THE REDESIGN OF THE MATCHING MARKET FOR AMERICAN PHYSICIANS: SOME ENGINEERING ASPECTS OF ECONOMIC DESIGN

Alvin E. Roth Elliott Peranson*

Abstract

We report on the design of the new clearinghouse adopted by the National Resident Matching Program, which annually fills approximately 20,000 jobs for new physicians. Because the market has complementarities between applicants and between positions, the theory of simple matching markets does not apply directly. However, computational experiments show the theory provides good approximations. Furthermore, the set of stable matchings, and the opportunities for strategic manipulation, are surprisingly small. A new kind of "core convergence" result explains this; that each applicant interviews only a small fraction of available positions is important. We also describe engineering aspects of the design process.

The entry level labor market for new physicians in the United States is organized via a centralized clearinghouse called the National Resident Matching Program (NRMP). Each year, approximately 20,000 jobs are filled in a process in which graduating physicians and other applicants interview at residency programs throughout the country, and then compose and submit Rank Order Lists (ROLs) to the NRMP, each indicating an applicant's preference ordering among the positions for which she has interviewed. Similarly, the residency programs submit ROLs of the applicants they have interviewed, along with the number of positions they wish to fill. The NRMP processes these ROLs and capacities to produce a matching of applicants to residency programs.

The clearinghouse used in this market dates from the early 1950's. It replaced a decentralized process that suffered a market failure when residency programs and applicants started to seek each other out individually through informal channels, earlier and earlier in advance of employment, rather than

*We thank Aljosa Feldin for able assistance with the theoretical computations reported in section VI. Parts of this work were financed by the National Resident Matching Program, and parts by the National Science Foundation.

waiting to participate in the larger market. (By the 1940's, contracts were typically being signed two years in advance of employment.) Although the matching algorithm has been adapted over time to meet changes in the structure of medical employment, roughly the same form of clearinghouse market mechanism has been used since 1951 (see Roth, 1984). The kind of market failure that gave rise to this clearinghouse has since been seen in many markets (Roth and Xiaolin Xing, 1994), a number of which have also organized clearinghouses in response.

In the mid 1990's, in an environment of rapidly changing health care financing with many implications for the medical labor market, the market began to suffer a crisis of confidence concerning whether the matching algorithm was unreasonably favorable to employers at the expense of applicants, and whether applicants could "game the system" by strategically manipulating the ROLs they submitted. The controversy was most clearly expressed in an exchange in *Academic Medicine* (Kevin J. Williams 1995a,b, Peranson and Richard R. Randlett 1995a,b). In reaction to this exchange, groups such as the American Medical Student Association together with Ralph Nader's Public Citizen Health Research Group (1995), and the Medical Student Section of the AMA (1995) advocated that the matching algorithm be changed and/or that the description of the match be changed to give applicants more accurate advice about how to participate.[1]

Medical school personnel responsible for advising students about the job market began to report that many students believed the NRMP did not function in the best interest of students, and that students were discussing the possibility of different kinds of strategic behavior. Given the prior history of market failure due to lack of confidence in the market in this and other entry level professional labor markets, these reports deserved and received the most serious attention.

In this atmosphere, in the Fall of 1995 the Board of Directors of the NRMP commissioned the design of a new algorithm for conducting the annual match, and a study comparing it to the existing NRMP algorithm. The present paper reports how the new algorithm was designed, how the two algorithms were compared, and what was learned about the market in the process. (In May of 1997 the NRMP Board of Directors decided to switch to the new algorithm, and the first match using the new algorithm was successfully completed in March 1998.)

In the course of designing, testing, and evaluating the new clearinghouse algorithm, some surprising properties of large labor markets emerged. The high transaction costs involved in interviewing place a practical limit on how many interviews are conducted, and one consequence of this is that the set of stable outcomes is very small, and there are very few opportunities for participants to engage in strategic manipulation of their stated preferences when it comes to making and accepting offers. (Neither of these would be the case in the absence

[1] At around the same time, the Antitrust Division of the Department of Justice initiated a wide ranging discovery process concerning these markets. This ultimately gave rise to a fairly narrowly focused consent decree involving the practices of the Association of Family Practice Residency Directors (U. S. District Court for the Western District of Missouri, (1996)).

of transaction costs.)

