Market design and maintenance

By Alvin E. Roth, Stanford University

1. Introduction

Market design is an ancient human activity, but only in the last three decades or so have economists begun to apply insights from game theory to the practical design of marketplaces. Market design has brought into focus some of the things that successful marketplaces accomplish (Roth, 2008). They must help the market become *thick*, so that enough transactions are available to attract participants to the marketplace. They must avoid *congestion* arising from the time it takes to identify, evaluate and complete transactions when there are many potential transactions available, and limited time. They must make the marketplace itself *safe* and (relatively) simple for participants, and sometimes they must also be acceptable to non-participants (Roth, 2007).

Notice that I speak not only of markets but also of *marketplaces*. Practical market designers are typically designing a marketplace situated in a larger economic environment. Potential marketplace participants have large strategy sets that may include actions outside of the marketplace (e.g. before or after transacting in the marketplace), and even strategies that may involve transacting entirely outside the marketplace. So the designer needs to make the marketplace more attractive than alternative venues, in addition to being concerned with equilibrium (and other) behavior in the game induced by the mechanism inside the marketplace.

This means that market design is different from theoretical mechanism design. Loosely speaking, the theoretical mechanism design literature conceives of marketplaces as if they were the entire market. The mechanism designer in this theoretical literature thus designs the entire universe that participants will inhabit, the entire strategy set of every participant, and how their choices interact to produce the market outcome.

But not only must actual marketplaces operate within a larger economic environment, this larger environment is likely to change over time, which may change the relationship of the marketplace to the larger market. By analogy, when you build a new house, a lot of construction is required. Decades later, some features of the house are no longer as well adapted to the inhabitants as they used to be, while other things need to be updated because of developments outside the house—e.g. you need to upgrade the electrical system after buying an electric car. So maintenance and renovation may supplement or even supplant some of the original construction.

---

1 Acknowledgment: This paper began as a presentation at the NBER conference on New Directions in Market Design, in the Spring of 2023. My work on this paper was partially supported by the National Science Foundation under Grant DMS-1928930 and by the Alfred P. Sloan Foundation under grant G-2021-16778, while I was from time to time at the Simons Laufer Mathematical Sciences Institute (formerly MSRI) in Berkeley, California, during the Fall of 2023.
In this spirit, I’ll discuss several venerable market design projects in which I have been privileged to play a role over a period of years. I’ll describe design changes in response to changing environments, as well as some current issues those markets are facing that may call for continuing redesign. Each of these markets had long intervals in which participants adjusted their behavior to the incentives created by the existing marketplace rules, followed by focused periods of market design, and then again by periods of behavioral response, and technological and social change in the larger market. My goal in this paper is to focus on how the design of marketplaces has interacted with developments in the larger market, rather than with the technical details of design (about which I’ll simply cite references).²

In Section 2 I’ll discuss the American market for new doctors, and the marketplace called the Match. That market, initially quite decentralized, began with the development of medical licensure around 1900, when it became important for new medical school graduates to obtain ‘postgraduate’ medical education as interns/residents in hospitals, where they gained experience under the supervision of more senior (and already licensed/board certified) physicians. It went through a long period of adjustments in how market participants behaved. From year to year, offers were made earlier and at more diffuse times in medical students’ careers, for positions that would be filled only upon graduation after four years of medical school. This made the market less thick: students were often faced with a single offer, which would expire before any other offers might be made. By 1944, these offers eventually came as early as students’ second year in medical school. When uniform times were established before which offers shouldn’t be made, congestion in the process of making and accepting/rejecting offers of employment also occurred, with verbal acceptances sometimes demanded immediately, in the same phone call in which the offer was made. This eventually led, with some trial and error, to the construction of a centralized clearinghouse in the early 1950’s, which solved that congestion problem and survived with relatively minor changes until the mid-1990’s, when accumulated changes required a redesign. That in turn led to a long period of orderly operation, which is presently troubled by congestion, not in making and responding to offers, but, before that, in processing the many applications (and subsequent interviews) that have developed with the introduction of computerized application services.

In Section 3 I’ll describe the decentralized process(es) by which new Ph.D. economists get their first positions. Being decentralized, that market offers many paths for economists to find employment, among a wide set of employers (not just economics departments at universities). A publication of the American Economic Association (AEA) advertising Job Openings for Economists (JOE) helped make the market thick in the 1970’s, and a widely attended January conference run by the AEA helped coordinate the timing of the market, since a very large percentage of job market interviews were conducted there. The interviewing process became congested, and a signaling mechanism was introduced to help coordinate interviews, without causing other major changes in the marketplace. Then came the Covid

² Other veteran market designers could give you a different perspective, but by focusing on markets I have lived with for a long time I can draw on information that may not have made it into published reports. It also allows me to indicate how market design can offer the satisfactions (and frustrations) of intimate involvement with particular markets over enough time to observe their evolution and to be able to diagnose and occasionally help address some of their problems.
pandemic, and in-person interviews were replaced by Zoom interviews, which proved very popular. Today we are seeing more early interviews, and also a limited number of early offers (with short times during which offers remain open), and the AEA has offered guidance in the hope of constructing a new equilibrium that will keep the market thick.

