What Have We Learned from Market Design?

In the centennial issue of the *Economic Journal*, I wrote (about game theory) that the real test of our success will be not merely how well we understand the general principles that govern economic interactions, but how well we can bring this knowledge to bear on practical questions of microeconomic engineering...” (Roth, 1991)

Since then, economists have gained significant experience in practical market design. One thing we learned from this experience is that transactions and institutions matter at a level of detail that economists have not often had to deal with, and, in this respect, all markets are different. But there are also general lessons. This essay will consider some ways in which markets succeed and fail by looking at some common patterns we see of market failures, and how they have been fixed.

This is a big subject, and I will only scratch the surface by concentrating on markets my colleagues and I helped design in the last few years. My focus will be different than in Roth (2002), where I discussed some lessons learned in the 1990s. The relevant parts of that discussion, which I’ll review briefly in the next section, gathered evidence from a variety of labor market clearinghouses to determine properties of successful clearinghouses, motivated by the redesign of the clearinghouse for new American doctors.¹ Other big market design lessons from the 1990s concern the design of auctions for the sale of radio spectrum and electricity.²

As we have dealt with more market failures, it has become clear that the histories of the American and British markets for new doctors, and the market failures that led to their reorganization into clearinghouses, are far from unique. Other markets have failed for similar reasons, and some have been fixed in similar ways. I’ll discuss common market failures we have seen in recent work on more senior

Alvin E. Roth is George Gund Professor of Economics and Business Administration in the Department of Economics at Harvard University and the Harvard Business School.

This paper was prepared to accompany the Hahn Lecture delivered by Professor Roth at the Royal Economic Society meetings on April 11, 2007, at the University of Warwick. The paper previously appeared in the Economic Journal, 118 (March), 2008, 285-310.

© 2008 Blackwell Publishing Ltd.
innovations / winter 2008 119
medical labor markets, and also on allocation procedures that do not use prices, for school choice in New York City and Boston, and for the allocation of live-donor kidneys for transplantation. These problems were fixed by the design of appropriate clearinghouses. I'll also discuss the North American labor market for new economists, in which related problems are addressed by marketplace mechanisms that leave the market relatively decentralized.

The histories of these markets suggest a number of tasks that markets and allocation systems need to accomplish to perform well. The failure to do these things causes problems that may require changes in how the marketplace is organized.

I'll argue that, to work well marketplaces, need to:

1. Provide thickness—that is, they need to attract a sufficient proportion of potential market participants to come together ready to transact with one another.
2. Overcome the congestion that thickness can bring, by providing enough time, or by making transactions fast enough, so that market participants can consider enough alternative possible transactions to arrive at satisfactory ones.
3. Make it safe to participate in the market as simply as possible, as opposed to transacting outside of the marketplace, or engaging in strategic behavior that reduces overall welfare.

I'll also remark in passing on some other lessons we have started to learn, namely that:

4. Some kinds of transactions are repugnant, and this can be an important constraint on market design.

And, on a methodological note, that:

5. Experiments can play a role, in diagnosing and understanding market failures and successes, in testing new designs, and in communicating results to policy makers.

This paper will be organized as follows. Section 1 will describe some of the relevant history of markets for new doctors, which at different periods had to deal with each of the problems of maintaining thickness, dealing with congestion, and making it safe to participate straightforwardly in the market. In the subsequent sections I'll discuss markets in which these problems showed up in different ways.

Section 2 will review the recent design of regional kidney exchanges in the United States, in which the initial problem was establishing thickness, but in which problems of congestion, and lately, making it safe for transplant centers to participate, have arisen. This is also the market most shaped by the fact that many people find some kinds of transactions repugnant: In particular, buying and selling kidneys for transplantation is illegal in most countries. So, unlike the several labor markets I discuss in this essay, this market operates entirely without money, which will cast into clear focus how the “double coincidence of wants” problems that are most often solved with money can be addressed with computer technology (and will highlight why these problems are difficult to solve even with money, in markets like labor markets in which transactions are heterogeneous).

Section 3 will review the design of the school choice systems for New York City high schools (in which congestion was the immediate problem to be solved), and the design of the new public school choice system in Boston, in which making it
safe to participate straightforwardly was the main issue. These allocation systems also operate without money.

Section 4 will discuss recent changes in the market for American gastroenterologists, who wished to adopt the kind of clearinghouse organization already in place for younger doctors, but who were confronted with some difficulties in making it safe for everyone to change simultaneously from one market organization to another. This involved making changes in the rules of the decentralized market that would precede any clearinghouse even once it was adopted.

This will bring us naturally to a discussion of changes recently made in the decentralized market for new economists in the United States.

1. MARKETS FOR NEW DOCTORS IN THE UNITED STATES, CANADA, AND BRITAIN

The first job American doctors take after graduating from medical school is called a residency. These jobs are a big part of hospitals' labor force, a critical part of physicians' graduate education, and a substantial influence on their future careers. From 1900 to 1945, one way that hospitals competed for new residents was to try to hire residents earlier than other hospitals. This moved the date of appointment earlier, first slowly and then quickly, until by 1945 residents were sometimes being hired almost two years before they would graduate from medical school and begin work.

When I studied this in Roth (1984) it was the first market in which I had seen this kind of “unraveling” of appointment dates, but today we know that unraveling is a common and costly form of market failure. What we see when we study markets in the process of unraveling is that offers not only become increasingly early, but also become dispersed in time and of increasingly short duration. So not only are decisions being made early (before uncertainty is resolved about workers’ preferences or abilities), but also quickly, with applicants having to respond to offers before they can learn what other offers might be forthcoming. Efforts to prevent unraveling are venerable; for example, Roth and Xing (1994) quote Salzman (1931) on laws in various English markets from the 13th century concerning “forestalling” a market by transacting before goods could be offered in the market.

In 1945, American medical schools agreed not to release information about students before a specified date. This helped control the date of the market, but a new problem emerged: hospitals found that if some of the first offers they made were rejected after a period of deliberation, the candidates to whom they wished to make their next offers had often already accepted other positions. This led hospitals to make exploding offers to which candidates had to reply immediately, before they could learn what other offers might be available, and led to a chaotic market that shortened in duration from year to year, and resulted not only in missed agreements but also in broken ones. This kind of congestion also has since been seen in other markets, and in the extreme form it took in the American med-
tical market by the late 1940s, it also constitutes a form of market failure (cf. Roth and Xing 1997, and Avery, Jolls, Roth, and Posner 2007 for detailed accounts of congestion in labor markets in psychology and law).

Faced with a market that was working very badly, the various American medical associations (of hospitals, students, and schools) agreed to employ a centralized clearinghouse to coordinate the market. After students had applied to residency programs and been interviewed, instead of having hospitals make individual offers to which students had to respond immediately, students and residency programs would instead be invited to submit rank order lists to indicate their preferences. That is, hospitals (residency programs) would rank the students they had interviewed, students would rank the hospitals (residency programs) they had interviewed, and a centralized clearinghouse—a matching mechanism—would be employed to produce a matching from the preference lists. Today this centralized clearinghouse is called the National Resident Matching Program (NRMP).

Roth (1984) showed that the algorithm adopted in 1952 produced a matching of students to residency programs that is stable in the sense defined by Gale and Shapley (1962), namely that in terms of the submitted rank order lists, there was never a student and a residency program that were not matched to each other but would have mutually preferred to be matched to each other than to (one of) their assigned match(es). However changes in the market over the years made this more challenging.

For example, one change in the market had to do with the growing number of married couples graduating from American medical schools and wishing to be matched to jobs in the same vicinity. This hadn't been a problem when the match was created in the 1950's, when virtually all medical students were men. Similarly, the changing nature of medical specialization sometimes produced situations in which a student needed to simultaneously be matched to two positions. Roth (1984) showed that these kinds of changes can sometimes make it impossible to find a stable matching, and indeed, an early attempt to deal with couples in a way that did not result in a stable matching had made it difficult to attract high levels of participation by couples in the clearinghouse.

