DISCUSSION PAPER SERIES

No. 7641

THE MULTI-UNIT ASSIGNMENT PROBLEM: THEORY AND EVIDENCE FROM COURSE ALLOCATION AT HARVARD

Eric Budish and Estelle Cantillon

INDUSTRIAL ORGANIZATION

Centre for Economic Policy Research

www.cepr.org

Available online at: www.cepr.org/pubs/dps/DP7641.asp

ISSN 0265-8003

THE MULTI-UNIT ASSIGNMENT PROBLEM: THEORY AND EVIDENCE FROM COURSE ALLOCATION AT HARVARD

Eric Budish, University of Chicago Booth School of Business
Estelle Cantillon, FNRS, Université Libre de Bruxelles and CEPR

Discussion Paper No. 7641
January 2010

Centre for Economic Policy Research
53–56 Gt Sutton St, London EC1V 0DG, UK
Tel: (44 20) 7183 8801, Fax: (44 20) 7183 8820
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 7641

January 2010

ABSTRACT

The Multi-unit Assignment Problem: Theory and Evidence from Course Allocation at Harvard

This paper uses data consisting of students' strategically reported preferences and their underlying true preferences to study the course allocation mechanism used at Harvard Business School. We show that the mechanism is manipulable in theory, manipulated in practice, and that these manipulations cause meaningful welfare losses. However, we also find that ex-ante welfare is higher than under the strategyproof and ex-post efficient alternative, the Random Serial Dictatorship. We trace the poor ex-ante performance of RSD to a phenomenon specific to multi-unit assignment, "callousness". We draw lessons for the design of multi-unit assignment mechanisms and for market design more broadly.

JEL Classification: C78, C93 and D02
Keywords: course allocation, dictatorship, ex-ante efficiency, ex-post efficiency, field data, market design, multi-unit assignment, random serial dictatorship, strategic behaviour and strategyproofness

Eric Budish
University of Chicago
Booth School of Business
5807 S Woodlawn Avenue
Chicago IL 60637
USA

Email: eric.budish@chicagobooth.edu

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=171429

Estelle Cantillon
Université Libre de Bruxelles
ECARES
50, av. F.D. Roosevelt, CP 114
1050 Bruxelles
BELGIUM

Email: Estelle.Cantillon@ulb.ac.be

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=151115

Submitted 11 January 2010

We are extremely grateful to Sheila Connelly, Assistant Director, MBA Registrar Services, for providing us with the data and sharing her extensive knowledge about the HBS course allocation mechanism with us. For suggestions, we thank Susan Athey, Marianne Bertrand, Drew Fudenberg, Matt Gentzkow, Fuhito Kojima, Jordi Massó, Paul Milgrom, David Parkes, Ariel Pakes, Parag Pathak, Canice Prendergast, Al Roth, and seminar audiences at ASSA 2008, Dagstuhl 2007, Paris Sorbonne, Harvard, Toulouse School of Economics, Autònoma de Barcelona, and Columbia. Financial support from the European Research Council (grant MaDEM), the Belgian National Science Foundation (FNRS), the Division of Research at Harvard Business School, and the University of Chicago Booth School of Business is gratefully acknowledged. This paper previously circulated under the title "Strategic Behavior in Multi-Unit Assignment Problems."

1 Introduction

Educational institutions commonly place limits on the number of students in any particular class. We are interested in a market-design problem that arises from class-size limits: if it is not possible for all students to take their most desired schedule of courses, then how should seats in over-demanded courses be allocated?[1]

Economists have increasingly played an active role in designing solutions to real-world resource-allocation problems like course allocation. In the past, mechanisms suggested by the pure theory literature have often provided a useful starting point. One prominent example comes from the redesign of the institution that matches medical-school graduates to hospital residency positions, on which Roth (2002) remarks that "the simple theory [of Gale and Shapley (1962)] offered a surprisingly good guide to the design [of the Roth-Peranson (1999) algorithm]." Other important examples include the design of school-choice procedures, advertising markets on internet search engines, and combinatorial auctions.[2]

For course allocation, however, the extant theory literature is of little help. Course allocation is an example of a *multi-unit assignment problem*, in which a set of indivisible objects (seats in courses) is to be allocated amongst a set of agents (students), the agents have multi-unit demand (for schedules of courses), and there are exogenous restrictions against the use of monetary transfers.[3] One set of results on such problems shows that specific efficiency, fairness and incentives criteria that are compatible for the single-unit assignment problem are impossible to achieve for the multi-unit case (Sönmez, 1999; Konishi, Quint and Wako, 2001; Klaus and Miyagawa, 2001; Manea, 2007; Kojima, forthcoming). A second set of results shows, essentially, that the *only* multi-unit assignment mechanisms that are ex-post Pareto efficient and strategyproof are dictatorships, which intuitively are highly unfair: for any two students, one gets to choose *all* her courses before the other gets to choose *any* (Papai, 2001; Ehlers and Klaus, 2003; Hatfield, 2009).[4]

Given the lack of positive results from theory, a sensible starting point for design is to see what we can learn from mechanisms that are actually used in practice. In this paper, we study the

[1] Press coverage and anecdotal evidence suggest that the scarcity problem is particularly acute in higher education, especially at professional schools. See Bartlett (2008), Guernsey (1999), Lehrer (2008), and Neil (2008).

[2] New school choice procedures in New York and Boston incorporated aspects of Gale and Shapley's mechanism for two-sided matching and Shapley and Scarf's (1974) mechanism for single-unit assignment (Abdulkadiroglu and Sonmez 2003; Abdulkadiroglu et al 2005a, b). Google cites the influence of Vickrey's (1961) "Nobel Prize-winning economic theory" in the design of its auction for advertising slots (Edelman, Ostrovsky and Schwarz, 2008; Varian, 2007). On the relationship between theory and practice for combinatorial auction design, see Cramton et al (2007) and Milgrom's (2004) aptly named text "Putting Auction Theory to Work."

[3] Other examples of multi-unit assignment problems include the assignment of tasks within an organization, the allocation of shared scientific resources amongst their users, sports drafts, the division of heirlooms amongst heirs, and the allocation of airport takeoff-and-landing slots to airlines in many countries.

[4] Budish (2009) formalizes the sense in which dictatorships are unfair in multi-unit assignment.

mechanism used at Harvard Business School (HBS) since the mid-1990s to allocate roughly 9,000 elective course seats to about 900 second-year MBA students every year. We choose this mechanism for two reasons. First, it is a *prima facie* sensible mechanism, satisfying attractive efficiency and fairness properties and differing from the dictatorship in an intuitively attractive way.[5] Rather than choosing courses all at once, which will lead to highly unequal allocations, students take turns choosing a single course at a time. (More accurately, a computer chooses courses for them based on their reported preferences; the priority order is random in the first round, and is reversed each subsequent round.) Similar mechanisms have been used for a long time both by other educational institutions and in other multi-unit assignment contexts (Brams and Straffin, 1979; Brams and Taylor, 1999). In this sense, the mechanism passes a market endurance test.

Second, we have great data. In addition to students' actual (strategic) reports of their preferences, we have data on students' *underlying truthful preferences*, from a survey conducted by the HBS administration. (We describe in detail in Section 5 our argument that the survey data are indeed truthful). Whereas strategic reports are naturally recorded by market administrators and commonly made available to researchers, data on underlying preferences typically are only available to researchers in laboratory settings. The combination of truthful and stated preferences is powerful for two reasons. First, it means that we can directly observe students' strategic manipulations and quantify their effect on welfare. Second, we can use the truthful preferences to simulate equilibrium play of the strategyproof Random Serial Dictatorship (RSD) suggested by the extant theory,[6] and so can compare the two mechanisms.

The main results of our analysis can be summarized as follows. The HBS mechanism is simple to manipulate in theory (Section 3), is heavily manipulated by students in practice (Section 5), and these manipulations cause congestion and substantial inefficiency, assessed either ex-ante (i.e. before priority orders are drawn) or ex-post (Section 6). Yet, ex-ante welfare is higher under the HBS mechanism than under the strategyproof and ex-post efficient RSD (Section 7). We trace the poor ex-ante performance of RSD to a phenomenon we call "callousness" that is specific to multi-unit assignment and unrelated to risk attitudes (Section 8).

Thus, one thing we learn from looking to practice is that the HBS mechanism, though flawed, is a sensible choice relative to the extant theoretical alternative; RSD can be rejected on welfare grounds alone. In the Conclusion we describe how our results suggest "where to look" for new mechanisms that are better still, a direction taken up in subsequent work by Budish (2009). We also discuss

[5] By contrast, Sönmez and Unver (forthcoming) show that another widely used course-allocation mechanism, the Bidding Points Mechanism, has a serious conceptual flaw. See also Krishna and Unver (2008) and Budish (2009; Section 7.3).

[6] Amongst dictatorship mechanisms, what distinguishes the Random Serial Dictatorship is that it satisfies the procedural fairness property of anonymity (also called "equal treatment of equals"), because the choosing order is uniform random.

how our results contribute to two active debates in the broader literature on market design: (i) the importance of strategyproofness, and whether gaming of real-life mechanisms has first-order welfare effects; and (ii) how to analyze the efficiency of random mechanisms, and specifically the relationship between ex-post and ex-ante efficiency.

Our analysis begins in Section 3 by studying the theoretical properties of the HBS mechanism. It satisfies several criteria of fairness and yields outcomes that are consistent with ex-post Pareto efficiency if students reveal their preferences truthfully. However, the HBS mechanism is simple to manipulate: students should overreport how much they like popular courses and underreport how much they like unpopular courses, so they do not waste early-round draft picks on courses they can get in later rounds. We provide a partial characterization of equilibrium behavior and equilibrium run-out times for courses. We find that strategic behavior affects students' welfare through two distinct channels. First, strategic behavior leads to congestion (popular courses reach capacity faster) which hurts students whose preferred courses are popular amongst other students. Second, strategic behavior can lead to ex-post Pareto inefficient outcomes (i.e., mutually beneficial trades) if a student strategically underreports some class and then turns out not to get it. We show through an example that all students can be worse off ex-ante under equilibrium strategic play.

Section 4 describes our data, which cover the allocation of second-year courses to MBA students at Harvard Business School during the 2005-2006 academic year. We have student-level survey data on preferences over courses at two different points in time, as well as their officially-submitted preferences during a trial run and the real run of the HBS mechanism.

We argue in Section 5 that the survey data collected prior to the initial allocation correspond to students' truthful preferences and that the preferences submitted for the real run of the HBS mechanism correspond to equilibrium behavior. This joint hypothesis is natural given the context of our data; beyond context, we provide support for this hypothesis at both the aggregate level and the individual level. At the aggregate level, we show that preferences at these two points in time differ significantly and in a way that is consistent with equilibrium strategic behavior, and largely inconsistent with alternative explanations such as social learning or new information. At the individual level, we test whether students' submitted preferences rearrange popular and unpopular courses in a manner that is consistent with our characterization of equilibrium best responses. We find that most students (82.2%) submit preferences that are consistent with equilibrium behavior. Most of those who do not seem to have changed their preference for some single course, or made a strategic error based on plausible incorrect beliefs about a course's popularity.

In Section 6 we then use these preferences and equilibrium strategies to quantify the welfare consequences of strategic play, by comparing actual equilibrium play with non-equilibrium truthful behavior. In other words, our counterfactual exercise asks what would happen if the social planner knew students' preferences and used the HBS mechanism to allocate courses. We first document that strategic behavior causes congestion and ex-post inefficient allocations: on average our Pareto-improvement-seeking integer program is able to find beneficial trades involving 84% of students and 15% of course seats. Ex-ante welfare comparisons are more subtle: our data consist of ordinal preferences over individual courses, but welfare depends on von Neumann-Morgenstern preferences over bundles. We develop a novel computational method which allows us to draw welfare comparisons for many students based on this limited ordinal information: nearly half of students' are unambiguously harmed by strategic play, whereas only 5% unambiguously benefit. To reach comparisons for the other students, or to evaluate welfare from the perspective of a utilitarian social planner, we put more structure on preferences, i.e., on the map between ordinal preferences over individual courses and vNM preferences over bundles. We develop two such maps, motivated by the HBS administration's emphasis on the number of students who obtain their single favorite course and on the average rank of the ten courses students receive. If students have what we call lexicographic preferences or what we call average-rank preferences, we can conclude that a large majority of individual students and a utilitarian social planner regard strategic play as harmful to welfare. The magnitudes are meaningful: strategic play reduces the proportion of students who receive their favorite course from 83% to 60%, and increases the expected average rank of the ten courses in a student's schedule from 7.76 to 8.35 (higher is worse, 5.50 is bliss). Intuitively, there is a basic asymmetry between the costs and benefits of strategic play; e.g., it is impossible to strategically overreport one's *favorite* course.

Because strategic behavior hurts most students it is natural to compare the HBS mechanism with its strategyproof alternative, RSD. In Section 7, our counterfactual exercise asks what would happen *in equilibrium* if the HBS administration adopted RSD as its course-allocation mechanism. Ex-post, we know that RSD is Pareto efficient whereas we found in the previous section that HBS yields substantial inefficiencies. But when we compare the two mechanisms ex-ante, HBS looks much more attractive than RSD. If students have either average-rank preferences or lexicographic preferences, then both the large majority of individual students and a utilitarian social planner prefer the HBS mechanism to RSD.[7] Perhaps surprisingly, this is the case even if students are risk neutral. Using RSD would reduce the proportion of students who receive their favorite course from 60% to 47%, and increase the expected average rank from 8.35 to 8.94.

Why is the ex-post inefficient HBS mechanism more attractive ex-ante than the ex-post efficient RSD? In Section 8 we propose a simple theoretical explanation. Under RSD, fortunate students with good random draws make their second, third, ..., last choices independently of whether these courses would be some unfortunate students' *first* choices. That is, the lucky "callously disregard"

[7] One measure on which RSD outperforms HBS is the proportion of students who obtain their bliss bundle: 29% under RSD versus 1% under HBS. Students who care mainly about the probability of getting their bliss bundle prefer RSD.

the unlucky.[8] In any one random draw, there is no way to improve the allocation of the unlucky without harming the lucky, so outcomes are ex-post Pareto efficient. But in expectation the harm to the unlucky exceeds the benefit to the lucky, so RSD does very poorly on measures of ex-ante efficiency. Note that this explanation does not depend on risk preferences; aversion to risk only exacerbates the unattractiveness of RSD.

2 The course allocation problem

2.1 Environment

Courses. There is a finite set of C courses, $\mathcal{C}$.[9] Courses have capacities $\mathbf{q} = (q_1, ..., q_C)$.

Students and preferences. There is a continuum of students described by the interval $[0, 1]$ and endowed with the Lebesgue measure.[10] Each student s is endowed with a von Neumann-Morgenstern utility function u_s that indicates her utility from each bundle of courses, including singletons. Associated with each utility function u_s is an ordinal preference relation defined over permissible bundles of courses, $P_s \in \mathcal{P}$. We assume that the mapping from the set of students to the set of preference relations, $\mathcal{P}$, is measurable. We further assume that the utility functions are such that students' ordinal preferences over individual courses are strict, and that their ordinal preferences over bundles are responsive to their preferences for individual courses.[11]

Demand. Students are allowed to consume any bundle that consists of 0 or 1 seat of each course, and at most $m > 1$ courses in total. Let $r_s(c) \in \mathbb{N}$ denote course c's rank in student s's preferences over individual courses. Thus $r_s(c) < r_s(c')$ if and only if cP_sc', with $r_s(c) = 1$ if cP_sc' for all $c' \neq c$. This allows us to define the demand for individual courses:

Definition 1 (Demand for Courses). The demand for course c is defined as $D_c(\rho) = \frac{\int_0^1 1_{\{r_s(c) \leq \rho\}} ds}{q_c}$, $\rho = 1, ..., C$.

The allocation problem is non-trivial if at least one capacity constraint binds. Thus, in the rest

[8] Callousness is distinct from Bogomolnaia and Moulin (2001)'s critique of RSD in the context of single-unit assignment. See Section 8 for discussion.

[9] We use the terms "students" and "courses" because of our application. We could equally use the generic terms "agents" and "objects".

[10] The use of a continuum of students is a technical, rather than substantive, assumption. It simplifies proofs and helps clarify the key forces behind the results.

[11] Preferences are responsive if, for any student s, courses c, c', and bundle of courses X with $c, c' \notin X$ and $|X| < m$, $cP_sc' \iff (X \cup c)P_s(X \cup c')$. Also, $cP_s\emptyset \iff (X \cup c)P_s(X \cup \emptyset)$ (Brams and Straffin, 1979, Roth, 1985). While restrictive, survey evidence suggests that responsiveness is a reasonable assumption in our context. This is in part due to the explicit design of the HBS elective curriculum to avoid overlap or interdependence amongst courses. Also, it is important to note that the dictatorship theorems mentioned in the introduction still obtain when preferences are responsive; see Hatfield (2009).

of our analysis we assume that there exists at least one course c such that $D_c(m) > 1$.

Feasible Allocations. An allocation in this environment is an assignment of courses to students. We denote by $a_s \subset \mathcal{C}$ student s's allocation of courses. An allocation is feasible if $|a_s| \leq m$ for all s and $\int_0^1 1_{\{c \in a_s\}} ds \leq q_c$ for all c. We denote by $\mathcal{A}$ the set of feasible allocations. A random allocation is a probability distribution over feasible allocations. We denote by $L(\mathcal{A})$ the set of random allocations.

Ex-Ante and Ex-Post Efficiency. A random allocation is ex-ante Pareto efficient if there is no other random allocation that all students weakly prefer and a strictly positive measure of students strictly prefers. A feasible allocation is ex-post Pareto efficient if there is no other feasible allocation that all students weakly prefer and a strictly positive measure of students strictly prefers. Ex-ante efficiency is strictly stronger than ex-post, in the sense that a necessary but not sufficient condition for a random allocation to be ex-ante efficient is that every realization of this allocation is ex-post efficient.

Information. We assume that students' preferences are common knowledge. Since we are working with a continuum, this is equivalent to assuming that students' preferences are private information but that the distribution over preferences is common knowledge.

2.2 Allocation Mechanisms

We focus attention on two specific course-allocation mechanisms, the HBS mechanism and the Random Serial Dictatorship (RSD).[12] Both mechanisms are ordinal in the sense that they take as inputs ordinal information about students' preferences (Bogomolnaia and Moulin, 2001). In both mechanisms, each student s reports a rank-order list (ROL) $\widehat{P}_s$ indicating her ordinal preferences over individual courses. We write $\widehat{P}_s : c_1, c_2, c_3, ...$ to describe that student s puts course c_1 ahead of c_2, course c_2 ahead of c_3, and so on (with a slight abuse of notation, we will also write P_s : $c_1, c_2, c_3, ...$, to describe her true preferences over individual courses). Then, the mechanism uniform randomly selects a priority order over students, which is a bijection from the set of students onto itself. Let $\lambda(t)$ denote the student who has priority t, and $\lambda^{-1}(s)$ the priority of student s. The set of priority orders is $\mathcal{L}$. We assume that all elements of $\mathcal{L}$ have a measurable inverse.