In Section 4 I’ll discuss a very different kind of matching market, involving kidney exchange to increase the availability of living donor kidney transplants, in the face of almost universal laws forbidding the compensation of donors. To work well, kidney exchange needs coordination among hospitals, which demanded from the outset that transplant centers would make substantial changes in their previously customary procedures. And unlike the labor markets for new physicians or Economics Ph.Ds., it operates continually rather than once a year. Consequently, it has experienced changes frequently rather than annually, and its ongoing design still requires substantial innovation.

2. Matching new doctors to positions

Today, and since the 1950’s almost all graduates of American medical schools participate in the National Resident Matching Program (NRMP), also called the Match®, the centralized clearinghouse that matches applicants to their first clinical jobs as medical residents (upon completion of which they can become licensed physicians). In the course of their senior year in medical school they will apply to many residency programs and be invited by some of them to participate in an interview, after which both the applicants and residency programs will fill out rank order lists of who they would like to be matched to, in order of their preferences. Then a centralized matching mechanism, powered by an algorithm that is in large part a deferred acceptance algorithm, will produce a match that will determine who fills which positions, and almost all positions will be filled in this way.

The use of a centralized clearinghouse began in the 1950s to solve a series of marketplace failures in the prior decentralized market for new doctors, which played out in the first half of the 20th Century.

---

3 This section draws on some of the market history recounted in Roth (1984, 2003). To engage with some of the current issues facing the Match that I will describe, I joined the Board of Directors of the NRMP for a term from June 2020-2024.

4 In 2023, about 5,500 residency programs offered approximately 40,000 postgraduate positions to about 48,000 registered applicants, of whom nearly 20,000 were new MD graduates and over 7,000 were new DO graduates of American medical schools, with most of the other applicants being graduates of foreign medical schools (NRMP, 2023).

5 The deferred acceptance algorithm came to wide attention through the work of Gale and Shapley (1962), who proposed and analyzed it from the point of view of stable matching. Their paper launched the market design literature concerned with matching. Roth (1984) discovered that the algorithm used by the resident match in the 1950’s was different from but equivalent to deferred acceptance (with hospitals proposing). The algorithm presently used by the Match was designed by Roth and Peranson (1999). It is roughly a hybrid of deferred acceptance combined with an algorithm for fixing instabilities (Roth and Vande Vate, 1990) that arise from the need to address a number of features of the medical labor market that were not standard in the theoretical study of matching, including the presence of couples looking for two positions, and reversions of unfilled positions from one residency program to another at the same hospital (about which see Niederle 2007).
Internships and residencies became a standard part of medical school graduates’ careers around 1900, and, in the early days, job searches were conducted around the end of a student’s last year of medical school. But competition among hospitals for good students, and among students for desirable jobs, first gradually and then quickly made the dates of appointment earlier, and more diffuse in time. Around the 1930’s positions had come to be mostly filled around the beginning of the calendar year, i.e. half way through students’ fourth year of medical school, and by the 1940’s the hiring date had moved to the summer after their second year. Because the timing of job offers had substantial variance, students often had to respond to an offer with a decision before knowing what other jobs might be offered to them later if they rejected the present offer, and knowing that many desirable positions would indeed be filled two years in advance.

Lots of information needed for a good match is missing two years in advance, and this was a widely recognized problem that was addressed when the medical schools agreed not to release information about students before fixed dates. This succeeded, in the sense that the hiring date for medical graduates was moved later, first to the third year of medical school, and in subsequent years back to the second half of the fourth and final year of school. But a new difficulty emerged, as hospitals came to demand immediate responses to their offers. So, once again, students had to respond to offers without knowing what other jobs might be offered to them if they could wait even a little longer.

In the period when offers were early and diffuse in time, the market lacked thickness, and applicants and employers weren’t able to consider many possible matches at the same time. And multiple offers also couldn’t be considered in parallel in the period of exploding offers, when too many offers had to be made and considered in too little time, i.e. when the process of offers and responses was congested.

The introduction of a centralized clearinghouse in the early 1950’s (after a little trial and error) solved these timing problems for offers and acceptances. Because both applicants and residency programs submitted rank order preference lists of one another, the clearinghouse could consider all the preferences simultaneously, i.e. applicants could indicate their preferences for all the positions for which they had applied and were interviewed, and offers from programs could be processed, sequentially if necessary, without fear that some applicants would be “captured” early by competitor programs that moved first.