In 1995, I was invited to direct the redesign of the medical match, in response to a crisis in confidence that had developed regarding its ability to continue to serve the medical market, and whether it appropriately served student interests. A critical question was to what extent the stability of the outcome was important to the success of the clearinghouse. Some of the evidence came from the experience of British medical markets. Roth (1990, 1991) had studied the clearinghouses that had been tried in the various regions of the British National Health Service, after those markets unraveled in the 1960s. A Royal Commission had recommended that clearinghouses be established on the American model, but since the American medical literature didn't describe in detail how the clearinghouse worked, each region of the NHS adopted a different algorithm for turning rank order lists into matches, and the unstable mechanisms had largely failed and been abandoned, while the stable mechanisms succeeded and survived.6
What Have We Learned from Market Design?

Of course, there are other differences between regions of the British health service than how they organized their medical clearinghouses, so there was also room for controlled experiments in the laboratory on the effects of stable and unstable clearinghouse. Kagel and Roth (2000) report a laboratory experiment that compared the stable clearinghouse adopted in Edinburgh with the unstable one adopted in Newcastle, and showed that, holding all else constant, the difference in how the two clearinghouses were organized was sufficient to account for the success of the Edinburgh clearinghouse and the failure of the unstable one in Newcastle.

Roth and Peranson (1999) report on the new clearinghouse algorithm that we designed, which aims to always produce a stable matching. It does so in a way that makes it safe for students and hospitals to reveal their preferences. The new algorithm has been used by the NRMP since 1998, and has subsequently been adopted by over three dozen labor market clearinghouses. The empirical evidence that has developed in use is that the set of stable matchings is very seldom empty.

An interesting historical note is that the use of stable clearinghouses has been explicitly recognized as part of a pro-competitive market mechanism in American law. This came about because in 2002, sixteen law firms representing three former medical residents brought a class-action antitrust suit challenging the use of the matching system for medical residents. The theory of the suit was that the matching system was a conspiracy to hold down wages for residents and fellows, in violation of the Sherman Antitrust Act.

Niederle and Roth (2003a) observed that, empirically, the wages of medical specialties with and without centralized matching in fact do not differ. The case was dismissed after the U.S. Congress passed new legislation in 2004 (contained in Public Law 108-218), noting that the medical match is a pro-competitive market mechanism, not a conspiracy in restraint of trade. This reflected modern research on the market failures that preceded the adoption of the first medical clearinghouse in the 1950s, which brings us back to the main subject of the present paper.

To summarize, the study and design of a range of clearinghouses in the 1980s and 1990s made clear that producing a stable matching is an important contributor to the success of a labor clearinghouse. For the purposes of the present paper, note that such a clearinghouse can persistently attract the participation of a high proportion of the potential participants, and when it does so it solves the problem of establishing a thick market. A computerized clearinghouse like those in use for medical labor markets also solves the congestion problem, since all the operations of the clearinghouse can be conducted essentially simultaneously, in that the outcome is determined only after the clearinghouse has cleared the market. And, as mentioned briefly, these clearinghouses can be designed to make it safe for participants to reveal their true preferences, without running a risk that by doing so they will receive a worse outcome than if they had behaved strategically and stated some other preferences.

In the following sections, we’ll see more about how the failure to perform these tasks can cause markets to fail.
2. KIDNEY EXCHANGE

Kidney transplantation is the treatment of choice for end-stage renal disease, but there is a grave shortage of transplantable kidneys. In the United States there are over 70,000 patients on the waiting list for cadaver kidneys, but in 2006 fewer than 11,000 transplants of cadaver kidneys were performed. In the same year, around 5,000 patients either died while on the waiting list or were removed from the list as “Too Sick to Transplant.” This situation is far from unique to the United States: In the UK at the end of 2006 there were over 6,000 people on the waiting list for cadaver kidneys, and only 1,240 such transplants were performed that year.

Because healthy people have two kidneys and can remain healthy with just one, it is also possible for a healthy person to donate a kidney, and a live-donor kidney has a greater chance of long-term success than does one from a deceased donor. However, good health and good will are not sufficient for a donor to be able to give a kidney to a particular patient: the patient and donor may be biologically incompatible because of blood type, or because the patient’s immune system has already produced antibodies to some of the donor’s proteins. In the United States in 2006 there were 6,428 transplants of kidneys from living donors (in the UK there were 590).

The total supply of transplantable kidneys (from deceased and living donors) clearly falls far short of the demand, but it is illegal in almost all countries to buy or sell kidneys for transplantation. This legislation is the expression of the fact that many people find the prospect of such a monetized market highly repugnant (see Roth 2007a).

So, while a number of economists have devoted themselves to the task of repealing or relaxing laws against compensating organ donors (see, e.g. Becker and Elias 2007, and the discussion of Elias and Roth 2007), another task that faces a market designer is how to increase the number of transplants subject to existing constraints, including those that forbid monetary incentives.

It turns out that, prior to 2004, in just a very few cases, incompatible patient-donor pairs and their surgeons had managed to arrange an exchange of donor kidneys (sometimes called “paired donation”), when the patient in each of two incompatible patient-donor pairs was compatible with the donor in the other pair, so that each patient received a kidney from the other’s donor. Sometimes a different kind of exchange had also been accomplished, called a list exchange, in which a patient’s incompatible donor donated a kidney to someone who (by virtue of waiting a long time) had high priority on the waiting list for a cadaver kidney, and in return the donor’s intended patient received high priority to receive the next compatible cadaver kidney that became available. Prior to December 2004 only five exchanges had been accomplished at the fourteen transplant centers in New England. Some exchanges had also been accomplished at Johns Hopkins in Baltimore, and among transplant centers in Ohio. So, these forms of exchange were feasible and non-repugnant. Why had so very few happened?

One big reason had to do with the (lack of) thickness of the market, i.e. the size
of the pool of incompatible patient-donor pairs who might be candidates for exchange. When a kidney patient brought a potential donor to his or her doctor to be tested for compatibility, donors who were found to be incompatible with their patient were mostly just sent home. They were not patients themselves, and often no medical record at all was retained to indicate that they might be available. And in any event, medical privacy laws made these potential donors’ medical information unavailable.

Roth, Sönmez, and Ünver (2004a) showed that in principle a substantial increase in the number of transplants could be anticipated from an appropriately designed clearinghouse that assembled a database of incompatible patient-donor pairs. That paper considered exchanges with no restrictions on their size, and allowed list exchange to be integrated with exchange among incompatible patient-donor pairs. That is, exchanges could be a cycle of incompatible patient-donor pairs of any size such that the donor in the first pair donated a kidney to the patient in the second, the second pair donated to the third, and so on, until the cycle closed with the last pair donating to the first. And pairs that would have been interested in a list exchange in which they donated a kidney in exchange for high priority on the cadaver waiting list could be integrated with the exchange pool by having them donate to another incompatible pair in a chain that would end with donation to the waiting list.

We sent copies of that paper to many kidney surgeons, and one of them, Frank Delmonico (the medical director of the New England Organ Bank), came to lunch to pursue the conversation. Out of that conversation, which grew to include many others (and led to modifications of our original proposals), came the New England Program for Kidney Exchange, which unites the fourteen kidney transplant centers in New England to allow incompatible patient-donor pairs from anywhere in the region to find exchanges with other such pairs.

For incentive and other reasons, all such exchanges have been done simultaneously, to avoid the possibility of a donor becoming unwilling or unable to donate a kidney after that donor’s intended patient has already received a kidney from another patient’s donor. So, one form that congestion takes in organizing kidney exchanges is that multiple operating rooms and surgical teams have to be assembled. (A simultaneous exchange between two pairs requires four operating rooms and surgical teams, two for the nephrectomies that remove the donor kidneys, and two for the transplantations that immediately follow. An exchange involving three pairs involves six operating rooms and teams, etc.) Roth et al. (2004a) noted that large exchanges would arise relatively infrequently, but could pose logistical difficulties.

