Matching "versus" Mechanism Design

ERIC BUDISH

University of Chicago Booth School of Business

In a typical mechanism design paper the goal is to find a mechanism that maximizes the designer's objective (e.g., social welfare) subject to technology and incentive constraints. In a typical applied matching paper, by contrast, the goal is to find a mechanism that satisfies various "good properties" (e.g., efficiency and fairness). This essay discusses the relationship and tradeoffs between these two approaches. An introductory example, school choice, represents a simple win for the good properties approach. The main example, course allocation, paints a more complicated picture.

Categories and Subject Descriptors: J.4 [**Social and Behavioral Sciences**]: Economics

General Terms: Economics, Theory

---

1. INTRODUCTION

I was recently invited to give a conference talk with the suggested title of "Matching versus Mechanism Design." Not being the combative sort, I both (i) was puzzled by the proposed title, and (ii) accepted the invitation without protest. This essay reflects my attempt at reconciliation – of a conflict that I don't think exists in the first place – mostly in the context of my own work on assignment problems with multi-unit demand (e.g., course allocation).

In a typical mechanism design paper the goal is to find a mechanism that maximizes some objective function (e.g., social welfare or auctioneer revenues) subject to various constraints (e.g., technology constraints, incentive constraints). Canonical examples of mechanisms produced according to this recipe include Vickrey's [1961] welfare-maximizing auction and Myerson's [1981] revenue-maximizing auction. In a typical matching paper, by contrast, the goal is to find a mechanism that satisfies various "good properties" (e.g., Pareto efficiency, stability, envy-freeness, strategyproofness). Canonical examples include the Gale and Shapley [1962] deferred acceptance algorithm, which satisfies the property of stability, and Gale's Top Trading Cycles algorithm [Shapley and Scarf 1974], which satisfies the property of ex-post Pareto efficiency.

These two approaches are not always as different as they first appear. Vickrey's auction has an axiomatic formulation as the unique mechanism that is Pareto efficient and strategyproof [Green and Laffont 1977]. Gale and Shapley's deferred acceptance algorithm can be obtained as the solution to a certain constrained optimization problem – maximize proposer-side welfare subject to stability, now in-

---

terpreted as a constraint – a point that Gale and Shapley [1962] make explicitly in their paper.

But, one often has to choose. In which case, the winner is ... matching! Actually, if anything, the reverse. It is almost a tautology that a market designer's objective is to maximize objectives subject to constraints. Nevertheless, I think there are at least three good reasons why an applied market design researcher might choose the good properties approach rather than the maximize objectives approach. First, it is sometimes difficult to pin down *the* objective. For instance, market designers often aim for measures of both efficiency and fairness. Second, it is sometimes difficult to pin down the true constraints of the problem. Or, the constraints turn out to be somewhat flexible, in that they can be violated without infinite cost. Third, there is a lack of tools. In most matching environments we do not yet know how to solve even the standard Myersonian problem of "maximize social welfare subject to technology and incentive constraints", let alone anything more nuanced. For all of these reasons, it can be attractive for an applied researcher to find one or more good solutions to the problem at hand, rather than pursue the optimal solution.

I will now discuss two specific examples in detail. An introductory example, school choice, represents a simple win for the good properties approach. The main example, combinatorial assignment (e.g., course allocation), paints a more complicated picture. I will then conclude with a few observations based on this brief inquiry into matching "versus" mechanism design.

2.   INTRODUCTORY EXAMPLE: SCHOOL CHOICE

In a seminal paper, Abdulkadiroğlu and Sönmez [2003] initiated the market design literature on school choice. They proposed two specific mechanisms that satisfy attractive properties. The first is a variant on Gale-Shapley, and is stable (i.e., no justified envy) and strategyproof (for the students). The second is a variant on Gale's Top Trading Cycles algorithm, and is ex-post Pareto efficient and strategyproof.

Abdulkadiroğlu and Sönmez [2003] then guide policy makers on how to choose between these two approaches:

> "In some applications, policy makers may rank complete elimination of justified envy before full [student] efficiency, and Gale-Shapley student optimal stable mechamism can be used in those cases ... In other applications, the top trading cycles mechanism may be more appealing ... In other cases the choice between the two mechanisms may be less clear and it depends on the policy priorities of the policy makers."