As mentioned earlier, the algorithm by which the early clearinghouse produced matches was equivalent to the deferred acceptance algorithm. In particular, Roth (1984) showed that the algorithm produced a matching that was “pairwise stable” in the sense defined by Gale and Shapley (1962): no doctor and residency program not matched to one another would prefer to be matched together. And empirically, there was good reason to think that this kind of stability was important for the success of the Match, since similarly organized clearinghouses that didn’t produce stable matchings tended to fail.6

---

6 For example, unstable matches tended to fail in the U.K.’s National Health Service (Roth 1990, 91), with some exceptions (Unver, 2001, 05) related to details of some markets there that involved a single hospital and medical school.
This worked well when the medical labor market looked like a simple many-to-one matching problem between residency programs and applicants, but changes in the demographics of the labor force, and in the organization of postgraduate education, created some features that were difficult to incorporate into a stable matching mechanism. Among these changes were changes in the larger economic environment, such as increased enrollment of women in medical schools, with the consequence that some graduates went on the market as part of a married couple looking for two positions.7

Amidst this and other stresses on the market, I was asked to direct the redesign of the match algorithm. I count this as the moment that I became committed to practical market design, which changed the way I think about economic theory. For example, in Roth (1984) I had shown that, when couples are present, the set of stable matchings may be empty, so that the presence of couples presented a hard problem for market design. But when I agreed to help redesign the match, I adopted that problem as one that I would have to address in a practical way, despite the fact that the available theory didn’t yet shed much light on stable matchings involving couples. So theory would have to be supplemented with computation and experimentation to make progress. (Roth and Peranson (1999) was able to show computationally that we could expect to find stable matchings reliably given the relatively small proportion of couples in the market. But it wasn’t until much later that we began to gain theoretical insight into this in Kojima, Pathak, and Roth (2013) and Ashlagi, Braverman and Hassidim (2014).)8

One issue for which the available theory was a considerable help involved the distribution of new doctors to rural hospitals. These are generally not the most preferred places at which to do a residency, and they often did not fill all the positions they offer. In addition, rural hospitals are often matched to graduates of foreign medical schools. One thing that I sometimes heard from medical administrators was that, if a new algorithm was to be designed, it would be useful if it could help rural hospitals hire more residents, and more graduates of American medical schools. However, given the importance of achieving a stable match, I could confidently answer that this was not something that we could do, because the Rural Hospital theorem (Roth, 1986) shows that the set of positions filled doesn’t vary from one stable matching to another, and employers who fail to fill all of their positions are matched with exactly the same employees at every stable matching.

Since going live in the late 1990’s, the Roth-Peranson algorithm has powered the Match, without incident.9 However, in the intervening quarter century the process of preparing for the match has become time consuming for both applicants and residency programs.

Growing congestion in applications and interviews before the opening of the Match has caused difficulties for both residency programs and applicants. Electronic application systems make it easy for

---

7 In 2023, 2,478 individuals participated in the match as part of 1,239 couples (NRMP 2023).
8 Another computational demonstration in Roth and Peranson (1999) was that the size of the set of stable matchings for new doctors and residency programs is very small. We began to gain a good theoretical understanding of that only with the theoretical work of Immorlica and Mahdian (2003); Kojima and Pathak (2009); and Ashlagi, Kanoria and Leshno (2013, 2017). (In 2023 Immorlica and Mahdian (2003) and Ashlagi, Kanoria and Leshno (2013) were awarded the Test of Time award by the Association for Computing Machinery’s Special Interest Group On Economics And Computation, https://www.sigecom.org/award-tot.html.)
9 See e.g. Agarwal (2015).
applicants to apply to many programs, but don’t make it quicker or cheaper to fully evaluate each application. And many applications by others cause applicants to want to submit more applications themselves, and cause programs to want to interview more applicants (so that they can rank them in the Match), as the growth in applications and interviews (by others) reduces the probability that any single application or interview will lead to a match.

Before the Covid pandemic, interviews were in person, and applicants were primarily responsible for paying their own travel expenses. Thus participating in many interviews was not only time consuming, but could add thousands of dollars to medical student debt. The scheduling of interviews started to exhibit some of the pathologies that unraveling of offers once exhibited. For example, some programs allocated a fixed number of interviews on a first come first serve basis after inviting a larger number of applicants to be interviewed, so that available slots were filled by those who could reply most quickly.

During the pandemic, interviews were primarily conducted by Zoom, and as Covid concerns have receded it appears that remote interviews will persist. So the out of pocket costs of interviews was substantially reduced, but the time requirements of many interviews remained a burden to both sides of the market.

Another concern about many applications and interviews was that programs were paying too much attention to easy to compare applicant attributes, such as their score on the first U.S. Medical Licensing Exam (USMLE-Step 1). In reaction, the USMLE-Step 1 stopped reporting numerical grades, and (since 2022) now reports only pass/fail results. This of course doesn’t make it easier to quickly (although perhaps inappropriately) eliminate some applicants from further consideration, indeed its goal was to promote more ‘holistic’ evaluation of all applicants. It is too soon to know how this will influence concentration on other easy to observe attributes such as the school each applicant has attended, and how this will influence the mix of residents at different programs.

Not all proposals considered seriously by the medical community had much potential upside. One that gathered early support was to match 50% of OBGYN positions early in a single-specialty match, with the other 50% being in the regular NRMP match later. It’s easy to see that this would produce instabilities. (Consider a program with 2 positions, and 2 applicants it ranks highly and who rank it first. At any stable match those applicants would be matched to that program, but in an early match at which only one of the two positions is available, one of those applicants can be matched early only to another program, that he or she likes less.)

This proved to be a difficult argument to convey, particularly since the American Medical Association had funded and promoted this proposal. But a simulation with preference data from previous years showed that such a two-stage match limited to a subset of the available positions in the first stage would not only produce instabilities, but would also harm many more applicants than it would help (Ashlagi et al. 2023). This demonstration seems to have eliminated this proposal from further consideration. Playing defense is sometimes a part of market maintenance.

