Strategy-proofness in the Large[*]

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

July 16, 2018

Abstract

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

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

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

1   Introduction

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

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

---

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

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

advantages of SP design.

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

The combination of the large market and the interim perspective is powerful for the following reason: it causes each participant to regard the societal distribution of play as exogenous to his own report (more precisely, the distribution of the societal distribution of play; see Section 3.1.1 and Lemma A.1). Regarding the societal distribution of play as exogenous to one's own play can be interpreted as a generalization of regarding prices as exogenous, i.e., of price taking. In some settings, such as multi-unit auctions or Walrasian exchange, the two concepts are economically equivalent. In other settings, such as school choice or two-sided matching, regarding the societal distribution of play as exogenous is equivalent to regarding certain price-like summary statistics of the mechanism as exogenous.

SP-L thus draws a distinction between two ways a mechanism can fail to be SP. If a mechanism is manipulable by participants who can affect prices (or price-like summary statistics), but is not manipulable by participants who regard the societal distribution of play as exogenous, the mechanism is SP-L. If a mechanism is manipulable even by participants who regard the societal distribution of play as exogenous – if even a price taker, or a taker of price-like statistics, wishes to misreport – then the mechanism, in addition to not being SP, is not SP-L.

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

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

The last part of the paper identifies a precise sense in which, in large markets, SP-L is

---

[3]For instance, the double auction and uniform-price auctions are SP-L but not approximately strategy-proof in an ex post sense; even in a large economy it is always possible to construct a knife-edge situation where a single player, by shading her demand, can have a large discontinuous influence on the market-clearing price. For this reason, papers such as Rustichini et al. (1994) and Swinkels (2001) have studied approximate or exact Bayes-Nash equilibria as opposed to a dominant strategy notion.

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

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

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

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

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

numerous impossibility theorems for SP.[4]

We emphasize that the idea that market size can ease incentive problems is quite old, with some of the earliest contributions being Roberts and Postlewaite (1976) and Hammond (1979), and rich and active literatures in both implementation theory and market design since these early contributions. We discuss the relevant literatures in detail in the body of the paper. Our paper makes three substantive contributions relative to this literature. First, the criterion of SP-L itself is new. There are many other criteria of approximate incentive compatibility in the literature (see footnote 12). SP-L is perhaps closest to the two criteria it lands in between: the traditional notion of approximate SP (which evaluates incentives to manipulate from the ex post perspective) and the traditional notion of approximate Bayes-Nash (which evaluates incentives to manipulate from the interim perspective, like SP-L, but with respect only to the single probability distribution associated with equilibrium play). Second, our classification of mechanisms, into those that are SP-L and those that are manipulable even in large markets, is new. We expect few market design researchers will find the classification surprising ex post; for example, Kojima and Pathak (2009) noted that the Boston mechanism fails their approximate IC criterion (roughly, that truth-telling is approximate Bayes-Nash), and Friedman (1991) informally discusses the incentives difference between pay-as-bid and uniform-price auctions. Rather, our contribution is to produce a classification of mechanisms across all of the canonical market design problems using a single criterion and analytical approach, which is both useful per se and speaks to the applicability of the criterion and analytical methods. Third, the result that SP-L is approximately costless relative to Bayes-Nash is new, and is the first such result, to our knowledge, for any form of approximate strategy-proofness relative to Bayes-Nash; though there are some more restricted settings where SP itself is costless relative to Bayes-Nash (Manelli and Vincent, 2010; Gershkov et al., 2013).

The rest of the paper is organized as follows. Section 2 defines the environment. Section 3 defines and discusses SP-L. Section 4 presents the theorem with sufficient conditions for SP-L and the classification of non-SP mechanisms. Section 5 presents the theorem that imposing SP-L is approximately costless relative to Bayes-Nash. Section 6 discusses technical

---

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

extensions. Section 7 discusses related literature. Section 8 concludes. Proofs and other supporting materials are in the appendix.