These logistical difficulties loomed large in our early discussions with surgeons, and out of those discussions came the analysis in Roth, Sönmez, and Ünver (2005a) of how kidney exchanges might be organized if only two-way exchanges were feasible. The problem of two-way exchanges can be modeled as a classic problem in graph theory, and, subject to the constraint that exchanges involve no more than two pairs, efficient outcomes with good incentive properties can be found in
innovations / winter 2008

computationally efficient ways. When the New England Program for Kidney Exchange was founded in 2004 (Roth et al. 2005b), it used the matching software that had had been developed to run the simulations in Roth et al. (2005a,b), and it initially attempted only two-way matches (while keeping track of the potential three-way matches that were missed). This was also the case when Sönmez, Ünver, and I started running matches for the Ohio-based consortium of transplant centers that eventually became the Alliance for Paired Donation.12

However, some transplants are lost that could have been accomplished if three-way exchanges were available. In Saidman, Roth, Sönmez, Ünver, and Delmonico (2006) and in Roth, Sönmez, and Ünver (2007), we showed that to get close to the efficient number of transplants, the infrastructure to perform both two- and three-way exchanges would have to be developed, but that once the population of available patient-donor pairs was large enough, few transplants would be missed if exchanges among more than three pairs remained difficult to accomplish. Both the New England Program for Kidney Exchange and the Alliance for Paired Donation have since taken steps to be able to accommodate three-way as well as two-way exchanges. Being able to deal with the (six operating room) congestion required to accomplish three-way exchanges has the effect of making the market thicker, since it creates more exchange possibilities.

As noted above, another way to make the market thicker is to integrate exchange between pairs with list exchange, so that exchange chains can be considered, as well as cycles. This applies as well to how the growing numbers of non-directed (altruistic) donors are used. A non-directed (ND) donor is someone who wishes to donate a kidney without having a particular patient in mind (and whose donor kidney therefore doesn’t require another donor kidney in exchange). The traditional way to utilize such non-directed donors was to have them donate to someone on the cadaver waiting list. But as exchanges have started to operate, it has now become practical to have the ND donor donate to some pair that is willing to exchange a kidney, and have that pair donate to someone on the cadaver waiting list. Roth, Sönmez, Ünver, Delmonico and Saidman (2006) report on how and why such exchanges are now done in New England. As in traditional exchange, all surgeries are conducted simultaneously, so there are logistical limits on how long a chain is feasible. But we noted that, when a chain is initiated by a ND donor, it might be possible to relax the constraints that all parts of the exchange be simultaneous, because

if something goes wrong in subsequent transplants and the whole ND-chain cannot be completed, the worst outcome will be no donated kidney being sent to the waitlist and the ND donation would entirely benefit the KPD [kidney exchange] pool. (Roth et al. 2006, p. 2704).

That is, if a conventional exchange were done in a non-simultaneous way, and if the exchange broke down after some patient-donor pair had donated a kidney but before they had received one, then that pair would not only have lost the promised transplant, but also have lost a healthy kidney. In particular, the patient would no
What Have We Learned from Market Design?

longer be in position to exchange with other incompatible patient-donor pairs. But in a chain that begins with an ND donor, if the exchange breaks down before the donation to some patient-donor pair has been made (because the previous donor in the chain becomes unwilling or unable to donate), then the pair loses the promised transplant, but is no worse off than they were before the exchange was planned, and in particular they can still exchange with other pairs in the future. So, while a non-simultaneous ND chain of donations could create an incentive to break the chain, the costs of a breach would be less than in a pure exchange, and so the benefits (in terms of longer chains) are worth exploring. The first such non-simultaneous “Never Ending” Altruistic Donor (NEAD) chain was begun by the Alliance for Paired Donation in July 2007. A week after the first patient was transplanted from an altruistic (ND) donor, her husband donated a kidney to another patient, whose mother later donated her kidney to a third patient whose daughter donated (simultaneously) to a fourth patient, whose sister is, as I write, now waiting to donate to another patient whose incompatible donor will be willing to “pass it forward” (Rees et al. 2007).

To summarize the progress to date, the big problem facing kidney exchange prior to 2004 was the lack of thickness in the market, so that incompatible patient-donor pairs were left in the difficult search for what Jevons famously described as a double coincidence of wants (Jevons 1876; Roth et al. 2007). By building a database of incompatible patient-donor pairs and their relevant medical data, it became possible to arrange more transplants, using a clearinghouse to maximize the number (or some quality- or priority-adjusted number) of transplants subject to various constraints. The state of the art now involves both two- and three-way cyclical exchanges and a variety of chains, either ending with a donation to someone on the cadaver waiting list or beginning with an altruistic non-directed donor, or both. While large simultaneous exchanges remain logistically infeasible, the fact that almost all efficient exchanges can be accomplished in cycles of no more than three pairs, together with clearinghouse technology that can efficiently find such sets of exchanges, substantially reduces the problem of congestion in carrying out exchanges. And, for chains that begin with non-directed donors, the early evidence is that some relaxation of the incentive constraint that all surgeries be simultaneous seems to be possible.

There remain some challenges to further advancing kidney exchange that are also related to thickness, congestion, and incentives.

Some patients have many antibodies, so that they will need very many possible donors to find one who is compatible. For that reason and others, it is unlikely that purely regional exchanges, such as presently exist, will provide adequate thickness for all the gains from exchange to be realized. Legislation has recently been passed in the U.S. House and Senate to remove a potential legal obstacle to a national kidney exchange. Aside from expanding kidney exchange to a national scale, another way to increase the thickness of the market would be to make kidney exchange available not just to incompatible patient-donor pairs, but also to those who are compatible but might nevertheless benefit from exchange.
Regarding congestion, while some of the congestion in terms of actually conducting transplants has been addressed, there is still congestion associated with the time it takes to test for immunological incompatibility between patients and donors who (based on available tests) are matched to be part of an exchange. That is, antibody production can vary over time, and so a patient and donor who appear to be compatible in the database may not in fact be. Because it now sometimes takes weeks to establish this, during which time other exchanges may go forward, sometimes exchanges are missed that could have been accomplished if the tests for compatibility could be done more quickly, so that the overall pattern of exchanges could have been adjusted.

And as regional exchanges have grown to include multiple transplant centers, a new issue has come to the fore concerning how kidney exchange should be organized to give transplant centers the incentive to inform the central exchange of all of their incompatible patient-donor pairs. Consider a situation in which transplant center A has two pairs that are mutually compatible, so that it could perform an in-house exchange between these two pairs. If the mutual compatibilities are as shown in Figure 1a, then if these two pairs exchange with each other, only those two transplants will be accomplished. If instead the pairs from transplant center A were matched with the pairs from the other centers, as shown in Figure 1a, four transplants could be accomplished (via exchanges of pair A1 with pair B, and pair A2 with C).

Double-headed arrows indicate that the connected pairs are compatible for exchange, i.e. the patient in one pair is compatible with the donor in the other. Pairs A1 and A2 are both from transplant center A, pairs B and C are from different transplant centers. Transplant center A, which sees only its own pairs, can conduct an exchange among its pairs A1 and A2 since they are compatible, and, if it does so, this will be the only exchange, resulting in two transplants. However, if in Figure 1a transplant center A makes its pairs available for exchange with other centers, then the exchanges will be A1 with B and A2 with C, resulting in four transplants. However, in Figure 1b the suggested exchange might be A1 with B, which would leave the patient in A2 without a transplant. Faced with this possibility (and not knowing if the situation is as in 1a or 1b), transplant center A might choose to transplant A1 and A2 by itself, without informing the central exchange.