Motivated by the practice of signaling in the job market for new Economics Ph.D.’s, many medical specialties have begun offering to transmit signals of interest from applicants to residency programs, in
an effort to convey some of the information that used to be conveyed by applying for a position, before applicants began applying to so many. (In Economics, the American Economics Association offers each applicant the ability to submit up to two such signals to potential employers, which are meant to aid in employers’ decisions about which applicants to invite to interviews, and thus to help deal with the congestion involved in matching applicants to interviews (Coles et al. 2010).) Each medical specialty has acted independently in choosing the number of signals, and so there is a disorganized natural experiment of sorts presently playing out, with some very preliminary indications from the first year of signaling. Many medical specialties adopted essentially the Economics approach of having relatively few signals (I’m taking single digit numbers to be few). The first year of results seems to indicate that residency directors in these specialties are paying attention to signals, but that not receiving a signal from an otherwise interesting applicant does not make him or her less interesting. A notable exception is Orthopedic Surgery, which invites applicants to send 30 signals. It appears that this acts like a soft cap on the number of applications that a candidate can effectively submit, since residency directors report that they would not interview even a highly attractive applicant who didn’t send them a signal. But these observations are still largely anecdotal.\(^\text{10}\) It remains to collect and analyze data in coming years on how signals are used and how this influences which interviews are conducted, and which applicants are ultimately matched to which positions.

Thus, even in a market that operates largely on an annual basis, and which has a smoothly running labor market clearinghouse, it seems likely that, as the larger economic environment in which physicians work continues to change, there will continue to be new design challenges for various parts of the marketplace for new physicians.

3. **Navigating the job market for new Ph.D. economists**\(^\text{11}\)

Another set of labor markets that operate largely on an annual basis are the markets for newly graduating Ph.D.’s. Different disciplines organize (or fail to organize) themselves differently. Prior to 1970, the market for new economists was almost entirely decentralized, with academic departments announcing searches largely by word of mouth, and information passing through networks of colleagues and academic advisors, and in response to letters of inquiry. Starting around 1974, the American Economic Association (AEA) has helped make the market thicker by publishing *Job Openings for Economists* (JOE), which has been successful in gathering most U.S. academic job openings for new Ph.D.s, and many other (industrial, government, and foreign) job openings as well.

An annual conference, the Allied Social Science Associations (ASSA) meetings, organized each year by the AEA in early January, served as a venue for most interviews for academic positions for new Ph.Ds. The conference took place in cities whose convention centers offered enough hotel and meeting space

\(^\text{10}\)These (anecdotal) observations come in part from the session on “Advising residency applicants in the era of program signaling,” from the Transition to Residency conference held in Boston in October, 2023: [http://nrmpconference.org/agenda.html](http://nrmpconference.org/agenda.html).

\(^\text{11}\)The market history in this section draws on Coles et al. (2010), which reports on the work of the AEA’s ad hoc committee on the job market.
for thousands of economists (around 13,000 attended the year that I organized the conference), and recruiting committees invited candidates to interview, often in rooms in one of the conference hotels, during one of the two and a half days of the conference. So interviews were congested, both to attend (as perhaps a thousand well-dressed young economists thronged the hotel elevators to arrive at interviews that began on the hour or half hour), and to schedule. A busy recruiting committee might have time to interview no more than two dozen candidates in the time available, but had to select these from possibly several hundred applications. The limited number of interviews that a committee could conduct in the time available became an increasingly salient constraint as the number of applications grew, aided by the advent of electronic applications over the internet.12

Following the interviews, departments invited some candidates to visit their campus for day-long “flyouts,” after which departments began to make offers to selected candidates. These offers had varying durations, some long enough so that candidates could collect and consider offers from January to March, but some requiring answers within two weeks (or sometimes less). So candidates at times faced difficult decisions, and often asked for and were frequently (but not always) granted deadline extensions so that they could hear at least from employers with whom they had already interviewed and who were almost ready to make offers. And every year, around the end of March, there were departments that had not filled their positions, and candidates who had not received offers, but this part of the market was decentralized and thin, with little help outside of personal networks in connecting people still in need of jobs with departments that might still be seeking to fill openings.

In 2005 I began to chair a new AEA committee on the job market, tasked with investigating and perhaps designing ways to ameliorate the congestion in interviewing, and the lack of thickness and coordination in the late part of the market. In 2006 we rolled out the signaling mechanism discussed above in connection with how years later it influenced the market for new physicians. And we also initiated a late-stage “scramble” for economists, modeled on the aftermarket that had long existed for new physicians.13 Our scramble was simply a website maintained by the AEA, on which applicants and departments could indicate their continued availability after many positions had already been filled, and then contact one another directly.