2   Environment

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

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

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

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

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

We assume that mechanisms are **anonymous**. Anonymity requires that each agent's outcome depends only on her own action and the distribution of all actions, but not on her identity or the specific identities associated with specific opponent actions. Formally, a mechanism is anonymous if the allocation function $\Phi^n(\cdot)$ is invariant to permutations for all $n \in \mathbb{N}$. Anonymity is a natural feature of many large-market settings. In Supplementary

Appendix C we relax anonymity to **semi-anonymity** (Kalai, 2004). A mechanism is semi-anonymous if agents are divided into a finite set of groups, and an agent's outcome depends only on her own action, her group, and the distribution of actions within each group. Semi-anonymity accommodates applications in which there are asymmetries among classes of participants, such as double auctions in which there are distinct buyers and sellers and school choice problems in which students are grouped into different priority classes, as well as some models of matching markets.

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

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

3   Strategy-proof in the Large

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

3.1   The Large Market

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

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

where $\Phi_i^n(a_i, a_{-i})$ denotes the marginal distribution of the ith coordinate of $\Phi^n(a)$, i.e., the lottery over outcomes received by agent i when she plays a_i and the other $n-1$ agents play a_{-i}, and $\Pr(a_{-i} | a_{-i} \sim iid(m))$ denotes the probability that the action vector a_{-i} is realized

given $n-1$ independent and identically distributed (i.i.d.) draws from the action distribution $m \in \Delta A$. In words, $\phi^n(a_i, m)$ describes what an agent who plays a_i expects to receive, ex interim, if the other $n-1$ agents play i.i.d. according to action distribution m.

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

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

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

if this limit exists.

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

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

3.1.1 Interpretation of the Limit: Generalized Price Taking

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

We interpret treating the societal distribution of play as exogenous to one's own report as a generalized version of price taking. Suppose that a mechanism has prices that are a function of the empirical distribution of play. For example, in the uniform-price auction or Walrasian mechanism, price is determined based on where reported demand equals reported

---

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

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

3.2 Definition of SP-L

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

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

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

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

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

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

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

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

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

3.3   Discussion

3.3.1   Relationship to SP

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

3.3.2   Relationship to other Approximate IC criteria

SP-L lies in between the standard notion of asymptotic strategy-proofness and the standard notion of approximate Bayes-Nash incentive compatibility.

SP-L is weaker than the standard notion of asymptotic SP — which requires that reporting truthfully is approximately optimal, in a large enough market, for any *realization* of opponent reports[6] — because SP-L evaluates incentives to misreport ex interim and not ex post. This distinction is important for the results below. Theorems 1 and 2 are both false for the ex post notion of asymptotic SP. In the classification of mechanisms, nearly all of the mechanisms that are classified as SP-L would fail this stronger criterion, with the lone exception being the probabilistic serial mechanism.

At the same time, SP-L is stronger than approximate Bayes-Nash incentive compatibility. In its direct-revelation mechanism form — e.g., the direct-revelation mechanism version of the pay-as-bid auction or Boston mechanism — Bayes-Nash incentive compatibility requires that the distribution of types is common knowledge among both participants and

---

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

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

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

the mechanism designer, and that reporting truthfully is approximately a best response given this single common-knowledge distribution. In its indirect mechanism form — e.g., the versions of the pay-as-bid auction or Boston mechanism used in practice — Bayes-Nash incentive compatibility requires that the distribution of types is common knowledge among participants, and that participants can make the potentially complex strategic calculations required by equilibrium as a function of the distribution (e.g., how much to shade one's bid, or whether to take a risk for a popular school). In contrast, SP-L does not impose any common knowledge assumptions on either participants or the mechanism designer, and is strategically simple in that participants approximately optimize by reporting their type truthfully. Our Theorem 2 shows that this strengthening relative to Bayes-Nash is, in a precise sense, approximately costless in large markets, and our classification exercise in Section 4 suggests that this strengthening yields a compelling classification of non-SP mechanisms.