This is not as definitive a conclusion as Vickrey or Myerson. But, it nevertheless is a highly influential paper, which subsequently led to major policy successes [cf. Roth 2008]. The reason for this is that, before Abdulkadiroğlu and Sönmez [2003], the mechanisms in practical use were not just sub-optimal, but seriously flawed. The most common mechanism in practice was (and remains) the so-called Boston mechanism, which has important incentives problems that lead to inefficiency and unfair outcomes [Abdulkadiroğlu et al. 2006; Pathak and Sönmez 2008; Azevedo and Budish 2012]. There are also what we might call "non mechanisms", meaning informal

procedures with blurry rules, whose efficiency and fairness performance are therefore hard to evaluate. Thanks to Abdulkadiroğlu and Sönmez [2003] we now have two mechanisms that satisfy attractive properties, both of which have been adopted for use in practice [cf. Abdulkadiroğlu et al. 2006; Roth 2012]. Moreover, subsequent empirical work on New York City's adoption of the Gale-Shapley variant shows that the mechanism performs well on real preference data [Abdulkadiroğlu et al. 2012], suggesting that the good properties approach was a reliable guide to welfare. The optimal should not be the enemy of the good.

3.   MAIN EXAMPLE: COMBINATORIAL ASSIGNMENT

School choice is a win for the "good properties" approach to market design. I now want to turn to a problem where the story is a bit more complicated, and on which I have done a lot of work: assignment with multi-unit demand. To fix ideas, consider course allocation at universities and professional schools. There is a set of agents (students), with preferences defined over bundles of indivisible objects (seats in courses), and monetary transfers are prohibited (even at the University of Chicago!). This problem is quite similar to the school choice problem, the main difference being that here students demand multiple courses, whereas in school choice the students demand a single school. As we will see, this shift from single-unit demand to multi-unit demand causes the good properties approach to point us, at first, in a very misleading direction.

Efficiency Criteria. There are two increasingly demanding notions of Pareto efficiency one might define for this setting. A feasible allocation is *ex-post Pareto efficient* if there is no other feasible allocation that all students weakly prefer, with at least one strict. A lottery over feasible allocations is *ex-ante Pareto efficient* if there is no other lottery over allocations that all students weakly prefer, with at least one strict. It is easy to see that ex-ante Pareto efficiency implies ex-post Pareto efficiency: if a lottery over feasible allocations placed positive probability on an allocation that is not ex-post Pareto efficient, then executing the Pareto-improving trades on that particular allocation would result in an ex-ante Pareto improvement as well. As a simple example to show that ex-ante Pareto efficiency is strictly stronger than ex-post, suppose that there are two students and two good courses, and that each student's von Neumann-Morgenstern utility is 0 if they get no good courses, 1 if they get one good course, and 1.5 if they get two good courses. The lottery in which with probability 0.5 the first student gets both good courses and with probability 0.5 the second student gets both good courses is Pareto efficient ex-post but not ex-ante: this lottery gives each student expected utility of $0.5 \cdot 0 + 0.5 \cdot 1.5 = 0.75$, which is less than the utility each student gets from the (degenerate) lottery in which they each get one good course for sure.

If one is willing to assume that utilities are inter-personally comparable, then one can define an efficiency criterion that is stronger still than ex-ante Pareto efficiency, namely *social-welfare maximization*. A feasible allocation maximizes social welfare if it maximizes the sum of student utilities amongst all feasible allocations. The easiest way to think about the relationship between social-welfare maximization and ex-ante Pareto efficiency is that welfare maximization requires the researcher to take a stand on the appropriate "weights" to place on each student's utility,

or, equivalently, to take a stand on the appropriate normalization of each student's utility. Any ex-ante Pareto efficient allocation is social-welfare maximizing for some set of welfare weights.

Impossibility Theorems. Unfortunately, impossibility theorems quickly establish that there is no perfect solution to this problem. Zhou [1990] shows that there is no symmetric mechanism that is both ex-ante Pareto efficient and strategyproof (even in a simpler version of the problem with single-unit demand), which implies that there is no mechanism that maximizes social welfare and is strategyproof. To see why there is a tension between welfare maximization and strategyproofness, consider a student who reports that her utility from her favorite schedule of courses is $+\infty$. The social-welfare maximizing allocation certainly will allocate this student her favorite schedule; but, unlike in the combinatorial auction setting, the mechanism cannot charge her for it!