Under RSD, students are allocated their courses all-at-once in ascending priority order. The algorithm has a single round that takes place from time $t = 0$ to time $t = 1$. At time t, student $\lambda(t)$ is allocated a seat in her m most-preferred courses on $\widehat{P}_s$ that still have remaining capacity.[13]

Under the HBS mechanism, students are allocated courses one-at-a-time over a series of m rounds. In odd rounds, which occur during time intervals $[0, 1], [2, 3], ...$, students are allocated

[12] Sönmez and Unver (forthcoming) and Budish (2009) describe other course-allocation mechanisms used in practice.

[13] A course has capacity remaining at time t if the measure of students allocated a seat in that course during time $[0, t)$ is strictly less than the course's capacity.

courses one-at-a-time in ascending priority order. In even rounds, which occur during time intervals $[1,2],[3,4],...$, students are allocated courses one-at-a-time in descending priority order. When it is student s's turn, she is allocated her most-preferred course on $\widehat{P}_s$ that (i) she has not already received in a previous round; and (ii) still has remaining capacity. Following the m rounds of the HBS mechanism, students have one additional opportunity to modify their schedule. Students can drop courses they obtained in the initial allocation and add courses that have excess capacity. This is conducted using a multi-pass algorithm that cycles over students (using a new random priority order) until no more add-drop requests can be satisfied. In particular, the algorithm satisfies a student's add-drop request only if the course that the student requests has spare capacity. It does not look for Pareto-improving trades amongst students. For modeling purposes, we model the add-drop phase as a random serial dictatorship where the only courses that can be requested are those with spare capacity at the end of the initial allocation. Students have the opportunity to modify their reported preferences, and a new random priority order is drawn. Thus, each student in turn creates the best possible schedule out of the courses they got in the initial allocation and those still with excess capacity.

2.3 Equilibrium

A Nash equilibrium in this setting is a measurable mapping from the set of students to the set of mixed strategies $\Delta\mathcal{P}$. There exists a pure strategy Nash equilibrium in both mechanisms:

1. RSD is dominant-strategy incentive compatible and we focus on this pure strategy equilibrium in the remainder.

2. The add-drop phase of the HBS mechanism is equivalent to RSD and thus we also restrict attention to the equilibrium where students report truthfully in this stage. Existence of a pure strategy Nash equilibrium is guaranteed in the initial allocation phase of the HBS mechanism because the action space is finite and students' expected utilities are continuous in the strategies of the other students and only depend on the fraction of students who report each preference profile (Schmeidler, 1973).

3 Properties of the HBS mechanism

The HBS mechanism satisfies several attractive efficiency and fairness properties if we ignore incentives and assume that students report their preferences truthfully. In terms of efficiency, it yields allocations that are ex-post Pareto possible in the sense of Brams et al. (2003). This means that there exist preferences over bundles of courses that are responsive to the reported preferences over individual courses, and for which the allocation is ex-post Pareto efficient. With respect to fairness, it is procedurally fair in the ex-ante sense of equal treatment of equals, and also in an interim sense, in that no students' set of choosing times dominates any others'. It also satisfies attractive criteria of outcome fairness, as described in Budish (2009).

These attractive properties explain the HBS administration's decision to adopt this mechanism, and may explain the widespread use of similar draft mechanisms in practice. However, as the following example illustrates, truthful play is not the expected outcome in the HBS mechanism.

Example 1 (Over-reporting and Congestion) Let $m = 2$ and suppose there are 4 courses with capacity of $\frac{2}{3}$ seats each. Preferences are as follows:

Proportion of Population	Type	Preferences
$\frac{1}{3}$	P_1	c_1, c_2, c_3, c_4
$\frac{1}{3}$	P_2	c_2, c_1, c_3, c_4
$\frac{1}{3}$	P_3	c_1, c_3, c_4, c_2

Truthful play is not a best response for the P_2 types. Indeed, suppose that student s is of type P_2. If all other students play truthfully, student s gets $\{c_2, c_3\}$. If, instead, he submits preferences $\widehat{P}_2 : c_1, c_2, c_3, c_4$, then he gets his first choice bundle $\{c_1, c_2\}$ for sure.

In fact, the P_1 and P_3 types reporting truthfully and the P_2 types reporting $\widehat{P}_2 : c_1, c_2, c_3, c_4$ is a Nash equilibrium, independently of risk attitudes or cardinal information about preferences. In this equilibrium, the P_1 and P_2 types get $\{c_1, c_2\}$ with probability $\frac{2}{3}$ and $\{c_2, c_3\}$ with probability $\frac{1}{3}$ and the P_3 types get $\{c_1, c_3\}$ with probability $\frac{2}{3}$ and $\{c_3, c_4\}$ with probability $\frac{1}{3}$. More students request c_1 (the most popular course based on true demand) in the first round than under truthful play, making c_1 fill up earlier in the round than under truthful play.

The story that Example 1 tells is that students in the HBS mechanism will have a tendency to overreport their preferences for popular courses, and that this causes those courses to reach capacity sooner. A P_2 type should not waste his first-round choice on c_2, since he can get it for sure in the second round, and if he waits until round two to ask for c_1 he is sure not to get it. Instead, he should attempt to obtain the popular c_1 in the first round.

In this section, we explore the incentives in the HBS mechanism further and provide a partial characterization of equilibrium outcomes. In particular, we argue that detecting profitable deviations from truthful behavior is very easy and that there will indeed be a tendency in equilibrium to overreport preferences for popular courses. This leads to congestion and earlier run-out times for courses. In the last subsection, we explore the welfare properties of the Nash equilibrium in the HBS mechanism. One consequence of strategic behavior is redistributive: some students, especially those whose top choices are not very popular, are better off; others, especially those whose top choices are very popular, are worse off. We also show that strategic behavior can lead to ex-post inefficiency and pin down conditions when this is not the case.

3.1 Incentives in the HBS mechanism

The basic trade-off that students face in the HBS mechanism is that upgrading a course relative to their truthful preferences increases the chance of getting that course, at the risk of missing another, possibly preferred, course. This trade-off is intrinsic to the HBS mechanism and does not depend on market size. Indeed, Example 1 would work just as well with three students.

To be able to say something about incentives in the HBS mechanism without making any assumption on risk attitudes, we analyze best responses for a fixed strategy profile $\widehat{\mathbf{P}} = \{\widehat{P}_s\}_{s\in[0,1]}$ and a fixed priority order λ. The outcome, then, is determinate (and, in particular, courses are characterized by their run-out times). A property of best responses will hold irrespective of risk attitudes, if we can show that it holds for all λ.

Consider student s and relabel courses such that student s's strategy during the initial allocation phase of the mechanism reads $\widehat{P}_s : c_1, c_2, ..., c_C$. Because of the continuum assumption, his strategy does not affect run-out times and thus does not affect the outcome and timing of requests by the other students. The only thing it does affect is whether student s gets a seat in the courses he requests. He will do so if he requests a particular course before that course runs out. He will not do so otherwise.

Our workhorse for most of the proofs in this section is the comparison between two strategies by student s that differ from one another by the position of a single course. We show that allocations for student s under those two strategies differ at most by one course.

For example, denote by $\widehat{P}_s^{c_k\downarrow l}$, the strategy that corresponds to $\widehat{P}_s$ except that course c_k has been downgraded to position l in the ROL, i.e. $\widehat{P}_s^{c_k\downarrow l} : c_1, ..., c_{k-1}, c_{k+1}, ..., c_l, c_k, c_{l+1}, ...$ Until the procedure reaches the k^{th} position in student s's ROL, the timing of requests and the outcomes are identical under both strategy profiles. From then on, $\widehat{P}_s^{c_k\downarrow l}$ asks for c_{k+1} one round earlier than $\widehat{P}_s$, and if either both strategies get c_{k+1} or both do not, then $\widehat{P}_s^{c_k\downarrow l}$ asks for c_{k+2} one round earlier than does $\widehat{P}_s$, and so on. If, before $\widehat{P}_s^{c_k\downarrow l}$ requests c_k, it gets a course that $\widehat{P}_s$ does not, then the two strategies get back in synch, requesting (and getting) the same courses at the same times. They can get back out of synch if $\widehat{P}_s^{c_k\downarrow l}$ turns out also to get c_k, only now $\widehat{P}_s$ is making requests for c_{l+1} one round earlier than does $\widehat{P}_s^{c_k\downarrow l}$. Clearly, the exact details on the timing and outcome for student s depend on the availability of courses when he requests them (they are developed in the proofs). The key aspect to note here is that requests by student s under both strategies are either in synch or out of synch by a maximum of one round. When requests are in synch, he must get the same outcome under both strategies. When requests are not in synch, he may get a course under one strategy that he does not get under the other strategy. Because requests become in synch after such event, the difference in outcomes between the two strategies is one course at most.

Let $a_s(\widehat{\mathbf{P}}, \lambda) \subset \mathcal{C}$ indicate student s's final allocation, including what happens in the add-drop phase, when students use the strategy profile $\widehat{\mathbf{P}}$ and λ is the realized priority order. Let $a_s(\widehat{\mathbf{P}})$ refer to student s's final (random) allocation under profile $\widehat{\mathbf{P}}$.

Definition 2 (Popularity): Course c is $\widehat{\mathbf{P}}$-popular if there exists a positive measure of students for whom $\Pr(c \notin a_s(\widehat{\mathbf{P}})$ and $c' \in a_s(\widehat{\mathbf{P}})) > 0$ for some c' such that $c\widehat{P}_s c'$. Course c is $\widehat{\mathbf{P}}$-unpopular otherwise.

In words: the popularity of a course is defined relative to a strategy profile. A $\widehat{\mathbf{P}}$-popular course runs out with positive probability under strategy profile $\widehat{\mathbf{P}}$. A $\widehat{\mathbf{P}}$-unpopular course never runs out. Note that, by the continuum assumption, $\widehat{\mathbf{P}}$-popular courses are also $(\widehat{P}'_s, \widehat{\mathbf{P}}_{-s})$-popular for any $\widehat{P}'_s$, so that we can as well talk about $\widehat{\mathbf{P}}_{-s}$-popular courses.

The next result shows that it is easy for students to find profitable deviations from truthful play: the deviations that Theorem 1 identifies only require that students know which courses are likely to run out, and which aren't.

Theorem 1 (Simple Manipulations): Fix $\widehat{\mathbf{P}}_{-s}$. Form the strategy $\widehat{P}_s^{\text{simple}}$ by taking the first m courses in P_s and rearranging them so that $c\widehat{P}_s^{\text{simple}}c'$ whenever:

(i) cP_sc' and both are $\widehat{\mathbf{P}}_{-s}-$popular or both are $\widehat{\mathbf{P}}_{-s}-$unpopular

(ii) c is $\widehat{\mathbf{P}}_{-s}-$popular and c' is $\widehat{\mathbf{P}}_{-s}$-unpopular

The strategy $\widehat{P}_s^{\text{simple}}$ generates weakly greater utility than truthful play P_s.

All proofs are in Appendix A.

Clearly, this simple strategy leads to overreporting of preferences for popular courses. Moreover, if everyone adopts this simple strategy starting from truthful play, these courses are likely to run out earlier than under truthful play. This is in line with Example 1. However, Theorem 1 is only indicative that this may be a general feature of equilibrium. Indeed, it is not yet an equilibrium characterization, nor even a characterization of best responses. We turn to this question next.

3.2 Equilibrium

We can identify two environments where truthful play in the HBS mechanism is always a Nash equilibrium.

Theorem 2 (Truthful Play in Special Cases): Consider the following two environments:

(i) Identical preferences: $P_s = P_{s'}$ for all s, s'

(ii) Independent preferences: for any two $\mathbf{P}$-popular courses c, c', $D_c(\rho) = D_{c'}(\rho)$ for $\rho = 1, ..., C$, and all students prefer every $\mathbf{P}$-popular course to every $\mathbf{P}$-unpopular course

In either environment, $\widehat{\mathbf{P}}^* = \mathbf{P}$ is a Nash equilibrium of the HBS mechanism.

The intuition is the following. When students have identical preferences, if a student prefers one course to another, then so do all other students and thus it is not possible to *over*report one's preferences for more popular courses. Likewise, when all courses for which demand exceeds supply are equally popular, there is no basis for misreporting. Theorem 2 indicates that *partial* correlation of preferences is what drives strategic misreporting in the HBS mechanism.

We next provide a necessary condition for best responses in the HBS mechanism.

Lemma 1 (Best-response Characterization): Consider any equilibrium strategy profile $\widehat{\mathbf{P}}$. Suppose c is $\widehat{\mathbf{P}}-$popular, and that $r_s(c) \leq m$. Then, one of the three following properties must hold:

(i) c appears before all $\widehat{\mathbf{P}}-$unpopular courses in $\widehat{P}_s$,

(ii) $\Pr(c \in a_s(\widehat{\mathbf{P}})) = 1$, or

(iii) $\Pr\left(c \in a_s\left(\widehat{P}_s^{c\uparrow k}, \widehat{\mathbf{P}}_{-s}\right)\right) = 0$ where k is the position of the first (highest) $\widehat{\mathbf{P}}-$unpopular course on $\widehat{P}_s$.

It is interesting to compare and contrast Lemma 1 and Theorem 1. Like Theorem 1, Lemma 1 suggests that popular courses will be ahead of unpopular courses in ROLs. There are two exceptions. A first exception is when student s is sure to get the popular course even if he requests it later. A second exception is when student s would not get the course even if he placed it before all unpopular courses. In that case, we cannot rule out that, at equilibrium, a student puts that course in another irrelevant position, that is, after some unpopular courses.

Another difference between Theorem 1 and Lemma 1 is that Lemma 1 is silent about the relative ordering of popular courses in students' ROLs. How students should optimally rank popular courses depends on the relative run-out times of these courses under $\widehat{\mathbf{P}}$, on the cardinal utilities they attach to these courses and on their attitudes towards risk. We will come back to this point later in the section. In the meantime, we state the properties of the equilibrium in the HBS mechanism that we can establish, just based on ordinal preferences.

Theorem 3 (Equilibrium Characterization): Suppose that $\widehat{\mathbf{P}}$ is a Nash equilibrium, and that $D_c(m) > 1$ for some c. Then,

(i) c runs out with probability one

(ii) the supremum of run-out times for course c over the different realizations of λ, $\bar{t}_c$, is weakly less than the number of $\widehat{\mathbf{P}}-$popular courses.

Theorem 3 says that courses for which demand exceeds supply based on truthful reports will run out in any equilibirum. Moreover, such courses will reach capacity in the first k rounds, where k is the number of $\widehat{\mathbf{P}}$-popular courses.

However, Theorem 3 remains silent concerning the relative run-out times of popular courses. The reason is that equilibrium in the HBS mechanism has a coordination feature: if a course is expected to be ranked high by others then students will tend to rank it higher too. This yields equilibrium multiplicity as the next example illustrates.

Example 2 (Multiple Equilibria) Let $m = 2$. Courses have 0.4 capacity. Courses c_1, c_2, c_3 have excess demand, all other courses do not. Students' preferences are as follows (where "other" stands for courses other than c_1, c_2, c_3) :

Proportion of Population	Type	Preferences
.25	P_1	c_1, c_2, other
.25	P_2	c_2, c_1, other
.30	P_3	c_3, other
.10	P_4	c_3, c_1, other
.10	P_5	c_3, c_2, other

Truthful play is always an equilibrium. In round 1, the P_1 and P_2 types get their first choice and the P_3, P_4 and P_5 types get their first choice with probability 0.8, else they get their second choices. The remaining capacity for courses c_1 and c_2 at the beginning of round 2 depends on the priority order over students in round 1. In expectation, these courses have 0.13 remaining capacity $(0.40 - 0.25 - 0.02)$. The expected demand for these courses in round 2 is equal to 0.33 $(0.25 + 0.08)$ so a request is satisfied with probability $0.13/0.33 \approx 0.4$.[14] Thus, the P_1 types get their best bundle $\{c_1, c_2\}$ with probability 0.4 and otherwise they get their second best bundle $\{c_1, \text{other}\}$ (the outcome for the P_2 types is symmetric). The P_3 types get their best bundle, $\{c_3, \text{other}\}$, with probability 0.8, and their second best bundle otherwise. The P_4 types get their best bundle, $\{c_3, c_1\}$, with probability 0.32, their second best bundle, $\{c_3, \text{other}\}$, with probability 0.48 and $\{c_1, \text{other}\}$ otherwise (the outcome for the P_5 types is symmetric). Clearly, no type has a profitable deviation.

If the P_4 and P_5 types' intensity of preference for c_3 versus c_1 and c_2, respectively, is not too large, there exists another equilibrium in which types P_1, P_2 and P_3 play truthfully, the P_4 types submit $\widehat{P}_4 : c_1, c_3$, other, and the P_5 types submit $\widehat{P}_5 : c_2, c_3$, other. In this equilibrium, the P_4 types get their best bundle, $\{c_1, c_3\}$ with probability 0.5 and $\{c_1, \text{other}\}$ otherwise. If they deviate from this equilibrium to play truthfully (the only deviation to consider given their preferences), they receive their best bundle with probability 0.2 (c_1 is available in round 2 with probability $\frac{.4-.1-.25}{.25}$) and get $\{c_3, \text{other}\}$ with probability 0.8. This is less preferable if the intensity of preference for c_3 is not too large. The analysis for the P_5 types is analogous.

In Example 2, course c_3 runs out in equilibrium in round 1 or 2, i.e., at the same time or later than under truthful play, even though more students have c_3 as their top choice relative to other

[14] Recall that a property of expected utilities is that preferences over compound lotteries correspond to the preferences over the reduced lottery. Thus it is sufficient to look at expected probabilities of outcomes.

courses. This is consistent with Theorem 3, which says that c_1, c_2 and c_3 should all run out with probability 1, but otherwise does not pin down the relative timing of these events.

To go beyond Theorem 3, we make additional assumptions on the relationship between students' ordinal preferences over courses and their utility functions. In Example 2, c_3 sells out at equilibrium in the first round if the P_4 and P_5 types students prefer the lottery $[0.2 : \{c_1, c_3\}; 0.8 : \{c_3, \text{other}\}]$ to the lottery $[0.5 : \{c_1, c_3\}; 0.5 : \{c_1, \text{other}\}]$. This motivates the following restriction:

Definition 3 (Lexicographic Preferences): Consider two lotteries over final allocations, L_1 and $L_2 \in L(\mathcal{A})$. Fix an arbitrary s, and label courses so that $P_s : c_1, c_2, ..., c_C$. Let $p_1(c)$, $p_2(c)$ denote the probability of getting c under lottery L_1 and L_2 respectively. We say that student s has lexicographic preferences if he prefers L_1 to L_2 whenever there exists any $k \in \mathbb{N}$ such that $p_1(c_i) \geq p_2(c_i)$ for all $i = 1, 2, ..., k$ with at least one strict inequality.