3.3.3 Clarifying Example: Multi-Unit Auctions

Appendix B provides a detailed example, multi-unit auctions, to illustrate several of the key definitions of this section. The two most common multi-unit auction formats are uniform-price auctions and pay-as-bid auctions. While neither mechanism is SP (Ausubel and Cramton, 2002), Milton Friedman famously argued in favor of the uniform-price auction on incentives grounds (Friedman, 1960, 1991). The example shows that the uniform-price auction is SP-L whereas the pay-as-bid auction is manipulable in the large. The example also illustrates the large-market limit, the role of the full-support requirement, and the contrast between SP-L and the traditional notion of approximate strategy-proofness based on ex post realizations of others' play rather than interim distributions of others' play. Notably, while the uniform-price auction is SP-L, it fails to satisfy the traditional ex post notion of approximate strategy-proofness.

4 Classification of Non-SP Mechanisms

This section classifies a number of non-SP mechanisms into SP-L and manipulable in the large (Table 1 in the Introduction), and discusses how the classification organizes the evidence on manipulability in large markets. Specifically, all of the known mechanisms for which there is a detailed theoretical case that the mechanism has approximate incentives for truthtelling in large markets are SP-L (Section 4.2), and all of the known mechanisms for which there is empirical evidence that non-strategy-proofness causes serious problems even in large markets

are manipulable in the large (Section 4.3). In particular, the classification of mechanisms based on whether or not they are SP-L predicts whether misreporting is a serious problem in practice better than the classification of mechanisms based on whether or not they are SP. These results suggest that, in large markets, SP-L versus not SP-L is a more relevant dividing line than SP versus not SP.

Before proceeding, we make three brief observations regarding the classification. First, both the SP-L and the manipulable in the large columns of Table 1 include mechanisms that explicitly use prices (e.g., multi-unit auctions), as well as mechanisms that do not use prices (e.g., matching mechanisms). For the mechanisms that do use prices, the SP-L ones are exactly those where an agent who takes prices as given wishes to report truthfully, such as the uniform-price auction. Second, the table is consistent with both Milton Friedman's (1960; 1991) argument in favor of uniform-price auctions over pay-as-bid auctions, and Alvin Roth's (1990; 1991; 2002) argument in favor of deferred acceptance over priority-match algorithms. Notably, while both Friedman's criticism of pay-as-bid auctions and Roth's criticism of priority-match algorithms were made on incentives grounds, the mechanisms they suggested in their place are not SP but are SP-L. Third, with the exception of probabilistic serial, none of the SP-L mechanisms satisfy the stronger, ex post, notion of approximate strategy-proofness that has been used in previous literature (cf. footnote 6 for the formal definition). That is, the classification would not conform to the existing evidence, nor to Friedman's and Roth's arguments, without the ex interim perspective in the definition of SP-L.

4.1 Obtaining the Classification

To show that a mechanism is not SP-L it suffices to identify an example of a distribution of play under which agents may gain by misreporting, even in the limit. For SP-L mechanisms, this section gives two easy-to-check sufficient conditions for a mechanism to be SP-L, which directly yield the classification for all of the SP-L mechanisms in Table 1. Formal definitions of each mechanism and detailed derivations are in Supplementary Appendix D.[7]

The first sufficient condition is envy-freeness, a fairness criterion which requires that no

---

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

player i prefers the assignment of another player j, for any realization of the reported payoff types t.

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

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

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

The mathematical intuition for why EF implies SP-L is as follows. In anonymous mechanisms, the gain to player i from misreporting as player j can be decomposed as the sum of the gain from receiving j's bundle, holding fixed the aggregate distribution of reports, plus the gain from affecting the aggregate distribution of reports (expression (A.2) in Appendix A). Envy-freeness directly implies that the first component in this decomposition is non-positive. Lemma A.1 then shows that the second component becomes negligible in large markets. More precisely, even though there may exist realizations of the other players' reports where player i's gain from affecting the aggregate distribution is large (e.g., if by misreporting he affects the clearing price in the uniform-price auction), his effect on the *interim distribution of the empirical distribution* of reports vanishes with market size, at a rate of essentially $\sqrt{n}$. This yields both that EF implies SP-L and the convergence rate for EF mechanisms as stated in Theorem 1.[8]

Most of the mechanisms in the SP-L column of Table 1 are EF, with the only exceptions being approximate CEEI and deferred acceptance.[9] To classify these mechanisms, we introduce a weakening of EF that we call envy-free but for tie breaking (EF-TB). A mechanism is envy-free but for tie breaking if, after reports are submitted, the mechanism runs a tie-breaking lottery, and allocations depend on reports and on the lottery. After the