A series of papers then studied what kinds of strategyproof mechanisms are possible if, instead of looking for ex-ante Pareto efficiency or welfare maximization, one settles for ex-post Pareto efficiency. Papai [2001], Ehlers and Klaus [2003], and Hatfield [2009] each show, essentially, that the only mechanisms that are ex-post Pareto efficient and strategyproof are dictatorships. The canonical dictatorship mechanism is called *random serial dictatorship* (RSD): students are randomly ordered, and then take turns choosing their entire bundle of courses, according to the specified serial order.

What should we make of these results? In a sense, we are left in a position that is similar to school choice after Abdulkadiroğlu and Sönmez [2003]. We have a mechanism that is strategyproof and ex-post Pareto efficient, and do not know much about how to achieve ex-ante Pareto efficiency or maximize social welfare. Papai [2001], Ehlers and Klaus [2003], and Hatfield [2009] thus each reach the conclusion that the random serial dictatorship is perhaps the best we can do for this problem. For instance, Hatfield [2009] writes (p. 514): "Although unfortunate, it seems that in many of these applications, the best procedure ... may well be a random serial dictatorship."

There is cause, however, to worry about this conclusion. Strategyproofness and ex-post Pareto efficiency are certainly attractive properties. But does the dictatorship stray too far from the underlying problem of maximizing social welfare subject to constraints, or the problem of finding a mechanism that is attractive with respect to both ex-ante efficiency and fairness? That is, does it stray too far from the problems that we would like to solve, but do not yet know how to solve?

A Mechanism from Practice: The Draft [Budish and Cantillon 2012]. In practice we rarely (never?) observe the random serial dictatorship, in which agents take turns choosing their entire bundle of objects. But we frequently observe draft mechanisms, in which agents take turns choosing one object at a time, over a series of rounds. Budish and Cantillon [2012] study the draft mechanism used at Harvard Business School (HBS) to allocate elective courses to MBA students. In the HBS course draft, students submit a rank-ordered list of their preferences over individual courses (an implicit assumption is that preferences are additive separable but for scheduling constraints). The students are then randomly ordered by a computer,

and are allocated courses one at a time in order, based on their reported preferences and remaining availability. In odd rounds $(1, 3, 5, \ldots)$ the allocation proceeds in ascending order of random priority, and in even rounds $(2, 4, 6, \ldots)$ in descending order.

It is easy to show that the draft is not strategyproof (cf. Example 1 of Budish and Cantillon [2012]). Students should overreport how much they like the popular courses that are likely to reach capacity in the early rounds, and underreport how much they like unpopular courses that they can get with ease in the later rounds. It is also straightforward to show that the draft is not ex-post Pareto efficient in Nash equilibrium. So, on the properties emphasized by Abdulkadiroğlu and Sönmez [2003], Papai [2001], Ehlers and Klaus [2003], and Hatfield [2009] we might conclude that

$$\text{dictatorship} > \text{draft}$$

Budish and Cantillon [2012] ask a different question about efficiency: how well does the draft do at the problem of maximizing social welfare? All we know from the failure of ex-post Pareto efficiency is that the draft does not achieve the unconstrained maximum. And we know that RSD does not achieve the unconstrained maximum either, from at least as early as Hylland and Zeckhauser [1979]. We use data from HBS consisting of students' actual submitted ROL's, which are potentially strategic, and their underlying true preferences, from an administration survey. (There are several robustness checks in the paper concerning the interpretation of the administration survey as true preferences). Because we have truthful and strategic preferences, we can look directly at how well the draft does at the social-welfare maximization problem. We can also use the truthful preferences to simulate equilibrium play of the counterfactual of interest, RSD. It turns out that on some simple measures of welfare, the draft outperforms the dictatorship:

	E(Avg. Rank)	% Who Get #1 Choice	% Who Get All Top 10
No Scarcity	5.50	100%	100%
Draft - Truthful Play	7.66	82%	1.5%
Draft - Strategic Play	7.99	63%	2.2%
RSD - Truthful Play	8.74	49%	29.7%

Table I.   Summary Statistics from Budish and Cantillon [2012].