Both of these interventions operated as part of the existing market procedures, which themselves continued to organize the market up until the Covid pandemic. However one concern that surfaced before Covid, and has become more salient since, is that there was increasing objection to interviews being conducted in hotel rooms. While some departments could afford to reserve suites with separate space that could be used for interviews, many departments conducted interviews in a hotel room in which some faculty member was staying while attending the conference. This led to crowded interviews, with some committee members sometimes seated on a bed, and the objection was that this created an unprofessional atmosphere that could be hostile to women candidates. The AEA issued

---

12 Coles et al. reported that the survey the AEA conducted on behalf of our committee showed an average of 80 applications per candidate in 2006-8. My sense is that number has grown further since then.

13 The aftermarket for new physicians has since undergone substantial change, but that is a story for another time.
statements suggesting that interviews in hotel rooms were undesirable and should be moved to more public parts of the conference venue.

But before this guidance could have much effect, in-person interviews had to be abandoned during the pandemic. Instead, interviews, and often flyouts as well, were conducted remotely, by Zoom (or similar software). Zoom interviews were effective and popular, and by the time in-person meetings could resume, it no longer seemed obvious that interviews should return to being in-person. In addition, the 2023 meetings were held in New Orleans, and following the 2022 U.S. Supreme Court decision in Dobbs v. Jackson Women's Health Organization overturning prior U.S. law on abortion, many AEA members felt it would be unfair and possibly unsafe to require women candidates to attend meetings in a state with restricted reproductive care. These concerns and the earlier concerns about hotel room interviews led the AEA to suggest that interviews be conducted remotely, and prior to the meetings so as not to conflict with them.  

Following three years of Zoom meetings it became apparent that allowing interviews to be conducted earlier than usual had sometimes facilitated earlier flyouts, followed by earlier offers, more often with deadlines that compelled applicants to make difficult decisions before they could receive offers or even flyouts from other departments at which they had interviewed. So immediately after the 2023 meetings, the AEA’s job market committee (now chaired by John Cawley) formed a subcommittee on job market design to consider the possibility that the market for new economists might substantially unravel in the coming years.

In the absence of comprehensive data yet to become available about the timing of events in the previous years’ job markets, we conducted focus group interviews of recruiting committees from a variety of academic and nonacademic employers. These revealed that many employers felt compelled to act earlier than in previous years, and would be unreceptive to guidelines from the AEA if they were to try to prohibit early actions. Some employers, such as small liberal arts colleges, felt competitive pressure to act early, from other early actors, such as Federal agencies, or from early moving overlapping markets like those for agricultural economists.

The AEA guidelines for the 2023-24 job market simply suggested dates and time intervals: interview invitations shouldn’t be issued before the AEA transmitted applicant signals around December 1, and offers should remain open until January 31, and for at least two weeks, at least in the early part of the market (Rousseau, 2023).

Earlier, the American Finance Association (AFA) had published guidelines for the 2022-23 job market, with a much stronger suggestion. Their guidelines (AFA, 2022) say in part: “the AFA promotes the

---

14 The location of AEA meetings is customarily arranged several years in advance, as it involves negotiation with convention centers in cities big enough to host the conference in hotels located closely enough together to allow applicants to move from one interview to another. As I write this in 2023, the 2024 meeting is scheduled for San Antonio, also in a state that the AEA is reluctant to require candidates to visit because of its restrictive laws about reproductive care. So at least another year will pass in which the meetings don’t serve in any way to organize interviews or coordinate their timing.

15 See https://www.aeaweb.org/about-aea/committees/job-market
following professional norm: If a job candidate receives and accepts a coercive exploding offer (i.e., one that expires before February 20), the AFA does not consider such an acceptance to be binding.”

The AFA is hoping to promote an equilibrium in which early exploding offers will not be made because they will neither effectively bind applicants (who can later change their mind) nor bind (other) employers, who can continue to woo applicants who have accepted early offers. Establishing norms like this has been successful in the market for Ph.D. admissions, in which very many universities subscribe to the Council of Graduate Schools’ “April 15 resolution,” which empowers applicants who have accepted an offer earlier to change their minds before that date and accept a different offer.16

The experience of other markets suggests that establishing this kind of norm as an equilibrium is sensitive to details of market culture. For example, this kind of policy has been successfully implemented in the medical labor market for Gastroenterology Fellows, but proved impractical in the apparently similar market for Orthopedic Surgery Fellows, in which it was felt that applicants couldn’t effectively be empowered to change their minds after accepting an early offer.17 It is still too early to tell how the markets for Finance and Economics Ph.D.s are going to play out in this respect. Market culture is at issue here, as it is in many markets with potentially multiple equilibria.

I anticipate that the Economics job market may see more early offers before we reach a new equilibrium.

4. Kidney Exchange18

Kidney failure is one of the top causes of death around the world, and the best treatment is a kidney transplant. A person with failed kidneys can receive a new one from a deceased donor, or from a living donor, because healthy people have two kidneys and can remain healthy with one. But most people with kidney failure will die without receiving a transplant.

In the U.S., for example, there are approximately 800,000 people with kidney failure,19 among whom not quite 90,000 are on a nationally maintained waiting list to receive a deceased donor kidney. But in 2022

---

16 The latest version of which (CGS 2019) states “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. If a student accepts an offer before April 15 and subsequently desires to withdraw that acceptance, the student may submit a written resignation of the appointment at any time through April 15. Applicants are not required to obtain a formal release from the program whose offer they accepted, either before or after the April 15 deadline. Once applicants have informed the program that they are withdrawing their acceptance of the offer, they then can accept any other offers.”