Aside from describing these new facts, and presenting some theoretical computation to explain them, we also describe in this paper the *process* by which the new clearinghouse algorithm was designed, evaluated, and compared with the existing algorithm. At each stage this process involved computational experiments. This process resembles engineering practice rather than theorem proving or hypothesis testing. But, despite the fact that economists are increasingly called upon to design markets, there is little or no economic literature devoted to the engineering aspects of economic design, and the practical problems of moving from theory about simple markets to workable institutions for complex markets. Yet if we fail to develop such an "engineering" literature, we will fail to profit from design experience in a cumulative way. The present paper then, in addition to presenting some new results, is intended to take a step in the direction of an engineering literature as well, by describing how those facts were learned, and how they impacted on design decisions.[2]

A rough analogy may be helpful for thinking about how the different parts of this paper hang together. Consider the design of suspension bridges. The Newtonian physics they embody is beautiful both in mathematics and in steel, and college students can be taught to derive the curves that describe the shape of the supporting cables. But no bridge could be built based only on this elegant theoretical treatment, in which the only force is gravity, and all beams are perfectly rigid. Real bridges are built of steel, and rest on rock and soil and water, and so bridge design also concerns metal fatigue, soil mechanics, and the force of waves and wind. Many design questions concerning these real world complications cannot be answered analytically, but must be explored using physical or computational models. Often these involve estimating magnitudes of phenomena missing from the simple Newtonian model, some of which are small enough to be of little consequence, while others will cause the bridge to fall down if not adequately addressed. And so, just as no suspension bridges could be built without an understanding of the underlying physics, neither could any be built without understanding many additional features, also physical in nature, but more varied and complex than addressed by the simple model. These additional features, and how they are related to and interact with that part of the physics captured by the simple model, are the concern of the scientific literature of engineering. Some of this is less elegant than the Newtonian model, but it is what makes bridges stand. Just as important, it allows bridges designed on the same basic Newtonian model to be built longer, stronger, and lighter over time, as the complexities and how to deal with them become better understood.

For the design of the medical labor market clearinghouse, the underlying theory is the theory of two-sided matching. Simple models of two-sided matching markets have proven to be elegant and tractable, and very useful in un-

[2]Some beginnings of such a literature can also be found in connection with the design of electricity markets (Wilson, 1993), and the auction of radio spectrum; see e.g. Ausubel et al (1997), Cramton (1997), Ledyard et al. (1997), McMillan (1994, 1995), McAfee and McMillan (1996), Milgrom (1997), Plott (1997), Salant (1997). There is of course already something of an engineering oriented literature in finance; for an innovative example see Shiller (1993).

derstanding the organization and evolution of many markets. But the theory concentrates on simple models in which no worker needs more than one job, and there are no married couples or other connections between workers or between positions. There is a large body of theory relevant to design problems (see e.g. Roth and Marilda Sotomayor, 1990), but none of the theorems apply directly to the medical market, although many of the counterexamples do. That is, many of the existing theorems rest on assumptions not met in the complex medical market, and many of the medical market's complexities are known to open the door to the possibility of serious design problems. But the counterexamples do not give any guidance to the magnitude of these problems, and for this we will have to rely on computational exploration, both of the data from the medical market itself, and of simpler models which will help explain what is going on in the complex market. In both cases, the computational explorations will be guided by the theory, which will make possible computational experiments that would be impossible to conduct by brute force on such large markets. It seems likely that as game theory moves from simple conceptual problems to complex design problems, we will need to make more general use of this interaction among theory, computational investigation of market data, and theoretical computation, and that this in turn will produce new problems and directions for traditional theory.