In addition, when we look at the societal distribution of the "average rank" statistic – this is the average rank of the courses in the student's assigned bundle, based on their true preferences, where the best possible outcome is to get their 10 favorite courses for an average rank of 5.50 – the distribution under the draft second-order stochastically dominates the distribution under the dictatorship (see Figure 1). The one measure on which RSD excels is the number of students who get their bliss point: around 30% under the dictatorship versus 2% under the draft. This is a reflection of the risk that RSD exposes students to: those early in the choosing order get their bliss point, those late in the choosing order get the leftovers. So
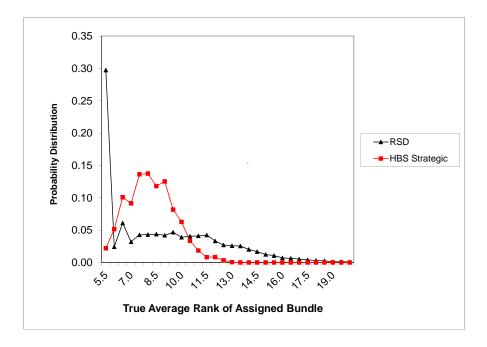
Fig. 1. Distribution of the average rank of the courses students receive: truthful play of the random serial dictatorship versus strategic play of the draft [Budish and Cantillon 2012].

long as students are weakly risk averse (which seems reasonable) we may conclude that, on empirical measures of social welfare:

$$\text{draft} > \text{dictatorship}$$

Why does the dictatorship, which is ex-post Pareto efficient, perform so poorly on simple measures of welfare such as average rank and % who get 1st choice? Consider the following example from Budish and Cantillon [2012]. There are N students who require 2 courses each, and $M = 4$ courses ($\mathcal{C} = \{a, b, c, d\}$) with capacity of $\frac{N}{2}$ seats each. Students' ordinal preferences over individual courses are as follows

$$\frac{N}{2} \text{ are } P_1 : a, b, c, d$$

$$\frac{N}{2} \text{ are } P_2 : b, a, d, c$$

In the draft, all students get their 1st and 3rd favorites, all the time: the P_1 types get $\{a, c\}$ while the P_2 types get $\{b, d\}$. (Note that truthful play is a Nash equilibrium of the HBS draft given these preferences). In RSD, by contrast, students who are in the first half of the choosing order do a bit better – they get their 1st and 2nd favorites ($\{a, b\}$), while students who are in the second half of the choosing order do much worse – they get their 3rd and 4th favorites ($\{c, d\}$). Mechanically, what drives this is that students with good lottery numbers use their *last* choices to take the course that would have been the *first* choices of students with poor lottery

numbers. Budish and Cantillon [2012] term this phenomenon "callousness", to reflect that students callously disregard the preferences of those who choose after them in line. Ex-post, since there are no transfers, these allocations are Pareto efficient: there are no Pareto improving trades between the lucky and unlucky students. But ex-ante, callous behavior is bad for welfare: the benefit to the lucky is small, whereas the harm to the unlucky is large.

Lessons from Budish and Cantillon [2012]. There are three lessons to take from Budish and Cantillon [2012]'s study of the HBS draft. The first lesson concerns ex-post Pareto efficiency. A sensible prior is that switching from the all-at-once RSD to the one-at-a-time HBS draft would be good for fairness but bad for welfare. After all, RSD is ex-post Pareto efficient and the HBS draft is not. But in non-transferable utility settings there can be many ex-post Pareto efficient allocations. And the lottery over efficient allocations induced by RSD is very unattractive when assessed ex-ante, so much so that the HBS lottery over *in*efficient allocations looks more appealing. The punchline is that if the real objective is to maximize social welfare subject to constraints, the HBS draft is a better solution than the RSD. The mistake in the dictatorship papers was to conclude that, just because we do not know how to maximize social welfare, we should settle for ex-post Pareto efficiency. In this instance, it would have been better to admit that the true objective is welfare, but that we simply do not know how to maximize it yet!

A second lesson concerns the role of strategyproofness in practical market design. The Budish and Cantillon [2012] field data directly document that students at HBS – real-life participants in a one-shot high-stakes setting – figured out how to manipulate the non-strategyproof HBS mechanism. Furthermore, Budish and Cantillon [2012] show that this manipulability harms welfare, and that the magnitudes are large. These findings are strongly consistent with the view that strategyproofness is an important desideratum in practical market design. However, constraints often have costs. And, in this setting, the welfare costs of using a strategyproof dictatorship are much larger than the welfare costs of manipulability. Overall, the results suggest a nuanced view of the role of strategyproofness in design, and the need for second-best alternatives to strategyproofness [cf. Azevedo and Budish 2012; Lubin and Parkes 2012].