17 On gastroenterologists, see Niederle et al. (2006), and on orthopedic surgeons see Harner et al. (2008), but note that both of those design efforts involved discouraging exploding offers in the decentralized market that preceded a new centralized matching marketplace, rather than (as is the case for economists) to coordinate timing in a decentralized labor market. See Niederle and Roth (2009) for a laboratory experiment designed to clarify the role of different market cultures regarding exploding offers.

18 For a technical account see Ashlagi and Roth (2021).

19 U.S. Renal Data Service (USRDS) 2023 Annual Data Report, Figure 1.5. About 500,000 of them receive hemodialysis at a dialysis center, while substantially smaller numbers receive home dialysis or receive transplants.
only 19,636 kidneys from deceased donors were transplanted. So the wait is long and hazardous: 4,423 died while on the waiting list and 4,396 were removed from the waiting list when they became too sick to transplant in 2022.\textsuperscript{20} To be clear, the opportunity to receive a transplant slips away each year from about 10% of even those formally enrolled on the deceased donor waiting list. Kidneys from living donors are also in short supply, in 2022 there were 5,864 kidney transplants from living donors in the U.S.

When economists see long queues of people waiting for a scarce, rationed good, we suspect that the price mechanism must not be working to increase supply. And indeed, it is against the law almost everywhere in the world to compensate a living donor or the family of a deceased donor, largely out of concern that allowing compensation could lead to the exploitation of the vulnerable. In the U.S., the National Organ Transplant Act (1984) forbids payments to donors, and similar laws are widespread.\textsuperscript{21} There are black markets, and vigorous enforcement of laws against compensating donors have in many countries driven those black markets out of hospitals and into venues that offer dangerous, substandard medical care to both paid donors and recipients.\textsuperscript{22} The only country with a recognized legal monetary market for kidneys is the Islamic Republic of Iran. That market is controversial both inside and outside Iran, but the resulting nephrectomies and transplants and postsurgical care are apparently delivered in established medical institutions meeting international standards of care.\textsuperscript{23}

I think economists need to understand better how and when repugnance to certain kinds of transactions leads to legal bans, how bans lead to black markets, and how we might better conduct and regulate controversial markets with an eye to both equity and efficiency.\textsuperscript{24}

\textit{Kidney exchange} is a way to increase the supply of living donor kidney transplants while (largely) avoiding the repugnance with which financial transactions for organs are widely viewed. Not everyone who is healthy enough to donate one of their kidneys can actually donate to their intended recipient, since kidneys need to be medically compatible. Until recently, a patient with an incompatible living donor had no choice but to remain on the waiting list in the hope that a compatible deceased donor kidney would become available, while their incompatible donor was sent home. But kidney exchange allows incompatible patient-donor pairs to exchange kidneys with one another, so that each patient receives a compatible kidney from another patient’s donor.

\textsuperscript{20} See https://optn.transplant.hrsa.gov/data/view-data-reports/national-data/
\textsuperscript{21} The U.S. National Transplant Act of 1984 contains the following: “TITLE III—PROHIBITION OF ORGAN PURCHASES SEC. 301. (a) It shall be unlawful for any person to knowingly acquire, receive, or otherwise transfer any human organ for valuable consideration for use in human transplantation if the transfer affects interstate commerce.” (A later paragraph refers to the broad interpretation of interstate commerce.)
\textsuperscript{22} See e.g. Columb (2020)
\textsuperscript{23} See e.g. Roth et al. (2022) and the references there.
\textsuperscript{24} It appears from survey experiments that there isn’t a simple connection between legal bans and repugnance among the general population. See e.g. Leider and Roth (2010), Elias et al. (2019), Roth and Wang (2020).
Kidney exchange was proposed conceptually by Rapoport (1986), but the first exchange in the U.S., between two patient-donor pairs, wasn’t carried out until 2001 in the Rhode Island Hospital. Roth, Sönmez, and Ünver (2003,2004) proposed how kidney exchange could scale beyond two-pair exchange in a single hospital. We sent the 2003 working paper to a number of transplant surgeons, but the only one who replied was my Harvard colleague Frank Delmonico, and we helped him and his transplant colleagues start the New England Program for Kidney Exchange (Delmonico et al., 2004; Roth, Sönmez, and Ünver, 2005a). But our surgical colleagues weren’t yet prepared to consider exchanges involving more than two patient-donor pairs, and so we proposed a simpler system for pairwise exchanges (Roth, Sönmez, and Ünver, 2005b), with which NEPKE began operations. Each exchange was to be done simultaneously, to make sure that no pair gave a kidney and subsequently didn’t receive one in return.

Actually implementing inter-hospital kidney exchange involved addressing many issues related to the larger environment of hospital workflows and healthcare payments.