But, note that if the situation had been that of Figure 1b, then transplant center A runs the risk that if it informs the central exchange of its pairs, then the recommended exchange will be between A1 and B, since B has high priority (e.g. B is a child). This would mean that pair A2 did not get a kidney, as they would have if A1 and A2 had exchanged inhouse. So, the situation facing transplant center A, not knowing what pairs will be put forward for exchange by the other transplant centers, is that it can assure itself of doing two transplants for its patients in pairs A1 and A2, but it is not guaranteed two transplants if it makes the pairs available for exchange and the situation is as in Figure 1b. If this causes transplant centers to withhold those pairs they can transplant by themselves, then a loss to society results in case the situation is as in Figure 1a. (In fact, if transplant centers with-
What Have We Learned from Market Design?

Figures 1a and 1b. Double headed arrows indicate that the connected pairs are compatible for exchange, i.e. the patient in one pair is compatible with the donor in the other. Pairs A1 and A2 are both from transplant center A, pairs B and C are from different transplant centers. Transplant center A, which sees only its own pairs, can conduct an exchange among its pairs A1 and A2 since they are compatible, and, if it does so, this will be the only exchange, resulting in two transplants. However, if in Figure 1a transplant center A makes its pairs available for exchange with other centers, then the exchanges will be A1 with B and A2 with C, resulting in four transplants. However in Figure 1b the suggested exchange might be A1 with B, which would leave the patient in A2 without a transplant. Faced with this possibility (and not knowing if the situation is as in 1a or 1b) transplant center A might choose to transplant A1 and A2 by itself, without informing the central exchange.

hold those pairs they can exchange in-house, then primarily hard-to-match pairs will be offered for exchange, and the loss will be considerable.)

One remedy is to organize the kidney exchange clearinghouse in such a way that guarantees center A that any pairs it could exchange in-house will receive transplants.

This would allow the maximal number of transplants to be achieved in situation 1a, and it would mean that in situation 1b the exchange between A1 and A2 would be made (and so the high priority pair B would not participate in exchange, just as they would not have if pairs A1 and A2 had not been put forward). This is a bit of a hard discussion to have with surgeons who find it repugnant that, for example, the child patient in pair B would receive lower priority than pairs A1 and A2 just because of the accident that they were mutually compatible and were being treated at the same transplant center. (Needless to say, if transplant center A withholds its pairs and transplants them in-house, they effectively have higher priority than pair B, even if no central decision to that effect has been made.) But this is an issue that will have to be resolved, because the full participation of all transplant centers substantially increases the efficiency of exchange.

innovations / winter 2008

129
Note that, despite all the detailed technical particulars that surround the establishment of kidney exchange programs, and despite the absence of money in the kidney exchange market, we can recognize some of the basic lessons of market design that were also present in designing labor market clearinghouses. The first issue was making the market thick, by establishing a database of patient-donor pairs available to participate in exchange. Then issues of congestion had to be dealt with, so that the clearinghouse could identify exchanges involving sufficiently few pairs (initially two, now three) so that they could be accomplished simultaneously. Simultaneity is related to making sure that everyone involved in an exchange never has an incentive not to go forward with it, but as exchanges have grown to include multiple transplant centers, there are also incentive issues to be resolved in making it safe for a transplant center to enroll all of its eligible pairs in the central exchange.

3. SCHOOL CHOICE

Another important class of allocation problems in which no money changes hands is the assignment of children to big-city public schools, based both on the preferences of students and their families, and on the preferences of schools, or on city priorities. Because public school students must use whatever system local authorities establish, establishing a thick market is not the main problem facing such systems. (Although how well a school choice system works may influence how many children ultimately attend city schools). But how well a school-choice system works still has to do with how effectively it deals with congestion, and how safe it makes it for families to straightforwardly reveal their preferences.

My colleagues and I were invited to help design the current New York City high school choice program chiefly because of problems the old decentralized system had in dealing with congestion. In Boston we were invited to help design the current school choice system because the old system, which was itself a centralized clearinghouse, did not make it safe for families to state their preferences. In both Boston and New York City the newly designed systems incorporate clearinghouses to which students (and, in New York City, schools) submit preferences. Although another alternative was considered in Boston, both Boston and New York City adopted clearinghouses similar to the kinds of stable clearinghouses used in medical labor markets (powered by a student-proposing deferred acceptance algorithm), adapted to the local situations. For my purpose in the present paper I’ll skip any detailed discussion of the clearinghouse designs, except to note that they make it safe for students and families to submit their true preferences. Instead, I’ll describe briefly what made the prior school choice systems congested or risky.

In New York City, well over 90,000 students a year must be assigned to over 500 high school programs. Under the old system, students were asked to fill out a rank order list of up to five programs. These lists were then copied and sent to the schools. Subject to various constraints, schools could decide which of their applicants to accept, waitlist, or reject. Each applicant received a letter from the New
York City Department of Education with the decisions of the schools to which she had applied, and applicants were allowed to accept no more than one offer, and one waitlist. This process was repeated: after the responses to the first letter were received, schools with vacant positions could make new offers, and after replies to the second letter were received, a third letter with new offers was sent. Students not assigned after the third step were assigned to their zoned schools, or assigned via an administrative process. There was an appeals process, and an “over the counter” process for assigning students who had changed addresses, or were otherwise unassigned before school began.

Three rounds of processing applications to no more than five out of more than 500 programs by almost 100,000 students was insufficient to allocate all the students. That is, this process suffered from congestion (in precisely the sense explored in Roth and Xing 1997): not enough offers and acceptances could be made to clear the market. Only about 50,000 students received offers initially, about 17,000 of whom received multiple offers. And when the process concluded, approximately 30,000 students had been assigned to a school that was nowhere on their choice list.

Three features of this process particularly motivated the NYCDOE desire for a new matching system. First were the approximately 30,000 students not assigned to a school they had chosen. Second, students and their families had to be strategic in their choices. Students who had a substantial chance of being rejected by their true first choice school had to think about the risk of listing it first, since, if one of their lower choice schools took students’ rankings into account in deciding on admissions, they might have done better to list it first. (More on this in a moment, in the discussion of Boston schools.) Finally, the many unmatched students, plus those who may not have indicated their true preferences (and the consequent instability of the resulting matching), gave schools an incentive to be strategic: a substantial number of schools managed to conceal capacity from the central administration, thus preserving places that could be filled later with students unhappy with their assignments.

As soon as New York City adopted a stable clearinghouse for high school matching (in 2003, for students entering high school in 2004), the congestion problem was solved; only about 3,000 students a year have had to be assigned administratively since then, down from 30,000 (and many of these are students who for one reason or another fail to submit preference lists). In addition, in the first three years of operation, schools learned that it was no longer profitable to withhold capacity, and the resulting increase in the availability of places in desirable schools resulted in a larger number of students receiving their first choices, second choices, and so forth from year to year. Finally, as submitted rank order lists have begun to more reliably reflect true preferences, these have begun to be used as data for the politically complex process of closing or reforming undesirable schools (Abdulkadiroglu, Pathak, and Roth, 2005 and 2007).

In Boston, the problem was different. The old school choice system there made it risky for parents to indicate their true first choice school if it was not their local
school. The old system was simple in conception: parents ranked schools, and the old Boston algorithm tried to give as many families as possible their first choice school. In case the capacity of a school was less than the number of students who ranked it first, ties were broken by giving priority to students who had siblings in the school, or who lived within walking distance, or, finally, who had been assigned a good lottery number. After these assignments were made, the old Boston algorithm tried to match as many remaining students as possible with their second choice school, and so on. The difficulty facing families was that, if they ranked a popular school first and weren’t assigned to it, they might find that by the time they were considered for their second choice school, it was already filled with people who had ranked it first. So, a family that had a high priority for their second choice school (e.g. because they lived close to it), and could have been assigned to it if they had ranked it first, might no longer be able to get in if they ranked it second.