---

[8]The assumption of a finite number of types is important for the proof of Theorem 1. With a finite number of types and full support, the number of participants reporting a type t_j is random from the perspective of a participant who knows her own report and believes that the other reports are i.i.d. Moreover, the participant cannot have a large effect on the distribution of this number as the market grows large. However, with an infinite set of types and an atomless distribution, the expected number of other participants reporting exactly t_j will be zero with probability one even in arbitrarily large finite markets, so a single participant can have an economically meaningful impact on the number of reports of t_j. An interesting direction of future research is to consider infinite sets of types, but where there is a topological notion of types that are close to one another, and results like Theorem 1 can be established.

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

lottery is realized, no participant envies another participant with a worse lottery number. The simplest example is the random serial dictatorship mechanism for allocating objects without using money. Random serial dictatorship orders participants according to a lottery, and participants then take turns picking their favorite object out of the objects that are still available. This mechanism has envy ex post, because a participant may prefer the allocation of another participant who got a better lottery number. Bogomolnaia and Moulin (2001) have shown that the mechanism can also have envy before the lottery is drawn. However, after the lottery is drawn, no participant envies another participant with a worse lottery number, which means that this mechanism is envy-free but for tie breaking. Formally, the definition is as follows.[10]

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

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

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

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

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

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

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

is bounded above by

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

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

---

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

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

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

4.2 Relationship to the Theoretical Literature on Large Markets

The SP-L column of Table 1 organizes a large literature demonstrating the approximate incentive compatibility of specific mechanisms in large markets. Our results show that a number of mechanisms for which the literature established approximate incentive compatibility results are SP-L. This includes Walrasian mechanisms (Roberts and Postlewaite, 1976, Hammond, 1979 and Jackson and Manelli, 1997), double auctions (Rustichini et al., 1994 and Cripps and Swinkels, 2006), uniform-price auctions (Swinkels, 2001), deferred acceptance mechanisms (Immorlica and Mahdian, 2005 and Kojima and Pathak, 2009),[11] and the probabilistic serial mechanism (Kojima and Manea, 2010). We also obtain new results on the Hylland and Zeckhauser (1979) pseudomarket mechanism, approximate CEEI (Budish, 2011), and the generalized Hylland-Zeckhauser mechanism (Budish et al., 2013), each of whose large-market incentive properties had not previously been formally studied.

The single concept of SP-L and Theorem 1 classifies all of these mechanisms. In contrast, the prior literature has employed different notions of approximate incentive compatibility and different analysis techniques, often with common knowledge assumptions, tailored for

---

[11]Our assumption of a finite number of types is a limitation in certain matching models. As the market grows, we can only have a finite number of preference orderings over partners. This contrasts with some other large-market matching models, such as Kojima and Pathak (2009), Kojima et al. (2013), and Ashlagi et al. (2014). An interesting question for future research (see also fn. 8) is whether our analysis can be generalized to accommodate infinite sets of types with a topological notion of types that are close to one another, as in the large-market matching model of Lee (2017).

each mechanism.[12] Of course, analyses that are tailored to specific mechanisms can yield a more nuanced understanding of the exact forces pushing players away from truthful behavior in finite markets, as in the first-order condition analysis of Rustichini et al. (1994) or the rejection chain analysis of Kojima and Pathak (2009).

4.3 Relationship to Empirical Literature on Manipulability

For each of the manipulable in the large mechanisms in Table 1, there is explicit empirical evidence that participants strategically misreport their preferences in practice. Furthermore, misreporting harms design objectives such as efficiency or fairness. In this section we briefly review this evidence.[13]

