsilon = 2\epsilon. \quad (A.28)
\]

Finally, by the definition of the limit we may take \( n_3 \) such that for all \( n > n_3 \)

\[
\left| \phi^n(\sigma^*_k(t_i), \sigma^*_h(\mu_k)) - \phi^\infty(\sigma^*_k(t_i), \sigma^*_h(\mu_k)) \right| < \epsilon. \quad (A.29)
\]

If we take \( n_0 = \max\{n_1, n_2, n_3\} \) the claim of Step 1 then follows from Inequalities (A.25), (A.28) and (A.29).

\[\square\]
The next step shows that the probability that a vector \((t_i, t_{-i})\) falls within region \(A_k\), 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 says an individual agent cannot appreciably change the probability that \(t\) falls within each \(A_k\), and therefore cannot have a large effect on the aggregate allocation.

**Step 2.**

There exists \(n_0\) such that, for all \(n > n_0\) there exist weights \(\pi^n_1, \ldots, \pi^n_K\) such that \(\sum_k \pi^n_k = 1\) and

\[
|\Pr((t_i, t_{-i}) \in A_k | t_{-i} \sim \mu_0) - \pi^n_k| < \epsilon/K
\]

for all \(k\) and all \(t_i\).

*Proof.* We begin by constructing numbers which will be approximately equal to the \(\pi^n_k\) in the statement of this step. Let

\[
\bar{\pi^n}_k = \Pr(t' \in A_k | t' \in T^n, t' \sim \mu_0)
\]

be the probability that a vector of \(n\) types drawn independently according to \(\mu_k\) is in \(A_k\). We will show that for large \(n\) these \(\bar{\pi^n}_k\) are very close to the probability \(\Pr\{(t_i, t_{-i}) \in A_k | t_{-i} \sim \mu_0\}\), for any type \(t_i\). To see this, consider the difference between the probability of a vector of types falling within region \(A_k\) when \(i\)'s type is fixed as \(t_i\), versus when \(i\)'s type is drawn randomly. This difference equals

\[
\Pr((t_i, t'_{-i}) \in A_k | t'_{-i} \in T^{n-1}, t'_{-i} \sim \mu_0) - \bar{\pi^n}_k = \\
\Pr((t_i, t'_{-i}) \in A_k, t' \notin A_k | t' \in T^n, t' \sim \mu_0) - \Pr((t_i, t'_{-i}) \notin A_k, t' \in A_k | t' \in T^n, t' \sim \mu_0).
\]

This expression equals the probability of choosing a vector \(t'\) where changing a single type \((i)'s\) from \(t'_i\) to \(t_i\) moves the vector of types from inside \(A_k\) to outside \(A_k\), minus the probability of choosing a vector where changing \(i\)'s type from \(t'_i\) to \(t_i\) moves the vector from outside \(A_k\) to inside \(A_k\). We now show that the probability of such vectors being drawn is very small in a sufficiently large market.

Consider the case where \((t_i, t'_{-i}) \notin A_k\), but \((t'_i, t'_{-i}) \in A_k\). 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 \((t_i, t'_{-i}) \notin A_k\). In that case, the segment \([t_i, t'_{-i}), t']\) must have a point that lies in \(\mathcal{B}\), as we
assumed \( \mathcal{N} \) to be convex. This means that the distance between \( t' \) and \( \mathcal{B} \) is at most \( 1/n \). By Condition 1 of Definition 9, we may take \( n_0 \) such that this probability is less than \( \epsilon/8 \). This argument then yields that

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

An analogous argument proves that we may assume 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 in Equation (A.30) yields that

\[
|\Pr((t_i, t'_{-i}) \in \mathcal{A}_k | t'_{-i} \in T^{n-1}, t'_{-i} \sim \mu_0) - \pi^n_k| < \epsilon/4 + \epsilon/4 = \epsilon/2. \quad (A.31)
\]

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

\[
\bar{\pi}^n_k = \pi^n_k / \sum_{k'} \bar{\pi}^n_{k'} . \quad (A.32)
\]

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 \( n > n_0 \)

\[
|1 - 1/\sum_{k'} \bar{\pi}^n_{k'}| < \epsilon/2. \quad (A.33)
\]

Putting this together, we may finish the proof of Step 2.

\[
|\Pr((t_i, t'_{-i}) \in \mathcal{A}_k | t'_{-i} \in T^{n-1}, t'_{-i} \sim \mu_0) - \pi^n_k| \\
\leq |\Pr((t_i, t'_{-i}) \in \mathcal{A}_k | t'_{-i} \in T^{n-1}, t'_{-i} \sim \mu_0) - \bar{\pi}^n_k| + |\pi^n_k - \bar{\pi}^n_k| \\
< \epsilon/2 + |\pi^n_k - \bar{\pi}^n_k| \\
= \epsilon/2 + |\bar{\pi}^n_k / \sum_{k'} \bar{\pi}^n_{k'} - \bar{\pi}^n_k| \\
= \epsilon/2 + |1 - 1/\sum_{k'} \bar{\pi}^n_{k'}| \cdot |\bar{\pi}^n_k| \\
< \epsilon/2 + \epsilon/2 < \epsilon.
\]

The series of steps in the above derivation were as follows. The second line derives from the triangle inequality. The third line uses the bound from Inequality (A.31). The fourth
line uses the definition of $\pi^n_k$ from equation (A.32). Finally, the fifth line is algebra, and the sixth line comes from the bound in inequality (A.33).

\[ |f^n(t_i, \mu_0) - \sum_{k=1}^{K} \pi^n_k \cdot z_k(t_i)| < 6\epsilon. \]

\[ f^n(t_i, \mu_0) - \sum_k \pi^n_k \cdot z_k(t_i) = \sum_{t_i} \Pr(t_{i-1} | t_i \sim \mu_0) \cdot \Phi^n_i(\sigma^n_{\bar{\mu}}(t_i), \sigma^n_{\mu}(t_{i-1})) - \sum_k \pi^n_k \cdot z_k(t_i). \]

This sum can be decomposed depending on whether $\hat{\mu}$ is in each of the $A_k$ sets or outside the union of the $A_k$ sets. We have

\[ f^n(t_i, \mu_0) - \sum_k \pi^n_k \cdot z_k(t_i) = \sum_{t_i : \hat{\mu} \notin \bigcup_k A_k} \Pr(t_{i-1} | t_i \sim \mu_0) \cdot |\Phi^n_i(\sigma^n_{\bar{\mu}}(t_i), \sigma^n_{\mu}(t_{i-1})) - z^n_i(t_i)| + \sum_{t_i : \hat{\mu} \in A_k} \Pr(t_{i-1} | t_i \sim \mu_0) \cdot \Phi^n_i(\sigma^n_{\bar{\mu}}(t_i), \sigma^n_{\mu}(t_{i-1})). \]

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

\[ |\sum_{t_i : \hat{\mu} \in A_k} \Pr(t_{i-1} | t_i \sim \mu_0) \cdot \Phi^n_i(\sigma^n_{\bar{\mu}}(t_i), \sigma^n_{\mu}(t_{i-1})) - z^n_i(t_i)| \leq \sum_{t_i : \hat{\mu} \in A_k} \Pr(t_{i-1} | t_i \sim \mu_0) \cdot |\Phi^n_i(\sigma^n_{\bar{\mu}}(t_i), \sigma^n_{\mu}(t_{i-1})) - z^n_i(t_i)| + |\sum_{t_i : \hat{\mu} \in A_k} \Pr(t_{i-1} | t_i \sim \mu_0) \cdot \Phi^n_i(\sigma^n_{\bar{\mu}}(t_i), \sigma^n_{\mu}(t_{i-1}))| \cdot \sum_{t_i : \hat{\mu} \in A_k} \Pr(t_{i-1} | t_i \sim \mu_0) \cdot |z^n_i(t_i)| \leq \max_{t_i : \hat{\mu} \in A_k} |\Phi^n_i(\sigma^n_{\bar{\mu}}(t_i), \sigma^n_{\mu}(t_{i-1})) - z^n_i(t_i)| \cdot \sum_{t_i : \hat{\mu} \in A_k} \Pr(t_{i-1} | t_i \sim \mu_0) \cdot |z^n_i(t_i)| \leq 1. \]

The first inequality follows from the triangle inequality. The second inequality bounds each term $|\Phi^n_i(\sigma^n_{\bar{\mu}}(t_i), \sigma^n_{\mu}(t_{i-1})) - z^n_i(t_i)|$ by the maximum value of these terms, and it bounds $|z^n_i(t_i)|$ by 1.
Consider now the right side of the last inequality. By step 1, we may take \( n_0 \) such that for all \( n \geq n_0 \),
\[
\max_{t_{-i}:i \in A_k} |\Phi^* (\sigma^*_\mu(t_i), \sigma^*_\mu(t_{-i})) - z_k(t_i)| < 4\epsilon.
\]

By step 2, \( n_0 \) may be taken such that, for \( n \geq n_0 \), the second term is bounded by
\[
| \sum_{t_{-i}:i \in A_k} \Pr(t_{-i} | t_{-i} \sim \mu_0) - \pi_k^n | < \frac{\epsilon}{K}.
\]

Substituting these two bounds in inequality (A.35) we have that for all \( n \geq n_0 \)
\[
| \sum_{t_{-i}:i \in A_k} \Pr(t_{-i} | t_{-i} \sim \mu_0) \cdot \Phi^* (\sigma^*_\mu(t_i), \sigma^*_\mu(t_{-i})) - \pi_k^n \cdot z_k^n(t_i) | \leq 5\epsilon.
\]