As a consequence, many families were faced with difficult strategic decisions, and some families devoted considerable resources to gathering relevant information about the capacities of schools, how many siblings would be enrolling in kindergarten, etc. Other families were oblivious to the strategic difficulties, and sometimes suffered the consequences; if they listed popular schools for which they had low priority, they were often assigned to schools they liked very little.

In Boston, the individual schools are not actors in the school choice process, and so there was a wider variety of mechanisms to choose from than in New York. My colleagues and I recommended two possibilities that were strategy-proof (in the sense that they make it a dominant strategy for students and families to submit their true preferences), and which thus would make it safe for students to submit their true preferences (Abdulkadiroglu, Pathak, Sönmez, and Roth 2005 and 2007). This proved to be decisive in persuading the Boston School Committee to adopt a new algorithm. Then Superintendent of Schools Thomas Payzant wrote, in a 2005 memo to the School committee:

The most compelling argument for moving to a new algorithm is to enable families to list their true choices of schools without jeopardizing their chances of being assigned to any school by doing so.

Superintendent Payzant further wrote:

A strategy-proof algorithm levels the playing field by diminishing the harm done to parents who do not strategize or do not strategize well.

Making the school choice system safe to participate in was critical in the decision of Boston Public Schools to move from a clearinghouse that was not strategy-proof to one that was. Different issues of safety were critical in the market for Gastroenterologists, discussed next.

4. GASTROENTEROLOGISTS

An American medical graduate who wishes to become a gastroenterologist first
What Have We Learned from Market Design?

completes three years of residency in internal medicine, and then applies for a job as a fellow in gastroenterology, a subspecialty of internal medicine. The market for gastroenterology fellows was organized via a stable labor market clearinghouse (a “match”) from 1986 through the late 1990s, after which the match was abandoned (following an unexpected shock to the supply and demand for positions in 1996; see Mckinney, Niederle, and Roth 2005). This provided an opportunity to observe the unraveling of a market as it took place. From the late 1990s until 2006, offers of positions were made increasingly far in advance of employment (moving back to almost two years in advance, so that candidates were often being interviewed early in their second year of residency). Offers also became dispersed in time, and short in duration, so that candidates faced a thin market. One consequence was that the market became much more local than it had been, with gastroenterology fellows more likely to be recruited at the same hospital at which they had worked as a resident (Niederle and Roth 2003; Niederle, Proctor, and Roth 2006).

Faced with these problems, the various professional organizations involved in the market for gastroenterology fellows agreed to try to resume using a centralized clearinghouse, to be operated one year in advance of employment. However, this raised the question of how to make it safe for program directors and applicants to wait for the clearinghouse, which would operate almost a year later than hiring had been accomplished in the immediate past. Program directors who wanted to wait for the match worried that if their competitors made early offers, then applicants would lose confidence that the match would work and consequently would accept those early offers. That is, in the first year of a match, applicants might not yet feel safe to reject an early offer in order to wait for the match. Program directors who worried about their competitors might thus be more inclined to make early offers themselves.

The gastroenterology organizations did not feel able to directly influence the hiring behavior of programs that might not wish to wait for the match. Consequently, we recommended that policies be adopted that would allow applicants who wished to wait for the match to more effectively deal with early offers themselves (Niederle, Proctor, and Roth 2006). We modeled our recommendation on the policies in place in the American market for graduate school admission. In this market, a policy (adopted by the large majority of universities) states that offers of admission and financial support to graduate students should remain open until April 15:

Students are under no obligation to respond to offers of financial support prior to April 15; earlier deadlines for acceptance of such offers violate the intent of this Resolution. In those instances in which a student accepts an offer before April 15, and subsequently desires to withdraw that acceptance, the student may submit in writing a resignation of the appointment at any time through April 15. However, an acceptance given or left in force after April 15 commits the student not to accept another
offer without first obtaining a written release from the institution to which a commitment has been made. Similarly, an offer by an institution after April 15 is conditional on presentation by the student of the written release from any previously accepted offer. It is further agreed by the institutions and organizations subscribing to the above Resolution that a copy of this Resolution should accompany every scholarship, fellowship, traineeship, and assistantship offer.

(See http://www.cgsnet.org/portals/0/pdf/CGSResolutionJune2005.pdf.)

This of course makes early exploding offers much less profitable. A program that might be inclined to insist on an against-the-rules early response is discouraged from doing so, because they can’t “lock up” a student to whom they make such an offer, because accepting such an offer does not prevent the student from later receiving and accepting a preferred offer.21

A modified version of this policy was adopted by all four major Gastroenterology professional organizations, the American Gastroenterological Association (AGA), the American College of Gastroenterology (ACG), the American Society for Gastrointestinal Endoscopy (ASGE), and the American Association for the Study of Liver Diseases (AASLD), regarding offers made before the (new) match. Their resolution states, in part,

The general spirit of this resolution is that each applicant should have an opportunity to consider all programs before making a decision and be able to participate in the Match. … It therefore seeks to create rules that give both programs and applicants the confidence that applicants and positions will remain available to be filled through the Match and not withdrawn in advance of it. This resolution addresses the issue that some applicants may be persuaded or coerced to make commitments prior to, or outside of, the Match. … Any applicant may participate in the matching process … by … resigning the accepted position if he/she wishes to submit a rank order list of programs … The spirit of this resolution is to make it unprofitable for program directors to press applicants to accept early offers, and to give applicants an opportunity to consider all offers.

The gastroenterology match for 2007 fellows was held June 21, 2006, and succeeded in attracting 121 of the 154 eligible fellowship programs (79%). 98% of the positions offered in the match were filled through the match, and so it appears that the gastroenterology community succeeded in making it safe to participate in the match, and thus in changing the timing and thickness of the market, while using a clearinghouse to avoid congestion.

The policies adopted by gastroenterologists prior to their match make clear that market design in this case consists not only of the “hardware” of a centralized clearinghouse, but also rules and understandings that constitute elements of “market culture.” This leads us naturally to consider how issues of timing, thickness, and congestion are addressed in a market that operates without any centralized clearinghouse.
The North American market for new Ph.D.s in economics is a fairly decentralized market, with some centralized marketplace institutions, most of them established by the American Economics Association (AEA). Some of these institutions are of long standing, while others have only recently been established. Since 2005 the American Economic Association has had an Ad Hoc Committee on the Job Market, charged with considering ways in which the market for economists might be facilitated.

Roughly speaking, the main part of this market begins each year in the early fall, when economics departments advertise for positions. Positions may be advertised in many ways, but a fairly complete picture of the academic part of the market can be obtained from the AEA’s monthly publication, Job Openings for Economists (JOE), which provides a central location for employers to advertise and for job seekers to see who is hiring (http://www.aeaweb.org/joe/). Graduate students nearing completion of their Ph.D.s answer the ads by sending applications, which are followed by letters of reference, most typically from their faculty advisors.

Departments often receive several hundred applications (because it is easy for applicants to apply to many schools), and junior recruiting committees work through the late fall to read applications, papers, and letters, and to seek information through informal networks of colleagues, to identify small subsets of applicants they will invite for half-hour preliminary interviews at the annual AEA meeting in early January. This is part of a very large annual set of meetings of the Allied Social Science Associations, which consist of the AEA and almost 50 smaller associations. Departments reserve suites for interviewing candidates at the meeting hotels, and young economists in new suits commute up and down the elevators, from one interview to another, while recruiting teams interview candidates one after the other, trading off with their colleagues throughout long days. While the interviews in hotel suites are normally pre-arranged in December, the meetings also host a spot market, in a large hall full of tables, at which both academic and non-academic employers can arrange at the last minute to meet with candidates. The spot market is called the Illinois Skills Match (because it is organized in conjunction with the Illinois Department of Employment Security).