This paper is organized as follows. Section I gives an overview of the medical market and the design problem, and presents some necessary background by discussing stable matchings and why they are important, how complex markets differ from simple markets with respect to stable matchings, and how the algorithm used by the NRMP prior to this study is structured. Section II presents statistics describing the market and previous match results. These demonstrate that three of the four match variations that make the NRMP a complex market are present in substantial numbers. Section III describes how the new algorithm was designed, including the role of computational experiments. Section IV compares the performance of the two algorithms on the data from recent matches, and Section V looks at the possibilities for strategic behavior when each of the two algorithms is employed. In studying the possibilities for strategic behavior, we will first treat the ROL data as if they were the true preferences of the agents, and then (in Section VI) show why this is justified, and also explain why the set of stable matchings turns out to be so small. Section VII presents some thoughts on the interplay among theory, computational experiments, and theoretical computation in the design of market mechanisms. The theory of simple markets framed the questions that needed to be answered in the course of this design, and suggested how to construct and evaluate computational experiments on the complex system to answer these questions. The magnitudes determined by the computational experiments were then explained with theoretical computations on simple markets, providing results which, with the aid of theory, could be unambiguously interpreted. This interplay was what gave the present design effort its "engineering" flavor, and we suspect that this will generalize to other design efforts. Section VIII concludes.

I. Background to the Present Study

The considerable body of theory that has been developed for two-sided matching markets, together with multiple opportunities to observe empirically both the successful and unsuccessful clearinghouse organization of other entry level labor markets, provided a general road map for both the design and evaluation of a new clearinghouse algorithm. Specifically, there was strong empirical evidence that successful clearinghouses are generally those that produce matchings that are stable in the sense that they do not create "blocking pairs" of agents, not matched to one another, who would mutually prefer to be matched to one another than to accept the matching produced by the clearinghouse. And the theory clearly shows that, in sufficiently simple markets (simple in a way that will shortly be made precise), systematic welfare comparisons can be made between different stable matchings, with some being relatively favorable to firms and unfavorable to workers, and some the reverse. In addition, for sufficiently simple markets the theory allows strong conclusions to be drawn about the opportunity and scope for strategic behavior. (For an overview of the theory, relevant parts of which will be reviewed below, see Roth and Sotomayor 1990).

The goal of the design was to construct an algorithm that would produce stable matchings as favorable as possible to applicants, while meeting the specific constraints of the medical market. The comparisons between the new and existing algorithms were to focus both on how many applicants and residency programs could be expected to receive more or less preferred matches under the two algorithms, and on how the different algorithms might influence the opportunity or need for strategic behavior by applicants and programs. Closely related issues were what advice could be given to participants in the match when it is conducted with one or the other of the algorithms, and what kinds of changes in the behavior of match participants might be anticipated if the matching algorithm were changed.

These questions were at the heart of the controversy that spilled into the medical journals in 1995. Much of that discussion referred to results in the theoretical literature concerning simple two-sided matching markets. But although the NRMP originated as a simple market, it has become more complex particularly since the early 1980's, as it has developed complementaries and linkages between positions and between applicants. These arise through four kinds of "match variations," introduced to accommodate the changing structure of the medical labor market, namely:

(i) couples in the applicant pool who seek two positions close to one another;

(ii) applicants who seek second year positions in the match and, if they are successful, have supplemental rank order lists which must be consulted to match them to prerequisite first year positions;

(iii) residency programs with positions that revert to other programs if they remain unfilled;[3] and

[3]Typically these reversions arise when e.g. the director of a second year postgraduate residency program arranges with the director of a prerequisite first year program that his

(iv) programs that wish to fill an even number of positions if they cannot fill all their positions.

These linkages can be shown to allow situations in which many of the conclusions reached about simpler markets no longer apply.

It was therefore necessary, both in designing the new algorithm and in making comparisons to the existing algorithm, to first conduct computational experiments to determine the extent to which the predictions of the theory of simple matching markets applied to the NRMP. These computational experiments, as well as those employed to compare the two algorithms, were conducted on the Rank Order Lists submitted by all applicants and residency programs in the four most recent matches (1993, 1994, 1995, and 1996) and in the 1987 match. The recent matches were selected to have contemporary patterns of preferences among applicants and residency programs, and 1987 was selected for a comparison over a longer period, and specifically because it had the lowest rate of unmatched US seniors in the available data set (6.0 percent, as opposed to the historically high rate of 7.5 percent for 1996).