Summing over all \( k \) we get
\[
\sum_k | \sum_{t_{-i}:i \in A_k} \Pr(t_{-i} | t_{-i} \sim \mu_0) \cdot \Phi^* (\sigma^*_\mu(t_i), \sigma^*_\mu(t_{-i})) - \pi_k^n \cdot z_k^n(t_i) | \leq 5\epsilon.
\]

Using the triangle inequality the sum operator can be brought into the norm, yielding the inequality
\[
| \sum_k ( \sum_{t_{-i}:i \in A_k} \Pr(t_{-i} | t_{-i} \sim \mu_0) \cdot \Phi^* (\sigma^*_\mu(t_i), \sigma^*_\mu(t_{-i})) - \pi_k^n \cdot z_k^n(t_i) ) | \leq 5\epsilon. \tag{A.36}
\]

The argument above bounds the terms in equation (A.34) that correspond to \( t \) within the sets \( A_k \). To bound the other term, note that we may take \( n_0 \) to be large enough so that for all \( n \geq n_0 \) the total probability that \( t \notin \bigcup_k A_k \) is strictly less than \( \epsilon \). That is,
\[
\sum_{t_{-i}:i \notin \bigcup_k A_k} \Pr(t_{-i} | t_{-i} \sim \mu_0) < \epsilon. \tag{A.37}
\]

Plugging equations (A.36) and (A.37) into equation (A.34) we obtain
\[
| f^n(t_i, \mu_0) - \sum_{k=1}^K \pi_k^n \cdot z_k(t_i) | < 6\epsilon,
\]
completing the proof of Step 3, and hence the Lemma. \( \square \)
References


Supplementary Appendix (Not Intended for Publication)

B Details for Table 1

This Section provides references on the mechanisms in Table 1, and formally classifies them as SP-L or not SP-L. Multi-unit auctions are described in the main text.

B.1 Anonymous Mechanisms.

B.1.1 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.$^{30}$ 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 null object $\emptyset$, in supply $n - \sum_{x_0' \neq \emptyset} \{q_{x_0'} \cdot n\} \geq 0$, so that there 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.

Probabilistic Serial Mechanism

The probabilistic serial mechanism has been proposed as a solution to the assignment problem by Bogomolnaia and Moulin (2001). The main advantage of this mechanism is that it satisfies an efficiency property known as ordinal efficiency. Subsequently, Kojima and Manea (2010) have shown that the mechanism has good incentive properties in large markets.

The probabilistic serial 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. Their setting formally encapsulates ours.

Bogomolnaia and Moulin (2001) show that the mechanism is EF. Consequently, Proposition 1 guarantees that it is SP-L. The fact that this mechanism is SP-L is a particular case

$^{30}$A bracketed expression denotes the nearest integer to the real number within brackets.
of Kojima and Manea’s Theorem 1. In fact, Kojima and Manea show that an agent with a fixed utility function has no incentives to deviate in a market with sufficiently many copies of each object.

**Boston Mechanism**

The Boston mechanism is a mechanism used in many cities to allocate seats in public schools. It was documented by Abdulkadiroğlu and Sönmez (2003). They demonstrate that the mechanism is not strategyproof, and Abdulkadiroğlu et al. (2006) document that the mechanism 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 agent 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 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 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 $m \in \Delta T$ 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

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

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.

**Hylland and Zeckhauser Pseudo-Market Mechanism**

Hylland and Zeckhauser (1979) proposed a pseudo-market mechanism, that always produces a Pareto efficient allocation. Unlike the probabilistic serial mechanism, the Hylland and Zeckhauser mechanism takes cardinal preference information into account. Note that, formally, the setting considered by Hylland and Zeckhauser (1979) is strictly more general than ours. They allow for example for indifferences, whereas we are restricting our attention to strict preferences. For that reason, we do not repeat formal definitions here, referring the reader to the original paper for further details.

The Hylland and Zeckhauser mechanism uses a market for probability shares of each object. Agents are initially endowed with equal budgets of an imaginary currency, with which they can purchase probability shares of the objects. Hylland and Zeckhauser show that there always exists a market clearing price, in a pseudo-market where agents are allowed to trade probability shares. This defines a vector of equilibrium probability shares to be given to each agent. Fix one such symmetric equilibrium, and define the allocation \(\Phi_i^n(t)\) as being a distribution over \(X^n_0\) such that the marginal distributions \(\Phi_i^n(t)\) equal the probability shares of each object agent \(i\) receives in the equilibrium.

Since the Hylland and Zeckhauser procedure assigns probability shares in a competitive equilibrium, each agent’s bundle \(\Phi_i^n(t)\) is optimal given prices. Therefore, as Hylland and Zeckhauser observe in page 307, the mechanism is EF. It follows that it is SP-L.

**B.1.2 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.\(^{31}\) 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.

**Competitive Equilibrium from Equal Incomes (CEEI)**

Budish (2011) proposed a pseudo-market mechanism to solve the course allocation problem. The mechanism addresses several shortcomings of previous course allocation mechanisms, and has good ex post fairness properties. 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 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 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 CEEI mechanism selects one such equilibrium, and gives each agent his equilibrium allocation. This defines a function \( x^n(t, l) \) giving an assignment of bundles for each vector of types \( t \) and lottery draws \( l \), which we take to be symmetric over coordinates. Finally, the 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 Proposition 1. By the definition of a competitive equilibrium, 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.

**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

\(^{31}\) \( \mathcal{P}(J) \) denotes the power set of \( J \).
of courses. Instead, students submit a preference ordering over single courses. To examine the possibility of truthful reporting, we restrict our attention to additive preferences over bundles, and strict over courses. We will say that a student of type \( t_i \) prefers course \( A \) to course \( B \) if she prefers a bundle consisting only of course \( A \) to a bundle consisting only of course \( B \). We will see that, even if \( T \) only contains additive preferences, students still have strong incentives to misreport.

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 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, \( J = \{A, B, C, D\} \), 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 \( j \in \{A, B, C, D\} \). Consider a probability distribution over students’ reports where \( \frac{1}{3} \) of the population lists courses in the order \( A, B, C, D \), \( \frac{1}{3} \) lists courses in the order \( B, A, C, D \), and \( \frac{1}{3} \) lists courses in the order \( A, C, D, B \). Given this distribution of reports, the probability that course \( A \) reaches capacity in the first round of the draft converges to 1 as the market grows large. The probability that course \( B \) is available in the second round is also 1 in a large market. For this reason, a student whose true preference order is \( B, A, C, D \) profits by misreporting as \( A, B, C, D \): by doing so, the student receives both \( A \) and \( B \), her two favorite courses, rather than courses \( B \) and \( C \) if she reports truthfully.\(^\text{32}\)

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 Krishna and Ünver (2008), who demonstrated that the mechanism is flawed in many 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

\(^{32}\)This particular profitable misreport is valid for any cardinal preferences consistent with the ordinal preferences \( B, A, C, D \). In other examples the profitability of a particular misreport might depend on cardinal preference information.
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. We will show that the mechanism does not have a weakly dominant strategy, even if preferences are additive. To focus on this setting, let $T$ be the set of additive preferences over courses, such that the vector $(u_t(g))_{g \in G}$ is in $A$. In this case, there is a natural bijection between $T$ and $A$. We will see that not only students do not wish to report the bids corresponding to their types, but also that there is no strategy that is weakly dominant in the limit.

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,

$$u_t(j_A) = u_t(j_B) = u_t(j_C) = B/3,$$

$$u_t(\emptyset) = 0.$$  \hfill (B.2)

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 (B.2), then the agent does not receive any course with very high probability. 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 in the sense defined in the text, as students would rather not play the action associated with their types.

Moreover, the mechanism does not even have a strategy that is weakly dominant in the limit. Consider an alternative distribution of play, $m'$, such that the last accepted bid for courses $j_A, j_B, j_C$ is $B/4$ with very high probability. Then reporting true preferences, that is, bidding $a_i(j_A) = a_i(j_B) = a_i(j_C) = B/3$, will give the agent all three desired courses, as opposed to only one of them. Therefore, no strategy is weakly dominant in the limit.

**The Generalized Hylland and Zeckhauser Pseudo-Market**

Budish et al. (Forthcoming) have proposed an extension of the Hylland and Zeckhauser (1979) pseudo-market mechanism that can be used for multi-unit assignment problems. Their mechanism can accommodate important elements of real-life problems, such as scheduling and curricular constraints in course allocation. At the same time, they show that the mechanism is ex-ante Pareto efficient and envy-free, like the original Hylland and Zeckhauser
(1979) pseudo-market mechanism.

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. 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 allocation that gives each agent her equilibrium probability share. Budish et al.’s Theorems 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 Proposition 1, this implies that the mechanism is SP-L.

B.1.3 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. We refer the reader to Jackson and Manelli for an overview and references.