These meetings make the early part of the market thick, by providing an easy way for departments to quickly meet lots of candidates, and by allowing candidates to efficiently introduce themselves to many departments. This largely controls the starting time of the market. Although a small amount of interviewing goes on beforehand, it is quite rare to hear of departments that make offers before the meetings, and even rarer to hear of departments pressing candidates for replies before the meetings.

But while the preliminary-interviewing part of the market is thick, it is congested. A dedicated recruiting committee might possibly be able to interview thirty candidates, but not a hundred, and hence can meet only a small fraction of the

innovations / winter 2008
available applicants. Thus the decision of who to interview at the meetings is an important one, and for all but elite schools a strategic one as well. That is, while a relatively few departments at the top of the pecking order can simply interview the candidates they like best, a lower ranked department that uses all its interview slots to interview the same candidates who are interviewed by the elite schools is likely to find that it cannot convert its initial interviews into new faculty hires. Thus, most schools have to give at least some thought not only to how much they like each candidate, but to how likely it is that they can successfully hire that candidate. This problem is only made more difficult by the fact that students can easily apply for many positions, so the act of sending an application does not itself send a strong signal of how interested the candidate might be. The problem may be particularly acute for schools in somewhat special situations, such as liberal arts colleges, or British and other non-American universities in which English is the language of instruction, since these may be concerned that some students who strongly prefer positions at North American research universities may apply to them only as insurance.

Following the January meetings, the market moves into a less organized phase, in which departments invite candidates for “flyouts,” day-long campus visits during which the candidate will make a presentation and meet a substantial portion of the department faculty and perhaps a dean. Here too, the market is congested, and departments can fly out only a small subset of the candidates they have interviewed at the meetings, because of the costs of various sorts.27 This part of the market is less well coordinated in time: some departments host flyouts already in January, while others wait until later. Some departments try to complete all their flyouts before making any offers, while others make offers while still interviewing. And some departments make offers that come with moderate deadlines of two weeks or so, which may nevertheless force candidates to reply to an offer before knowing what other offers might be forthcoming.28

By late March, the market starts to become thin. For example, a department that interviewed twenty people at the meetings, invited six for flyouts, made offers to two, and was rejected by both may find that it is now difficult to assess which candidates who it did not interview may still be on the market. Similarly, candidates whose interviews and flyouts did not result in job offers may find it difficult to know which departments are still actively searching. To make the late part of the market thicker, the first thing our AEA job market committee did was to institute a “scramble” webpage through which departments with unfilled positions and applicants still on the market could identify each other (see Guide to the Economics Job Market Scramble at http://www.aeaweb.org/joe/scramble/guide.pdf.) For simplicity, the scramble webpage was passive (i.e. it didn’t provide messaging or matching facilities), it simply announced the availability of any applicant or department who chose to register. The scramble webpage operated for the first time in the latter part of the 2005-2006 job market, when it was open for registrants between March 15 and 20, and was used by 70 employers and 518 applicants (of which only about half were new,
What Have We Learned from Market Design?

2006 Ph.D.s). It was open only briefly, so that its information provided a snapshot of the late market, which didn’t have to be maintained to prevent the information from becoming stale.

The following year our committee sought to alleviate some of the congestion surrounding the selection of interview candidates at the January meetings by introducing a signaling mechanism through which applicants could have the AEA transmit to no more than two departments a signal indicating their interest in an interview at the meetings. The idea was that by limiting applicants to two signals, each signal would have some informational value that might not be contained merely in the act of sending a department an application, and that this information might be helpful in averting coordination failures.29

The signaling mechanism operated for the first time in December 2006, and about 1,000 people used it to send signals.30

Both the scramble and the signaling facility attracted many users, although it will take some time to assess their performance. Like the JOE and the January meetings, they are marketplace institutions that attempt to help the market provide thickness and deal with congestion.

6. DISCUSSION

In the tradition of market design, I have concentrated on the details of particular markets, from medical residents and fellows to economists, and from kidney exchange to school choice. But, despite their very different details, these markets, like others, struggle to provide thickness, to deal with the resulting congestion, and to make it safe and relatively simple to participate. While the importance of thick markets has been understood by economists for a long time, my impression is that issues of congestion, safety, and simplicity were somewhat obscured when the prototypical market was thought of as a market for a homogeneous commodity.31

Thickness in a market has many of the properties of a public good, so it is not surprising that it may be hard to provide it efficiently, and that free riders have to be resisted, whether in modern markets with a tendency to unravel, or in medieval markets with rules against “forestalling.” Notice that providing thickness blurs the distinction between centralized and decentralized markets, since marketplaces—from traditional farmers’ markets, to the AEA job market meetings, to the New York Stock Exchange—provide thickness by bringing many participants to a central place. The possibility of having the market perform other centralized services, as clearinghouses or signaling mechanisms do, has only grown now that such central places can also be electronic, on the Internet or elsewhere. And issues of thickness become if anything more important when there are network externalities or other economies of scope.32

Congestion is especially a problem in markets in which transactions are heterogeneous, and offers cannot be made to the whole market. If transactions take even a short time to complete, but offers must be addressed to particular participants (as in offers of a job, or to purchase a house), then someone who makes an
offer runs the risk that other opportunities may disappear while the offer is being considered. And even financial markets (in which offers can be addressed to the whole market) experience congestion on days with unusually heavy trading and large price movements, when prices may change significantly while an order is being processed, and some orders may not be able to be processed at all. As we have seen, when individual participants are faced with congestion, they may react in ways that damage other properties of the market, e.g. if they try to gain time by transacting before others.33

Safety and simplicity may constrain some markets differently than others. Parents engaged in school choice may need more of both than, say, bidders in very high value auctions of the sort that allow auction experts to be hired as consultants. But even in billion dollar spectrum auctions, there are concerns that risks to bidders may deter entry, or that unmanageable complexity in formulating bids and assessing opportunities at each stage may excessively slow the auction.34

Somewhere in between, insider trading laws with criminal penalties help make financial markets safe for non-insiders to participate. And if it is risky to participate in the market, individual participants may try to manage their risk in ways that damage the market as a whole, such as when transplant centers withhold patients from exchange, or employers make exploding offers before applicants can assess the market, or otherwise try to prevent their trading counterparties from being able to receive other offers.35

In closing, market design teaches us both about the details of market institutions and about the general tasks markets have to perform. Regarding details, the word “design” in “market design” is not only a verb, but also a noun, so economists can help to design some markets, and profitably study the design of others. And I have argued in this essay that among the general tasks markets have to perform, difficulties in providing thickness, dealing with congestion, and making participation safe and simple are often at the root of market failures that call for new market designs.

I closed my 1991 *EJ* article (quoted in the introduction) on a cautiously optimistic note that, as a profession, we would rise to the challenge of market design, and that doing so would teach us important lessons about the functioning of markets and economic institutions. I remain optimistic on both counts.

**Acknowledgements**

The work I report here is a joint effort of many colleagues and coauthors. I pay particular attention here to work with Atila Abdulkadiroglu, Muriel Niederle, Parag Pathak, Tayfun Sönmez, and Utku Ünver. I’ve also benefited from many conversations on this topic with Paul Milgrom (including two years teaching together a course on Market Design). This work has been supported by grants from the NSF to the NBER.
What Have We Learned from Market Design?

Endnotes


2. See e.g. Cramton (1997), Milgrom (2000), Wilson (2002), and, particularly, Milgrom (2004) Following that literature to the present would involve looking into modern designs for package auctions, see e.g. Cramton, Shoham, and Steinberg (2006), and Milgrom (2007).