Third, the Budish and Cantillon [2012] analysis suggests "where to look" for mechanisms that are better still than the HBS draft: seek an incentives middle ground between strategyproof dictatorships and the simple-to-manipulate draft; and, seek a mechanism that yields a relatively equal distribution of outcomes, like the draft and unlike the dictatorship. We turn to this next.

Approximate Competitive Equilibrium from Equal Incomes. Budish [2011] designs a new mechanism for combinatorial assignment, informed by the lessons from Budish and Cantillon [2012]. The mechanism is based on an old idea from general equilibrium theory, the Competitive Equilibrium from Equal Incomes ("CEEI"). CEEI is well known to be an efficient and fair solution to the problem of allocating divisible goods [Varian 1974]. In our environment with indivisibilities, CEEI would mean the following. First, agents report their preferences over bundles. Second, agents are given equal budgets b^* of an artificial currency. Third, we find an item

price vector $\mathbf{p}^* = (p_1, \ldots, p_M)$ such that, when each agent is allocated his favorite affordable bundle – that is, his most preferred permissible schedule from the budget set $\{x \in 2^{\mathcal{C}} : \mathbf{p}^* \cdot x \leq b^*\}$ – the market clears. Last, we allocate each agent their favorite affordable bundle.

It is easy to see that existence of CEEI prices is problematic. Just consider the case in which all agents have the same preferences. Then, at any price vector, all N agents have the same demand, so demand for each course j is either 0 or N, which could be very far from its supply q_j. The main theorem in Budish [2011] is an existence theorem for an approximation to CEEI in which (i) agents are given approximately equal instead of exactly equal budgets (all budgets are in the interval $[b^*, b^* + \epsilon]$ for arbitrarily small $\epsilon > 0$); and (ii) the market clears approximately instead of exactly. Interestingly, the RSD mentioned above can also be interpreted as a competitive equilibrium mechanism, but from highly unequal budgets: each student's budget dwarfs that of anybody after him in the serial order, and is in turn dwarfed by that of anybody prior to him in the serial order.

As a simple example, suppose that there are two agents with additive preferences over four objects. Two of the objects are valuable diamonds (Big and Small) and two of the objects are ordinary rocks (Pretty and Ugly). Each agent has capacity for at most two objects. In the RSD, whichever agent gets to choose first will get both of the diamonds, so the allocation is very unfair ex-post. Furthermore, we know from Budish and Cantillon [2012] that the lottery over allocations will be unappealing ex-ante. Exact CEEI does not exist. If the Big Diamond costs weakly less than b^* then both agents will demand it, whereas if it costs strictly more than b^* neither agent can afford it. So demand is either 2 or 0, but supply is 1. However, if we give agents approximately equal instead of exactly equal budgets, we can recover existence. If budgets are, say, b^* and $b^* + \epsilon$ (randomly assigned), set the price of the Big Diamond to $b^* + \epsilon$, so that whichever agent randomly gets the slightly larger budget gets the Big Diamond (and the Ugly Rock, which we can price at 0), and set the prices of the Small Diamond and the Pretty Rock to $\frac{b^*}{2}$ each, so the agent with the smaller budget gets {Small Diamond, Pretty Rock}.

Approximate CEEI satisfies attractive second-best properties of efficiency, fairness and incentives. It is ex-post Pareto efficient but for small market-clearing error. (In the Diamond-Rock example market clearing is perfect, but this is not the case in general.) It bounds envy to within the level of a single indivisible good: the poorer agent envies the richer agent who gets the Big Diamond, but at least the richer agent does not get both diamonds while the poorer agent gets neither. And, while not strategyproof, it satisfies a large-market notion of approximate strategyproofness [Azevedo and Budish 2012], just as the Walrasian mechanism is not strategyproof but is approximately so [Roberts and Postlewaite 1976].

There are two possible interpretations of the role of ex-post fairness in Approximate CEEI. One interpretation is that ex-post fairness is an explicit design objective alongside efficiency. The second interpretation is that ex-post fairness is a means to an end, namely ex-ante welfare. More specifically, Budish and Cantillon [2012] showed us that the ex-post unfairness of the RSD has a cost in terms of ex-ante welfare, and Approximate CEEI avoids this concern by imposing ex-post fairness as a design objective. Indeed, on the HBS preference data, using the Othman et al.