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 vector \( t \) with a finite number of types, there always exists at least one competitive equilibrium where all agents of the same type receive identical bundles. 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

\[
 u_{t_i} = v_{t_i}(e_{t_i} + x_0) \text{ if } e_{t_i} + x_0 \in \mathbb{R}^J_+ \\
 -\infty \text{ if } e_{t_i} + x_0 \notin \mathbb{R}^J_+.
\]

We let \( X_0 \) be a compact cube in \( \mathbb{R}^J_+ \) that includes all possible vectors of net trades received by an agent in a competitive equilibrium. Formally, \( X_0 = [-\max_{j,t_i} e_{t_i}^j, \max_{j,t_i} e_{t_i}^j]^J \).
Having defined $X_0$ and $T$, we now define the mechanism. For all $n, t$, $\Phi^n(t)$ 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, by definition. However, this mechanism does not fall within our framework in the body of the paper for two reasons. First, $X_0$ includes an infinite number of points. Second, some although utility of all net trades is bounded above by 1, some net trades have yield utility of $-\infty$ to some players.

Nevertheless, it is simple to show that Proposition 1 still holds in this slightly more general setting. To see this, note that utility of any bundle is bounded above by 1. This implies that the argument in the proof of Proposition A1 leading to the bound A.4 follows from identical arguments. Moreover, since there is still a finite number of types, Lemma A1 also holds. Note that the term $E(t_i', t, n)$ is bounded above by 1, as a player who reports truthfully is guaranteed a minimum utility of 0. Along with Lemma A1, and inequality A.4, this implies the bound in A.5, and therefore that the Walrasian mechanism is SP-L.

B.2 Semi-Anonymous Mechanisms.

See Appendix C.1 for a definition of semi-anonymous mechanisms.

B.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 side, which is $s_{t_i} = b$(uyer) or $s$(eller), and her value for the object, which is $v_{t_i}$. Sellers are endowed with an unit of the object, while buyers are not. The set of types is $T = S \times V$, with $S = \{b, s\}$ and $V = \{1, \cdots, \bar{v}\}$. Buyers and sellers are in different groups. A bundle $x_0$ specifies whether the agent trades or not, $\tau_{x_0} = 0$ 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 $p_i - v_i$ 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, so therefore cannot 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.\(^{33}\) Therefore, double auctions are SP-L.

### B.2.2 Matching

Since these mechanisms are semi-anonymous the setting is defined formally in Section C.1. That Section also defines stable matching mechanisms, which are SP-L because stability implies 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 clearing-houses 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 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 have no preferences over one man or the other. In this

\(^{33}\)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.
case, the priority match mechanism coincides with the Boston mechanism, which we know 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.

C Extensions and Robustness

C.1 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).

Assume that agents belong to groups $g$ in a finite set $G$. Each group has a different set of possible types and actions, so that

\[ T = T_{g_1} \cup T_{g_2} \cup \cdots \cup T_{g_G} \]
\[ A = A_{g_1} \cup A_{g_2} \cup \cdots \cup A_{g_G}. \]

A semi-anonymous mechanism is defined as \( \{(\Phi^n)_{n \in \mathbb{N}}, (A_g)_{g \in G}\} \). As before, the \( \Phi^n \) are functions

\[ \Phi^n : A^n \rightarrow \Delta(X_0^n). \]

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.

**Example C1.** (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

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

That is, there are $M$ groups of men and $W$ groups of women. Men and women within each group have the same traits, and hence 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 either their sex or their 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 used in practice (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 over its coordinates, to conform to the semi-anonymity assumption. The mechanism is then defined as

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

The extension of the definition of SP-L is straightforward. A semi-anonymous mechanism is SP-L if no agent wants to misreport to a type within the same group.

**Definition C1.** Semi-anonymous mechanism $\{(\Phi^n)_M, (T_g)_{g \in G}\}$ is _strategyproof in the_
large, or SP-L, if, for all $g \in G$, $t_i, t'_i \in T_g$, and $m \in \Delta T$:  

$$u_{t_i}[\phi^\infty(t_i, m)] \geq u_{t_i}[\phi^\infty(t'_i, m)].$$  \hfill (C.1)

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 C2.** A direct semi-anonymous mechanism $\{(\Phi^n)_n, T\}$ is envy-free modulo tie-breaking (EF-TB) if for each $n$ there exists a function $x^n : (T \times [0,1])^N \rightarrow \Delta(X^n_0)$, 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

$$u_{t_i}[x^n_i(t, l)] \geq u_{t_i}[x^n_j(t, l)].$$

With this definition, an extension of Proposition 1 to semi-anonymous mechanisms follows from essentially the same proof. This implies that the stable matching procedure in example C1 is SP-L, as an agent envying another agent with a lower lottery number would violate the stability condition.

We will now extend the definition of limit BNE to this setting, and state and prove an extension of Theorem 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 1. The proof follows from noting that the argument in the anonymous case implies that the approximation formulas in Theorem 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 $\sigma : T \rightarrow \Delta A$ such that if $t_i \in T_g$ then the support of $\sigma(t_i)$ is contained in $A_g$.

**Definition C3.** Given a semi-anonymous mechanism $\{(\Phi^n)_{n\in\mathbb{N}}, (A_g)_{g\in G}\}$, the strategy $\sigma^*_\mu(\cdot)$
is a limit $\mu$-BNE 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 continuous and quasi-continuous families of limit equilibria are then identical to the anonymous case, as is the definition of a direct mechanism to approximate outcomes of limit equilibria of a mechanism. With these definitions, the statement of Theorem 1 for the semi-anonymous case is identical, save for the broader class of mechanisms, to that for the anonymous case.

**Theorem C1.** Consider a semi-anonymous mechanism $\{(\Phi^\mu)_N, (A_g)_{g \in G}\}$ with a quasi-continuous family of limit equilibria $(\sigma^*_\mu)_{\mu \in \Delta T}$. Define the direct mechanism $\{(F^n)_N, (T_g)_{g \in G}\}$ as in Equation (5.1). Then,

1. The direct mechanism is SP-L.

2. If at some prior $\mu_0 \in \bar{\Delta} T$ the family $(\sigma^*_\mu)_{\mu \in \Delta T}$ is continuous, then outcomes of the direct mechanism $\{(F^n)_N, T\}$ approximate equilibrium outcomes of $\{(\Phi^\mu)_N, A\}$ under $(\sigma^*_\mu)_{\mu \in \Delta T}$ at $\mu_0$.

3. If at some prior $\mu_0 \in \bar{\Delta} T$ the family $(\sigma^*_\mu)_{\mu \in \Delta T}$ is quasi-continuous, then outcomes of the direct mechanism $\{(F^n)_N, T\}$ approximate a convex combination of equilibrium outcomes of $\{(\Phi^\mu)_N, A\}$ under $(\sigma^*_\mu)_{\mu \in \Delta T}$ at $\mu_0$.

**Proof.** Parts 2 and 3 of the Theorem follow from the same argument as in the proof of Theorem 1. This is the case since the argument deriving the approximation formulas in the proof of Theorem 1 does not use the fact that agents play strategies restricted to be in the action set of their groups.

To establish the first part of the Theorem, we employ a small modification of the original argument, as now the $\sigma^*_\mu$ are assumed to be limit equilibria of a semi-anonymous mechanism. Given $\epsilon > 0, g \in G, t_i, t'_i \in T_g, and m \in \bar{\Delta} T$, by Lemma A3 there exists $n_0$ such that for all $n \geq n_0$

$$u_{t_i}[f^n(t'_i, m)] - u_{t_i}[f^n(t_i, m)] < \sum_{k=1}^K \pi^n_k \cdot \{u_{t_i}[x^n_k(t'_i)] - u_{t_i}[x^n_k(t_i)]\} + \epsilon/2$$


where the weights $\pi^*_n$ sum to 1 and for any $t'_i \in T_g$

$$x^*_n(t''_i) = \phi^n(\sigma^*_{\mu_k}(t''_i), \sigma^*_{\mu_k}(\mu_k)).$$

Moreover, since $\sigma^*_{\mu_k}$ are limit equilibria, $n_0$ may be taken such that

$$u_{t_i}[x^n_{t_i}(t')] - u_{t_i}[x^n_{t_i}(t)] < \epsilon/2.$$  

Therefore,

$$u_{t_i}[f^n(t', m)] - u_{t_i}[f^n(t, m)] < \epsilon,$$

and the first part of the Theorem follows.

\[\square\]

C.2 Complete Information Nash Equilibria

Our construction in Section 4 takes as input a mechanism that has a family of Bayes-Nash equilibria. The same idea can be applied to a mechanism that has a family of complete-information Nash equilibria.

A complete information Nash equilibrium, in an economy where the type profile is known to be $t$, is defined in the usual way.

**Definition C4.** A (symmetric) t-Complete-Information Nash Equilibrium (t-CINE) of n-mechanism $\{\Phi^n, A\}$ is a strategy $\sigma^n_{\text{emp}[t]}(\cdot)$ such that, for all $t_i \in t$, $a'_i \in A$:

$$u_{t_i}[\Phi^n_i(\sigma^n_{\text{emp}[t]}(t_i), \sigma^n_{\text{emp}[t]}(t_{-i}))] \geq u_{t_i}[\Phi^n_i(a'_i, \sigma^n_{\text{emp}[t]}(t_{-i}))]$$

It is important to highlight that, in Definition C4, the strategy played when the vector of types is $t$ depends only on $\text{emp}[t]$. That is, we restrict attention to symmetric CINE, which are invariant to permutations of $t$. The fact that strategies $\sigma^n_{\text{emp}[t]}$ only depend on the empirical distribution of types allows us to define quasi-continuity in a manner analogous to Definition 9, which then yields a theorem statement analogous to Theorem 1.