An immediate and formidable obstacle was the need for coordination between hospitals. Before kidney exchange, transplant centers seldom had to coordinate with one another: each center transplanted its own patients. But efficient exchange, which requires a pool of eligible patient-donor pairs, often involved patients at different hospitals. In New England alone, there were fourteen transplant centers. Because they were relatively close to each other, inter-hospital exchange in New England between two patient-donor pairs was arranged by having each patient remain at the hospital where he or she was being treated, and admitting the relevant donor to that hospital for the nephrectomy. But operating rooms in a hospital are a scarce resource that needs to be scheduled, and a living donor kidney transplant is an elective surgery, which can be scheduled in advance. So, at each hospital, there was a customary day when living donor kidney transplants were done. Although these were mostly Tuesdays, Wednesdays or Thursdays, it nevertheless often meant that if two hospitals wished to engage in a simultaneous exchange, one of them had to operate on a day that was irregular for them. This involved not just a negotiation between the two kidney transplant programs (each of which would prefer their customary day), but also with colleagues doing other kinds of surgeries at the same hospital, whose operations might also have to be rescheduled.

This was one reason our surgical colleagues initially felt that they could perform only exchanges between two patient-donor pairs. Exchanges involving three or more pairs could require coordination

25 For early kidney exchange history see Wallis et al. (2011); for an account of the first U.S. exchange see Zarsadias (2010).
26 Roth, Sönmez, and Ünver (2004) was based largely on the top trading cycles algorithm described in Shapley and Scarf (1974) and studied in Roth and Postlewaite (1977) and Roth (1982b) and extended in Abdulkadiroğlu and Sönmez (1999).
27 There was some early opposition to kidney exchange phrased in terms of the “valuable consideration” wording in the NOTA, but in 2007 both houses of Congress unanimously passed Public Law 110-144 - the Charlie W. Norwood Living Organ Donation Act, “to amend the National Organ Transplant Act to provide that criminal penalties do not apply to human organ paired donation…” (The phrase “paired donation” avoids use of the word “exchange” that made some people think of organ sales.)
28 Since deceased-donor transplants are emergency surgeries, there was some customary flexibility that allowed more coordination than otherwise might have been achievable.
among three or more hospitals (while a three-way simultaneous exchange at a single hospital would involve six operating rooms and surgical teams, which would strain the resources of all but the biggest transplant centers). As experience was gained, three-way exchanges became feasible, and short chains initiated by nondirected donors and also involving three transplants and nephrectomies could be scheduled simultaneously.

A nondirected donor (NDD) donates without having a particular recipient in mind. A short, simultaneous chain of transplants involves the NDD donating a kidney to a patient in a patient-donor pair, whose donor donates to another pair, whose donor passes it forward again, to someone who does not have a willing living donor and would otherwise languish on the deceased donor waiting list (Saidman et al. 2006).

However, unlike an exchange involving only patient-donor pairs, a chain could be conducted nonsimultaneously, with every patient-donor pair receiving a kidney before donating their own, which would lessen the cost should a link be broken, since no pair would be at risk of donating a kidney without receiving one (Roth et al. 2006). The first long, nonsimultaneous NDD chain was reported in Rees et al. (2009), and today (after overcoming considerable controversy) such chains account for the majority of American kidney exchange transplants, which recently number well over a thousand a year. Because NDD chains can be conducted non-simultaneously they can be long, since fewer resources need to be simultaneously available. And because they ease the coordination problems between hospitals, they open up kidney exchange to a larger pool of patients and donors, including hard to match patient-donor pairs who are unlikely to match when exchanges are restricted to be among small numbers of pairs.

As kidney exchange has grown, and exchange chains extend across the country, donors travel less often and kidneys are shipped, so coordination of inter-hospital kidney exchange has become complex in different ways. Now surgeries sometimes are arranged to take account of airline schedules. And hospitals in different Medicare regions levy different charges, so that hospitals needed to be reimbursed for nephrectomies that may have been charged at vastly different amounts. (This obstacle has now been largely overcome by the introduction of standard acquisition charges, see Rees et al. 2012).

Reflecting these difficulties in inter-hospital coordination, Agarwal et al. (2019) report that as of 2014 more than

---

29 Nonsimultaneous nondirected donor chains didn’t become standard practice until after the publication of Ashlagi et al. (2011a, 2011b). NEPKE never adopted nonsimultaneous chains, and ceased organizing kidney exchange in July, 2011, having completed only 83 exchange transplants. It’s operations were folded into a pilot program operated by the United Network for Organ Sharing (UNOS), which joined national kidney exchange networks already operated by the Alliance for Paired Kidney Donation (APKD) and the National Kidney Registry (NKR), as well as exchanges operated within single hospitals and hospital chains.

30 In 2020 there was a pandemic-related drop to 835 kidney exchange transplants. I don’t think that the numbers reported for kidney exchange ("paired donation") take account of the nondirected donors ("Non-Biol,unrel: Anonymous Donation") in the OPTN data https://optn.transplant.hrsa.gov/data/view-data-reports/national-data/, who now number in the hundreds.

31 Direct airline flights also facilitate sharing of deceased donor organs, see Wang et al. (2022).

32 Melcher et al. (2013) report the results of a ‘consensus conference’ aimed at solving other coordination problems.
60% of all exchange transplants in the U.S. were (still) internal exchanges made within a single hospital.\textsuperscript{33} That is, the demands of inter-hospital coordination, which was such a large obstacle to kidney exchange when it began, has taken years to overcome in the U.S., although we have slowly made considerable progress.