[2010] computational procedure, Approximate CEEI outperforms both RSD and the HBS draft on the welfare measures emphasized in [Budish and Cantillon 2012]. Approximate CEEI also performed well in laboratory tests of its performance on MBA students from the Wharton School at the University of Pennsylvania [Budish and Kessler 2012].

Under the first interpretation, Approximate CEEI is a mechanism in the "good properties" tradition, whereas under the second interpretation Approximate CEEI is a mechanism in the "maximize welfare subject to constraints" tradition. As I stated at the outset of this essay, I do not think these two approaches are necessarily that distinct.

4. MATCHING "VERSUS" MECHANISM DESIGN: SOME OBSERVATIONS

Let me conclude with a few observations based on this brief inquiry into matching "versus" mechanism design.

Observation 1: New tools are needed. In both the school choice environment of Abdulkadiroğlu and Sönmez [2003] and the course-allocation environment of Budish [2011], the problem of maximizing social welfare subject to technology and incentive constraints remains open. The basic reason why social-welfare maximization is difficult in these problems is the lack of a numeraire good like money. One somehow needs to avoid incentivizing agents to report that their utility from their favorite bundle is $+\infty$. The mechanisms proposed in Abdulkadiroğlu and Sönmez [2003] and Budish [2011] work around this difficulty by asking only for ordinal preference information (over individual objects and bundles, respectively). The Hylland and Zeckhauser [1979] mechanism for single-unit assignment, as well as its generalization to multi-unit assignment by Budish et al. [2012], work around this difficulty by asking only for marginal rates of substitution across individual objects. But perhaps there is a better, more direct approach to social-welfare maximization?

Recent work by Nguyen and Vohra [2012] seems promising in this regard. Nguyen and Vohra [2012] formulate a linear program for combinatorial assignment in which social-welfare maximization is the objective and, instead of incentive constraints there are (interim) envy-freeness constraints. Finding an interim envy-free allocation involves assigning agents lotteries over bundles, as in Hylland and Zeckhauser [1979] or Budish et al. [2012]. An agent who reports a value of $+\infty$ for her favorite bundle will get a lottery that places as much probability as possible, given envy freeness, on her getting that bundle, and zero probability of getting any other bundle. The paper then exploits recent advances in combinatorial optimization to show that the resulting lottery over bundles can be implemented in an approximate sense, with a "small" amount of market-clearing error as in Budish [2011]. It will be interesting to learn more about the empirical and computational performance of this approach.

Other interesting approaches include Procaccia and Tennenholtz [2009] in the context of facility location problems and Guo and Conitzer [2010] in the context of assignment problems with just two agents. These problem environments are sufficiently specialized that the $+\infty$ issue can be circumvented.

Observation 2: Know thy objectives. To the extent that the axiomatic literature on multi-unit assignment made a "mistake", it was to conclude that, because it is difficult or impossible to maximize social welfare (or even obtain exact ex-ante efficiency), we should instead seek exact ex-post Pareto efficiency. In assignment problems with single-unit demand, such as the school choice problem discussed above, ex-post Pareto efficiency has been a reliable guide to good market design. But, in multi-unit assignment, ex-post Pareto efficiency admits severely unfair allocations, which is both per se undesirable and significantly harms social welfare.

Sometimes we do not yet know how to maximize the true objective subject to the true constraints because of limitations of the theory. In my view, such situations call for attacking the true objective using both theory and either data or computational simulations. For instance, Approximate CEEI gives theoretical worst-case performance guarantees for both approximate ex-post Pareto efficiency and approximate envy freeness, and then shows in the data that average-case performance on both of these dimensions is considerably better than the worst-case bound (e.g., in the data only 1% of students have any envy). The data is also used to show that ex-ante welfare performance – which I was not able to study directly (cf. Observation 1) – is attractive. This combination of worst-case performance guarantees and empirical analysis of average-case performance is of course familiar to computer scientists; to give just one recent example, see Ashlagi et al. [2012]. See also Hartline [2012] for a recent survey on approximation methods.

Observation 3: Know thy constraints. Sometimes it is tolerable to satisfy constraints approximately instead of exactly. Such approximations represent a challenge for both methodologies. The mechanism design methodology of "max objectives subject to constraints" treats the constraints as lexicographically more important than the objectives. The matching methodology of seeking a mechanism that satisfies attractive properties treats the stated axioms or properties as lexicographically more important than other properties.