4. On the costs of such unraveling in some markets for which unusually good data have been available, see Niederle and Roth (2003b) on the market for gastroenterology fellows, and Fréchette, Roth, and Unver (2007) on the market for post-season college football bowls. For some other recent unraveled markets, see Avery, Fairbanks, and Zeckhauser (2003) on college admissions; and Avery, Jolls, Posner, and Roth (2001) on appellate court clerks. For a line of work giving theoretical insight into some possible causes of unraveling, see Li and Rosen (1998), Li and Suen (2000), Suen (2000), and Damiano et al. (2005).

5. “Thus at Norwich no one might forestall provisions by buying, or paying ‘earnest money’ for them before the Cathedral bell had rung for the mass of the Blessed Virgin; at Berwick-on-Tweed no one was to buy salmon between sunset and sunrise, or wool and hides except at the market-cross between 9 and 12; and at Salisbury persons bringing victuals into the city were not to sell them before broad day.” Unraveling could be in space, as well as in time. Salzman also reports (p132) that under medieval law markets could be prevented from being established too near to an existing market, and also, for markets on rivers, nearer to the sea. “Besides injury through mere proximity, and anticipation in time, there might be damage due to interception of traffic... “Such interception was more usual in the case of water-borne traffic. In 1233 Eve de Braose complained that Richard fitz- Stephen had raised a market at Dartmouth to the injury of hers at Totnes, as ships which ought to come to Totnes were stopped at Dartmouth and paid customs there. No decision was reached, and eight years later Eve’s husband, William de Cantelupe, brought a similar suit against Richard’s son Gilbert. The latter pleaded that his market was on Wednesday and that at Totnes on Saturday; but the jury said that the market at Dartmouth was to the injury of Totnes, because Dartmouth lies between it and the sea, so that ships touched there and paid toll instead of going to Totnes; and also that cattle and sheep which used to be taken to Totnes market were now sold at Dartmouth; the market at Dartmouth was therefore disallowed.”

6. The effects of instability were different in Britain than in the U.S., because positions in Britain were assigned by the National Health Service, and so students were not in a position to receive other offers (and decline the positions they were matched to) as they were in the U.S. Instead, in Britain, students and potential employers acted in advance of unstable clearinghouses. For example Roth (1991) reports that in Newcastle and Birmingham, it became common for students and consultants (employers) to reach agreement in advance of the match, and then submit only each other’s name on their rank order lists.

7. Abstracting somewhat from the complexities of the actual market, the Roth-Peranson algorithm is a modified student-proposing deferred acceptance algorithm (Gale and Shapley, 1962, see Roth, 2007b). In simple markets, this makes it a dominant strategy for students to state their true preferences (see Roth, 1982, 1985, Roth and Sotomayor, 1990). Although it can’t be made a dominant strategy for residency programs to state their true preferences (Roth, 1985; Sonmez, 1997), the fact that the medical market is large turns out to make it very unlikely that residency programs can do any better than to state their true preferences. This was shown empirically in Roth and Peranson (1999), and has more recently been explained theoretically by Immorlica and Mahdian (2005) and Kojima and Pathak (2007).

8. Bulow and Levin (2006) sketch a simple model of one-to-one matching in which a centralized clearinghouse, by enforcing impersonal wages (i.e. the same wage for any successful applicant) could cause downward pressure on wages (see also Kamecke 1998). Subsequent analysis suggests more skepticism about any downward wage effects in actual medical labor markets. See, for example, Kojima (2007) which shows that the Bulow-Levin results don’t follow in a model in which hospitals can employ more than one worker, and Niederle (forthcoming) who shows that the
results don’t follow in a model that includes the facility that the medical match offers to hospitals that wish to fill more of one kind of position if they fail to fill enough positions of another kind. Crawford (forthcoming) considers how the deferred acceptance algorithm of Kelso and Crawford (1982) could be adapted to adjust personal wages in a centralized clearinghouse, see also Artemov (forthcoming).

9. See Roth (2003). The law states in part: “Congress makes the following findings: For over 50 years, most United States medical school seniors and the large majority of graduate medical education programs (popularly known as ‘residency programs’) have chosen to use a matching program to match medical students with residency programs to which they have applied. ... “Before such matching programs were instituted, medical students often felt pressure, at an unreasonably early stage of their medical education, to seek admission to, and accept offers from, residency programs. As a result, medical students often made binding commitments before they were in a position to make an informed decision about a medical specialty or a residency program and before residency programs could make an informed assessment of students’ qualifications. This situation was inefficient, chaotic, and unfair and it often led to placements that did not serve the interests of either medical students or residency programs.

“The original matching program, now operated by the independent non-profit National Resident Matching Program and popularly known as ‘the Match’, was developed and implemented more than 50 years ago in response to widespread student complaints about the prior process. ... “The Match uses a computerized mathematical algorithm... to analyze the preferences of students and residency programs and match students with their highest preferences from among the available positions in residency programs that listed them. Students thus obtain a residency position in the most highly ranked program on their list that has ranked them sufficiently high among its preferences. ... “Antitrust lawsuits challenging the matching process, regardless of their merit or lack thereof, have the potential to undermine this highly efficient, pro-competitive, and long-standing process. The costs of defending such litigation would divert the scarce resources of our country’s teaching hospitals and medical schools from their crucial missions of patient care, physician training, and medical research. In addition, such costs may lead to abandonment of the matching process, which has effectively served the interests of medical students, teaching hospitals, and patients for over half a century. “... It is the purpose of this section to—confirm that the antitrust laws do not prohibit sponsoring, conducting, or participating in a graduate medical education residency matching program, or agreeing to do so; and ensure that those who sponsor, conduct or participate in such matching programs are not subjected to the burden and expense of defending against litigation that challenges such matching programs under the antitrust laws.”


12. The New England Program for Kidney Exchange has since integrated our software into theirs, and conducts their own matches. The Alliance for Paired Donation originally used our software, and as the size of the exchange pool grew larger, the basic (integer programming) algorithms were rewritten in software that can handle much larger numbers of pairs, by Abraham, Blum, and Sandholm (2007). Roth et al. (2005a,b) were also widely distributed to transplant centers (as working papers in 2004). The active transplant program at Johns Hopkins has also begun to use software similar in design to that in Roth et al. (2004b, 2005a), to optimize pairwise matches, see Segev et al. (2005).

13. Increasing the number of patients who benefit from the altruism of a non-directed donor may also increase the willingness of such donors to come forward. After recent publicity of the first NEAD chain on ABC World News Tonight [see http://utoledo.edu/utcommcenter/kidney/], the Alliance for Paired Donation has had over 100 registrations on its website of people who are offering to be altruistic living non-directed donors (Rees, personal communication).
What Have We Learned from Market Design?

14. The proposed bill (H.R. 710 introduced on 1/29/07 and passed in the House on 3/7/07 and S. 487 introduced on 2/1/07 and passed in the Senate February 15, 2007) is “To amend the National Organ Transplant Act to clarify that kidney paired donations shall not be considered to involve the transfer of a human organ for valuable consideration.”. Kidney exchange is also being organized in the UK; see <http://www.uktransplant.org.uk/ukt/about_transplants/organ_allocation/kidney_(renal)/living_donation/paired_donation_matching_scheme.jsp>. The first British exchange was carried out on July 4, 2007 (see the BBC report at <http://news.bbc.co.uk/1/hi/health/7025448.stm>).

15. For example, a compatible middle aged patient-donor pair, and an incompatible patient-donor pair in which the donor is a 25 year old athlete could both benefit from exchange. Aside from increasing the number of pairs available for exchange, this would also relieve the present shortage of donors with blood type O in the kidney exchange pool, caused by the fact that O donors are only rarely incompatible with their intended recipient. Simulations on the robust effects of adding compatible patient-donor pairs to the exchange pool are found in Roth, Sönmez and Ünver (2004a and 2005b), and in Gentry et al. 2007.