In Example 2, if the P_4 and P_5 types have lexicographic preferences, they prefer any lottery in which they obtain c_3 for sure to any lottery in which they don't, and so truthful play is the unique equilibrium. This yields sharper predictions about the structure of students' best responses in the HBS mechanism.

Lemma 2 (Best-response Characterization with Lexicographic Preferences): Suppose students have lexicographic preferences and consider any equilibrium strategy profile $\widehat{\mathbf{P}}$. Suppose c is $\widehat{\mathbf{P}}-$popular, and $r_s(c) \leq m$. Then, one of the three properties must hold:

(i) c appears before all $\widehat{\mathbf{P}}-$unpopular courses in $\widehat{P}_s$ and, for all c', $cP_sc' \Rightarrow c\widehat{P}_sc'$

(ii) $\Pr(c \in a_s(\widehat{\mathbf{P}})) = 1$, or

(iii) $\Pr\left(c \in a_s\left(\widehat{P}_s^{\text{simple}}, \widehat{\mathbf{P}}_{-s}\right)\right) = 0.$

In words, Lemma 2 says that students only downgrade a course when they are sure to get it (or when they cannot hope to get it anyways). Theorem 4 provides a tighter characterization of equilibrium run-out times:

Theorem 4 (Equilibrium with Lexicographic Preferences): Suppose students have lexicographic preferences. Suppose further that $\widehat{\mathbf{P}}$ is a Nash equilibrium and that $D_c(m) > 1$. Then:

(i) $\bar{t}_c < \rho_c \equiv \inf\{\rho : D_c(\rho) > 1\} \leq m$

(ii) $\arg\min_r\{\widehat{D}_c(r) \geq 1\} \leq \arg\min_r\{D_c(r) \geq 1\}$ where $\widehat{D}_c(r)$ corresponds to the reported demand under $\widehat{\mathbf{P}}$.

Theorem 4 provides an upper bound on the times by which courses run out, based on their true demand, and a prediction on the relationship between truthful demand and reported demand at equilibrium. We revisit these predictions in section 5.

3.3 Welfare

Strategic behavior in the HBS mechanism has redistributive consequences. On the one hand, it helps students who, by overreporting their preferences for popular courses, increase their chances of getting them. On the other hand, it creates congestion for these courses and hurts students who value these courses highly. This can be seen in Example 1 where the P_1 and P_3 types are better off ex-post and ex-ante from truthful play, and the P_2 types are better off under strategic play.[15]

Strategic behavior also has efficiency consequences because of the intrinsic trade-off that students face between upgrading a less preferred but very popular course at the risk of potentially missing a seat in a preferred but less popular course. The next example illustrates this ex-post inefficiency of equilibrium in the HBS mechanism.

Example 3 (Ex-post Inefficiency of Strategic Play) Let $m = 2$. Courses c_1, c_2 have excess demand with respective capacity 0.6 and 0.8. All other courses do not. Suppose preferences are as follows (where "other" stands for courses other than c_1 or c_2)

Proportion of Population	Type	Preferences
0.3	P_1	c_1, c_2, other
0.4	P_2	c_2, c_1, other
0.3	P_3	c_2, other

Consider the strategy profile where all students of type P_1 play $\widehat{P}_1 : c_2, c_1$, other, and the P_2 and P_3 types submit truthful ROLs. Under this strategy profile, all students request c_2 in round 1, which means that only the first 0.8 are successful. Those who do not get their first choice get their second choice in round 1. At the beginning of round 2, 0.46 seats $(0.6 - (0.7)(0.2))$ remain in c_1 in expectation, whereas 0.56 students request it. This means 82% of these students are successful. Thus, the P_1 and P_2 types face the following lottery:

$$[0.66 : \{c_1,\ c_2\}; 0.20 : \{c_1,\ \text{other}\}; 0.14 : \{c_2,\ \text{other}\}] \tag{1}$$

To check whether this strategy profile is an equilibrium, we only need to look at the opportunity for a P_1 student to deviate and submit his truthful preferences. If he does, he gets the deterministic outcome $\{c_1, \text{other}\}$. Thus $\widehat{P}_1, P_2, P_3$ is an equilibrium if all P_1 students prefer the lottery in (1) over the deterministic outcome $\{c_1, \text{other}\}$. This equilibrium is ex-post inefficient because it is possible that a P_1 student ends up with $\{c_2, \text{other}\}$ and that a P_2 student ends up with $\{c_1, \text{other}\}$. Those students would prefer to trade.

[15] The P_1 and P_3 types got their best bundle under truthful play and now only get it with probability 1/3. The P_2 types who never got their top bundle now get it with probability 1/3, and otherwise they get what they used to get for sure under truthful play.

Because the ex-post inefficiency of the HBS mechanism is due to risk-taking behavior by students who deviate from truthful behavior, we can expect that no such inefficiency occurs in settings where, at equilibrium, students do not deviate from truthful behavior or when risk aversion is so high or preferences so extreme that they do not take any risk. Theorem 5 confirms this.

Theorem 5 (Ex-post Efficiency in Special Cases) (i) Nash equilibria in truthful strategies are always ex-post efficient possible; (ii) All Nash equilibria are ex-post efficient possible when preferences are lexicographic.

By contrast, ex-post inefficiencies are likely when preferences are correlated, students are not very risk averse and cardinal utilities attached to each course are not very different.

Not surprisingly, the HBS mechanism is also ex-ante inefficient, independently of whether it is ex-post inefficient or not, for the very same reasons why RSD if inefficient when students require a single course (Bogomolnaia and Moulin, 2001). More interestingly, the next example shows that strategic behavior can hurt all students ex-ante, relative to truthful behavior. In other words, strategic behavior has consequences beyond redistribution. We will revisit this issue in section 6.

Example 3 (cont'd) Strategic Behavior May Hurt All Students Ex-Ante Consider again Example 3. The next table compares the lotteries that students face under truthful behavior and under the equilibrium identified in Example 3 above.

Type	Lottery under truthful play	Lottery under strategic behavior
P_1	$[0.33 : \{c_1, c_2\}; 0.67 : \{c_1, \text{other}\}]$	$[0.66 : \{c_1, c_2\}; 0.20 : \{c_1, \text{other}\}; 0.14 : \{c_2, \text{other}\}]$
P_2	$[0.75 : \{c_1, c_2\}; 0.25 : \{c_2, \text{other}\}]$	$[0.66 : \{c_1, c_2\}; 0.14 : \{c_2, \text{other}\}; 0.20 : \{c_1, \text{other}\}]$
P_3	$[1 : \{c_2, \text{other}\}]$	$[0.80 : \{c_2, \text{other}\}; 0.20 : \{\text{other}, \text{other}\}]$

Clearly, the P_2 and P_3 types are worse off under strategic behavior, independently of cardinal information about their preferences. The P_1 types are worse off if, for example $u_s(\{c_1, c_2\}) = 4$, $u_s(\{c_1, \text{other}\}) = 3.4$ and $u_s(\{c_2, \text{other}\}) = 0.8$. (These numbers also ensure that strategic behavior is indeed an equilibrium.)

4 Description of Data

Our dataset covers the allocation of second-year elective courses to second-year MBA students at Harvard Business School during the 2005-2006 academic year. Students choose 10 elective courses each, five for each semester (Fall 2005, Winter 2006). Courses for both semesters are allocated together in a single allocation process. The HBS mechanism was described formally in Section 2.2.

4.1 Timing of actions and information

Students are asked to report their ranking over individual courses at three separate times: in early May, in mid-May and in mid-July. Prior to this, students have information on the past enrollment of each course and they receive both official and unofficial course evaluation information.

In early May, they are asked to participate in a survey in which they rank their top five favorite courses. Participation is voluntary. The results are used to aggregate information about demand and adjust some course capacities.[16] The students have access to the full results, except for the student identities which are removed.

In mid-May, students participate in a trial run of the allocation mechanism. Participation is compulsory and students can rank up to 30 courses (rankings can be section-specific for courses offered in different sections). The administration reports the resulting course enrollments based on one single run of the algorithm. For courses at capacity, students are told how many times a course was overenrolled based on the submitted preferences. In addition, the administration reports the 10 courses most often ranked first in the submitted rank order lists (ROLs) with the number of times each was ranked first. Students do not receive any feedback on their individual assignment of courses from the trial run.

Finally, students submit their ROLs for the real run of the mechanism in mid-July. Some changes are possible at the beginning of each semester during an add-drop phase, as described in Section 2.2.

4.2 Course Characteristics

Our data contain all course characteristics, including section, capacity, term and scheduling information as they were available at the time of the May trial run and the July run of the algorithm. Seats in 71 courses and 21 half-courses (147 sections) were offered in May for a total capacity of 11,871 seats. Course capacities ranged from 12 to 404 students. The numbers for July were 67 courses and 22 half-courses (141 sections) for a total of 10,898 seats.[17] The capacities range was the same. A total of 9,269 seats were allocated in the July 2005 run of the algorithm.[18]

[16] The exact text is the following: "This poll has been set up to gauge current interest in 2005-06 courses. Be sure you enter your top 5 course selections for the coming year, with #1 the course you want most. Your selections will be anonymous to others. As a participant, you'll be able to view anonymous results on Friday, May 6."

[17] Two courses were added between the May poll and the trial run. Between the trial run and the July run, one course was added, four courses were cancelled, one full semester course was changed into a half course, and several courses had their capacities increased or decreased slightly.

[18] The reason this sums to a bit more than 10 courses per student is the half courses.

4.3 Submitted Preferences

Our data contain students' submitted ordinal preferences over individual courses (with student identifiers) in the May poll, May trial run and the July run of the algorithm. In addition, we conducted an auxiliary survey in January 2006 in which we asked students to rank their 30 favorite courses. The poll was conducted after the add-drop phase for the second semester but before courses started. In the poll, we explicitly asked students to rank the courses according to their true preferences, independently of whether they got the course or not. The stated objective of the poll was to collect data on preferences to investigate potential improvements to the HBS allocation mechanism.[19]

Table 1: Descriptive statistics − submitted preferences

	May poll	May trial run	July run	Jan poll
# students	460	922	916	163
avg # courses per ROL	5	22.33	21.96	17.46
std dev # courses per ROL	0	5.13	4.86	7.31
# courses listed at least once	85	92	89	92

Table 1 summarizes the number of students and courses covered by the data on each occasion. Because participation was compulsory, the May trial run data and the July run data cover the entire population. The small discrepancy in numbers is due to students leaving for or returning from military duty, maternity leave or any other leave of absence. The table also reports the number of courses ranked by the students. For the May trial run and the July run, the submitted rank order lists can be section-specific. When constructing a student's rank order list over courses, we kept the first time a course appeared in the original rank order list.[20]

[19] The exact text was the following: "Please use the following pull down menus to rank your top 30 most preferred EC courses for 2005-2006, irrespective of whether you were assigned the course or not. Courses are not section-specific. If you have fewer than 30 courses that you would like to rank, please select "Finished Ranking Courses" from the pull-down menu and move on to Question #2. It is critical that the ranking you submit completely reflects your preferences. In particular, do not feel the need to rank courses that fill up quickly first. Alternatively, do not ignore courses just because you perceived that they would be difficult to get. You should rank the courses according to how you actually feel about them." The interface was identical to the interface used for the May poll.

[20] This convention affects very few observations. Out of the 20,279 student-course observations in the July run, 14,296 observations are for courses that have multiple sections but for most of them the student listed the different sections of the course in consecutive order. Requests for different sections of the same course were non consecutive for only 282 student-course observations (2%). In our robustness checks we considered alternative conventions for treating those non-consecutive requests.

5 Evidence of Strategic Behavior

In this section, we provide evidence that students understand the strategic incentives of the HBS mechanism and act accordingly. Our analysis lends support to the joint hypothesis that we use in the rest of the paper according to which: (1) students' May poll responses represent their truthful preferences, and (2) students' July run preferences represent equilibrium behavior.

This joint hypothesis is natural given the context of our data. In the May poll, students were explicitly asked by the administration to state their true preferences, and we see no compelling reason for them to disobey this request. Submitted preferences in the July run are those used for the initial allocation of courses. We have argued in section 3 that profitable deviations from truthful behavior are easy to detect in the HBS mechanism: they only require the kind of information that the HBS administration provides to the students through the May trial run. In a high-stakes environment with sophisticated and informed players that have had some opportunity to learn, equilibrium play is a natural hypothesis.

However, we may worry about two issues. First, preferences may change over time and May poll preferences may no longer represent preferences in July. Second, students may fail to play a best response despite the information and learning opportunities they have had.[21] To address these concerns, we show that May poll preferences and July run ROLs differ systematically, in a way that is by and large consistent with our joint hypothesis, and inconsistent with alternative accounts.

5.1 Evidence based on aggregate data

Four distinct reasons may explain why a student submits different preferences for courses at different points in time: idiosyncratic preference change, new information, social learning and strategic consideration. By definition, idiosyncratic preference changes should not affect aggregate demand for individual courses over time.[22] This leaves three potential drivers of changes in the aggregate demand for individuals courses over time:

1. **New information about courses** (whether good or bad) should lead to a correlated and persistent shock to preferences and thus a persistent shock to the aggregate demand for these courses. These shocks should not, however, be related to the popularity of these courses.

[21] If we used the May trial-run preferences instead of the July actual-run preferences, the first issue would be less of a concern because less time would have elapsed since the May poll, and the second issue would be more of a concern, since the May trial run was an opportunity for students to learn about best responses. We focus on the July run since that is what actually mattered for students' welfare. For robustness, we performed all of the analyses of Sections 5-7 using the May trial-run data instead; the results change very little. See Appendix B for this and other robustness checks.

[22] An additional benefit of using aggregate demand is that it also removes any small randomness in submitted preferences due to students' "carelessness" or near indifference.

While very little observable information was provided to students between May and July, we cannot a priori rule out the possibility of unobservable new information.

2. **Social learning**, i.e. the updating of one's own valuation for a course, through word-of-mouth or the observation of its popularity among students (for example, through the feedback from the May poll or the trial run), should lead to correlated and persistent shocks to individual preferences and thus to aggregate demand. These shocks may be correlated to the initial popularity of the course.

3. **Strategic behavior** should lead to a correlated but temporary shock to submitted preferences and thus to aggregrate demand during the initial allocation. As shown in Section 3, this effect will be related to the popularity of courses.

Our working hypothesis implies that strategic behavior, not social learning or new information, drives the aggregate differences between the May poll and July run preferences. To distinguish strategic behavior from the effect of new information or social learning, we compare the aggregate demands for courses at three different points in time: May poll, July run and January poll. Consider course c and a sample of students. Course c's distribution of ranks in that sample, $D_c(r)$, is the proportion of students in the sample that rank course c on or before r. To test for the equality or inequality of aggregate demand for a course across time, we use Gehan (1965)'s extension of the Wilcoxon rank-based test for discrete and censored data (censoring in our data arises from the fact that students only rank five courses in the May poll and that some students rank less than 30 courses in the July run).[23]

Table 2 reports the results of the Gehan test for the 83 courses that appear in both the May poll and the July run, at the 5% significance level. The theory suggests that students will overrreport courses they believe will be popular in equilibrium, and vice versa. As a proxy for students' beliefs about equilibrium popularity, we use the popularity information provided by the administration

[23] The idea behind this non parametric test is the following. Fix a course, say course c, and consider two independent samples of students of size n_1 and n_2. An observation is a student's rank for course c or, if the student did not rank that course, the rank of the last course she ranked, which will be taken as the censoring point for that observation (in words, we do not know how where student ranks course c but we know that it must be below this censoring point). Pair every observation in the first sample with each observation in the second sample. This creates $n_1 n_2$ pairs. To each pair, we assign a value of -1 if the observation in sample 1 is definitely before the observation in sample 2 (this will be the case if the rank in sample 1 is smaller than the rank in sample 2 or if the observation in sample 2 is censored and the censoring point is higher than the observation in sample 1). Similarly, we assign a value of +1 if the observation in sample 2 is definitely before the observation in sample 1. We assign a value of zero to the pair otherwise. Gehan (1965) shows that the resulting sum over each pair is distributed according to a normal distribution which can be used to test the null hypothesis that $D_c^1(r) = D_c^2(r)$ for $r \leq R$ where D_c^1 and D_c^2 are the distributions of ranks in sample 1 and sample 2 respectively and $R = \min\{$highest censoring point in sample 1, highest censoring point in sample 2$\}$.

after the May trial run (see section 4.1). Specifically, the administration reports the enrollment in each course based on a single run of the algorithm. For courses at capacity, they also report very explicity by how much the course was oversubscribed: courses at capacity are marked in bold, with n asterisks added if the course was n times oversubscribed. In the Table, we categorize courses not at capacity as "unpopular" and courses at capacity according to their level of oversubscription.[24]

Table 2: Comparison between May poll demand and July run demand

	N	July demand lower	No difference	July demand higher
Unpopular	47	26	21	0
Popular, 0×oversubscribed	16	7	9	0
Popular, 1×oversubscribed	8	1	7	0
Popular, 2×oversubscribed	6	0	4	2
Popular, 3×oversubscribed	3	0	0	3
Popular, 4×oversubscribed	1	0	0	1
Popular, 5×oversubscribed	2	0	1	1

Table 2 shows that the null hypothesis of unchanged demand between the May poll and the July run is rejected for 41 out of the 83 courses (49%), i.e. submitted preferences in the May poll and in the July run differ significantly. For courses that are multiple times oversubscribed, the reason for rejection is always because demand is higher in July. For courses that are unpopular or only somewhat popular, rejection occurs because demand is lower in July. This pattern is consistent with strategic behavior. It is also consistent with social learning.

To distinguish between strategic behavior and social learning, we use the same test to check for differences in aggregate demand between the May poll and the January poll. Despite the fact that 8 months elapsed between the May poll and the January poll,[25] we are only able to reject the null hypothesis of unchanged demand for 16% of courses, as compared with 49% for the comparison between the May poll and the July run (Table 2). This suggests that most of the observed change in submitted preferences in July is due to short-term strategic considerations rather than long-lasting social learning.

Additional evidence comes from the pattern of rejections in Table 3. Unlike in Table 2, there is no systematic relationship between the reason for rejection and the level of popularity.[26]

[24] We also considered alternative definitions of popularity based on the May poll demand and the July actual run. In both cases, the results in Tables 2 and 3 are essentially unchanged.

[25] During this time period students took Fall 2005 classes, and their experience of these courses may have affected their preferences. Table 3 would look similar if we focused on just the Winter 2006 courses.