Consider first multi-unit auctions for government securities. Empirical analyses have found considerable bid shading in discriminatory auctions (Hortaçsu and McAdams, 2010), but negligible bid shading in uniform-price auctions, even with as few as 13 bidders (Kastl, 2011; Hortaçsu et al. (2015)). Friedman (1991) argued that the need to play strategically in pay-as-bid auctions reduces entry of less sophisticated bidders, giving dealers a sheltered market that facilitates collusion. In uniform-price auctions, by contrast, "You do not have to be a specialist" to participate, since all bidders pay the market-clearing price. Consistent with Friedman's view, Jegadeesh (1993) shows that pay-as-bid auctions depressed revenues to the US Treasury during the Salomon Squeeze in 1991, and Malvey and Archibald (1998) 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 Bren-

---

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

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

ner et al. (2009) find a positive relationship between a country's using uniform-price auctions and indices of ease of doing business and economic freedom, whereas pay-as-bid auctions are positively related with indices of corruption and of bank-sector concentration.

Next, consider the Boston mechanism for school choice. Abdulkadiroğlu et al. (2006) find evidence of a mix of both sophisticated strategic misreporting and unsophisticated naive truthtelling; see also recent empirical work by Agarwal and Somaini (2014), Hwang (2014), De Haan et al. (2015) and Kapor et al. (2017). Sophisticated parents strategically misreport their preferences by ranking a relatively unpopular school high on their submitted preference list. Unsophisticated parents, on the other hand, frequently play a dominated strategy in which they waste the highest positions on their rank-ordered list on popular schools that are unattainable for them. In extreme cases, participants who play a dominated strategy end up not receiving any of the schools they ask for.

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

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

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

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

5.1 Social Choice Functions and Type Spaces

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

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

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

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

Note, importantly, that we allow social choice functions to depend on both preferences and beliefs, not only on preferences. The reason is that the outcomes of many commonly used mechanisms vary with both preferences and beliefs. For example, outcomes of the pay-as-bid auction or the Boston mechanism depend on what participants believe about other

participants' preferences and information; holding an agent's preferences fixed, their beliefs affect how much they will shade their bid, or whether they will take a risk and rank a popular school first. To discuss the social choice functions implemented by these mechanisms, we need to include beliefs in the definition of a social choice function. Our definition differs from that in Bergemann and Morris (2005), where social choice functions only depend on payoff types. Our definition is similar to that in the literature on implementation with incomplete information (as in Postlewaite and Schmeidler, 1986 or Jackson, 1991), where social choice functions may depend on payoff-irrelevant information.[14] We say that a social choice function **depends only on payoff types** if, for all n, and all ω and ω' in $\Omega_{n,i}$ such that $\hat{t}(\omega_i) = \hat{t}(\omega'_i)$ for all i, we have $F^n(\omega) = F^n(\omega')$.

Much applied work in mechanism design considers Bayesian equilibria where agents have a common, i.i.d. prior μ about payoff types, and know their own type.[15] To describe the outcomes of such equilibria for a range of priors μ, we need a type space that includes the union of these simple type spaces for a range of values of μ.

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

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

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

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

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

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

---

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

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

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

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

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

5.2   Limit Bayes-Nash and SP-L Implementability

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

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

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

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

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

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

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

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

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

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

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

The theorem also requires the following regularity condition.

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

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

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

That is, a social choice function defined over Ω^* is continuous if, in large enough markets, social outcomes vary continuously with beliefs. In the working paper version of this article we established a version of Theorem 2 that used a weaker condition, called quasi-continuity. See Section 6.1 for further discussion.

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

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

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

5.3 Construction Theorem

We now state the main result of this section.

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

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

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

with a limit Bayes-Nash equilibrium σ^* such that

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

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

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

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

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

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

5.4   Discussion of Theorem 2

5.4.1   Relationship to the Revelation Principle

The construction used in the proof of Theorem 2 is related to the traditional Bayes-Nash direct revelation mechanism construction (Myerson, 1979). In a traditional Bayes-Nash direct revelation mechanism, the mechanism designer and participants have a common knowledge

prior about payoff types, say μ_0. The mechanism announces a Bayes-Nash equilibrium strategy $\sigma^*(\cdot, \mu_0)$, and plays $\sigma^*(t_i, \mu_0)$ on behalf of an agent who reports t_i. Truthful reporting is a Bayes-Nash equilibrium.