16. The invitation to meet with Boston Public Schools came after a newspaper story recounted the difficulties with the Boston system, as described in Abdulkadiroğlu and Sönmez 2003. For subsequent explorations of the old Boston system, see Chen and Sonmez 2006, Ergin and Sonmez 2006, Pathak and Sonmez 2007, Abdulkadiroğlu, Pathak, Roth, and Sönmez (2007).


18. In addition to the student proposing deferred acceptance algorithm that was ultimately adopted, we proposed a variation of the “top trading cycles” algorithm originally explored in Shapley and Scarf (1974), which was shown to be strategy-proof in Roth (1982b), and which was extended, and explored in a school choice context, in Abdulkadiroğlu and Sönmez (1999, 2003).

19. A much more thorough treatment of the material in this section is given in Niederle and Roth (2008).

20. The American system of residents and fellows is similar but not precisely parallel to the system in the UK of house officers and registrars, which has also recently faced some problems of market design.

21. Niederle and Roth (2007) study in the laboratory the impact of the rules that govern the types of offers that can be made (with or without a very short deadline) and whether applicants can change their minds after accepting an early offer. In the uncongested laboratory environments we studied, eliminating the possibility of making exploding offers, or making early acceptances non-binding, prevents the markets from operating inefficiently early.

22. This is not a closed market, as economics departments outside North America also hire in this market, and as American economics departments and other employers often hire economists educated elsewhere. But a large part of the market involves new American Ph.D.s looking for academic positions at American colleges and universities. See Cawley (2006) for a description of the market aimed at giving advice to participants, and Siegfried and Stock (2004) for some descriptive statistics.

23. Its members are Alvin E. Roth (chair), John Cawley, Philip Levine, Muriel Niederle, and John Siegfried, and the committee has received assistance from Peter Coles, Ben Greiner, and Jenna Kutz.

24. These applications are usually sent through the mails, but now often also via email and on web pages set up to receive them. Applicants typically apply to departments individually, by sending a letter accompanied by their curriculum vitae and job market paper(s) and followed by their letters of reference.

25. The situation is different in Europe, for example, where hiring is more dispersed in time. In an attempt to help create a thicker European market, the Royal Economic Society held a “PhD presentations event” for the first time in late January (2006), Felli and Sutton (2006) remark that “The issue of timing, unsurprisingly, attracted strong comment...”
26. While the large scale interviewing at the annual meetings has not been plagued by gradual unraveling, some parts of the market have broken off. In the 1950’s, for example, the American Marketing Association used to conduct job market meetings at the time of the ASSA meetings, but for a long time it had held its job market in August, a year before employment will begin, with the result that assistant professors of marketing are often hired before having made as much progress on their dissertations as is the case for economists (Roth and Xing, 1994).

27. These costs arise both not only because budgets for airfares and hotels may be limited, but also because faculties' faculties quickly become fatigued after too many seminars and recruiting dinners.

28. In 2002 and 2003 Georg Weizsacker, Muriel Niederle, Dorothea Kubler and I conducted surveys of economics departments regarding their hiring practices, asking in particular about what kinds of deadlines, if any, they tended to give when they made offers to junior candidates. Loosely speaking, the results suggested that departments that were large, rich, and elite often did not give any deadlines (and sometimes were able to make all the offers they wanted to make in parallel, so that they would not necessarily make new offers upon receiving rejections). Less well endowed departments often gave candidates deadlines, although some were in a position to extend the deadline for candidates who seemed interested but needed more time.

29. For a simple conceptual example of how a limited number of signals can improve welfare, consider a market with two applicants and two employers, in which there is only time for each employer to make one offer, and each applicant can take at most one position. Even if employers and applicants wish only to find a match, and have no preference with whom they match, there is a chance for signals to improve welfare by reducing the likelihood of coordination failure. In the absence of signals, there is a symmetric equilibrium in which each firm makes an offer to each worker with equal probability, and at this equilibrium, half the time one worker receives two offers, and so one worker and one employer remain unmatched. If the workers are each permitted to send one signal beforehand, and if each worker sends a signal to each firm with equal probability, then if firms adopt the strategy of making an offer to an applicant who sends them a signal, the chance of coordination failure is reduced from one half to one quarter. If workers have preferences over firms, the welfare gains from reducing coordination failure can be even larger. For recent treatments of signaling and coordination, see Coles and Niederle (2007), Lee and Schwarz (2007a,b), Lien (2007), and Stack (2007). See also Abdulkadiroglu, Che, and Yasuda (2007), who discuss allowing applicants to influence tie-breaking by signaling their preferences in a centralized clearinghouse that uses a deferred acceptance algorithm.

30. The document “Signaling for Interviews in the Economics Job Market,” at http://www.aeaweb.org/joe/signal/signaling.pdf includes the following bits of advice: “Advice to Departments: Applicants can only send two signals, so if a department doesn’t get a signal from some applicant, that fact contains almost no information. (See advice to applicants, below, which suggests how applicants might use their signals). But because applicants can send only two signals, the signals a department does receive convey valuable information about the candidate’s interest. A department that has more applicants than it can interview can use the signals to help break ties for interview slots, for instance. Similarly, a department that receives applications from some candidates who it thinks are unlikely to really be interested (but might be submitting many applications out of excessive risk aversion) can be reassured of the candidate’s interest if the department receives one of the candidate’s two signals. A department that receives a signal from a candidate will likely find it useful to open that candidate’s dossier and take one more look, keeping in mind that the candidate thought it worthwhile to send one of his two signals to the department. Advice to Applicants: The two signals should not be thought of as indicating your top two choices. Instead, you should think about which two departments that you are interested in would be likely to interview you if they receive your signal, but not otherwise (see advice to departments, above). You might therefore want to send a signal to a department that you like but that might otherwise doubt whether they are likely to be able to hire you. Or, you might want to send a signal to a department that you think might be getting many applications from candidates somewhat similar to you, and a signal of your particular interest would help them to break ties.
What Have We Learned from Market Design?

You might send your signals to departments to whom you don’t have other good ways of signaling your interest.

31. Establishing thickness, in contrast, is a central concern even in financial markets, see for example the market design (“market microstructure”) discussions of how markets are organized at their daily openings and closings, e.g. Biais, Hillion and Spatt 1999 on the opening call auction in the Paris Bourse and Kandel, Rindi, and Bosetti 2007 on the closing call auctions in the Borsa Italiana and elsewhere.

32. Thickness has received renewed attention in the context of software and other “platforms” that serve some of the functions of marketplaces, such as credit cards, which require large numbers of both consumers and merchants (see e.g. Evans and Schmalensee, 1999 and Evans, Hagiu and Schmalensee, 2006; and see Rochet and Tirole 2006, who concentrate on how the price structure for different sides of the market may be an important design feature.

33. The fact that transactions take time may in some markets instead inspire participants to try to transact very late, near the market close, if that will leave other participants with too little time to react. See e.g. the discussion of very late bids (“sniping”) on eBay auctions in Roth and Ockenfels (2002) and Ariely, Ockenfels, and Roth (2005).

34. Bidder safety lies behind discussions both of the “winner’s curse” and collusion (cf. Kagel and Levin 2002, Klemperer, 2004) as well as of the “exposure problem” that faces bidders who wish to assemble a package of licences in auctions that do not allow package bidding (see e.g. Milgrom 2007). And simplicity of the auction format has been addressed in experiments prior to the conduct of some FCC auctions, see e.g. Plott (1997). Experiments have multiple uses in market design, not only for investigation of basic phenomena, and small-scale testing of new designs, but also in the considerable amount of explanation, communication, and persuasion that must take place before designs can be adopted in practice.

35. For example, Roth and Xing (1994) report that in 1989 some Japanese companies scheduled recruiting meetings on the day an important civil service exam was being given, to prevent their candidates from also applying for government positions.

Bibliography


What Have We Learned from Market Design?


What Have We Learned from Market Design?