[26] To give a rough sense of magnitudes, the two 5×oversubscribed courses constituted 38.6% of first choices in the May poll, and 42.2% of first choices in the January poll, which is suggestive of long-lasting social learning. But they represented 50.0% of first choices in the July run, which is suggestive of short-run strategic behavior.

Table 3: Comparison between May poll demand and January poll demand

	N	January demand lower	No difference	January demand higher
Unpopular	47	3	43	1
Popular, 0×oversubscribed	16	2	14	0
Popular, 1×oversubscribed	8	2	6	0
Popular, 2×oversubscribed	6	0	5	1
Popular, 3×oversubscribed	3	1	1	1
Popular, 4×oversubscribed	1	0	0	1
Popular, 5×oversubscribed	2	0	1	1

As a final piece of evidence for strategic behavior by students, we check whether the predictions of Theorems 3 and 4 are borne out in our data. Theorem 3 predicts that in any equilibrium, those courses for which truthful demand exceeds supply will reach capacity during the initial allocation. We use the May poll preferences to construct the set of such courses. Because only 456 students filled in the poll we scale course capacities accordingly. A conservative estimate is that any course whose demand restricted to the top 5 ranks exceeds adjusted capacity should belong to the set of courses that run out at equilibrium. Six courses satisfy this definition, and they all run out during the initial allocation. As an alternative we considered any course whose demand in the poll exceeded 50% of adjusted capacity; all 16 such courses run out during the initial allocation.

Theorems 3 and 4 also provide predictions on equilibrium run-out times. Because $m = 10$ and 44 courses run out at equilibrium, Theorem 3 is automatically satisfied in our data. Theorem 4 has more bite. For each course for which $D_c(5) > 1$ based on the poll data and adjusted capacities, we check whether this course always runs out before the round at which true demand exceeds supply. All six such courses satisfy this stronger test.

5.2 Evidence based on individual data

We now turn to individual data. We first run the HBS mechanism using the July run submitted preferences and 10,000 randomly drawn priority orders over students to identify popular courses from unpopular courses using Definition 2. We then focus on the 456 students who submitted their preferences for the May poll and in July. Out of the 2,280 courses that appear as their 5 most preferred courses according to the poll, 1,744 are popular.

For each of these courses, we check whether Lemma 1's necessary conditions for a best response are satisfied in the July run preferences submitted by these students. Specifically, if a popular course is after an unpopular course on a student's ROL, we check whether this student gets it for sure (condition (ii)), and if not, whether moving it up to the position of the first unpopular course on his ROL would secure a positive probability of getting it (condition (iii)).[27] Out of the

[27] For the purpose of the test, we did not consider a multiple sections course as unpopular if one of its sections was

1,744 popular course entries, 97 (5.6%) violate the necessary conditions for a best response. These violations involve 81 students (17.8%).

Violations of Lemma 1 mean that the relationship between the May poll preferences and the July run preferences does not correspond to the predicted relationship between a student's true preferences and his reported preferences in the equilibrium of the HBS mechanism. In our data, such discrepancy can be explained by genuine preference changes between May and July so that the submitted preferences in the May poll are actually no longer the true preferences of the students in July, or by strategic mistakes during the play of the HBS mechanism. We investigate each hypothesis in more detail.

For every popular course that violates Lemma 1, we first look at its position in the student's preferences in May and in July. All Lemma 1 violations are caused by courses that were strictly downgraded relative to the student's reported preferences in May. Out of the 97 violations, 59 correspond to a course that appears in a student's top 5 courses in May and no longer appears in the submitted preferences in July. These involve 49 students. A likely interpretation of these violations is that those students changed their preferences between May and July.

Another 20 violations can be explained by courses that appeared to be unpopular based on the trial run but happened to be popular in the real run of the mechanism.[28] Students slightly downgraded these courses, thinking they were "safe", and failed to get some of them as a result. Likewise, 10 violations can be explained by courses that seemed to be popular based on the trial run but ended up not being popular in July. Some students needlessly placed these courses ahead of other popular courses. In total, these (slight) strategic mistakes explain 30 violations. They involve 28 students.

Twenty two violations are not explained by either obvious preference changes or strategic mistakes of the kind just described, involving 22 distinct students. Ten of these mistakes involve the 1×oversubscribed course that is the outlier in Table 2; we suspect that students received unobservable information about this course that caused a meaningful number of them to downgrade their preference for it between May and July. There is no discernible pattern to the remaining Lemma 1 violations.

To conclude, the analysis of the individual data confirms the overall picture from the aggregate data. 82.2% of students submit ROLs that are consistent with the joint hypothesis that May poll preferences correspond to their true preferences in July and that they play a best response in July; 94.4% of popular-course requests are consistent. About 60% of violations can be traced to likely

filled. This choice affects nine requests.

[28] Recall that students receive information on course enrollment and oversubscriptions based on one single trial run of the algorithm in May. If we define courses as popular based on the trial run if all their sections reach their capacity (courses are unpopular otherwise), there are five courses that are popular based on the trial run but are no longer popular in the July run, and nine courses that are unpopular based on the trial run and are popular in July.

preference changes and about a third can be traced to slight strategic mistakes due to incorrect beliefs about the popularity of courses.

A robustness check reported in Appendix B computes the main welfare results of Sections 6 and 7 for a sub-economy consisting of the 82.2% of students who are Lemma 1 compliers. The results are substantially similar.

5.3 Constructed Truthful Preferences

So far, we have argued that the May poll preferences can be taken as an approximation of truthful preferences in July. However, May poll preferences have two limitations for the analysis of the welfare properties of the HBS mechanism. First, they are restricted to students' top five courses. Second, they are available for only 456 students out of the 916 students who participated in the July run.

To address the first limitation, we construct an extension of students' truthful preferences as follows. We assume that students' truthful top five courses correspond to their top five courses in the May poll.[29] Other courses are moved down to position six and below in a way that preserves their relative ranking in the July ROLs. We call the result "constructed truthful preferences." To illustrate, suppose a student submitted the ROL c_1, c_2, c_3, c_4, c_5 in the May poll but submitted $c_4, c_3, c_6, c_1, c_2, c_7, c_8$ in the July run. His constructed truthful preference is given by $c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8$. Note that the way we construct truthful preferences will cause us to underestimate the extent of strategic behavior, because we assume that the relative ranking of courses not in the top five is truthful.[30]

To address the second limitation, we restrict attention to the 456 students for whom we have May poll preferences and adjust course capacities accordingly. For each course, we used the Gehan test to compare the distribution of course ranks in July for the students who did answer the poll to the distribution for those who did not. At the 5% level, we found significant differences across these two samples of students for 7 of 89 courses (8%). We take this as evidence that the 456 students who answered the poll are representative of the entire population of students.

These constructed preferences and the scaled down course capacities are the main input in our welfare analysis in the next two sections.

6 Welfare Consequences of Strategic Play

The purpose of this section is to quantify the welfare consequences of students' strategic play in the HBS mechanism. Section 3.3 identified two channels through which strategic behavior affects

[29] Courses that were offered in the May poll but were no longer available in the July run are dropped.

[30] In the other direction, the presence of unobserved preference changes between May and July for May-poll-ranked courses will cause us to overestimate the extent of strategic behavior.

welfare: increased congestion and ex-post inefficiencies due to risk-taking behavior by students. In this section, we first document that strategic behavior has indeed an effect through both channels, and that these effects are large.

We then turn to the analysis of students' ex-ante welfare (i.e. before priority orders are drawn). Specifically, we compare welfare under the actual play of the HBS mechanism to welfare under a non-equilibrium counterfactual in which students report their preferences truthfully. Congestion is expected to have redistributive consequences (some students benefit, some students lose). Ex-post inefficiencies due to risk-taking behavior on the other hand hurt all students. Our results show that the net effect is that many more students are harmed by strategic behavior than benefit. Welfare comparisons at the aggregate level confirm this: a utilitarian social planner unambiguously prefers truthful play over strategic play, as long as students are risk neutral or risk averse.

The technical challenge we face in this section is that our data consist of students' ordinal preferences over individual courses, yet ex-ante welfare depends on their von Neumann-Morgenstern preferences over bundles. We address this challenge in two ways. First, we introduce a new computational method, inspired by the maximum flow problem in operations research, that allows us to compare ex-ante welfare levels, just based on students' ordinal preferences over individual courses and the assumption of responsiveness. This method allows us to pin down students' preference between truthful and strategic play for 51.5% of them.

Second, for the remaining students, we develop a series of comparison results that indicate conditions under which we can say that a particular student, or society as a whole, prefers truthful play or strategic play. These conditions place constraints on the way students' preferences over bundles of courses relate to their preferences over individual courses and on their risk attitude.

For all results in this section, our economy consists of the 456 students who submitted preferences in the May poll and in July, with course capacities scaled accordingly. When calculating students' outcomes we include an add-drop phase as described in Section 2.2.[31]

6.1 Effect of Strategic Behavior on Congestion

To evaluate the effect of strategic behavior on congestion, we run the HBS algorithm for 10,000 random priority orders using both the constructed truthful preferences and the July run preferences and we record the time at which each course reaches capacity (Time = # of rounds elapsed, including fractions of rounds). We say that a course reaches capacity earlier under strategic (truthful) play if the time it reaches capacity is earlier than it is under truthful (strategic) play for at least 99% of the priority orders. Table 4 reports the results of this analysis, with courses categorized as in Section 5.1.

[31] Note that this add-drop phase will only be used if students strategically misreport; if students report truthfully there will be no activity in the add-drop phase.

Table 4: Effect of Strategic Behavior on Congestion

		Course Reaches Capacity			
	N	Never	Truthful earlier	Indeterminate	Strategic earlier
Unpopular	47	35	1	9	2
Popular, 0×oversubscribed	16	4	3	8	1
Popular, 1×oversubscribed	5	0	2	1	5
Popular, 2×oversubscribed	6	0	0	0	6
Popular, 3×oversubscribed	3	0	0	0	3
Popular, 4×oversubscribed	1	0	0	0	1
Popular, 5×oversubscribed	2	0	0	0	2

As suggested by our theoretical analysis, congestion is related to popularity. Nearly all the courses that were at least one time oversubscribed based on the trial run (17 out of 20) reach capacity earlier. Very few of the other courses experience such congestion (3 of 63).

The magnitudes are large, especially amongst the most popular courses. For the 37 courses that reach capacity under both truthful play and strategic behavior, the round by which they do so goes from 6.73 to 6.35 on average (6% faster). If we focus on the 12 courses that are at least twice oversubscribed, the average round at which they reach capacity goes from 3.78 under truthful play to 2.61 under strategic play (31%), and the least congested amongst these still reaches capacity 14% faster on average under strategic play.

6.2 Ex-Post Efficiency Consequences of Strategic Play

To assess the magnitude of ex-post inefficiency in the HBS mechanism, we run the HBS algorithm for 100 random priority orders using students' July run preferences. For each priority order, we compute the number of ex-post Pareto improving trades we can find based on students' constructed truthful preferences.

Because our data consist only of ordinal preferences over individual courses, there are some profitable trades that we will not be able to find. For instance, if a student's truthful ROL is $P_s : c_1, c_2, c_3, c_4$ and his allocation is $\{c_1, c_4\}$ then we know that he is willing to trade c_4 for c_2 or c_3, but we do not know whether he is willing to trade the bundle $\{c_1, c_4\}$ for the bundle $\{c_2, c_3\}$.

Subject to this caveat, it is without loss of generality to restrict attention to trades in which each participant gives and receives a single course seat: whatever many-to-many trades we are able to find can be found using multiple one-for-one trades. For instance, student s above would be willing to trade $\{c_2, c_4\}$ for $\{c_1, c_3\}$, but this can be executed using two one-for-one trades of $\{c_2\}$ for $\{c_1\}$ and $\{c_4\}$ for $\{c_3\}$.

Let $x_{scc'}$ indicate whether we execute the one-for-one trade in which student s gives c and gets c'. For each priority order, we seek to maximize the number of pareto-improving trades subject to the constraint that students trade a course at most once. The result is the following binary integer program:

$$\max_{\mathbf{x}\in\{0,1\}^{SC^2}} \sum_{s,c,c'} x_{scc'}$$
such that

$$\sum_s \sum_{c'} x_{scc'} - x_{sc'c} = 0, \forall c \tag{2}$$
$$\sum_{c'} x_{scc'} + x_{sc'c} \leq 1, \forall s, c \tag{3}$$

$$x_{scc'} = 1 \Rightarrow c \in a_s(\widehat{\mathbf{P}}, \lambda), c' \notin a_s(\widehat{\mathbf{P}}, \lambda), c' P_s c \tag{4}$$

where constraint (2) captures the condition that each course must be given as often as it is received, constraint (3) prevents a student from trading the same course twice, both to ensure feasibility and to avoid double-counting, and constraint (4) ensures that the trade is both feasible and desirable, in the sense that student s's original allocation must include c, not include c', and he must prefers c' to c. Table 5 reports the results. It suggests that the level of ex-post inefficiency under the HBS mechanism is substantial: on average 15% of course seats can be profitably reallocated, involving 84% of students.

Table 5. Ex-Post Pareto Improving Trades

	Mean	Std. Dev.
# of Executed Trades per Student	1.54	(0.04)
% of Allocated Course Seats Traded	15.4%	(0.31%)
% of Students Executing		
0 Trades	16.4%	(1.1%)
1 Trade	35.4	(1.7)
2 Trades	30.5	(1.6)
3+ Trades	17.8	(1.3)

Given the magnitude of this inefficiency we run several robustness checks.[32] Each places additional restrictions on what constitutes a desirable trade (i.e., on (4)). First, we eliminate from

[32] Note that the magnitudes in Table 5 are consistent with the results of an informal survey conducted by two HBS students in the Spring of 2005. As part of a class project related to the HBS mechanism, these students surveyed 160 of their classmates. One of the questions was: “Did you know of a trade with another student that could have made you both better off?” 58.1% responded yes. This is suggestive of both the magnitude and students’ awareness of ex-post Pareto inefficiency. This figure likely includes some double counting. For example, if many students want to trade A for B, and only one wants to trade B for A, many more students might know of a Pareto-improving trade than could actually execute them. Of course, it also likely excludes lots of trades that students aren’t aware of, including multi-way trades. For many random priority orders we are able to find 43-way trades involving one seat in each of the 43 popular courses.

consideration any trade involving a student whose submitted preferences violate Lemma 1 because of one of the two courses involved in that trade. This reduces the volume of Pareto-improving trades to roughly 12% of all course seats. Next, we restrict attention to trades that are true to the spirit of the equilibrium ex-post inefficiency of Example 3. Specifically, we restrict attention to trades in which the course received is either overreported by the student or is a course that he downgrades but still receives with strictly positive probability (see type P_1 in Example 3). This reduces the volume of Pareto-improving trades to roughly 7% of all course seats. Combining the two tests reduces the level to 6%. This magnitude remains economically meaningful. For comparison, if students played truthfully, Theorem 5 tells us that there would be no pareto-improving one-for-one trade.

6.3 Consequences on Ex-Ante Individual Welfare

Both congestion and ex-post inefficiencies due to risk-taking behavior will impact students' welfare, measured before priority orders are drawn. In this section, we evaluate the impact of strategic behavior by comparing students' ex-ante welfare under the actual play of the HBS mechanism and the non-equilibrium counterfactual where students submit their true preferences.

The challenge we face is that our data consist of students' ordinal preferences over individual courses, whereas such ex-ante welfare comparisons require information about students' preferences over random allocations of bundles of courses. Students' preferences over individual courses together with responsiveness generate, for each student, a partial order over bundles of courses and thus, *a fortiori* a partial order over random allocations. This partial order is enough however to yield the following comparison criterion:

Comparison Result 1 (Responsive Preferences) Suppose that student s has responsive preferences. Then student s prefers truthful play $\mathbf{P}$ to strategic play $\widehat{\mathbf{P}}$ if, for any complete order over bundles consistent with the partial order implied by P_s and responsiveness, the lottery over bundles he receives under $\mathbf{P}$ first-order stochastically dominates that under $\widehat{\mathbf{P}}$. He prefers $\widehat{\mathbf{P}}$ if the reverse relationship holds. He is indifferent if the two distributions are equivalent.

We implement this criterion as follows. First, we run the HBS algorithm for both truthful play and strategic play for 100 randomly drawn priority orders, $\lambda_1, ..., \lambda_{100}$. Second, for each student s, we form a bipartite graph where one set of nodes is her outcomes from truthful play, $a_s(\mathbf{P},\lambda_1), ..., a_s(\mathbf{P},\lambda_{100})$, and the other set of nodes is her outcomes from strategic play, $a_s(\widehat{\mathbf{P}},\lambda_1), ..., a_s(\widehat{\mathbf{P}},\lambda_{100})$. Third, we draw an edge from node $a_s(\mathbf{P},\lambda_j)$ to node $a_s(\widehat{\mathbf{P}},\lambda_k)$ if we know from P_s and responsiveness that she weakly prefers $a_s(\mathbf{P},\lambda_j)$ to $a_s(\widehat{\mathbf{P}},\lambda_k)$. We do this for all pairs j, k. Fourth, we check if the resulting bipartite graph has a perfect matching, i.e., a subset of edges such that each node in the truthful-play node set is connected to exactly one node in the strategic-play edge set. If there is a perfect match, this means that the set of outcomes under $\mathbf{P}$ first-order stochastically dominates that under $\widehat{\mathbf{P}}$ *for any complete order consistent with the responsiveness partial order*. To check if $\widehat{\mathbf{P}}$ dominates $\mathbf{P}$ we reverse the way the edges are drawn in the third step.

We find that 45% of students are unambiguously harmed by strategic play, versus just 5.5% that unambiguously benefit. 1% of students are indifferent. For the remaining 48.5% the comparison is indeterminate. Ignoring the indeterminate cases for now, we note that these results confirm our intuition about the asymmetry between the costs and benefits of strategic play: the students who are harmed by strategic play are those whose preferred courses are popular, and by definition there are more of them than students whose preferred courses are not popular and who then gain from the ability to overreport popular courses. The fact that ex-post inefficiencies hurt everyone further exacerbates this asymmetry. The data show that the net result is sizable: there is a 1:9 ratio between the students who unambiguously gain from strategic play and those who lose.

To confirm these results however we need to pin down the indeterminate cases. These arise from two distinct sources. First, we have only a partial order over bundles; mechanically, this means that fewer edges are drawn in the bipartite graph than would be the case if we knew students' complete ordinal preferences. Second, even if we knew students' ordinal preferences over bundles, we still would need to know their preferences over lotteries to always reach a comparison; first-order stochastic dominance is a demanding order. In the rest of this section, we impose additional assumptions on preferences that fill in this information gap and pin down the indeterminate cases.