The mechanism we construct to prove the CINE version of Theorem 1 is

$$F^n(t) = \Phi^n(\sigma^n_{\text{emp}[t]}(t)). \quad (C.2)$$

In words, agents report their types to the mechanism, which then computes a symmetric complete information Nash equilibrium strategy in the economy induced by the reports. Note that in general it is not a Nash equilibrium for players to report their preferences truthfully.
to this mechanism in finite markets. The reason is that, by changing one’s report from say $t_i$ to $t'_i$, one changes the profile of reported types from say $t$ to $t'$, and this in turn changes the strategies used by the proxy agents.

An interesting feature of the construction given by equation (C.2) is that, if agents tell the truth in finite markets, then (C.2) produces outcomes that are identical to the outcomes under the complete information Nash equilibria of the original mechanism. By contrast, with Bayes-Nash equilibria our constructed mechanism only approximates the finite market outcomes. Quasi-continuity is defined analogously to Definition 9, with the key difference being Condition 3).

**Definition C5.** Consider a mechanism $\{(\Phi^n)_N, A\}$ with limit $\phi^\infty$, and a family of Complete Information Nash equilibria $(\sigma^n_{emp[t]}(i \in T^n, n \in N \cup \infty)$. The family of equilibria is quasi-continuous at a prior $\mu_0 \in \Delta T$ if, for every $\epsilon > 0$, there exists a neighborhood $N$ of $\mu_0$ that can be decomposed as $N = \cup_{1 \leq k \leq K} A_k \cup B$ with each $A_k$ open, such that:

1. If types are drawn iid according to $\mu_0$, then the probability that the empirical distribution of types lands within distance $1/n$ of $B$ goes to zero as $n$ grows large. Formally,

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

2. For each $A_k$, there exists $n_0$ such that for any $n > n_0$, and any $\text{emp}[t_i, t_{-i}], \text{emp}[t'_i] \in A_k$, we have:

$$|\Phi^n_i(\sigma^n_{emp[t]}(t_i), \sigma^n_{emp[q]}(t_{-i})) - \Phi^n_i(\sigma^n_{emp[t]}(t_i), \sigma^n_{emp[q]}(t_{-i}))| < \epsilon.$$ 

The family of equilibria is continuous at $\mu_0$ if, for the prior $\mu_0$, Conditions 1 and 3 hold with $K = 1$ and $B = \emptyset$. The family is quasi-continuous if it is quasi-continuous at every prior in $\Delta T$.

The main theorem statement is analogous to Theorem 1.

**Theorem C2.** Consider a mechanism $\{(\Phi^n)_N, A\}$ with a quasi-continuous family of complete information Nash equilibria $(\sigma^n_{emp[t]}(i \in T^n, n \in N \cup \infty)$. Define the direct mechanism $\{(F^n)_N, T\}$ as in equation (C.2). Then,

1. The direct mechanism is SP-L.

2. For any size market $n$, truthful play of $\{(F^n)_N, T\}$ and Complete Information Nash equilibrium play of $\{(\Phi^n)_{n \in N}, A\}$ give agents the same utilities.
Part 2 of the Theorem statement is immediate from inspection of (C.2).

For Part 1 of the Theorem statement, fix arbitrary \( \mu_0 \in \bar{\Delta}T \). Initially, suppose that \((\sigma^n_{\text{emp}[t]}(t_i), n \in \mathbb{N} \cup \infty)\) is continuous at \( \mu_0 \). Choose arbitrary \( t_i, t'_i \), and \( n > n_0 \), and then choose arbitrary \( t_{-i} \) so that \( \text{emp}[t_i, t_{-i}] \) and \( \text{emp}[t'_i, t_{-i}] \) are in the neighborhood of \( \mu_0 \) as defined in Definition C5. Then, Nash equilibrium implies that

\[
 u_{t_i}(\Phi^n_{(\sigma_{\text{emp}[t]}(t_i), \sigma_{\text{emp}[t]}(t_{-i}))}) \geq u_{t_i}(\Phi^n_{(\sigma_{\text{emp}[t]}(t'_i), \sigma_{\text{emp}[t]}(t_{-i}))}) \tag{C.3} \]

Moreover, continuity implies that

\[
 u_{t_i}(\Phi^n_{(\sigma_{\text{emp}[t]}(t'_i), \sigma_{\text{emp}[t]}(t_{-i}))}) \geq u_{t_i}(\Phi^n_{(\sigma_{\text{emp}[t'_i, t_{-i}]}, \sigma_{[t'_i, t_{-i}]}(t_{-i}))}) - \epsilon \tag{C.4} \]

Combining inequalities (C.3) and (C.4) yields

\[
 u_{t_i}(\Phi^n_{(\sigma_{\text{emp}[t]}(t_i), \sigma_{\text{emp}[t]}(t_{-i}))}) \geq u_{t_i}(\Phi^n_{(\sigma_{\text{emp}[t'_i, t_{-i}]}, \sigma_{[t'_i, t_{-i}]}(t_{-i}))}) - \epsilon \]

Hence, for realizations of \( t_{-i} \) with empirical distributions close enough to \( \mu_0 \), type \( t_i \)'s gain from misreporting as \( t'_i \) is bounded above by \( \epsilon \). Additionally, by the law of large numbers, there exists \( n_0 \), such that, for \( n \geq n_0 \), the probability that \( t_{-i} \) is close enough to \( \mu_0 \) (i.e., that both \( \text{emp}[t_{i, t_{-i}}] \) and \( \text{emp}[t'_{i, t_{-i}}] \) are in the relevant neighborhood) is greater than \( 1 - \epsilon \). Hence, \( t_i \)'s total gain from misreporting, if his opponents’ reports are distributed according to \( m \), is at most \( 2\epsilon \) in a large enough market.

To complete the argument that \( \{(F^n)_N, T\} \) is SP-L, we need to address the case where \( \{(\Phi^n)_N, A\} \) is not continuous at \( m \). In this case, the same argument as given above for the continuous case works within each of the open sets \( A_k \) in the neighborhood of \( \mu_0 \), as defined in Definition C5. That is, within each set \( A_k \), we can bound \( t_i \)'s gain from misreporting by \( 2\epsilon \) in a large enough market. With an argument analogous to the proof of Theorem 1, it can be shown that agent \( i \) regards the probability that \( \text{emp}[t_{i, t_{-i}}] \in A_k \) as approximately exogenous to her report in a large enough market. By Condition 2 of Definition C5, the probability that by misreporting as \( t'_i \) he can change which of the sets \( A_k \) or \( B \) the empirical distribution lands on goes to zero as the number of players grows. Therefore, agent \( i \)'s gain from misreporting converges to zero as \( n \) grows.
C.3 Finite-Economy Bayes-Nash Equilibria

Theorem 1 starts from a family of limit Bayes-Nash equilibria \((\sigma^\mu_n)_{\mu \in \Delta T}\), and constructs a direct SP-L mechanism that implements approximately the same outcome. This construction could also be obtained based on a family of finite economy Bayes-Nash equilibria. This Section makes this point precise, giving an alternative version of Theorem 1.

The reason why we focus on limit equilibria for the statement of the main result is that finite-economy equilibria are often analytically less tractable. This is discussed in detail by Bodoh-Creed (2010), who gives conditions under which limit equilibria correspond to the limits of finite-economy equilibria. In multi-unit auctions, for example, closed form solutions for equilibria are often unavailable, and even showing basic properties of equilibria is a difficult problem (Swinkels (2001); Engelbrecht-Wiggans et al. (2006)). A particular analyst might thus find it more convenient to work with limit equilibria. Moreover, an analyst might even find limit equilibria more compelling in their own right. One argument in favor of limit equilibria is that in a game where exact Nash equilibria are cognitively and computationally very complex, it may be more likely that players reason through an approximate model. On the other hand, depending on the application, an analyst might see exact Nash equilibria of finite economies as a more appropriate solution concept. For these reasons, we do not take a view on which solution concept is in general more appropriate and useful, and provide versions of our main result for either solution concept.

We begin by defining the relevant continuity assumptions given a family of (finite-economy) Bayes-Nash equilibria \((\sigma^\mu_n)_{\mu \in \Delta T, n \in \mathbb{N} \cup \{\infty\}}\).

**Definition C6.** Consider a mechanism \(\{(\Phi^n)_{n}, A}\) with limit \(\phi^\infty\), and a family of Bayes-Nash equilibria \((\sigma^\mu_n)_{\mu \in \Delta T, n \in \mathbb{N} \cup \{\infty\}}\). The family of equilibria is **quasi-continuous** at a prior \(\mu_0 \in \Delta T\) if, for every \(\epsilon > 0\), there exists a neighborhood \(N\) of \(\mu_0\) that can be decomposed as \(N = \bigcup_{1 \leq k \leq K} A_k \cup B\) with each \(A_k\) open, such that:

1. The probability

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

   where \(t \sim \text{iid}(\mu_0)\) denotes a vector of \(n\) types \(t\) with each component drawn iid according to \(\mu_0\).