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

5.4.2   Relationship to other Equivalence Results for Bayes-Nash and SP

Theorem 2 can be understood in relation to recent papers by Manelli and Vincent (2010) and Gershkov et al. (2013), which find striking *exact* equivalences between Bayes-Nash and dominant-strategy incentive compatibility in finite markets.

In Myerson's (1981) celebrated analysis of optimal auction design, the seller's maximal revenue can be achieved either in a Bayes-Nash mechanism (a first-price sealed bid auction) or in a dominant-strategy mechanism (a Vickrey auction). That is, there is no loss to the seller from moving from Bayes-Nash incentive compatibility to SP. Manelli and Vincent (2010) show that an equivalence between Bayes-Nash and SP obtains, in Myerson's environment, for the expected, ex interim allocations in *any* Bayes-Nash incentive compatible mechanism. Gershkov et al. (2013) extend the Manelli and Vincent (2010) result to a broader class of mechanism design environments, including auctions, public goods, bilateral trade, and screening, provided that, as in Myerson, types are one-dimensional and agents have quasi-linear utility.

Gershkov et al. (2013) then show, however, that their exact equivalence result no longer obtains once assumptions like one-dimensional types and quasi-linear utility are relaxed. These limitations are important when thinking about practical market design problems, because multi-dimensional types and non quasi-linear preferences are prevalent. For instance,

in all of the mechanisms listed in Table 1, agents' types are multi-dimensional, e.g., types are rank-order lists or demand functions. Our Theorem 2 recovers a form of *approximate* equivalence between Bayes-Nash and dominant-strategy incentive compatibility, for a richer class of environments without assumptions such as one-dimensional types and quasi-linear utility.[16]

5.4.3   The Debate on the Boston Mechanism

Theorem 2 contributes to a debate in the market design literature concerning the Boston mechanism for student assignment. (See Appendix E for a formal description of the Boston mechanism and the formal application of Theorem 2 to the mechanism).

The earliest papers on the Boston mechanism, Abdulkadiroğlu and Sönmez (2003) and Abdulkadiroğlu et al. (2006), criticized the mechanism on the grounds that it is not SP, and proposed that the strategy-proof Gale-Shapley deferred acceptance algorithm be used instead.[17] These papers had a major policy impact as they led to the Gale-Shapley algorithm's eventual adoption for use in practice (cf. Roth (2008)).

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

However, these second-generation papers rely on students being able to reach the attractive Bayes-Nash equilibria. This raises 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?

Theorem 2 shows that, in a large market, it is possible to obtain the attractive welfare properties of the Bayes-Nash equilibria identified by these second-generation papers on the Boston mechanism, but using a mechanism that is SP-L. The proof also shows how to

---

[16]While our environment does not impose one-dimensional types or quasi-linearity, it is not strictly speaking a generalization of the Manelli and Vincent (2010) and Gershkov et al. (2013) environments, because we require finite type and outcome spaces.

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

construct such a mechanism.[18]

6   Extensions and Discussion

6.1   Extensions

Semi-Anonymity   Our analysis focuses on mechanisms that are anonymous. In Supplementary Appendix C we generalize key definitions and results to the case of semi-anonymous mechanisms, as defined in Kalai (2004). A mechanism is semi-anonymous if each agent belongs to one of a finite number of groups, and her outcome is a symmetric function of her own action, her group, and the distribution of actions within each group. This generalization is useful for two reasons. First, it allows our analysis to cover more mechanisms. For instance, double auctions are semi-anonymous if buyers and sellers belong to distinct groups; two-sided matching markets are semi-anonymous under the assumption that the number of possible types of match partners is finite (cf. footnote 11); and school choice mechanisms are semi-anonymous if there are multiple priority classes (e.g., sibling priority). Second, it allows results and concepts stated for i.i.d. distributions to be extended to more general distributions.