We say that student s has *additive* preferences if there exist numbers $v_s(c)$ for all courses in C, such that $u_s(a_s) > u_s(a'_s) \iff \sum_{c \in a_s} v_s(c) > \sum_{c \in a'_s} v_s(c)$ where u_s is student s's vNM utility function and a_s and a'_s are allocations. Additive preferences are a special case of responsive preferences. By itself, the additivity assumption does not yield new results, but it provides a structure onto which we can layer additional assumptions. Specifically, if, in addition, student s is risk neutral, then his expected utility under strategy profile $\widehat{\mathbf{P}}$ can be expressed as $\sum_\lambda \sum_{c \in a_s(\widehat{\mathbf{P}},\lambda)} v_s(c)$. This yields our second comparison result:

Comparison Result 2 (Additive Preferences, Risk Neutral). Suppose that student s is risk neutral and has additive preferences. Student s prefers truthful play $\mathbf{P}$ to strategic play $\widehat{\mathbf{P}}$ if, for any j, the expected number of top-j courses he gets under $\mathbf{P}$ exceeds that under $\widehat{\mathbf{P}}$. He prefers $\widehat{\mathbf{P}}$ if the reverse relationship holds. He is indifferent if he gets each course with equal probability under both strategy profiles.

Note the difference with Comparison Result 1. The combination of additivity and risk neutrality allows us to aggregate a distribution over $\binom{C}{m}$ bundles into a distribution over just C courses.

A special case of additive preferences that allows us to compress this information further, and gives us a complete order over bundles, is when the difference in utilities derived from the student's 1^{st} and 2^{nd} favorite courses is the same as that between the n^{th} and $n+1^{\text{st}}$ favorite, for any n. Such a student cares about the *average rank* of the courses in his allocation. Average rank is a measure of mechanism performance emphasized by the HBS administration. Combined with different assumptions on risk attitudes, the average-rank assumption yields the following comparison results:

Comparison Result 3 (Average-rank Preferences). Assume student s has average-rank preferences and let $\overline{r}_s(\mathbf{P},\lambda)$ ($\overline{r}_s(\widehat{\mathbf{P}},\lambda)$) denote the average rank of the courses that student s get under strategy profile $\mathbf{P}$ ($\widehat{\mathbf{P}}$) for the priority order λ :

(i) Independently of his attitude towards risk, student s prefers truthful play $\mathbf{P}$ to strategic play $\widehat{\mathbf{P}}$ if $-\overline{r}_s(\mathbf{P},\cdot)$ first-order stochastically dominates $-\overline{r}_s(\widehat{\mathbf{P}},\cdot)$. He prefers $\widehat{\mathbf{P}}$ to $\mathbf{P}$ if the converse holds.

(ii) If student s is risk averse, he prefers truthful play $\mathbf{P}$ to strategic play $\widehat{\mathbf{P}}$ if $-\overline{r}_s(\mathbf{P},\cdot)$ second-order stochastically dominates $-\overline{r}_s(\widehat{\mathbf{P}},\cdot)$. He prefers $\widehat{\mathbf{P}}$ to $\mathbf{P}$ if the converse holds.[33]

(iii) If student s is risk neutral, he prefers truthful play $\mathbf{P}$ to strategic play $\widehat{\mathbf{P}}$ if $\sum\limits_{\lambda}\overline{r}_s(\mathbf{P},\lambda) \leq \sum\limits_{\lambda}\overline{r}_s(\widehat{\mathbf{P}},\lambda)$. He prefers $\widehat{\mathbf{P}}$ to $\mathbf{P}$ if the converse holds.

Another special case of additive preferences is lexicographic preferences (defined in Section 3) which puts a high premium on getting one's favorite courses. Lexicographic preferences can be seen as the other extreme from average-rank preferences. The HBS administration implicitly assumes lexicographic preferences when they evaluate the performance of the mechanism by the number of students who get their favorite course. Lexicographic preferences also generate a complete order over random allocations, and yield the following comparison result.

Comparison Result 4 (Lexicographic Preferences). Assume student s has lexicographic preferences. He prefers truthful play $\mathbf{P}$ to strategic play $\widehat{\mathbf{P}}$ if he gets his first choice course more often under $\mathbf{P}$ than under $\widehat{\mathbf{P}}$ or if he gets each of his n favorite courses as often under both profiles but gets his $n+1^{\text{th}}$ favorite course more often under $\mathbf{P}$, for some n. He prefers $\widehat{\mathbf{P}}$ if the reverse relationship holds.

To implement Comparison Results 2-4, we ran the HBS algorithm for 10,000 random priority orders over students using both the constructed truthful and the July-run preferences, and record each student's distribution over outcomes. Table 6 reports the results. As a benchmark, we include

[33] For two cumulative distributions of average ranks, say F and G, with ranks distributed on $[\underline{\rho},\overline{\rho}]$, F second-order stochastically dominates G iff $\int_x^{\overline{\rho}}(1-F(x))dx \leq \int_x^{\overline{\rho}}(1-G(x))dx$ for all $x \in [\underline{\rho},\overline{\rho}]$. The difference versus the usual formula (Gollier, 2001; Section 3.2) is due to the fact that lower is better.

the results from CR1.

Table 6. Individual preferences over play of the HBS Mechanism using CR1–4

	Assumption on Preferences					
	Responsive	Additive	Average-Rank			Lexicographic
	Any Risk Attitude	Risk Neutral	Any Risk Attitude	Risk Averse	Risk Neutral	Risk Neutral
Outcome	(1)	(2)	(3)	(4)	(5)	(6)
Prefers HBS Truthful	45%	46%	56%	68%	73%	90%
Prefers HBS Strategic	5%	5%	13%	17%	26%	9%
Indifferent	1%	1%	1%	1%	1%	1%
Indeterminate	48%	47%	30%	14%	0%	0%

The results confirm the basic asymmetry between the benefits and costs of strategic play: by each comparison criterion, strategic play harms more students than it benefits. This asymmetry is especially acute when preferences are lexicographic (comparison between column (6) and columns (3)-(5)) because it is impossible to overeport one's favorite course.

6.4 Consequences on Ex-Ante Social Welfare

We now turn to social welfare. Clearly, we cannot Pareto rank truthful play and strategic play based on the assumption of responsive preferences alone: some students prefer strategic play and others prefer truthful play. So in this section, we impose some inter-personal comparison of utilities and assume additive preferences and a utilitarian social planner. An alternative interpretation is that we take the perspective of an individual student who does not know his preferences but knows the distribution of preferences in the population; that is, a student behind a veil of ignorance in the sense of Harsanyi (1953). The "social" analogues of Comparison Results 2-4 are as follows:

Comparison Result 5 (Additive Preferences, Risk Neutral). Assume that students are risk neutral and have additive preferences. Society prefers play truthful play $\mathbf{P}$ to strategic play $\widehat{\mathbf{P}}$, if, for any j, the expected number of top n courses allocated to students under $\mathbf{P}$ exceeds that under $\widehat{\mathbf{P}}$. Society prefers $\widehat{\mathbf{P}}$ if the converse holds.

Comparison Result 6 (Average-rank Preferences). Assume students have average-rank preferences:

(i) Independently of students' attitude towards risk, society prefers truthful play $\mathbf{P}$ to strategic play $\widehat{\mathbf{P}}$ if $-\overline{r}.(\mathbf{P},\cdot)$ first-order stochastically dominates $-\overline{r}.(\widehat{\mathbf{P}},\cdot)$. Society prefers $\widehat{\mathbf{P}}$ if the converse holds. (The notation $\overline{r}.(\mathbf{P},\cdot)$ indicates that the distribution is taken over priority orders and students.)

(ii) If students are risk averse, society prefers truthful play $\mathbf{P}$ to strategic play $\widehat{\mathbf{P}}$ if $-\overline{r}.(\mathbf{P},\cdot)$ second-order stochastically dominates $-\overline{r}.(\widehat{\mathbf{P}},\cdot)$. Society prefers $\widehat{\mathbf{P}}$ if the converse holds.

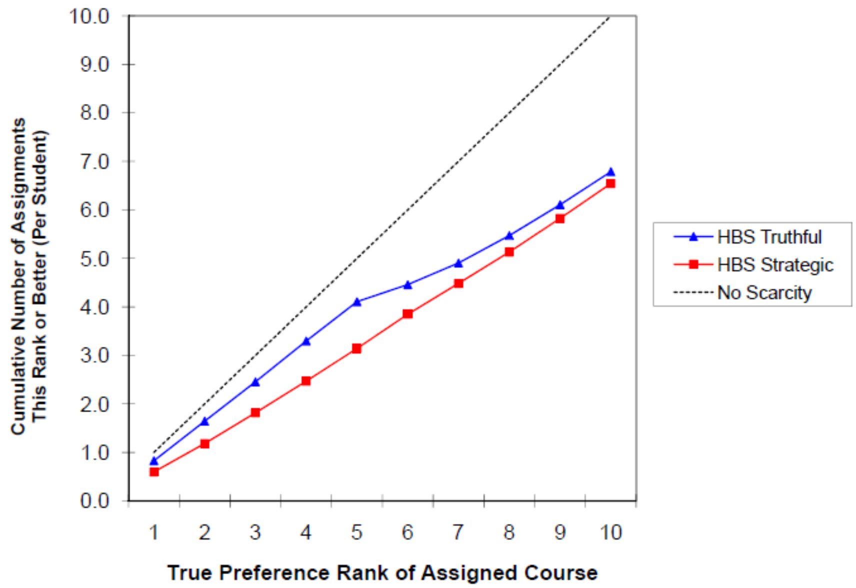

Figure 1: Cumulative distribution of the true preference rank of students' assigned courses: truthful versus strategic play of the HBS mechanism. The distribution under truthful play first-order stochastically dominates that under strategic play, so CR5, CR6(iii) and CR7 obtain.

(iii) If students are risk neutral, society prefers truthful play $\mathbf{P}$ to strategic play $\widehat{\mathbf{P}}$ if $\sum_{\lambda}\sum_{s}\overline{r}_s(\mathbf{P},\lambda) < \sum_{\lambda}\sum_{s}\overline{r}_s(\widehat{\mathbf{P}},\lambda)$. Society prefers $\widehat{\mathbf{P}}$ if the converse holds.

Comparison Result 7 (Lexicographic Preferences). Assume students have lexicographic preferences. Society prefers truthful play $\mathbf{P}$ to strategic play $\widehat{\mathbf{P}}$ if the expected number of students who get their first choice course is higher under $\mathbf{P}$ than under $\widehat{\mathbf{P}}$ or if the expected number of students who get their 1^{st},..., n^{th} favorite courses is the same under both strategy profiles, for some n, but the expected number of students who get their $n+1^{\text{th}}$ favorite course is higher under $\mathbf{P}$. Society prefers $\widehat{\mathbf{P}}$ if the reverse relationship holds.

Figure 1 shows the average number of courses that students get among their top n choices. There is a first order stochastic dominance relationship between the distribution of outcomes under truthful and strategic play: students get more of their favorite course, more of their top two favorite courses and so on under truthful play than under strategic play.[34] Thus CR5 obtains and by consequence CR6(iii) and CR7 obtain as well since both are special cases of risk-neutral additive

[34] The kink in the HBS Truthful line at rank 6 is a mechanical effect due to the way we construct truthful preferences (see Section 5.3). Students report their top-5 truthful preferences in the May Poll. Their 6th favorite course is the first course they rank in the strategic rank order list that they didn't rank in the May Poll. If this course is rated highly by many other students in the May Poll, then the student will never obtain it under Truthful play, but might obtain it under Strategic play if he ranks it highly enough.

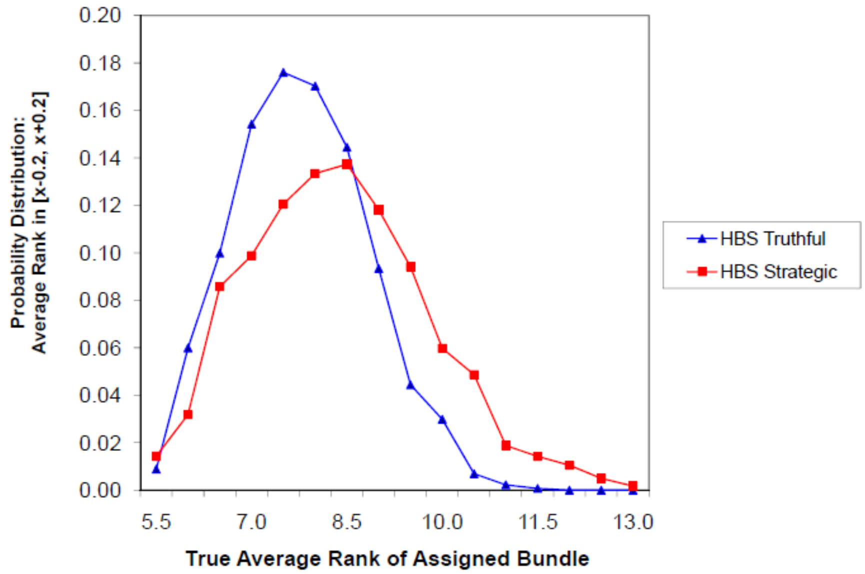

Figure 2: Probability distribution of the true-preference average rank of students' assigned bundles: truthful versus strategic play of the HBS mechanism. The distribution under truthful play second-order stochastically dominates that under strategic play, so CR6(ii) obtains.

preferences. In other words, if students are risk neutral, a utilitarian social planner unambiguously prefers truthful play of the HBS mechanism.

The difference is economically meaningful. 83% of students receive their favorite course under truthful play, and they receive 2.46 of their top three courses, versus 60% and 1.82 under strategic play. What is driving the result is that some of the most popular courses go to students for whom they are not the most preferred courses. For example, the two most popular courses in our data account for 50% of all truthful first choices, and 68% of all strategic first choices. These two courses reach capacity in the first round of strategic play, so, on average, around 26% of the seats in these courses go to students for whom it is not their true first choice.[35]

To investigate the role of students' risk attitudes, Figure 2 plots the distribution of the average rank of course allocations in the population over all 10,000 trials. There is a bit more mass at the very best outcomes under strategic play than under truthful play; this is due to the targeted opportunism of students who are fortunate to mainly like unpopular courses. Otherwise, the distribution under strategic play is dominated by that from truthful play, reflecting the negative consequences of congestion and ex-post Pareto inefficiencies. There is no first-order stochastic dominance relationship between the two distributions (CR6(i) is not satisfied by a tiny margin),

[35] That is, (68%-50%)/68%. These two courses alone account for around 100 fewer students (11% of the student body) obtaining their first-choice course under strategic play.

but second-order stochastic dominance (CR6(ii)) does obtain. Thus, strategic behavior hurts ex-ante social welfare if students have average-rank preferences and are weakly risk averse.

7 Comparison of the HBS Mechanism to a Strategyproof Alternative

In the previous section we showed that strategic play of the HBS mechanism harms efficiency, assessed either ex-ante or ex-post. This section asks the logical next question: should HBS switch to a strategyproof mechanism? To answer this, we perform a welfare comparison between actual play of the HBS mechanism and truthful play of its strategyproof alternative, Random Serial Dictatorship (RSD). We use the same methodology as in Sections 6, except that the counterfactual is (equilibrium) truthful play of RSD, as opposed to (non-equilibrium) truthful play of HBS.

The first thing to note is that RSD is ex-post efficient, whereas we found in Section 6.2 that the HBS mechanism is highly inefficient ex-post.

In order to assess ex-ante efficiency, we will need to impose additional structure on preferences beyond responsiveness. Under RSD, students will often obtain their ideal bundle of courses, but will also often obtain a very poor bundle. The responsiveness assumption does not rule out the possibility that a student only places value on obtaining his ideal bundle, nor does it rule out that the student only cares about maximizing the minimum bundle he obtains. So comparisons based on CR1 are entirely indeterminate.

As soon as we put additional structure on preferences we find that the HBS mechanism is more attractive ex-ante than RSD. RSD's ex-ante unattractiveness is surprising since ex-post it is efficient. Furthermore, this result does *not* depend on risk aversion. We provide a novel theoretical explanation of RSD's unattractiveness in the following section.

7.1 Ex-Ante Individual Welfare

We repeat the methodology of Section 6.3. Table 7 compares HBS to RSD under additive, average rank, and lexicographic preferences using Comparison Results 1-4:

Table 7. Individual preferences between HBS and RSD: CR1–4

	Assumption on Preferences					
	Responsive	Additive	Average-Rank			Lexicographic
Outcome	Any Risk Attitude	Risk Neutral	Any Risk Attitude	Risk Averse	Risk Neutral	Risk Neutral
	(1)	(2)	(3)	(4)	(5)	(6)
Prefers RSD	0%	0%	0%	0%	19%	25%
Prefers HBS Strategic	0%	26%	2%	81%	81%	75%
Indifferent	0%	0%	0%	0%	0%	0%
Indeterminate	100%	74%	98%	19%	0%	0%

We begin by comparing columns (3), (4) and (5). Without any structure on students' risk preferences, the comparison is almost entirely indeterminate. This is because RSD induces such extreme outcomes. If we assume that students are risk-averse, then the indeterminacies are resolved in favor of the HBS mechanism. No student unambiguously prefers RSD and the vast majority unambiguously prefers HBS. This reflects the well-known criticism of RSD that it exposes students to risk.

There is however more in Table 7 than risk. Indeed, compare now columns (5) and (6). In both cases, and even though students are risk neutral, the large majority of students prefer the HBS mechanism to RSD. Interestingly, and unlike in the comparison in Table 6, the ratio does not vary much between the two columns. This suggests that preference intensity is not what drives students' ex-ante preference for HBS over RSD.

7.2 Ex-Ante Social Welfare

We repeat the methodology of Section 6.4. Figure 3 compares the aggregate rank distributions of HBS and RSD. The distribution under strategic play of the HBS mechanism first-order stochastically dominates that under truthful play of RSD. Thus, a utilitarian social planner prefers HBS to RSD when students are risk neutral and have any additive preferences (CR5, CR6(iii), CR7). This result confirms the picture at the individual level but is more suprising at the social level, given RSD's ex-post efficiency. It suggests that RSD's ex-post efficiency is not a good proxy for social welfare.