2. For each set \(A_k\), there exists \(n_0\) such that, for any \(n > n_0\), any \(\mu, \mu'\), \(\text{emp}[t_i, t_{-i}]\), \(\text{emp}[t_i, t'_{-i}] \in \)
\[ \mathcal{A}_k \text{ we have:} \]
\[ |\Phi^n_i(\sigma^n_\mu(t_i), \sigma^n_\mu(t_{-i}))-\Phi^n_i(\sigma^n_{\mu'}(t_i), \sigma^n_{\mu'}(t'_{-i}))| < \epsilon. \]

3. In addition, for all \( \mu \in \Delta T, t_i \in T \)
\[ \lim_{n \to \infty} \phi^n(\sigma^n_\mu(t_i), \sigma^n_\mu(\mu)) = \phi^\infty(\sigma^\infty_\mu(t_i), \sigma^\infty_\mu(\mu)). \]

The family is continuous at \( \mu_0 \) if conditions 2 and 3 hold with \( K = 1 \) and \( \mathcal{B} = \emptyset \). The family is quasi-continuous if it is quasi-continuous at every prior in \( \overline{\Delta T} \).

The Definition of a direct mechanism to approximate outcomes of the original mechanism is as follows. Consider a mechanism \( \{(\Phi^n)_{n \in \mathbb{N}}, A\} \) with a family of BNE \( (\sigma^n_\mu)_{\mu \in \Delta T, n \in \mathbb{N} \cup \{\infty\}} \). Outcomes of the direct mechanism \( \{(F^n)_{n \in \mathbb{N}}, T\} \) approximate outcomes of the original mechanism at \( \mu_0 \) if, for any \( \epsilon > 0 \), there exists \( n_0 \) such that, for all \( n > n_0 \) and all \( t_i \),
\[ |f^n(t_i, \mu_0) - \phi^n(\sigma^n_{\mu_0}(t_i), \sigma^n_{\mu_0}(\mu_0))| < \epsilon, \]
where \( f^n(\cdot) \) is constructed from \( F^n(\cdot) \) according to equation (2.4).

The direct mechanism approximates a convex combination of equilibrium outcomes of the original mechanism at \( \mu_0 \) if, given \( \epsilon > 0 \), there exists \( n_0 \), and integer \( K \), numbers \( \pi^n_k \) with \( \sum_{k=1, \ldots, K} \pi^n_k = 1 \), and priors \( \mu_k \) with \( |\mu_k - \mu_0| < \epsilon \) such that, for all \( n \geq n_0 \) and \( t_i \in T \),
\[ |f^n(t_i, \mu_0) - \sum_{k=1, \ldots, K} \pi^n_k \cdot \phi^n(\sigma^n_{\mu_k}(t_i), \sigma^n_{\mu_k}(\mu_k))| < \epsilon. \]

The Theorem statement is as follows.

**Theorem C3 (Alternative Statement of Theorem 1).** Consider a mechanism \( \{(\Phi^n)_{n \in \mathbb{N}}, A\} \) with a quasi-continuous family of equilibria \( (\sigma^n_\mu)_{\mu \in \Delta T, n \in \mathbb{N} \cup \{\infty\}} \). Define the direct mechanism \( \{(F^n)_{n \in \mathbb{N}}, T\} \) as
\[ F^n(t) = \Phi^n(\sigma^n_{\text{emp}[t]}(t)). \]

Then,

1. The direct mechanism is SP-L.

2. If at some prior \( \mu_0 \in \overline{\Delta T} \) the family \( (\sigma^n_\mu)_{\mu \in \Delta T, n \in \mathbb{N} \cup \{\infty\}} \) is continuous, then outcomes of the direct mechanism \( \{(F^n)_{n \in \mathbb{N}}, T\} \) approximate equilibrium outcomes of \( \{(\Phi^n)_{n \in \mathbb{N}}, A\} \) under \( (\sigma^n_\mu)_{\mu \in \Delta T, n \in \mathbb{N} \cup \{\infty\}} \) at \( \mu_0 \).
3. If at some prior \( \mu_0 \in \tilde{\Delta}T \) the family \( (\sigma^n_{\mu})_{\mu \in \Delta T, n \in \mathbb{N} \cup \{\infty\}} \) is quasi-continuous, then outcomes of the direct mechanism \( \{(F^n)_{n}, T\} \) approximate a convex combination of equilibrium outcomes of \( \{(\Phi^n)_{n}, A\} \) under \( (\sigma^n_{\mu})_{\mu \in \Delta T, n \in \mathbb{N} \cup \{\infty\}} \) at \( \mu_0 \).

The proof of this Theorem is largely analogous to that of the Theorem for limit Nash equilibria. For conciseness, we will discuss the points where the proofs diverge, and how to adjust the proof of Theorem 1, instead of giving a complete proof. The proof is also based on an approximation Lemma. The statement of the Lemma differs slightly.

**Lemma C1.** Fix a prior \( \mu_0 \) and \( \epsilon > 0 \). Let \( \mathcal{N} \) be a neighborhood as in Definition C6. Let \( \mu_k \) be priors \( \mu_k \in A_k \) for each \( k = 1, \ldots, K \), with \( |\mu_k - \mu_0| < \epsilon \). Then there exists \( n_0 \), such that for all \( n > n_0 \): there exist positive weights \( \pi^n_k \) with \( \sum_{1 \leq k \leq K} \pi^n_k = 1 \), such that for all \( t_i \)

\[
|f^n(t_i, \mu_0) - \sum_{k=1}^{K} \pi^n_k \cdot z_k(t_i)| < 6\epsilon,
\]

where

\[
z_k(t_i) = \phi^\infty(\sigma^\infty_{\mu_k}(t_i), \sigma^\infty_{\mu_k}(\mu_k)).
\]

The proof of the alternate Lemma is largely similar to the proof of Lemma A3. In fact, the steps are basically the same, but replacing \( \sigma^* \) by \( \sigma^n \) or \( \sigma^\infty \) as appropriate. The only step of the proof that differs significantly is deriving the analogue of Inequality (A.29). Mutatis mutandis, this inequality would be showing that we may take \( n_3 \) large enough such that

\[
|\phi^n(\sigma^n_{\mu_k}(t_i), \sigma^n_{\mu_k}(\mu_k)) - \phi^\infty(\sigma^\infty_{\mu_k}(t_i), \sigma^\infty_{\mu_k}(\mu_k))| < \epsilon.
\]  

(C.5)

This is still true. However, it does not follow from the definition of the limit, as in the proof of Lemma (A3). Instead, it is a consequence of Condition 4 in the definition of a quasi-continuous family of equilibria. The rest of the proof follows straightforwardly, with the modifications we described.

### C.4 Aggregate Uncertainty

The definition of SP-L requires truth-telling to be close to optimal if a player regards opponents’ actions as independently identically distributed. The reason for this requirement is that, in many large market settings, it is not reasonable to assume that an agent knows exactly what opponents’ actions are. Knowing only the distribution of opponents’ actions
is often a reasonable upper bound to the information each particular player has. Indeed, in many settings players may have even coarser information, having aggregate uncertainty about opponents’ play.

This Section shows that reporting truthfully in an SP-L mechanism is still close to optimal, even in the presence of aggregate uncertainty. Formally we show that, if agents are given strictly less information than an iid belief over opponent types, in the Blackwell sense, then it is approximately optimal to report truthfully in an SP-L mechanism.

Consider the gain from misreporting for an agent who knows that other agents’ actions are distributed according to $m \in \bar{\Delta}T$. If $\{(\Phi^n)_n, T\}$ is SP-L, then this gain must be vanishingly small. That is, given $\epsilon, t_i, t'_i$, there exists $n_0$ such that, for all $n \geq n_0$,

$$u_{t_i}[\phi^n(t'_i, m)] - u_{t_i}[\phi^n(t_i, m)] \leq \epsilon.$$

Now consider an agent who knows strictly less than agents with any iid beliefs. Following Blackwell, we define a garbling of iid beliefs as a measure $\nu \in \Delta(\Delta A)$. The agent assigns probability $\nu(m)$ that opponents’ types are iid according to $m \in \bar{\Delta}T$.

For an agent with such beliefs, the gain from deviating is

$$\int u_{t_i}[\phi^n(t'_i, m)] - u_{t_i}[\phi^n(t_i, m)]d\nu(m).$$

We now show it is approximately optimal for such an agent to report truthfully. Given $\epsilon > 0$, from the definition of SP-L, we know that for each $m \in \Delta T$ there must exist $n_0(m)$ such that

$$u_{t_i}[\phi^n(t'_i, m)] - u_{t_i}[\phi^n(t_i, m)] < \epsilon/2$$

for all $n \geq n_0(m)$. Now take $n_1$ such that $n_1 \geq n_0(m)$ for all $m$ in a set $M \subseteq \Delta T$ with measure $\nu(M)$ at least $1 - \epsilon/2$. We then have

$$\int_{M} u_{t_i}[\phi^n(t'_i, m)] - u_{t_i}[\phi^n(t_i, m)]d\nu(m) =$$

$$\int_{M} u_{t_i}[\phi^n(t'_i, m)] - u_{t_i}[\phi^n(t_i, m)]d\nu(m) +$$

$$\int_{M^c} u_{t_i}[\phi^n(t'_i, m)] - u_{t_i}[\phi^n(t_i, m)]d\nu(m) <$$

$$\epsilon/2 + \epsilon/2 = \epsilon.$$

This result may be formally stated as follows.
**Proposition C1.** Consider an SP-L mechanism \( \{(\Phi^n)_n \in \mathbb{N}, T\} \). For any garbling \( \nu \) of iid beliefs over opponents’ types, and any \( \epsilon > 0 \), there exists \( n_0 \) such that an agent with beliefs \( \nu \) in a market of size \( n \geq n_0 \) cannot gain more than \( \epsilon > 0 \) by misreporting her type.