However much more progress is needed. One innovation that has yet to be widely accepted in the U.S. is to start chains with a nondirected deceased donor. (Almost all deceased donors are nondirected, but current practice is to use each deceased donor kidney for a single transplant to someone on the deceased donor waiting list.) However, a proposal for deceased donor originated chains in Melcher et al. (2016) has been productively implemented in Italy (in a collaboration between physicians and economists there; see Furian et al. 2020).\textsuperscript{34}

Big gains could still be made by organizing kidney exchange across national borders, which today happens only seldom, in exceptional circumstances. All the coordination difficulties that had to be overcome before kidney exchange could become a significant source of transplants in the U.S. are magnified when borders must be crossed. Some of these difficulties are financial, but these should be manageable because transplantation is much cheaper than dialysis, so transplants generate savings for national health systems when a patient who would otherwise languish on dialysis is transplanted (Rees et al. 2017). To date, only handfuls of foreign patients have been included in American kidney exchange chains, or involved in cross border exchange in Europe (Biró, 2019). This is partly because cross border exchange, particularly when funded by savings from foregone dialysis, was initially characterized by some opponents as being similar to illegal organ trafficking (but fortunately this is changing).\textsuperscript{35}

But there remain countries (such as Germany and Brazil) where even domestic kidney exchange faces legal bans, and other countries where it is practiced not at all (such as China), or only with difficulty (such as India). So there are big gains yet to be made domestically in these large countries as well.\textsuperscript{36}

Small countries, however, may not be able to operate kidney exchange at efficient scale domestically, and so may especially benefit from cross-border cooperation. For example, Israel and the UAE each have populations of around ten million, which isn’t big enough to reliably find matches for hard to

\textsuperscript{33} See also Agarwal et al. (2018).
\textsuperscript{34} In the U.S., deceased donation is regulated by a large national bureaucracy, while living donor transplants are much less regulated. Merging these two systems so that a deceased donor kidney could start a chain of transplants that began within the kidney exchange pool and ended with a transplant to someone on the deceased donor waiting list will involve changes in regulations that have now been under consideration for several years. (Early ideas about integrating deceased donors into exchange pools can already be found in Delmonico et al (2004) and Roth, Sönmez, and Ünver (2004).)
\textsuperscript{35} See e.g. Minerva et al. (2020), Ambagtsheer et al. (2020) for recent assessments.
\textsuperscript{36} China and India, while they are not internationally competitive in the number of transplants per million population, nevertheless do more total kidney transplants than any country other than the U.S. (And Brazil, where kidney exchange isn’t yet legal, is fourth in the world in total kidney transplants.) So the potential for increased transplantation in these large countries is enormous. In India, there is a thriving kidney exchange program in the state of Gujarat, despite the fact that nondirected donation is not allowed (amidst many other restrictions). Novel ways to organize kidney transplants in the face of these difficulties have been pioneered there; see e.g. Kute et al. (2021).
match pairs. In 2021 I had the privilege of visiting the UAE in connection with the first (and so far only) kidney exchange between those two nations, which gives an indication that all sorts of obstacles—medical, logistical, social and political—can be overcome to save lives.\textsuperscript{37} If Israel and the UAE can do it, it may eventually even be possible for the U.S. and Canada.

Kidney exchange thus remains a work in progress.

5. Conclusions

Effective market design often requires attention both to how participants interact with one another within a marketplace, and to how the marketplace interacts with a larger market and a still larger economy.

This dual mandate has contributed to a changed view of game theory. When I studied game theory in the 1970s, the received wisdom, largely following von Neumann and Morgenstern (1944), was that there were two kinds of game theory—cooperative and noncooperative-- for two kinds of games. But market design has called for combining those approaches. Dominant strategies and equilibrium (tools of noncooperative game theory) are helpful for thinking about behavior in rule-based interactions within a marketplace, within which participants’ strategies can be modeled well. Outside of the marketplace, strategy sets can be large and largely unobservable, and the tools of cooperative game theory (like coalitional models, stable matchings and the core) are helpful in modeling the incentives that may exist for transacting in the larger world (Roth and Wilson 2019).\textsuperscript{38}

And because no marketplace operates in isolation from the larger world, marketplace designs may need to adapt to changes in the larger environments.

In labor markets such as those in medicine and economics, these changes in the environment have included the growing number of two-career households, the rise in electronic applications, and lately the easy availability of remote interviewing. Market designs also need to take account of societal constraints, including reservations framed in moral or ethical terms, concerning for example interviews in hotel rooms or in States with gaps in reproductive health care, or payments to kidney donors, or kidney exchange across borders between countries with different medical systems and situations.

But while practical market design presents a host of challenges, it also offers many rewards. Among the rewards to market designers themselves is the opportunity to become intimately familiar with markets that shape the lives and careers of their participants.


\textsuperscript{38} The interaction between centralized marketplaces and the decentralized market opportunities that may precede them are worth special attention, on which see Benjamin Roth and Shorrer (2021).