The magnitudes are of the most economic importance in the tails. Students receive their favorite course with 60% probability under HBS, but with only 47% probability under RSD.[36] Students actually receive slightly more of their 2nd-10th favorite courses under RSD (6.02) than under HBS

[36] To give a sense of the magnitude of this difference, we note Pathak's (2006) findings in the context of a single–unit assignment problem, school choice. He finds that students receive their first-choice school 60.6% of the time

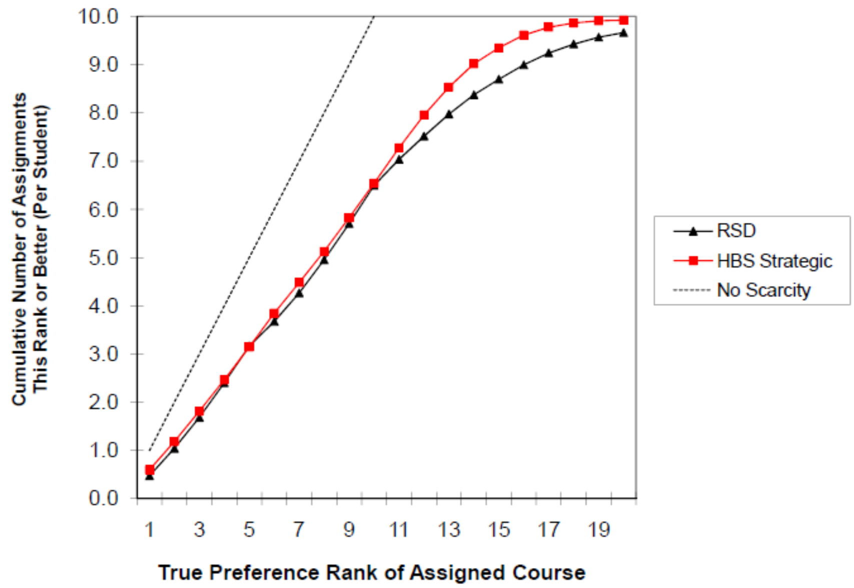

Figure 3: Cumulative distribution of the true preference rank of students' assigned courses: RSD versus strategic play of the HBS mechanism. The distribution under HBS first-order stochastically dominates that under RSD, so CR5, CR6(iii) and CR7 obtain.

(5.95). This is because students with lucky draws in RSD get all 10 of their favorite courses. The cost is that students receive twice as many courses they like less than 15th (1.30) under RSD than under HBS (0.65). As a result the mean average rank under RSD is 8.94, versus 8.35 under HBS. This is an economically meaningful difference, around that of the difference between truthful and strategic play of the HBS mechanism.

Finally, Figure 4 illustrates the risk to which RSD exposes students by examining the distribution of average ranks. RSD puts much more weight on the tails of the distribution, and indeed is second-order stochastically dominated by HBS (CR6 (ii)). So a utilitarian social planner prefers HBS to RSD if students are weakly risk averse and have average-rank preferences. Again, the magnitudes are significant. Under RSD, around 29% of students obtain their "bliss bundle" consisting of their 10 favorite courses, versus around 1% under HBS. But over 17% of students obtain a bundle with average rank worse than 12, versus just 1% under HBS.

under RSD, versus 60.8% in the counterfactual of interest (Bogomolnaia and Moulin (2001)'s Probabilistic Serial mechanism).

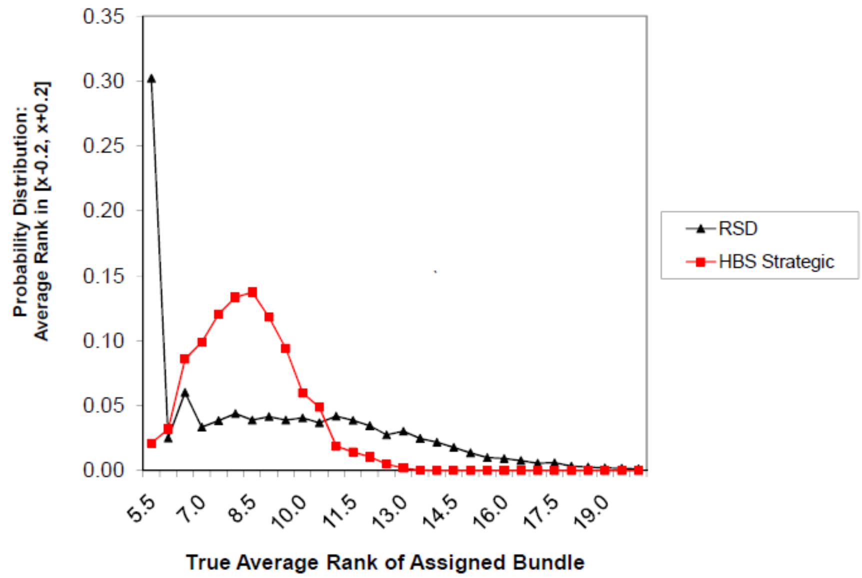

Figure 4: Probability distribution of the true-preference average rank of students' assigned bundles: RSD vs. strategic play of the HBS mechanism. The distribution under HBS second-order stochastically dominates that under RSD, so CR6(ii) obtains.

8 Callousness

Our intuition for RSD's poor ex-ante performance is simple. Under RSD, fortunate students with good random draws make their *last* choices independently of whether these courses would be some unfortunate students' *first* choices; students "callously disregard" the preferences of those who choose after them. The reason this callous behavior matters for welfare is that the ex-post utility benefit to the fortunate students from these last choices will generally be small relative to the ex-post harm these choices cause to the unfortunate students. Thus, RSD is unattractive when evaluated ex-ante. Notice that the unattractiveness of RSD does not depend on students being risk averse; even risk neutral agents regard a "win a little, lose a lot" lottery as unappealing.

We formalize this intuition with a simple example and a simple theorem.

Example 4 (Callousness of RSD). There are two students, a and b, and four courses each in unit supply. Students' ordinal preferences over singletons are drawn uniformly i.i.d., and they report their preferences truthfully.

Consider first the RSD choosing order $aabb$. Student a always gets his 1st and 2nd favorite courses, while b gets either his 1st/2nd, 1st/3rd, 1st/4th, 2nd/3rd, 2nd/4th, or 3rd/4th favorite courses, each with equal probability. a's average rank is 1.5 and b's is 2.5, so the societal mean is 2.0. a always gets his favorite course whereas b gets it with probability 0.5, so the societal mean is

0.75.

Now consider the HBS choosing order *abba*. Student a always gets his 1st favorite course. Then, b gets his 1st and 2nd favorites with probability one-half, and otherwise gets either his 1st/3rd or 2nd/3rd, each with probability one-quarter. Last, a gets either his 2nd, 3rd, or 4th, each with equal probability. a's average rank is 2.0 and b's is 1.875, for a societal mean of 1.9375. a always gets his favorite course, whereas b gets it with probability 0.75, so the societal mean is 0.875.

The following table summarizes the differences between HBS and RSD.

Table 8. Results of Example 4

	Pr(student gets (i^{th}, j^{th}) favorite courses)						Summary Stats	
	$(1,2)$	$(1,3)$	$(1,4)$	$(2,3)$	$(2,4)$	$(3,4)$	E(Avg Rank)	Pr(Get Favorite)
RSD	$\frac{7}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	2.0	0.75
HBS	$\frac{5}{12}$	$\frac{7}{24}$	$\frac{1}{8}$	$\frac{1}{6}$	0	0	1.9375	0.875

In this simple example, ex-ante welfare is higher under HBS than RSD for risk-neutral students with either average-rank or lexicographic preferences. The driving force behind both results is that it is harmful, in terms of these measures of welfare, to give a his second choice before b has made his first choice. Risk aversion only exacerbates the case against RSD, because of the increased probability of the very worst outcomes.

Simulations suggest that the average-rank finding in Example 4 generalizes to larger economies. For instance, in an HBS-sized version of Example 4 with 1000 students, 100 courses, 100 seats per course, and $m = 10$, the average rank under HBS is 5.72 versus 6.45 under RSD.[37]

The following simple theorem shows that the first-choice-course finding in Example 4 generalizes. It helps contrast callousness to existing critiques of RSD.

Theorem 6 (Callousness of RSD): Suppose there are S students, each of whom requires m courses, and mS courses each in unit supply. Students' ordinal preferences over courses are drawn uniformly i.i.d., and they report their preferences truthfully.[38] Then the expected proportion of students who obtain their first-choice course is $1 - \frac{(S-1)}{2Sm}$ under HBS which is strictly greater than the proportion $1 - \frac{(S-1)}{2S}$ under RSD whenever $m > 1$. As $S \to \infty$ the proportion converges to $1 - \frac{1}{2m}$ under HBS versus $\frac{1}{2}$ under RSD.

A direct consequence of Theorem 6 is that ex-ante welfare is lower under RSD even if students are risk neutral (and have lexicographic preferences). Thus callousness is distinct from the risk critique of RSD. Second, Theorem 6 (as well as the simulation evidence) shows that RSD's unattractiveness

[37] Further simulation results and code are available from the authors. We also are able to show theoretically that Example 4 generalizes to any number of students S, with $m = 2$ and Sm courses each in unit supply. The proof is somewhat involved and is available upon request.

[38] Theorem 2 indicates that truthful reporting is an equilibrium in the continuum version of this environment.

persists in large markets. This helps to illustrate that callousness is distinct from Bogomolnaia and Moulin's (2001) critique of RSD in the single-unit assignment setting, since the magnitude of the inefficiency they address goes to zero as the market grows large (Che and Kojima, forthcoming). Finally, note that callousness is specific to multi-unit assignment. If $m = 1$, the HBS and RSD mechanisms are equivalent.

9 Conclusion

This paper is useful at three levels: the mechanism, the problem, and the field.

We analyze two specific mechanisms: one used in practice to allocate courses to students at Harvard Business School, and the other extensively studied in theory. We show that the HBS mechanism is importantly flawed: it is simple to manipulate in theory and heavily manipulated in practice, with meaningful welfare consequences. However, we also show that on measures of ex-ante welfare, RSD is worse, due to an aspect of RSD that we call "callousness".

At the level of the problem, multi-unit assignment, our paper not only shows that the HBS mechanism is a sensible choice relative to RSD, but suggests "where to look" for mechanisms that are better still. First, one should seek a mechanism that is not a strategyproof dictatorship but that is not simple to manipulate like the HBS mechanism. The mechanism should aim to induce truthful reporting in realistic market environments. Second, one should seek a mechanism that more resembles HBS than RSD in its ex-ante efficiency characteristics. To avoid callousness, participants' realized resources (here, their choosing times) should not be highly unequal as in RSD, but rather roughly equal as in HBS. Budish (2009) proposes a mechanism directly inspired by the present analysis. His mechanism is strategyproof in an economy with a large number of students (unlike HBS) and exposes participants to an arbitrarily small amount of resource variance (unlike RSD). When evaluated on our data, his mechanism outperforms both the HBS mechanism and RSD.

Finally, our paper contributes to two active debates in the broader literature on market design. The first contribution concerns welfare analysis for random allocation mechanisms. The challenge we face of analyzing welfare on the basis of just ordinal preference data is not unusual for empirical market design research.[39] One approach to such data incompleteness is to limit attention to measures of ex-post efficiency, for which ordinal data is sufficient. On this basis, RSD is more attractive than HBS. Our contribution is a set of simple tools that allow one to make statements about ex-ante welfare based on just ordinal preference data. When we use these tools to compare HBS and RSD

[39] For example, hospitals in the National Resident Matching Program report only ordinal preferences over individual doctors, but their welfare depends in the end on the team they assemble (Roth and Peranson, 1999). Note that the challenge is less severe in unit-demand problems, because the dimensionality of vNM preferences equals the dimensionality of the ordinal preference data. Papers that exploit this fact include Bogomolnaia and Moulin (2001), Miralles (2009), and Featherstone and Niederle (2008).

ex-ante, the conclusion from the ex-post analysis is reversed. That is, ex-post efficiency need not even proxy for ex-ante efficiency, and care should be taken if analyzing the former while hoping for the latter.

The second contribution to the field of market design concerns the role of strategyproofness, and the welfare costs of using a manipulable mechanism. Strategyproofness has long been viewed as an important desideratum in practical market design, especially in the context of assignment and matching problems (Roth 2008).[40] Our field data allow us to directly document that students at HBS – real-life participants in a one-shot high-stakes setting – figure out how to strategically manipulate the non-strategyproof HBS mechanism. Further, we show that these manipulations have real welfare consequences. But while these findings are strongly consistent with the view that strategyproofness is an important desideratum, our finding that the HBS mechanism is more attractive than RSD suggests that this desideratum should be used with caution. It would have been a mistake for HBS to impose strategyproofness as an inflexible design requirement.

[40] There are at least three reasons to favor strategyproof mechanisms in such settings. First, strategyproof mechanisms are the ultimate robust mechanisms in the sense of Wilson (1987). Second, strategyproof mechanisms make it easy to advise market participants and help to level the playing field between sophisticated and naive players (Abdulkadiroglu et al (2009), Pathak and Sonmez (2008)). Third, strategyproof mechanisms generate preference information that can be used for ex-post policy evaluation and public decisions (Roth, 2008).

A Proofs

A.1 Proof of Theorem 1 (Simple Manipulations)

The strategy of the proof is to show that a sequence of deviations from P_s, that consist in downgrading the $\widehat{\mathbf{P}}_{-s}$-unpopular courses to the bottom half of the top m courses in student s's ROL while preserving the relative ordering of the $\widehat{\mathbf{P}}_{-s}$-popular and $\widehat{\mathbf{P}}_{-s}$-unpopular courses, leaves student s weakly better off for all λ. For ease of reference, relabel courses such that $P_s : c_1, c_2, c_3, ..., c_C$. To save on notations, in the remainder, we simply refer to popular and unpopular courses for $\widehat{\mathbf{P}}_{-s}$-popular and $\widehat{\mathbf{P}}_{-s}$-unpopular courses respectively.

Claim 1: Suppose c_k is unpopular. Then, for all λ, $\widehat{P}_s^{c_k \downarrow l}$ gets exactly the same courses as $\widehat{P}_s$ or exactly one more course in $\{c_{k+1}, ..., c_l\}$ than $\widehat{P}_s$, at the cost of a course in $\{c_{l+1}, ..., c_C\}$.

Proof of Claim 1: Fix an arbitrary λ. Because $\widehat{P}_s$ and $\widehat{P}_s^{c_k \downarrow l}$ only differ from position k onwards, the game proceeds identically until the time at which $\widehat{P}_s$ requests c_k (and $\widehat{P}_s^{c_k \downarrow l}$ requests course c_{k+1}). Let r_k be the round at which this happens. By construction, c_k is available when $\widehat{P}_s$ requests it in round r_k. Because student s has zero mass, the fact that his outcome in round r_k is different across the two strategies does not affect course seat availabilities and thus, a fortiori, the allocation and requests of other students in any given round.

From round $r_k + 1$ onwards, student s requests courses one round earlier under strategy $\widehat{P}_s^{c_k \downarrow l}$ than under strategy $\widehat{P}_s$, until we either reach a course, say $c_{k'}$, in $\{c_{k+1}, ..., c_l\}$ that student s gets under $\widehat{P}_s^{c_k \downarrow l}$ but not under $\widehat{P}_s$, or reach position l in student s's ROL. We consider each case in turn:

1. There exists $c_{k'}$ in $\{c_{k+1}, ..., c_l\}$ that student s gets under $\widehat{P}_s^{c_k \downarrow l}$ but not under $\widehat{P}_s$.

   Let $r_{k'}$ be the round at which student s requests but does not get this course under $\widehat{P}_s$. From round $r_{k'}$ onwards, student s's requests are in synch under both strategies and thus he gets the same outcome until the algorithm reaches position l in his ROL.

   When the algorithm reaches the request in position l, student s requests (and gets) course c_k under $\widehat{P}_s^{c_k \downarrow l}$. From then on, student s requests courses one round earlier under $\widehat{P}_s$. This has two possible consequences: either there exists a course that he gets under $\widehat{P}_s$ but not under $\widehat{P}_s^{c_k \downarrow l}$ (after which his requests are in synch and there is no more discrepancy between the two outcomes), or the algorithm reaches round m (and thus the course that the student requests in round m under $\widehat{P}_s$ is never requested by $\widehat{P}_s^{c_k \downarrow l}$). In both cases, there is a single course in $\{c_{l+1}, ..., c_C\}$ that student s gets under $\widehat{P}_s$ instead of $c_{k'}$ that he does not get under $\widehat{P}_s^{c_k \downarrow l}$.

2. The algorithm reaches position l in student s's ROL without any difference in allocations between the two strategies

At that round, $\widehat{P}_s^{c_k\downarrow l}$ requests c_k and student s's requests become in synch again. There is thus no more difference in outcomes.

Claim 2: Let c_k be the lowest-ranked unpopular course among the top m courses in P_s. Let $\widehat{P}_s^1 = P_s^{c_k\downarrow m}$. Student s is weakly better off using $\widehat{P}_s^1$ than P_s for all λ.

Proof of Claim 2: By claim 1, $\widehat{P}_s^1$ gets exactly the same courses or exactly one additional course in $\{c_{k+1}, ..., c_m\}$ than P_s, at the cost of a course in $\{c_{m+1}, ..., c_C\}$. Because all courses in $\{c_{k+1}, ..., c_m\}$ are strictly preferred to courses in $\{c_{m+1}, ..., c_C\}$, student s is either indifferent or strictly better off using $\widehat{P}_s^1$ (here we are using the fact that preferences are responsive and that students have vNM preferences over lotteries).

Claim 3: Let c_j be the n^{th} lowest unpopular courses among the top m courses in P_s. Let $\widehat{P}_s^n = \widehat{P}_s^{n-1 \ \ c_j\downarrow m-n+1}$ (student s downgrades course c_j just above all the other less preferred unpopular courses that he has already downgraded). Student s is weakly better off using $\widehat{P}_s^n$ than $\widehat{P}_s^{n-1}$ for all λ.

Proof of Claim 3: By claim 1, $\widehat{P}_s^n$ gets either exactly the same courses or exactly one additional course among the popular courses that were between c_j and position $m-n+1$ in $\widehat{P}_s^{n-1}$. This comes at the expense of a course in $\{c_{m+1}, ..., c_C\}$. Given that preferences are responsive and take the vNM form, student s is weakly better off using $\widehat{P}_s^n$ over $\widehat{P}_s^{n-1}$.

We continue until there is no further unpopular course to downgrade. At each deviation, student s is weakly better off for all λ. The claim then follows by transitivity. QED

A.2 Proof of Theorem 2 (Truthful Play in Special Cases)

Proof: Identical preferences: Let $P_s = P_{s'} : c_1, c_2, c_3, ...$ Under truthful play, course c_1 runs out earlier than c_2, which itself runs out earlier than c_3 and so on. Also note that c_1 runs out with probability 1 in round 1 for all strategy profiles $(\widehat{P}_s, \mathbf{P}_{-s})$ for any $\widehat{P}_s$.

Towards a contradiction, suppose $\widehat{P}_s \neq P_s$ constitutes a profitable deviation for student s when the other students play $\mathbf{P}_{-s}$. Let $\widehat{P}_s^{c\uparrow}$ equal $\widehat{P}_s$ except that c is moved to the first position. Similarly, let $\widehat{P}_s^{cc'\uparrow}$ equal $\widehat{P}_s$, except that c is moved to the first position and c' is moved to the second position, and so on for $\widehat{P}_s^{cc'c''\uparrow}$, $\widehat{P}_s^{cc'c''c'''\uparrow}$, ...