### C.5 Ex post Robustness of SP-L Mechanisms in Large Finite Markets

Kalai (2004) studies the ex post robustness of Bayes-Nash equilibria in large games. Under an equicontinuity assumption that we provide below, he shows that Bayes-Nash equilibria are ex post robust in the following sense: For any \( \epsilon > 0 \), the probability that any player has an ex post deviation that yields a gain of more than \( \epsilon \) converges to 0 exponentially in market size \( n \).

Here we show that, under Kalai’s equicontinuity assumption, SP-L mechanisms satisfy a stronger robustness property. Namely, equicontinuity implies that, for any \( \epsilon > 0 \), in a large enough market, no player has an ex post deviation which increases her payoff by more than \( \epsilon \). This property is stronger than Kalai’s for two reasons. First, in a large enough market, the probability of an \( \epsilon \) deviation is exactly zero rather than converging to zero. Second, the fact that the ex post gain from any deviation is small does not depend on a player knowing the distribution of opponent types, nor on agents playing in equilibrium.

Formally, we follow Kalai and define an equicontinuous mechanism as follows.

**Definition C7.** A mechanism \( \{(\Phi^n)_n \in \mathbb{N}, A\} \) is equicontinuous if, for all \( \epsilon > 0 \) there exists \( \delta > 0 \) such that for all \( n, n', a_{-i} \in A^{n-1}, a'_{-i} \in A'^{n'-1} \) with

\[
|\text{emp}[a_{-i}] - \text{emp}[a'_{-i}]| < \delta
\]

we have that for all \( a_i \)

\[
|\Phi^n_i(a_i, a_{-i}) - \Phi^{n'}_i(a_i, a'_{-i})| < \epsilon.
\]

We then define \( \epsilon \)-Strategyproofness as follows.

---

34 More specifically, Kalai (2004) considers a sequence of (semi-)anonymous games, with an increasing number of players, that satisfies an equicontinuity condition. He defines an \( \epsilon \)-ex post Nash equilibrium profile as a profile of types and actions such that no player may gain more than \( \epsilon \) by changing her action. He defines a profile of (possibly mixed) strategies to be an \( (\epsilon, \rho) \) ex post strategy profile if with probability at least \( 1 - \rho \) the realized profile of types and strategies is an \( \epsilon \) ex post Nash equilibrium. Kalai’s Theorem 1 shows that, for any sequence of Bayes Nash equilibria \( (\sigma^n)_{n \in \mathbb{N}} \) of the games, and any \( \epsilon > 0 \), there exist constants \( \alpha > 0, \beta < 1 \) such that \( \sigma^n \) is an \( (\epsilon, \alpha \beta^n) \) ex post strategy profile. We refer the interested reader to Kalai (2004) for more details.
Definition C8. A direct mechanism \( \{ (\Phi^n)_{n \in \mathbb{N}}, T \} \) is \((\epsilon, n)\)-strategyproof if for all \( t \in T^n, t' \in T \)
\[
u_t_i[\Phi^n_i(t)] \geq u_{t_i}[\Phi^n_i(t', t_{-i})] - \epsilon.
\]

It is then straightforward to prove the following result (see Appendix D).

Proposition C2. If \( \{ (\Phi^n)_{n \in \mathbb{N}}, T \} \) is SP-L and equicontinuous, then given \( \epsilon > 0 \) there exists \( n_0 \) such that for all \( n > n_0 \), the mechanism \( \{ (\Phi^n)_{n \in \mathbb{N}}, T \} \) is \((\epsilon, n)\) strategyproof.

Proof of Proposition C2.

Step 1. Given \( \epsilon > 0 \), there exists \( n_0 \) such that for all \( n, n' \geq n_0, t_i, t_{-i} \) we have
\[
|\Phi^n_i(t_i, t_{-i}) - \phi^n(t_i, \text{emp } t_{-i})| < \epsilon.
\]

Proof. Let \( \hat{\mu} = \text{emp } t_{-i} \). We may write
\[
\phi^n(t_i, \hat{\mu}) = \sum_{t'_{-i}} \Pr(t'_{-i} \mid t'_{-i} \in T^{n-1}, t'_{-i} \sim \hat{\mu}) \cdot \Phi^n_i(t_i, t'_{-i}). \tag{C.6}
\]

By the definition of equicontinuity, we may take \( \delta > 0 \) such that for all \( t'_{-i} \) with
\[
|\text{emp } t'_{-i} - \hat{\mu}| < \delta
\]
we have
\[
|\Phi^n_i(t_i, t_{-i}) - \Phi^n_i(t_i, t'_{-i})| < \epsilon/2. \tag{C.7}
\]

Moreover, we may take \( n_0 \) such that for all \( n \geq n_0 \), by the law of large numbers,
\[
\sum_{|\text{emp } t'_{-i} - \hat{\mu}| \geq \delta, t'_{-i} \in T^{n-1}} \Pr(t'_{-i} \mid t'_{-i} \in T^{n-1}, t'_{-i} \sim \hat{\mu}) < \epsilon/2. \tag{C.8}
\]

Consider now the difference
\[
|\Phi^n_i(t_i, t_{-i}) - \phi^n(t_i, \hat{\mu})|.
\]

From Equation (C.6), we have that
\[
|\Phi^n_i(t_i, t_{-i}) - \phi^n(t_i, \hat{\mu})| =
|\Phi^n_i(t_i, t_{-i}) - \sum_{t'_{-i}} \Pr(t'_{-i} \mid t'_{-i} \in T^{n-1}, t'_{-i} \sim \hat{\mu}) \cdot \Phi^n_i(t_i, t'_{-i})|,
\]
By the triangle inequality we have that
\[
|\Phi^n_i(t_i, t_{-i}) - \phi'(t_i, \hat{\mu})| \\
\leq \sum_{t'_{-i} \in T^{n'-1}, t'_{-i} \sim \hat{\mu}} \Pr(t'_{-i} \mid t_{-i}) \cdot |\Phi^n_i(t_i, t_{-i}) - \Phi_i^{n'}(t_i, t'_{-i})| \\
+ \sum_{t'_{-i} \in T^{n'-1}, t'_{-i} \sim \hat{\mu}} \Pr(t'_{-i} \mid t_{-i}) \cdot |\Phi^n_i(t_i, t_{-i}) - \Phi_i^{n'}(t_i, t'_{-i})|.
\]

Plugging in Inequalities (C.7) and (C.8) we have
\[
|\Phi^n_i(t_i, t_{-i}) - \phi'(t_i, \hat{\mu})| < \epsilon / 2 + \epsilon / 2 = \epsilon.
\]

Moreover, note that the above bounds in Inequalities (C.7) and (C.8) may be taken uniform in \(t_i, t_{-i}\). Therefore the overall bound is uniform. This completes this step. \(\square\)

**Step 2.** Given \(\epsilon > 0\), there exists \(n_0\) such that for all \(n \geq n_0, t_i, t_{-i}\) we have
\[
|\Phi^n_i(t_i, t_{-i}) - \phi^\infty(t_i, \text{emp } t_{-i})| < \epsilon. \tag{C.9}
\]

*Proof.* By Step 1, we may take \(n_0\) such that for all \(n, n' \geq n_0, t_i, t_{-i}\) we have
\[
|\Phi^n_i(t_i, t_{-i}) - \phi^\infty(t_i, \text{emp } t_{-i})| < \epsilon / 2.
\]

Taking the limit as \(n' \to \infty\) we have
\[
|\Phi^n_i(t_i, t_{-i}) - \phi^\infty(t_i, \text{emp } t_{-i})| \leq \epsilon / 2 < \epsilon.
\]

\(\square\)

**Step 3.** Use Step 2 to complete the proof of the Proposition.

*Proof.* By Step 2, we may take \(n_0\) large enough such that for all \(n \geq n_0, t:\)
\[
|\Phi^n_i(t_i, t_{-i}) - \phi^\infty(t_i, \text{emp } t_{-i})| < \epsilon / 2. \tag{C.10}
\]

Consider now the gain for agent \(i\) to, given \(t_{-i}\), perform an ex post deviation to \(\hat{t}_i\). We
have

\[ u_{t_i}[\Phi^n_i(\hat{t}_i, t_{-i})] - u_{t_i}[\Phi^n_i(t)] \leq |u_{t_i}[\Phi^n_i(\hat{t}_i, t_{-i})] - u_{t_i}[\phi^\infty(\hat{t}_i, \text{emp } t_{-i})]| + |u_{t_i}[\Phi^n_i(t_{-i})] - u_{t_i}[\phi^\infty(t_{-i})]| + (u_{t_i}[\phi^\infty(\hat{t}_i, \text{emp } t_{-i})] - u_{t_i}[\phi^\infty(t_{-i})]). \]

By the boundedness of \( u \) we have

\[ u_{t_i}[\Phi^n_i(\hat{t}_i, t_{-i})] - u_{t_i}[\Phi^n_i(t)] \leq |\Phi^n_i(\hat{t}_i, t_{-i}) - \phi^\infty(\hat{t}_i, \text{emp } t_{-i})| + |\phi^\infty(t_{-i}) - \Phi^n_i(t_{-i})| + (u_{t_i}[\phi^\infty(\hat{t}_i, \text{emp } t_{-i})] - u_{t_i}[\phi^\infty(t_{-i})]). \]

Plugging in Inequality (C.10), we have

\[ u_{t_i}[\Phi^n_i(\hat{t}_i, t_{-i})] - u_{t_i}[\Phi^n_i(t)] < \epsilon/2 + \epsilon/2 + (u_{t_i}[\phi^\infty(\hat{t}_i, \text{emp } t_{-i})] - u_{t_i}[\phi^\infty(t_{-i})]). \]

Since the original mechanism is SP-L, we have that the last term is nonpositive. Therefore,

\[ u_{t_i}[\Phi^n_i(\hat{t}_i, t_{-i})] - u_{t_i}[\Phi^n_i(t)] < \epsilon. \]

This completes the proof. \( \square \)

C.6 An Example without Quasi-Continuity

This Section gives an example of a sequence of families of BNE that is not quasi-continuous. Moreover, the construction used in the proof of Theorem C3 does not produce an SP-L mechanism. 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 1 and C3 does not work.