With respect to incentive constraints, there has been a lot of recent work on second-best notions of approximate strategyproofness. Parkes et al. [2001], Day and Milgrom [2008], Erdil and Klemperer [2010], Carroll [2011], Dutting et al. [2012] and others have proposed notions of approximate strategyproofness that are based on minimizing the magnitude of the most profitable manipulation (either in realization or in expectation). Lubin [2010] explores a measure of approximate strategyproofness based on maximizing the number of agents with zero ex-post regret, and argues that this in turn generates approximate interim incentives for truthful reporting in a Bayesian sense. Azevedo and Budish [2012] propose a criterion called strategyproof in the large (SP-L), which requires that any incentives to misreport one's preferences vanish towards zero as the market grows large. Heuristically, SP-L requires that a mechanism be strategyproof for the kinds of agents we might think of as "price takers". As an example, uniform-price auctions are not strategyproof [Ausubel and Cramton 2002] but are SP-L, whereas pay-as-bid auctions not only are not strategyproof, but are not even SP-L, since even in a large market in which agents are price takers agents have incentive to misreport their preferences. The concept can also be applied to mechanisms without explicit prices. For instance, the Gale-Shapley deferred acceptance algorithm is not SP but

is SP-L, whereas the Boston mechanism for school choice is not even SP-L.

Feasibility constraints, too, can sometimes be satisfied approximately instead of exactly. More precisely, in some settings violating feasibility constraints is not infinitely costly. Examples include course allocation [Budish and Kessler 2012], in which course capacity constraints are sometimes better thought of as targets rather than true feasibility constraints (though sometimes not), and affirmative action constraints in school choice [Ehlers et al. 2011].

Observation 4: Is Perfect the enemy of the Good?. As theorists we often ask for perfect solutions to problems. It is tempting to find *the* solution. The danger of course is that we end up only paying attention to the problems that we can solve beautifully. There is the old joke about searching for one's lost keys under the lamppost, because "that's where the light is".

I believe that there is value in designing mechanisms that are "better" even if not perfect. I also believe that data can play an important role in such theorizing to yield, as Roth [2002] says, "a sense of magnitudes." Comparing the good but imperfect mechanism to existing mechanisms gives a sense of the magnitude of the achievement. Comparing the good but imperfect mechanism to a performance upper bound gives a sense of how much work there is left to do.

REFERENCES

ABDULKADIROĞLU, A., PATHAK, P., AND AGARWAL, N. 2012. Sorting and welfare consequences of coordinated admissions: Evidence from New York City. Working Paper.

ABDULKADIROĞLU, A., PATHAK, P., ROTH, A., AND SÖNMEZ, T. 2006. Changing the Boston mechanism: Strategyproofness as equal access. Mimeo, Harvard University.

ABDULKADIROĞLU, A. AND SÖNMEZ, T. 2003. School choice: A mechanism design approach. *American Economic Review 93,* 3, 729–747.

ASHLAGI, I., FISCHER, F., KASH, I., AND PROCACCIA, A. 2012. Mix and Match: A strategyproof mechanism for multi-hospital kidney exchange. Working Paper.

AUSUBEL, L. AND CRAMTON, P. 2002. Demand reduction and inefficiency in multi-unit auctions. Working paper.

AZEVEDO, E. AND BUDISH, E. 2012. Strategyproofness in the large. Working Paper.

BUDISH, E. 2011. The combinatorial assignment problem: Approximate competitive equilibrium from equal incomes. *Journal of Political Economy 119(6)*, 1061–1103.

BUDISH, E. AND CANTILLON, E. 2012. The multi-unit assignment problem: Theory and evidence from course allocation at Harvard. *American Economic Review 102(5)*, 2237–71.

BUDISH, E., CHE, Y., KOJIMA, F., AND MILGROM, P. 2012. Designing random allocation mechanisms: Theory and applications. *American Economic Review.* Forthcoming.

BUDISH, E. AND KESSLER, J. 2012. An experimental test of a complex market design: Changing the course allocation system at Wharton. In Preparation.

CARROLL, G. 2011. A quantitative approach to incentives: Application to voting rules. Working Paper.

DAY, R. AND MILGROM, P. 2008. Core-selecting package auctions. *International Journal of Game Theory 36,* 3, 393–407.