We show that the sequence $\widehat{P}_s^{c_1\uparrow}, \widehat{P}_s^{c_1c_2\uparrow}, ..., \widehat{P}_s^{c_1...c_{C-1}\uparrow} = P_s$ constitutes a chain of profitable deviations so that student s is at least as well off under P_s as under $\widehat{P}_s$. This contradicts the hypothesis that $\widehat{P}_s$ was a profitable deviation from P_s.

Consider first $\widehat{P}_s^{c_1\uparrow}$ and $\widehat{P}_s$ and suppose c_1 is not first in $\widehat{P}_s$ (otherwise $\widehat{P}_s^{c_1\uparrow} = \widehat{P}_s$ and we are done).

Claim 1: Student s gets either exactly the same courses under $(\widehat{P}_s^{c_1\uparrow}, \mathbf{P}_{-s})$ and $(\widehat{P}_s, \mathbf{P}_{-s})$ or his two allocations differ by exactly one course: he gets c_1 under $(\widehat{P}_s^{c_1\uparrow}, \mathbf{P}_{-s})$ which he does not get under $(\widehat{P}_s, \mathbf{P}_{-s})$, in exchange for getting a course under $(\widehat{P}_s, \mathbf{P}_{-s})$ that he does not get under $(\widehat{P}_s^{c_1\uparrow}, \mathbf{P}_{-s})$.

Proof of Claim 1: We compare how the game plays out under the two strategies. Partition the set of priority orders $\mathcal{L}$ into $\mathcal{L}_1$ and $\mathcal{L}_0$ according to whether student s does or does not get c_1 in the first round when playing $\widehat{P}_s^{c_1\uparrow}$. Under all priority orders in $\mathcal{L}_0$ the two games play out exactly in the same fashion (since student s never gets c_1 under $\widehat{P}_s$), so we focus on priority orders in $\mathcal{L}_1$. Fix $\lambda \in \mathcal{L}_1$. Under $(\widehat{P}_s^{c_1\uparrow}, \mathbf{P}_{-s})$, student s gets c_1 which he does not get under the original strategy. From round 1 onwards until we reach a course that student s gets under one strategy but not under the other, student s requests each specific course exactly one round later under $\widehat{P}_s^{c_1\uparrow}$. Because of the continuum assumption, other students' requests and outcomes are otherwise not affected and courses run out at the same time under both strategy profiles. Thus if there is a course that student s gets under one strategy but not under the other it is a course that he does not get under $(\widehat{P}_s^{c_1\uparrow}, \mathbf{P}_{-s})$. Call this course c_l and let r be the round at which this happens. From round r, student s's requests are in synch again and so are other students' requests. This implies there are no additional discrepancies between the two allocations.

Claim 1, together with responsiveness ($c_1 P_s c_l$) and vNM preferences over uncertain outcomes, implies that student s is strictly better off playing $\widehat{P}_s^{c_1\uparrow}$ than $\widehat{P}_s$. We next show that $\widehat{P}_s^{c_1\ldots c_k\uparrow}$ is preferred to $\widehat{P}_s^{c_1\ldots c_{k-1}\uparrow}$.

Claim 2: Student s gets either exactly the same courses under $(\widehat{P}_s^{c_1\ldots c_k\uparrow}, \mathbf{P}_{-s})$ and $(\widehat{P}_s^{c_1\ldots c_{k-1}\uparrow}, \mathbf{P}_{-s})$ or his two allocations differ by exactly one course: he gets c_k under $(\widehat{P}_s^{c_1\ldots c_k\uparrow}, \mathbf{P}_{-s})$ which he does not get under $(\widehat{P}_s^{c_1\ldots c_{k-1}\uparrow}, \mathbf{P}_{-s})$, in exchange for getting c_l, $l > k$ under $(\widehat{P}_s^{c_1\ldots c_{k-1}\uparrow}, \mathbf{P}_{-s})$ that he does not get under $(\widehat{P}_s^{c_1\ldots c_k\uparrow}, \mathbf{P}_{-s})$ for $k \geq 2$

Proof of Claim 2: The proof proceeds along similar lines as the proof of claim 1. Without loss of generality, assume that $\widehat{P}_s^{c_1\ldots c_k\uparrow} \neq \widehat{P}_s^{c_1\ldots c_{k-1}\uparrow}$. Until student s requests c_k under $\widehat{P}_s^{c_1\ldots c_k\uparrow}$, the two games proceed identically. Partition the set of priority orders into $\mathcal{L}_{11}$ (s gets c_k under both strategies), $\mathcal{L}_{10}$ (s gets c_k only under $(\widehat{P}_s^{c_1\ldots c_k\uparrow}, \mathbf{P}_{-s})$) and $\mathcal{L}_{00}$ (s does not get c_k under either strategies). Clearly, for priority orders in $\mathcal{L}_{00}$, the two games proceed identically and s gets the same final allocation.

We claim that s gets also the same final allocation for priority orders $\mathcal{L}_{11}$. To show this, fix λ and let r be the round at which student s requests c_k under $\widehat{P}_s^{c_1\ldots c_{k-1}\uparrow}$and $r' < r$ the round at which he requests c_k under $\widehat{P}_s^{c_1\ldots c_k\uparrow}$. Because c_k fills up earlier than c_l for $l > k$, it means that all courses requested by student s between c_{k-1} and c_k under $\widehat{P}_s^{c_1\ldots c_{k-1}\uparrow}$ are still available at the time of student s's turn in round r under the alternative strategy $\widehat{P}_s^{c_1\ldots c_k\uparrow}$. Thus, by round r student s has the same allocation under both strategies. Because requests are identical across the two games from then on, so are allocations.

Finally, we argue that, under priority orders in $\mathcal{L}_{10}$, student s gets c_k under $\widehat{P}_s^{c_1\ldots c_k\uparrow}$ at the cost of c_l for some $l > k$. The argument here is identical to the argument in the proof of claim 1. There exists a course c_l that student s does not get under $\widehat{P}_s^{c_1\ldots c_k\uparrow}$. From the time of this unsuccessful request, student s's requests are identical across the two strategies. Thus so are his outcomes.

Claim 2, responsiveness and the assumption of vNM preferences over uncertain outcomes implies that student s prefers $\widehat{P}_s^{c_1 \ldots c_k \uparrow}$ to $\widehat{P}_s^{c_1 \ldots c_{k-1} \uparrow}$. Theorem 2(2) then follows from transivity.

Independent preferences: Let $\overline{r}$ be such that $D_c(\overline{r}-1) < 1$ and $D_c(\overline{r}) \geq 1$ for $\mathbf{P}$-popular course c. Under truthful play, all $\mathbf{P}$-popular courses run out exactly in round $\overline{r}$. This also holds for all strategy profiles $(\widehat{P}_s, \mathbf{P}_{-s})$ for any $\widehat{P}_s$. Truthful play guarantees that each student gets his top $\overline{r}-1$ courses. Moreover, it maximizes the chance that he gets $\overline{r}$ $\mathbf{P}$-popular courses, and conditional on getting $\overline{r}$ $\mathbf{P}$-popular courses, the probability distribution it generates on those $\overline{r}$-course bundles first order stochastically dominates the outcome from any alternative (here we use the assumption of responsiveness and the fact that all $\overline{r}$-course bundles differ by a single course, the one requested in round $\overline{r}$, to generate an order over them). Finally, the fact that $\mathbf{P}$-unpopular courses are listed in order of preferences ensures that he gets his $m-\overline{r}$ (or $m+1-\overline{r}$) most preferred courses among them, with no need to resort to the add-drop phase. The claim then follows from responsiveness and the assumption of vNM preferences over uncertain outcomes. QED

A.3 Proof of Lemma 1 (Best-Response Characterization)

Towards a contradiction, suppose that student s's best response $\widehat{P}_s$ involves ranking $\widehat{\mathbf{P}}$-popular course c lower than a $\widehat{\mathbf{P}}$-unpopular course despite the fact that (i) $r_s(c) \leq m$, (ii) $\Pr(c \in a_s(\widehat{\mathbf{P}})) < 1$ and (iii) $\Pr\left(c \in a_s\left(\widehat{\mathbf{P}}^{c \uparrow k}\right)\right) > 0$ where k is the position of the first $\widehat{\mathbf{P}}-$unpopular course on $\widehat{P}_s$.
Let $\widetilde{c}$ denote the last $\widehat{\mathbf{P}}$-unpopular course to appear before c on $\widehat{P}_s$. Moreover, let c' denote the lowest ranked $\widehat{\mathbf{P}}-$popular course in $\widehat{P}_s$ (c' could be c) and let l and l' be the position of c and c' in $\widehat{P}_s$.
We construct an alternative strategy, $\widetilde{P}_s$, by making two changes relative to $\widehat{P}_s$. First, switch the positions of c and $\widetilde{c}$. Second, downgrade $\widetilde{c}$ further down to position l' if there is a popular course below c in $\widehat{P}_s$. Thus, the two strategies can be written as:

$$\begin{array}{llllllll}
 & & & & l & & l' & \\
\underline{\text{If } c \neq c'}: & & & & & & & \\
\widehat{P}_s: & \ldots & \widetilde{c} & \ldots & c & \ldots & c' & \ldots \\
\widetilde{P}_s: & \ldots & c & \ldots & \ldots & c' & \widetilde{c} & \ldots \\
\underline{\text{If } c = c'}: & & & & & & & \\
\widehat{P}_s: & \ldots & \widetilde{c} & \ldots & c & \ldots & & \\
\widetilde{P}_s: & \ldots & c & \ldots & \widetilde{c} & \ldots & &
\end{array} \tag{5}$$

where the dots denote courses that do not change relative positions between $\widehat{P}_s$ and $\widetilde{P}_s$. Partition the set of priority orders into three:

$$\mathcal{L}_1 = \{\lambda | c \in a_s(\widehat{\mathbf{P}}, \boldsymbol{\lambda}) \text{ and } c \in a_s((\widetilde{P}_s, \widehat{\mathbf{P}}_{-s}), \lambda)\}$$
$$\mathcal{L}_2 = \{\lambda | c \notin a_s(\widehat{\mathbf{P}}, \lambda) \text{ and } c \notin a_s((\widetilde{P}_s, \widehat{\mathbf{P}}_{-s}), \lambda)\}$$
$$\mathcal{L}_3 = \{\lambda | c \notin a_s(\widehat{\mathbf{P}}, \lambda) \text{ and } c \in a_s((\widetilde{P}_s, \widehat{\mathbf{P}}_{-s}), \lambda)\}$$

Note that we do not need to consider the case where $c \in a_s(\widehat{\mathbf{P}}, \lambda)$ and $c \notin a_s((\widetilde{P}_s, \widehat{\mathbf{P}}_{-s}), \lambda)$ because this outcome is impossible: for all λ, $\widetilde{P}_s$ requests c strictly earlier than the original strategy.

Claim 1: Student s is weakly better off using $\widetilde{P}_s$ if $\lambda \in \mathcal{L}_1$.

Proof of Claim 1: Because $\widetilde{P}_s$ gets c when it requests it, and $\widehat{P}_s$ gets $\widetilde{c}$ in that round, the two strategies are in synch and obtain identical outcomes until we reach position l in student s's ROL. If $\underline{c \neq c'}$ $\widehat{P}_s$ requests c at that round and, by hypothesis, gets it, while the alternative strategy requests the next course on the list. In other words, $\widetilde{P}_s$ requests courses early from that round onwards until we reach position l' on the ROL or a course that $\widetilde{P}_s$ gets but that $\widehat{P}_s$ does not get. If we reach position l' first, $\widetilde{P}_s$ gets $\widetilde{c}$ and requests are again back in synch. In that case, there is no difference in outcomes between the two strategies. Otherwise, $\widetilde{P}_s$ gets a course that $\widehat{P}_s$ does not get (it could be a popular course or the unpopular course requested by $\widetilde{P}_s$ in round m and never requested under the original strategy because we never reach position l'). $\underline{\text{If } c = c'}$, $\widehat{P}_s$ requests and gets c and $\widetilde{P}_s$ requests and gets $\widetilde{c}$ when we reach position l in the ROL. Requests and outcomes are identical from then on.

To summarize, during the initial allocation, $\widetilde{P}_s$ gets every popular course that $\widehat{P}_s$ receives, plus possibly one additional popular course ranked lower than c on $\widehat{P}_s$ (at the expense of an unpopular course). Because unpopular courses are available with probability one during the add-drop phase, the final schedule that s is able to form using $\widetilde{P}_s$ is at least weakly preferred to that from using $\widehat{P}_s$.

Claim 2: Student s is weakly better off using $\widetilde{P}_s$ if $\lambda \in \mathcal{L}_2$

Proof of Claim 2: By a similar argument as above, we can again show that $\widetilde{P}_s$ gets every popular course that $\widehat{P}_s$ receives, plus possibly one additional. This additional course might be any popular course ranked lower than $\widetilde{c}$ on $\widehat{P}_s$. Because of the add-drop phase, $\widetilde{P}_s$ yields a weakly better outcome than does $\widehat{P}_s$.

Claim 3: Student s is strictly better off using $\widetilde{P}_s$ if $\lambda \in \mathcal{L}_3$.

Proof of claim 3: Again we distinguish two cases depending on whether $c = c'$. $\underline{\text{If } c \neq c'}$, the two strategies are essentially in synch until we reach their requests for c': indeed, except for the fact that $\widehat{P}_s$ gets $\widetilde{c}$ when $\widetilde{P}_s$ gets c, they get the same courses in exactly the same rounds. Once we reach the (successful) request for $\widetilde{c}$ by $\widehat{P}_s$, they become out of synch and $\widetilde{P}_s$ requests courses one round earlier. Because, by construction, all these courses are unpopular and thus available with probability one, the end result is that the allocations from the two strategies differ by exactly one course: $\widehat{P}_s$ gets popular course c that $\widetilde{P}_s$ does not get, at the cost of an unpopular course. Since $r_s(c) \leq m$, and unpopular courses are available with probability one in the add-drop phase, the final schedule that s is able to form using $\widetilde{P}_s$ is strictly preferred to that from using $\widehat{P}_s$. If $c = c'$, the two strategies become out of synch when we reach position l in the ROL. By assumption, $\widehat{P}_s$ is unsuccessful at getting c and $\widetilde{P}_s$ is successful at getting $\widetilde{c}$. From then on, $\widehat{P}_s$ requests courses one round earlier than $\widetilde{P}_s$. Because, by construction, these are unpopular courses, the end result is that the allocations from the initial phase differ by a single course: $\widehat{P}_s$ gets c at the cost of an unpopular course. Since $r_s(c) \leq m$, and unpopular courses are available with probability one in the add-drop phase, the final schedule that s is able to form using $\widetilde{P}_s$ is strictly preferred to that from using $\widehat{P}_s$.

To complete the argument and prove that student s is strictly better off using $\widetilde{P}_s$, we need to argue that $\Pr(\lambda \in \mathcal{L}_3) > 0$. If $\Pr(c \in a_s(\widehat{\mathbf{P}})) \in (0,1)$, this follows from the fact that $\widetilde{P}_s$ requests c strictly earlier than does $\widehat{P}_s$. If, instead, $\Pr(c \in a_s|\widehat{\mathbf{P}}) = 0$ but $\Pr\left(c \in a_s\left(\widehat{\mathbf{P}}^{c\uparrow k}\right)\right) > 0$ where k is the position of the first $\widehat{\mathbf{P}}-$unpopular course on $\widehat{P}_s$, we may need to reiterate the argument using the next unpopular course ahead of c in $\widetilde{P}_s$. Once no more unpopular courses are ahead of c then we can call upon the fact that $\Pr\left(c \in a_s\left(\widehat{\mathbf{P}}^{c\uparrow k}\right)\right) > 0$ where k is the position of the first $\widehat{\mathbf{P}}-$unpopular course on $\widehat{P}_s$. QED

A.4 Proof of Theorem 3 (Equilibrium Characterization)

(i) Fix an equilibrium $\widehat{\mathbf{P}}$. For any priority order λ, every student for whom c belongs to their top$-m$ favorite courses either requests c in the original allocation or requests it in the add-drop phase. Since $D_c(m) > 1$ there exists a positive measure set of students whose requests are rejected.

(ii) Let k denote the number of $\widehat{\mathbf{P}}$-popular courses. If $k > m$ the claim follows trivially. Suppose $k \leq m$ and $\bar{t}_c > k$. Then there exists a positive mass of students who (1) have c among their top-m courses, but who place it in position $k+1$ or below in their submitted ROLs and (2) get course c with probability strictly less than 1. Consider one such student, say s. $\widehat{P}_s$ must contain at least one $\widehat{\mathbf{P}}$-unpopular course, c^*, in the top k positions. This contradicts Lemma 1. QED

A.5 Proof of Lemma 2 (Best-response Characterization with Lexicographic Preferences)

We prove Lemma 2 in two steps. The first step is identical to the proof of Lemma 1.

Step 1: Move unpopular courses down

Suppose that none of (i)-(iii) hold for course c. Consider the deviation strategy $\widetilde{P}_s$, described in (5), possibly repeated so that no unpopular course appears before c in $\widetilde{P}_s$. Then, by the same arguments as claims 1-3 of the proof of Lemma 1, student s is weakly better off using $\widetilde{P}_s$ than using $\widehat{P}_s$. He is strictly better off if $\Pr(c \notin a_s(\widehat{\mathbf{P}}, \lambda)$ and $c \in a_s((\widetilde{P}_s, \widehat{\mathbf{P}}_{-s})) > 0$, i.e. if by doing so student s gets c more often. Note, that unlike in the proof of Lemma 1, we cannot guarantee that $\Pr(c \notin a_s(\widehat{\mathbf{P}}, \lambda)$ and $c \in a_s((\widetilde{P}_s, \widehat{\mathbf{P}}_{-s})) > 0$.

If there exists another popular course that is preceded by an unpopular course and violates conditions (ii) and (iii) of the Lemma repeat the operation until we reach a point where the deviation strategy increases the probability of getting that popular course or there is no more popular course preceded by an unpopular course. In the first case, student s is strictly better off and we are done. In the second case, he is weakly better off.

Step 2: Reorder popular courses

Without loss of generality (given step 1), consider a strategy $\widehat{P}_s$ where all unpopular courses appear after popular courses but where, for some popular course c, conditions (i)-(iii) are all violated. For ease of reference relabel courses such that $P_s : c_1, c_2, c_3, ...$

Consider first c_1. If conditions (i)-(iii) are violated for c_1, then move it up to position 1. Given that this strictly increases the probability of getting it and that student s has lexicographic preferences, student s is strictly better off using this strategy and so we are done.

Suppose next that conditions (i)-(iii) are not violated for c_1 but are violated for c_2. Consider the deviation where the pair of courses $\{c_1, c_2\}$ are moved up to the first and second positions of the ROL, *in the order in which they appear on* $\widehat{P}_s$. This does not decrease the probability that c_1 belongs to the final allocation but strictly increases the probability that c_2 belongs to the final allocation. Given lexicographic preferences, student s is strictly better off, and so we are done.