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 are fined 10 units. The set of is types \( T = O = \{o_1, o_2\} \), with an agent’s type denoting her favorite object. Utility is given by

\[ u_{t_i}(x_0) = 1\{x_0(1) = t_i\} + x_0(2). \]
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 \( a_i(1) \), 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 \( a_j(2) = nf, j = 1, \ldots, n \), then \( \Phi^n_i(a) = (a_i(1), 0) \). That is, if all agents choose the no fine option, then each agent receives her favorite object and no one is fined.

- If some \( a_j(2) = f, j = 1, \ldots, n \), 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:
    \[ \Phi^n_i(a) = (a_i(1), -10 \cdot 1\{a_i(1) = o_1\}). \]
  - If \( \#\{j : a_j(1) = o_1\} \) is even, then agents asking for object \( o_2 \) are fined:
    \[ \Phi^n_i(a) = (a_i(1), -10 \cdot 1\{a_i(1) = o_2\}). \]

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 \( \mu_0 \) be the distribution putting equal weight on \( o_1 \) and \( o_2 \). Define now the following subset of \( N \times \bar{\Delta}T \),

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

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 \( \mu \), the number of players with \( t_i = o_1 \) is odd. Moreover, \( n \) has to be larger than \( n_0 \) and \( \mu \) sufficiently close to \( \mu_0 \).

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

- If \( (n, \mu) \in S \), then \( \sigma^n_\mu \) 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 \( \delta \), this is a family of limit 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 sequence of families of BNE is not quasi-continuous at \( \mu_0 \). For \( \mu \) and \( \text{emp} t \) in any small neighborhood of \( \mu_0 \), the allocation \( \Phi^n_{\sigma} (t_i, \sigma^n_{\mu} (t_{-i})) \) varies discontinuously with \( t_{-i} \) and \( \mu \). To see this formally, take a neighborhood \( \mathcal{N} \) of \( \mu_0 \) small enough such that the set

\[
\{ \mu \in \mathcal{N} : \exists n \text{ with } (n, \mu) \in S \}
\]

is relatively dense with respect to \( \mathcal{N} \). Any open subset \( \mathcal{A}_k \) of this neighborhood therefore contains infinite points in this set, and infinite points outside this set. In particular, there exist \( n, \mu \) and \( \text{emp} [t] \) in \( \mathcal{A}_k \), where type \( o_1 \) agents are not fined, \( \Phi^n_{\sigma} (o_1, \sigma^n_{\mu} (t_{-i})) = (o_1, 0) \) and \( n, \mu \) and \( \text{emp} [t] \) where they are fined, \( \Phi^n_{\sigma} (o_1, \sigma^n_{\mu} (t_{-i})) = (o_1, -10) \). Therefore, the conditions for quasi-continuity are not satisfied.

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

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

This is the construction used in the proof of our main theorems. We now show that this mechanism is neither SP-L, nor does it approximate outcomes of the \( \sigma^n_{\mu} \) 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 \( \sigma^n_{\mu} \) 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. Note that, if types are distributed according to \( \mu_0 \), 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

\[
f^{\infty}_i(t_i, \mu_0) = \begin{cases} 
\frac{1}{2}(o_1, 0) + \frac{1}{2}(o_1, -10) & \text{if } t_i = o_1 \\
(o_2, 0) & \text{if } t_i = o_2.
\end{cases}
\]

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, as would be the case if the sequence of families of equilibria were quasi-continuous, by Theorem C3.

D A Quasi-Continuous Family of Equilibria of Multi-Unit Auctions

This Section derives a family of limit equilibria of the pay-as-bid auctions in Example 1, and shows that this family is quasi-continuous, despite not being continuous everywhere.

For a given prior \(\mu \in \bar{\Delta}T\) we define the equilibrium \(\sigma_\mu\) as follows. We begin with the case where \(\rho^*(\mu) > 1\). Bids \(\sigma_\mu(t_i)\) are

- \((\rho^*(\mu), q_i)\) for agents with \(v_i > \rho^*(\mu)\).
- \((\rho^*(\mu) - 1, q_i)\) for agents with \(v_i < \rho^*(\mu)\).
- Agents with \(v_i = \rho^*(\mu)\) mix so that on average the market clears exactly at prices \(\rho^*(\mu)\). That is, they play \((\rho^*(\mu), q_i)\) with probability \(\pi\) and \((\rho^*(\mu) - 1, q_i)\) with probability \(1 - \pi\), where

\[
\pi = \frac{D(\rho^*(\mu); \mu) - k}{D(\rho^*(\mu); \mu) - D(\rho^*(\mu) - 1; \mu)}.
\]

If agents follow these strategies, then the realized market clearing price will be \(\rho^*(\mu)\) or \(\rho^*(\mu) - 1\) with equal probability. Moreover, bids of \(\rho^*(\mu)\) are accepted with probability 1 in the limit, whereas bids of \(\rho^*(\mu) - 1\) are rejected with probability close to 1 in the limit. That is, since agents with valuation of exactly \(\rho^*(\mu)\) are randomizing, in the limit there is no rationing.

It follows that the strategy \(\sigma_\mu\) is a limit equilibrium. The bid necessary to purchase the goods is \(\rho^*(\mu)\), since any lower bid is rationed with very high probability. Therefore, all agents with valuations higher than \(\rho^*(\mu)\) will bid this value, and purchase the goods for sure.
Agents with lower valuations are happy to bid lower, and not purchase the good. Agents with valuations of exactly $\rho^*(\mu)$ are happy to randomize, as they are indifferent between buying the goods at these prices or not.

In the case where $\rho^*(\mu) = 1$, the minimum price, we define equilibrium strategies similarly. In this case bids $\sigma_\mu(t_i)$ are defined as

- $(1, q_i)$ for agents with $v_i > 1$.
- Agents with $v_i = \rho^*(\mu)$ mix so that on average the market clears exactly at prices $\rho^*(\mu)$. That is, they play $(1, q_i)$ with probability $\pi$ and $(1, 0)$ with probability $1 - \pi$, where $\pi$ is given as in (D.1).

An argument analogous to the $\rho^*(\mu) > 1$ case shows that this consists an equilibrium.

We now show that the family of equilibria described above is quasi-continuous. For simplicity, we focus on priors $\mu_0 \in \tilde{\Delta}T$ such that $\rho^*(\mu_0) > 2$. We consider the most complicated case, when $\rho^*(\mu_0)$ clears the market exactly. In this case, the equilibrium strategies can vary a lot with small perturbations of beliefs. For some priors $\mu$ close to $\mu_0$ we have $\rho^*(\mu) = \rho^*(\mu_0)$, while for some other priors $\rho^*(\mu) = \rho^*(\mu_0) - 1$.

Therefore, outcomes may change discontinuously, and the reason why we have quasi-continuity is that we can decompose any neighborhood of $\mu_0$ in components where outcomes do vary continuously. Take $\epsilon > 0$. Consider, for a small number $\delta > 0$, the sets

\[ A_1 = \{ \mu : \rho^*(\mu) = \rho(\mu_0), E[D(\rho^*(\mu_0); t_i)|t_i \sim \mu] > k, |\mu - \mu_0| < \delta \} \]
\[ A_2 = \{ \mu : \rho^*(\mu) = \rho(\mu_0) - 1, |\mu - \mu_0| < \delta \}. \]

We take $\delta$ to be small enough such that the market does not clear exactly in expectation at any prior $\mu$ in $A_2$. Moreover, $\delta$ is taken to be small enough, so that the ball $N'$ of radius $\delta$ around $\mu_0$ can be decomposed into $A_1, A_2$, and the set of priors where average demand clears exactly at $\rho^*(\mu_0)$, defined as

\[ B = \{ \mu : \rho^*(\mu) = \rho(\mu_0), E[D(\rho^*(\mu_0); t_i)|t_i \sim \mu] = k, |\mu - \mu_0| < \delta \}. \]

Note that the probability that a randomly drawn vector of types $t$ has an empirical distribution close to $B$ is converging to 0, as required by condition 1 of the definition of quasi-continuity. To see this, note that for the distance of $\text{emp}[t]$ to $B$ to be less than $1/n$ it is necessary that

\[ |D(\rho^*(\mu_0); t) - nk| < \bar{q}. \]
However, the random variable $D(\rho^*(\mu_0; t))$ has standard deviation proportional to $\sqrt{n}$. Therefore, this probability converges to 0.