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

In the working paper version of this article (Azevedo and Budish, 2016), we show that a weaker version of Theorem 2 obtains under a condition that we call quasi-continuity. Quasi-continuity allows for the desired social choice function to have discontinuities, with respect to both the prior and the empirical distribution of reports, but requires that the discontinuities

---

[18]To be clear, our results do not imply which is preferable between the SP-L mechanism constructed per Theorem 2 and the SP deferred acceptance mechanism. Our point is rather that *if* one believes that the analyses of Abdulkadiroğlu et al. (2011) and others imply that the Boston mechanism should be used in practice over deferred acceptance, then one should consider whether the SP-L implementation of these attractive equilibria is better still. Some recent empirical evidence (Agarwal and Somaini, 2014; Casalmiglia et al., 2014; Hwang, 2014) suggests that the magnitude of the welfare gains at stake, i.e., the difference in welfare between the Bayes-Nash equilibria of the Boston mechanism and the dominant strategy equilibria of deferred acceptance, may be small. If the gains are small, then the simpler deferred acceptance mechanism is likely more desirable than the SP-L implementation of the Boston mechanism's Bayes-Nash equilibria.

are in a certain sense knife-edge. Roughly, any discontinuity is surrounded by regions in which outcomes are continuous. Under this condition, the conclusion of the theorem is as follows. If the social choice function is continuous at a given prior μ_0, then, as before, there exists an SP-L mechanism that gives agents the same outcomes in the large-market limit. If the social choice function is not continuous at μ_0, then there exists an SP-L mechanism that gives agents a convex combination of the outcomes they would obtain under the desired social choice function, for a set of priors in an arbitrarily small neighborhood of μ_0.

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

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

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

We can formalize this question as follows. Consider a participant whose aggregate uncertainty about the distribution of others' play can be modeled by a finite set of states of the

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

It is straightforward to see that the answer to this question is affirmative. SP-L requires that reporting truthfully is approximately optimal, in a large enough market, for any full support i.i.d. distribution of play. This implies that reporting truthfully is approximately optimal, in a large enough market, for any finite mixture of full support i.i.d. distributions of play.

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

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

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

The key difference between voting and public goods problems, on the one hand, and the market design problems we have emphasized, is that in voting and public goods provision every agent gets the same outcome.[19] Technically, this can be accommodated in our frame-

---

[19]For expositional simplicity our discussion of public goods provision focuses on the case where agents' payments vary with the societal decision (e.g., whether or not to build a park, or whether to build an expensive

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

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

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

$$\lim \inf_{n \to \infty} u_{t_i}[\phi^n(t_i, m)] - u_{t_i}[\phi^n(t'_i, m)] > 0.$$

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

$$u_{t_i}[\phi^\infty(t_i, m)] > u_{t_i}[\phi^\infty(t'_i, m)].$$

In words, strict SP-L requires that truthful reporting is at least approximately as good as all other reports (as in Definition 4), and, in addition, is strictly preferred to at least some other report for at least some distribution. This is a modest additional requirement and it is easy to see that all of the SP-L mechanisms in Table 1 satisfy it. At the same time, voting

---

park, a cheap park, or no park) but not, conditional on the societal decision, with their specific report. For public goods mechanisms in which agents' payments do vary with their specific report (conditioning on the societal decision), and preferences are quasi-linear, the conclusions are as follows. Such mechanisms are SP-L if and only if, in the large market limit, all types make the same expected payment from the interim perspective, for any full-support i.i.d. distribution of others' reports. Such mechanisms are never strictly SP-L. If they satisfy the my play matters condition, it is because agents' reports can affect their payments in the large market limit, not the social decision; in such a case the mechanism is manipulable in the large. (To avoid potential confusion, note that in the Vickrey-Clarke-Groves mechanism, while agents' payments do vary with their reports in finite markets – e.g., if they are pivotal in the decision to provide the expensive park – in the large market limit all types' payments are zero because no agent is pivotal.)

and public goods mechanisms, in which all agents get the same outcome, are easily seen to fail this condition.[20]

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

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

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

In words, the my play matters condition requires that reporting truthfully affects one's utility as compared to at least some potential misreport and distribution. The following is immediate:

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

Remark 1 allows results for SP-L to be translated into results for strict SP-L by checking the my play matters condition.