DUTTING, P., FISCHER, F., JIRAPINYO, P., LAI, J., LUBIN, B., AND PARKES, D. 2012. Payment rules through discriminant-based classifiers. *Proceedings of the 13th ACM Conference on Electronic Commerce (EC'12).*

EHLERS, L., HAFALIR, I., YENMEZ, B., AND YILDIRIM, M. 2011. School choice with controlled choice constraints: Hard bounds versus soft bounds. Working Paper.

EHLERS, L. AND KLAUS, B. 2003. Coalitional strategy-proof and resource-monotonic solutions for multiple assignment problems. *Social Choice and Welfare 21(2)*, 265–80.

ERDIL, A. AND KLEMPERER, P. 2010. A new payment rule for core-selecting package auctions. *Journal of the European Economic Association*.

GALE, D. AND SHAPLEY, L. 1962. College admissions and the stability of marriage. *American Mathematical Monthly 69,* 1, 9–15.

GREEN, J. AND LAFFONT, J. 1977. Characterization of satisfactory mechanisms for the revelation of preferences for public goods. *Econometrica*, 427–438.

GUO, M. AND CONITZER, V. 2010. Strategy-proof allocation of multiple items between two agents without payments or priors. *Proceedings of the 9th International Joint Conference on Autonomous Agents and Multi Agent Systems (AAMAS'10)*.

HARTLINE, J. 2012. Approximation in mechanism design. *American Economic Review 102,* 3, 330–336.

HATFIELD, J. 2009. Strategy-proof, efficient, and nonbossy quota allocations. *Social Choice and Welfare 33(3)*, 505–15.

HYLLAND, A. AND ZECKHAUSER, R. 1979. The efficient allocation of individuals to positions. *Journal of Political Economy 87,* 2, 293–314.

LUBIN, B. 2010. Combinatorial markets in theory and practice: Mitigating incentives and facilitating elicitation. *PhD thesis, Computer Science, Harvard University*.

LUBIN, B. AND PARKES, D. 2012. Approximate strategyproofness. *Current Science 103(9)*, 1021–1032.

MYERSON, R. 1981. Optimal auction design. *Mathematics of Operations Research 6,* 1, 58–73.

NGUYEN, T. AND VOHRA, R. 2012. The allocation of indivisible objects via rounding. Working Paper.

OTHMAN, A., BUDISH, E., AND SANDHOLM, T. 2010. Finding approximate competitive equilibria: Efficient and fair course allocation. *Proceedings of the 9th International Joint Conference on Autonomous Agents and Multi Agent Systems (AAMAS'10)*.

PAPAI, S. 2001. Strategyproof and nonbossy multiple assignments. *Journal of Public Economic Theory 3,* 3, 257–271.

PARKES, D., KALAGNANAM, J., AND ESO, M. 2001. Achieving budget-balance with Vickrey-based payment schemes in exchanges. In *International Joint Conference on Artificial Intelligence*. Vol. 17. AAAI Press, 1161–1168.

PATHAK, P. AND SÖNMEZ, T. 2008. Leveling the playing field: Sincere and sophisticated players in the Boston mechanism. *The American Economic Review 98,* 4, 1636–1652.

PROCACCIA, A. AND TENNENHOLTZ, M. 2009. Approximate mechanism design without money. *Proceedings of the 10th ACM Conference on Electronic Commerce (EC'09)*.

ROBERTS, D. AND POSTLEWAITE, A. 1976. The incentives for price-taking behavior in large exchange economies. *Econometrica*, 115–127.

ROTH, A. 2002. The economist as engineer: Game theory, experimentation, and computation as tools for design economics. *Econometrica 70,* 4, 1341–1378.

ROTH, A. 2008. What have we learned from market design? *The Economic Journal 118*, 285–310.

ROTH, A. 2012. School choice in New Orleans with top trading cycles. *marketdesigner.blogspot.com*.

SHAPLEY, L. AND SCARF, H. 1974. On cores and indivisibility. *Journal of Mathematical Economics 1,* 1, 23–37.

VARIAN, H. 1974. Equity, envy and efficiency. *Journal of Economic Theory 9,* 1, 63–91.

VICKREY, W. 1961. Counterspeculation, auctions, and comptitive sealed tenders. *Journal of Finance 16(1)*, 8–37.

ZHOU, L. 1990. On a conjecture by Gale about one-sided matching problems. *Journal of Economic Theory 52,* 1, 123–136.