**Showing that condition 2 is satisfied.**

It only remains to show that condition 2 is satisfied. That is, that outcomes do vary continuously for priors and vectors of types with empirical distributions within each $A_k$. Consider $A_1$ for simplicity. We now fix $\epsilon > 0$, and will show that $\delta > 0$ and an $n_0$ can be taken so that condition 2 holds. To simplify the proof, we fix a number $\epsilon' > 0$, to be determined below as a function of $\epsilon$. We will begin by showing that the probability that a bid of $\rho^*(\mu_0)$ receives $q_i$ units is close to 1, while the probability that a bid of $\rho^*(\mu_0) - 1$ receives 0 units is close to 1.

Take $\mu$ and $t$ with $\mu, \text{emp}[t] \in A_1$. Given the vector of types $t$ and strategies $\sigma_\mu$, the outcomes of the mechanism for agent $i$ are

$$\Phi_i^n(\sigma_\mu(t_i), \sigma_\mu(t_{-i})) = \sum_a \Pr\{a | t, \sigma_\mu\} \cdot \Phi_i^n(a).$$

That is, the distribution of outcomes when players have types given by the vector $t$, and strategies distributed as $\sigma_\mu$. Since $\mu$ is in $A_1$, the only bids are $\rho^*(\mu_0)$ and $\rho^*(\mu_0) - 1$. Therefore, to determine the aggregate outcome, we only have to count the number of bids equal to $\rho^*(\mu_0)$. This number $N(t, \mu)$ is given by the random variable

$$N(t, \mu) = \sum_{i: v_i > \rho^*(\mu_0)} q_i + \sum_{i: v_i = \rho^*(\mu_0)} Z_i \cdot q_i,$$

where $Z_i$ is a random variable equal to 0 if player $i$ bids $\rho^*(\mu_0) - 1$ and 1 if player $i$ bids $\rho^*(\mu_0)$. By the description of the equilibrium $\sigma_\mu$, the random variables $Z_i$ are independently identically distributed, with mean equal to the probability $\pi(\mu)$ of bidding $\rho^*(\mu_0)$. The probability $\pi(\mu)$ is given by equation (D.1).

**Showing that $N(t, \mu)$ is close to $kn$ with high probability.**

Note that, for large $n$, $N(t, \mu)$ is approximately equal to $kn$ with high probability. To
establish this formally, we rewrite $N(t, \mu)$ as

$$N(t, \mu) = \sum_{i: v_i > \rho^*(\mu_0)} q_i + \sum_{i: v_i = \rho^*(\mu_0)} \pi(\mu) \cdot q_i + \sum_{i: v_i = \rho^*(\mu_0)} (Z_i - \pi(\mu)) \cdot q_i$$

$$= D(\rho^*(\mu_0) + 1; t) + \pi(\mu) \cdot [D(\rho^*(\mu_0); t) - D(\rho^*(\mu_0) + 1; t)] + \sum_{i: v_i = \rho^*(\mu_0)} (Z_i - \pi(\mu)) \cdot q_i$$

$$= D(\rho^*(\mu_0); t) - [1 - \pi(\mu)] \cdot [D(\rho^*(\mu_0); t) - D(\rho^*(\mu_0) + 1; t)] + \sum_{i: v_i = \rho^*(\mu_0)} (Z_i - \pi(\mu)) \cdot q_i.$$}

The first line follows from adding and subtracting the $\pi(\mu)$ terms. The second line follows from substituting the formulas for $D(\cdot)$. The third line is algebra. From the definition of $A_1$ we can take $\delta$ to be small enough such that

$$|D(\rho^*(\mu_0); t) - kn| < \epsilon'n/2$$

$$|1 - \pi(\mu)| < \epsilon'/2\bar{q}, \quad (D.2)$$

for any $t, \mu$ in $A_1$. Therefore, the difference between the number of bids $\rho^*(\mu_0)$ and the supply of the good is bounded by

$$|N(t, \mu) - kn| < \epsilon'n/2 + \epsilon'n/2 + |\sum_{i: v_i = \rho^*(\mu_0)} (Z_i - \pi(\mu)) \cdot q_i|.$$}

Note that the sum in the right term is composed of multiples of iid random variables with zero mean. Moreover, for sufficiently small $\delta$, the number of elements in the sum is bounded between two fixed positive fractions of $n$, due to $\text{emp}[t] \in A_1$. Therefore, by the law of large numbers, there exists $n_0$ such that for all $n \geq n_0$

$$\Pr\left\{\frac{|N(t, \mu) - kn|}{n} < 2\epsilon' \right\} \geq 1 - \epsilon'.$$ 

$$\quad (D.3)$$

**Showing that a bid of $\rho^*(\mu_0)$ has a high probability of being fulfilled.**

We can now show that the probability that a bid of $\rho^*(\mu_0)$ is fulfilled is close to 1. Bids of $\rho^*(\mu_0)$ are only rationed if $kn < N(t, \mu)$. Moreover, in this case the fraction of bids that
are fulfilled equals
\[
\frac{kn}{N(t, \mu)} = \frac{kn}{kn + (N(t, \mu) - kn)} = \frac{1}{1 + \frac{N(t, \mu) - kn}{kn}} \geq 1 - \frac{|N(t, \mu) - kn|}{kn}.
\]

Using the bound (D.3), the probability that a bid of \( \rho^*(\mu_0) \) is fulfilled is greater than
\[
(1 - 2\epsilon' / k)(1 - \epsilon') \geq 1 - 3\epsilon'. \tag{D.4}
\]
This implies that the probability that an agent bidding \( \rho^*(\mu_0) \) receives \( q_i \) units, conditional on \( t \) and \( \sigma_{\mu_i} \), is at least \( 1 - 3\bar{q}\epsilon' \).

**Showing that a bid of \( \rho^*(\mu_0) - 1 \) has a low probability of being fulfilled.**

Likewise, we can show that the probability that a bid of \( \rho^*(\mu_0) - 1 \) is fulfilled is close to 0. Conditional on \( N(t, \mu) \), the probability that a \( \rho^*(\mu_0) - 1 \) bid is fulfilled is 0 unless the number of \( \rho^*(\mu_0) \) bids is smaller than the supply of objects. That is, unless \( N(t, \mu) < kn \). Whenever this is the case, the probability of a \( \rho^*(\mu_0) - 1 \) bid being fulfilled equals the number of remaining objects, after all bids of \( \rho^*(\mu_0) \) are satisfied, divided by the number of remaining bids. That is,
\[
\frac{kn - N(t, \mu)}{D(1; t) - N(t, \mu)}. \tag{D.5}
\]
Using again the law of large numbers, we take \( n_0 \) large enough such that (D.3) is satisfied. Moreover, we can take a constant \( C > 0 \) independent of \( \mu, t \) and \( \epsilon' \), and \( n_0 \) such that, for all \( n \geq n_0 \),
\[
\Pr\{D(1, t) - N(t, \mu) \geq Cn\} \geq 1 - \epsilon'
\]
Take now \( n \geq n_0 \). With probability of at least \( 1 - 2\epsilon' \), we have that both
\[
|N(t, \mu) - kn| / n < 2\epsilon' \quad \text{and} \quad D(1, t) - N(t, \mu) \geq Cn.
\]
Whenever both these bounds are satisfied, the probability of a bid of \( \rho^*(\mu_0) - 1 \) being satisfied given by (D.5) is at most \( 2\epsilon' / C \). Therefore, the probability that a bid of \( \rho^*(\mu_0) - 1 \) is satisfied
conditioning only on $t$ and $\mu$ is bounded above by

$$2\epsilon' + \frac{2\epsilon'}{C} = 2(1 + \frac{1}{C})\epsilon'.$$

Therefore, the probability that an agent bidding $\rho^*(\mu_0) - 1$ receives no objects is greater than

$$1 - 2\bar{q}(1 + \frac{1}{C})\epsilon'.$$  \hfill(D.6)

**Completing the proof.**

To complete the proof, we show that $\Phi_n^n(\sigma_\mu(t))$ does not depend too much on $t_{-i}$ and $\mu$ as long as $\mu$ and $\text{emp}[t]$ are in $A_1$. Indeed, if $v_i < \rho^*(\mu_0)$, then agent $i$ bids $\rho^*(\mu_0) - 1$ for sure. Therefore, the probability that she receives the null bundle is bounded below by the expression in (D.6). Consider now an agent with $v_i \geq \rho^*(\mu_0)$. The probability that she bids $\rho^*(\mu_0)$ is bounded below by $1 - \epsilon'$ by equation (D.2), and if she does so she receives $q_i$ objects is bounded below by $1 - 3\bar{q}\epsilon'$. Therefore, the overall probability of receiving $q_i$ units is bounded below by $1 - \epsilon' - 3\bar{q}\epsilon'$. Therefore, we take $\epsilon'$ such that

$$\max\{2\bar{q}(1 - \frac{1}{C})\epsilon', \epsilon' + 3\bar{q}\epsilon'\} = \epsilon.$$

This implies that, for all $n \geq n_0$, and all $\mu, \mu' \in A_1$, and $t_i, t_{-i}, t'_{-i}$ with $\text{emp}[t_i, t_{-i}], \text{emp}[t_i, t'_{-i}] \in A_1$, condition 2 in the definition of quasi-continuity is satisfied. The argument for $A_2$ is analogous.