7   Related Literature

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

Large Markets   Our paper is most closely related to the large theory literature that has studied how market size can ease incentive constraints for specific mechanisms. We discussed this literature in detail in Section 4.2. It is important to highlight that the aim of our

---

[20]Carroll (2013) introduces an interesting approach to studying approximate incentive compatibility for voting rules. His criterion, like SP-L, evaluates incentives to misreport from the interim perspective with respect to i.i.d. distributions of others' play. But, since in voting all mechanisms are trivially SP-L, he studies the *rate* at which incentives to misreport vanish as the market grows large. Also of interest are d'Aspremont and Peleg (1988) and Majumdar and Sen (2004), who study ordinal Bayesian incentive compatibility (OBIC). OBIC strengthens BIC by requiring that reporting one's ordinal preferences truthfully is optimal for any cardinal representation of an agent's true ordinal preferences, but at the same time OBIC is weaker than SP because it evaluates incentives to misreport with respect to a common-knowledge prior about the distribution of others' reports. While mathematically unrelated to SP-L, OBIC is in a similar spirit in that it identifies a compelling criterion between BIC and SP.

paper is different from, and complementary to, this literature. Whereas papers such as Roberts and Postlewaite (1976) provide a defense of a *specific pre-existing mechanism* based on its approximate incentives properties in large markets, our paper aims to justify SP-L as a *general desideratum* for market design. In particular, our paper can be seen as providing justification for focusing on SP-L when designing *new* mechanisms. Another point of difference versus this literature is that our criterion itself is new. See fn. 12 for details of the approximate incentives criteria used in this prior literature.

Implementation Theory Our paper is related to the implementation theory literature, both in goals and in specific ideas. The goal of implementation theory is to determine what social choice rules can be implemented by some mechanism, under different solution concepts, and to find applicable necessary and/or sufficient conditions for implementation (Maskin and Sjöström, 2002). Our paper proposes an incentive compatibility concept, SP-L. Theorem 1 gives applicable sufficient conditions for a mechanism to be SP-L. Theorem 2 shows that the set of social choice functions that can be implemented is not much more restrictive than under Bayes-Nash equilibria. More specifically, our paper is related to the literature on partial implementation, which considers whether a social choice rule is implemented by at least one equilibrium of a mechanism, as in Hurwicz (1972) and Bergemann and Morris (2005). This is in contrast to full implementation, which considers whether a social choice rule is implemented by every equilibrium of a mechanism.[21]

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

---

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

Strategy-proofness in Market Design Our paper is related to three strands of literature on the role of strategy-proofness in market design. First, there is an empirical literature that studies how participants behave in real-world non-SP market designs. We discussed this literature in detail in Section 4.3. This literature supports the SP-L concept, because all of the examples in which there is evidence of harm from misreporting involve mechanisms that not only are not SP, but are not even SP-L.

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

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

cording to their partial order. We view our approach as complementary to these alternative approaches. Two important advantages of our approach are that it yields the classification of non-SP mechanisms as displayed in Table 1, and yields an explicit second-best criterion for designing new mechanisms, namely that they be SP-L.

8   Conclusion

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

Whether SP-L can also be viewed as sufficient depends upon the extent to which the large-market abstraction is compelling in the problem of interest. Unfortunately, even with convergence rates such as those stated in Theorem 1, there rarely is a simple bright-line answer to the question of "how large is large".[22] But – just as economists in other fields instinctively understand that there are some contexts where it is necessary to explicitly model strategic interactions, and other contexts where it may be reasonable to assume price-taking behavior – we hope that market designers will pause to consider whether it is necessary to restrict attention to SP mechanisms, or whether SP-L may be sufficient for the problem at hand.

---

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

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Gershkov, Alex, Jacob Goeree, Alexey Kushnir, Benny Moldovanu, and Xianwen Shi, "On the Equivalence of Bayesian and Dominant Strategy Implementation," *Econometrica*, 2013, *81(1)*, 197–220.

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

Haan, Monique De, Pieter A Gautier, Hessel Oosterbeek, and Bas Van der Klaauw, "The performance of school assignment mechanisms in practice," 2015.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Kapor, Adam, Christopher A Neilson, Seth D Zimmerman et al., "Heterogeneous beliefs and school choice mechanisms," Technical Report 2017.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

___ , "Optimal Auction Design," *Mathematics of Operations Research*, 1981, *6* (1), 58–73.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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