We can reiterate this argument if conditions (i)-(iii) are satisfied for $c_1, ..., c_k$ but not c_{k+1}, and show that moving up these courses leaves student s strictly better off.

To finish the proof, we need to argue that, if the result of the deviations in Step 1 satisfies conditions (i)-(iii), then these deviations must lead to student s being strictly better off. This follows directly given that conditions (ii)-(iii) were not satisfied by the original strategy. QED.

A.6 Proof of Theorem 4 (Equilibrium with Lexicographic Preferences)

(i). Towards a contradiction assume that $\bar{t}_c > \rho_c$. Then there exists a positive mass of students for whom $r_s(c) \leq \rho_c$ but who request the course later than round ρ_c with strictly positive probability and get rejected with positive probability. These students ranked c in position $\rho_c + 1$ or lower in their ROLs, in contradiction of Lemma 2.

(ii). From the first part of the Theorem, we know that all courses for which $D_c(m) > 1$ are $\widehat{\mathbf{P}}-$popular in any equilibrium. Moreover, by Lemma 2, we know that the only time when these courses are moved down on students' ROLs is when (1) the student nevertheless gets the course for sure, or (2) when even ranking them in their truthful position would yield the course with probability zero. In both cases, this downgrading of course c does not affect the timing of its run-out time relative to truthful play. On the other hand, whenever unpopular courses are moved down these high demand courses are requested weakly earlier. QED.

A.7 Proof of Theorem 5 (Ex-post Efficiency in Special Cases)

(i) An allocation is not ex-post efficient possible if there exist a chain of courses $c_1, c_2, ..., c_k$ and k students such that student s_1 prefers c_1 to c_2 but got c_2 and not c_1, student s_2 prefers c_2 to c_3 but got c_3 and not c_2,... etc, and student s_k prefers c_k to c_1 but got c_1 and not c_k. This means there exists a chain of one-for-one pareto improving trades among those students (given responsiveness, one-for-one trades are the only pareto improving trades we can detect). This implies that, for the particular priority order that generated this allocation, $t_{c_1} < t_{c_2} < ... < t_{c_k} < t_{c_1}$. An impossibility.

(ii) Suppose student s is part of a chain of one-for-one trades. Let c be the course he wants to give and c' the course he gets in return. Then either:

(a) the time of the successful request for $c <$ the time of the unsuccessful request for c', or

(b) the time of the successful request for $c >$ the time of the unsuccessful request for c'.

By Lemma 2, case (a) is impossible. If student s prefers c' to c then he only downgrades it if he is sure to get it. Thus everybody in the trade must be in the situation (b). But this is impossible: it cannot be that everyone gets a course that sells out earlier than the course they give up. QED

A.8 Proof of Theorem 6 (Callousness of RSD)

The probability that the jth student in the random priority order gets his first favorite course is $\frac{Sm-(j-1)}{Sm}$ under HBS, as $j-1$ of the Sm objects have been selected by other students, and which objects were selected is random due to the uniform i.i.d. assumption. For RSD the figure is $\frac{Sm-m(j-1)}{Sm}$, as $m(j-1)$ objects have been randomly selected by the time of j's turn. Taking the arithmetic average over all j yields the desired expressions. QED.

B Robustness Checks

The results of our welfare analysis in Sections 6 and 7 are robust to a variety of alternate specifications for truthful and strategic preferences.

Our first robustness check uses students' May trial-run preferences instead of their July actual-run preferences, both for students' strategic reports and for the construction of their truthful preferences beyond the top five from the May poll data. The advantage of using the May trial-run preferences is that just 10 days elapsed between the May poll and the May trial-run, so there is less reason to worry about social learning, new information and idiosyncratic preference changes. The disadvantage is that the July run is what actually mattered for welfare, and students may have used the May trial-run to learn about equilibrium best responses.

Our second robustness check considers a sub-economy consisting only of the 82% of students whose July strategic reports are consistent with our partial characterization of equilibrium best responses (Lemma 1). Students whose play violates Lemma 1 either are making a strategic error or their preferences changed between May and July. Ideally, our welfare calculations would exclude students whose preferences changed but include students who make strategic errors; this robustness check is likely to make the HBS mechanism look a bit better than it actually is because it excludes the possibility of human error.

Reassuringly, the results of Sections 6 and 7 do not move that much under either of these alternate specifications. Furthermore, they move in roughly the direction we would expect given the issues we are concerned about. For instance, eliminating the Lemma 1 violators somewhat narrows the difference in mean average ranks between truthful and strategic play of the HBS mechanism. And, using the May trial preferences instead of the July actual preferences slightly reduces the number of indeterminacies in Tables 6 and 7.

Another limitation of our analysis is that we rely on a survey to ascertain students' true preferences.[41] Therefore as another simple robustness check we consider a simulation economy in which we can directly control agents' preferences. Specifically, there are 1000 students each of whom require 10 courses, and 100 courses with 120 seats each, for 20% excess capacity as in the data. Students have additive risk-neutral preferences. Student s's value for course c is given by

$$v_{sc} = v_c + \epsilon_{sc}$$

where $v_c \sim U[0,1]$ is a common-value quality component for course c, and $\epsilon_{sc} \sim U[0,1]$ represents s's idiosyncratic taste for c. We then assume that students report their preferences under the

[41] Some of the advantages and disadvantages of using survey data for economic analysis are described in Bertrand and Mullainathan (2001). Fortunately, our context avoids some of the most important disadvantages (as compared e.g. to surveys of political attitudes).

HBS mechanism as if their preferences are

$$\widehat{v}_{sc} = 2v_c + \epsilon_{sc}$$

that is, students over-weight the common-value quality component. Our aim is not to model equilibrium behavior but rather to understand, in a simple and transparent way, the mechanical effects of strategic overreporting of popular courses on the welfare measures we care about.

The main patterns we find in Sections 6 and 7 emerge in this simple simulation as well. Thus, a reader who is skeptical of our use of survey data, but who is persuaded that students are likely to strategically overreport popular classes, should be somewhat willing to believe our basic results.

Table B1 reports the most salient moments of our analysis under each of the above-described specifications. Table B2 reports the social comparison results (CR5-7) for each of the specifications. We then re-run the individual-level comparison results (CR1-4), i.e., Tables 6 and 7 from the main text, under each of the specifications.

Table B1: Robustness Checks: Summary Statistics

	E(Average Rank)	Pr(Get 1st Favorite)	Pr(Get Top Ten)
Main Specification			
HBS - Truthful Play	7.76	82.7%	0.9%
HBS - Strategic Play	8.35	59.7	1.5
RSD - Truthful Play	8.94	47.3	29.4
May Trial Run instead of July Actual Run			
HBS - Truthful Play	7.56	83.7%	1.0%
HBS - Strategic Play	8.09	60.2	1.5
RSD - Truthful Play	8.61	48.3	30.8
Subeconomy with Lemma 1 Compliers			
HBS - Truthful Play	7.94	80.1%	0.4%
HBS - Strategic Play	8.39	58.4	1.6
RSD - Truthful Play	9.12	43.9	28.7
Simulation Economy			
HBS - Truthful Play	15.34	99.1%	0.0%
HBS - Strategic Play	16.38	86.5	0.0
RSD - Truthful Play	19.98	29.8	12.0

Table B2: Robustness Checks: Social Comparison Results

	CR5	CR6(i)	CR6(ii)	CR6(iii)	CR7
HBS Truthful vs. HBS Strategic					
Main Specification	HBS-T	Indet.	HBS-T	HBS-T	HBS-T
May Trial Run instead of July Actual Run	HBS-T	Indet.	HBS-T	HBS-T	HBS-T
Subeconomy with Lemma 1 Compliers	HBS-T	Indet.	HBS-T	HBS-T	HBS-T
Simulation Economy	HBS-T	HBS-T	HBS-T	HBS-T	HBS-T
HBS Strategic vs. RSD					
Main Specification	HBS-S	Indet.	HBS-S	HBS-S	HBS-S
May Trial Run instead of July Actual Run	HBS-S	Indet.	HBS-S	HBS-S	HBS-S
Subeconomy with Lemma 1 Compliers	HBS-S	Indet.	HBS-S	HBS-S	HBS-S
Simulation Economy	HBS-S	Indet.	HBS-S	HBS-S	HBS-S

Table 6 Robustness Check - May Trial Run instead of July Actual Run

	Assumption on Preferences					
	Responsive	Additive	Average-Rank			Lexicographic
	Any Risk Attitude	Risk Neutral	Any Risk Attitude	Risk Averse	Risk Neutral	Risk Neutral
Outcome	(1)	(2)	(3)	(4)	(5)	(6)
Prefers HBS Truthful	46%	47%	56%	64%	68%	88%
Prefers HBS Strategic	8%	8%	14%	21%	30%	11%
Indifferent	1%	1%	1%	1%	1%	1%
Indeterminate	45%	44%	29%	13%	0%	0%

Table 7 Robustness Check - May Trial Run instead of July Actual Run

	Assumption on Preferences					
	Responsive	Additive	Average-Rank			Lexicographic
	Any Risk Attitude	Risk Neutral	Any Risk Attitude	Risk Averse	Risk Neutral	Risk Neutral
Outcome	(1)	(2)	(3)	(4)	(5)	(6)
Prefers RSD	0%	0%	0%	0%	23%	23%
Prefers HBS Strategic	0%	30%	1%	76%	76%	77%
Indifferent	0%	0%	0%	0%	0%	0%
Indeterminate	100%	70%	98%	23%	0%	0%

Table 6 Robustness Check - Subeconomy with Lemma 1 Compliers

	Assumption on Preferences					
	Responsive	Additive	Average-Rank			Lexicographic
	Any Risk Attitude	Risk Neutral	Any Risk Attitude	Risk Averse	Risk Neutral	Risk Neutral
Outcome	(1)	(2)	(3)	(4)	(5)	(6)
Prefers HBS Truthful	45%	46%	53%	63%	67%	87%
Prefers HBS Strategic	9%	9%	16%	20%	31%	11%
Indifferent	1%	1%	1%	1%	2%	1%
Indeterminate	45%	44%	30%	15%	0%	0%

Table 7 Robustness Check - Subeconomy with Lemma 1 Compliers

	Assumption on Preferences					
	Responsive	Additive	Average-Rank			Lexicographic
	Any Risk Attitude	Risk Neutral	Any Risk Attitude	Risk Averse	Risk Neutral	Risk Neutral
Outcome	(1)	(2)	(3)	(4)	(5)	(6)
Prefers RSD	0%	0%	0%	0%	15%	22%
Prefers HBS Strategic	0%	28%	2%	85%	85%	78%
Indifferent	0%	0%	0%	0%	0%	0%
Indeterminate	100%	72%	98%	15%	0%	0%

Table 6 Robustness Check - Simulation Economy

	Assumption on Preferences					
	Responsive	Additive	Average-Rank			Lexicographic
	Any Risk Attitude	Risk Neutral	Any Risk Attitude	Risk Averse	Risk Neutral	Risk Neutral
Outcome	(1)	(2)	(3)	(4)	(5)	(6)
Prefers HBS Truthful	26%	30%	68%	85%	95%	96%
Prefers HBS Strategic	0%	0%	1%	2%	6%	4%
Indifferent	0%	0%	0%	0%	0%	0%
Indeterminate	74%	70%	31%	13%	0%	0%

Table 7 Robustness Check - Simulation Economy

	Assumption on Preferences					
	Responsive	Additive	Average-Rank			Lexicographic
	Any Risk Attitude	Risk Neutral	Any Risk Attitude	Risk Averse	Risk Neutral	Risk Neutral
Outcome	(1)	(2)	(3)	(4)	(5)	(6)
Prefers RSD	0%	0%	0%	0%	1%	2%
Prefers HBS Strategic	0%	42%	0%	10%	99%	98%
Indifferent	0%	0%	0%	0%	0%	0%
Indeterminate	100%	58%	100%	90%	0%	0%

References

[1] Abdukadiroğlu, Atila, Parag Pathak, and Alvin E. Roth (2009). "Strategyproofness versus Efficiency in Matching with Indifferences: Redesigning the NYC High School Match." Forthcoming, American Economic Review.

[2] Abdukadiroğlu, Atila and Tayfun Sönmez (2003). "School Choice: A Mechanism Design Approach." American Economic Review, 93(3), 729-47.

[3] Abdukadiroğlu, Atila, Parag Pathak, and Alvin E. Roth (2005a). "The Boston Public School Match" American Economic Review, Papers and Proceedings, 95(2), 368-71.

[4] Abdukadiroğlu, Atila, Parag Pathak, Alvin E. Roth, and Tayfun Sönmez (2005b). "The New York City High School Match" American Economic Review, Papers and Proceedings, 95(2), 364-67.

[5] Bartlett, Thomas (2008). "Class Warfare: When Getting in is the Hardest Part." Chronicle of Higher Education 54(23), Feb 15, Page A1.

[6] Bertrand, Marianne and Sendhil Mullainathan (2001). "Do People Mean What They Say? Implications for Subjective Survey Data." American Economic Review, 116(3), 901-932.

[7] Bogomolnaia, Anna and Hervé Moulin (2001), A New Solution to the Random Assignment Problem, Journal of Economic Theory, 100, 295-328.

[8] Brams, Steven J., Paul H. Edelman and Peter C. Fishburn (2003). "Fair Division of Indivisible Items." Theory and Decision 55: 147-80.

[9] Brams, Steven J. and Philip D. Straffin, Jr. (1979). "Prisoners' Dilemma and Professional Sports Drafts." The American Mathematical Monthly, 86(2): 80-88.

[10] Brams, Steven J., and Alan D. Taylor (1999). The Win-Win Solution: Guaranteeing Fair Shares to Everybody. W.W. Norton & Company.

[11] Budish, Eric (2009). "The Combinatorial Assignment Problem: Approximate Competitive Equilibrium from Equal Incomes." Mimeo.

[12] Che, Yeon-Koo and Fuhito Kojima (forthcoming). "Asymptotic Equivalence of Probabilistic Serial and Random Priority Mechanisms." Econometrica.

[13] Edelman, Benjamin, Michael Ostrovsky and Michael Schwarz (2007). "Internet Advertising and the Generalized Second-Price Auction: Selling Billions of Dollars Worth of Keywords." American Economic Review, 97(1), 242-59.

[14] Ehlers, Lars and Bettina Klaus (2003). "Coalitional Strategy-proof and Resource-Monotonic Solutions for Multiple Assignment Problems." Social Choice and Welfare, 21: 265-80.

[15] Featherstone, Clayton and Muriel Niederle (2008). "Ex-Ante Efficiency in School Choice Mechanisms: An Experimental Investigation." Mimeo.

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

[17] Gehan, Edmund A. (1965), "A Generalized Wilcoxon Test for Comparing Arbitrarily Singly-Censored Samples." Biometrika, 52 (1 and 2), 203-223.

[18] Gollier, Christian (2001). The Economics of Risk and Time. MIT Press.

[19] Guernsey, Lisa (1999). "Business School Puts Courses in Hands of an On-Line Market." New York Times. Sept 9.

[20] Harsanyi, John C. (1953). "Cardinal Utility in Welfare Economics and in the Theory of Risk-taking." The Journal of Political Economy, 61(5), 434-435.

[21] Hatfield, John William (2007). "Strategy-Proof, Efficiency, and Nonbossy Quota Allocations." Social Choice and Welfare, 33(3), 505-15.

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

[23] Klaus, Bettina and Eiichi Miyagawa (2001). "Strategy-proofness, Solidarity, and Consistency for Multiple Assignment Problems." International Journal of Game Theory, 30: 421-435.

[24] Konishi, Hideo, Thomas Quint and Jun Wako (2001). "On the Shapley-Scarf economy: the case of multiple types of indivisible goods." Journal of Mathematical Economics, 35: 1-15.

[25] Kojima, Fuhito (forthcoming). "Random Assignment of Multiple Indivisible Objects." Mathematical Social Sciences.

[26] Krishna, Aradhna and Utku Unver (2008). "Improving the Efficiency of Course Bidding at Business Schools: Field and Laboratory Studies." Management Science, 27(March/April), 262-82.

[27] Lehrer, Jim (2008). "College Students Squeezed by Rising Costs, Less Aid: Straining the System." NewsHour, aired December 9, 2008. Transcript accessed at http://www.pbs.org/newshour/bb/education/july-dec08/collegecosts_12-09.html.

[28] Manea, Mihai (2007). "Serial Dictatorship and Pareto Optimality." Games and Economic Behavior, 61(2): 316-330.

[29] Miralles, Antonio (2009). "School Choice: The Case for the Boston Mechanism." Mimeo.

[30] Neil, Martha (2008). "NYU Students Seek Coveted Law School Classes, Will Pay Cash." ABA Journal. Jul 28.

[31] Pápai, Szilvia (2001). "Strategyproof and Nonbossy Multiple Assignments." Journal of Public Economic Theory, 3(3): 257-71.

[32] Pathak, Parag (2006). "Lotteries in Student Assignment." Mimeo.

[33] Pathak, Parag and Tayfun Sönmez (2008). "Leveling the Playing Field: Sincere and Sophisticated Players in the Boston Mechanism." American Economic Review, 98(4), 1636-52.

[34] Roth, Alvin E. (1985). "The College Admissions Problem is not Equivalent to the Marriage Problem." Journal of Economic Theory, 36: 277-88.

[35] Roth, Alvin E. and Elliot Peranson (1999). “The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design.” American Economic Review, 89, 748-80.

[36] Roth, Alvin E. (2002). “The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics.” Econometrica, 70(4): 1341-1378.

[37] Roth, Alvin E. (2008). "What Have We Learned from Market Design?" Hahn Lecture, Economic Journal, 118 (March), 285-310.

[38] Roth, Alvin E. and Elliot Peranson (1999). "The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design." American Economic Review, 89, 748-80.

[39] Schmeidler, David (1973). "Equilibrium Points of Nonatomic Games." Journal of Statistical Physics, 7(4), 295-300.

[40] Shapley, Lloyd and Herbert Scarf (1974). “On Cores and Indivisibility.” Journal of Mathematical Economics, 1(1) 23-37.

[41] Sönmez, Tayfun (1999). "Strategy-Proofness and Essentially Single-Valued Cores." Econometrica, 67(3): 677-89.

[42] Sönmez, Tayfun and Utku Unver (2010). "Course Bidding at Business Schools." Forthcoming, International Economic Review.

[43] Varian, Hal R. (2007). "Position Auctions." International Journal of Industrial Organization, 25, 1163-78.

[44] Vickrey, William (1961). “Counterspeculation, Auctions, and Competitive Sealed Tenders.” Journal of Finance, 16, 8-37.

[45] Wilson, Robert (1987). "Game-Theoretic Analyses of Trading Processes," in Advances in Economic Theory: Fifth World Congress, ed. by Timothy Bewley. Cambridge, UK: Cambridge University Press, Chap. 2, 33-70.