A number of specialty matches are also run under the auspices of the NRMP, and these are largely free of the match variations which add complexity to the general resident match. The existing theory of simple matching markets therefore provides accurate predictions about the nature and direction of changes to be anticipated in these matches if the existing NRMP algorithm were replaced by the new algorithm. However the theory offers little guide to the *magnitude* of the changes to be expected, and for this purpose computational experiments on the data of past matches were also needed. These were conducted for the Thoracic Surgery match, for the five years 1991-1994 and 1996.

The design of the new algorithm and the comparisons of the two algorithms will be discussed in detail in the body of the paper. But the general conclusions can be summarized by noting that, both for the NRMP and the specialty matches, the effects of changing from the existing algorithm to the newly designed algorithm are in the directions predicted by the theory for simple markets, but the size of these changes is small, and the opportunities for profitable strategic behavior are comparably small for both applicants and programs under either algorithm.

In the course of explaining why the differences are so small, we will present a new kind of "core convergence" result, which shows that the size of the set of stable matchings becomes small as the size of the market increases, even when preferences are uncorrelated, provided that the number of positions for which an applicant can interview remains small (and not otherwise).

residents will spend their first year in that prerequisite program. However if the second year program then fails to match with as many residents as were anticipated, this leaves vacancies in the first year program that can be filled by other applicants.

A. Stable Matchings in Simple and Complex Matching Markets

Centralized matching mechanisms often arise to solve market failures due to unraveling of appointment dates. Perhaps the most important and least controversial empirical finding about centralized matching algorithms is that they are most often successful if the matchings they produce are stable (Roth '84, 90, 91, Roth and Xing '94, John Kagel and Roth '99). In a simple matching market, a matching between applicants and residency programs is *stable* if there is no applicant or program matched to an unacceptable (unlisted) partner, and if there are no applicant-program pairs such that the applicant prefers the program to his/her current match, and the program also prefers the applicant to one of its current matches (or vacant position).[4]

So this study, and the controversy which preceded it, focused on choices among algorithms which produce stable matchings. The reason for the controversy is that there can be systematic differences among stable matchings. Appendix C gives formal definitions of stability in simple and complex matching markets, but the basic ideas can be conveyed by considering the "deferred acceptance algorithm" first formally studied by David Gale and Lloyd Shapley (1962).[5] There are two basic versions of this algorithm, in each of which one side of the market (firms or workers) makes offers, which the other side can reject or hold to see if any better offers are forthcoming.

In the worker proposing version of the algorithm, each worker begins by applying for the position at the top of her preference list. Each firm rejects any unacceptable candidates, and if it has q positions it temporarily holds the (up to) q most preferred applications it has so far received, and rejects the rest. A candidate who is rejected at any step of the algorithm next applies to her next highest ranked position (if any remain) among those not yet applied to. The algorithm stops at any step in which no new applications are made, at which point each worker is matched to the firm (if any) holding her application.

In a simple market the resulting matching must be stable (i.e. there are no firm-worker blocking pairs) since, if a worker w prefers firm f to her final match, she must have applied to firm f and been rejected, and hence firm f does not

[4] Among the programs and applicants who have interviewed one another, programs do not list applicants with whom they are unwilling to match, and applicants do not list programs with whom they are unwilling to match. (Unmatched programs and applicants can be matched in the post-match secondary market called the "scramble," which takes place primarily in the twenty four hour period before the official public announcements of the match results.) Programs and applicants generally also do not list applicants or programs with whom they have not had interviews (and this is of course an equilibrium, since the clearinghouse produces a stable matching, at which you cannot be matched to someone who has not listed you, so there is no incentive to list him.) There is also a charge to applicants who list more than 15 residency programs, which may dissuade some applicants from listing some programs.

[5] Although Gale and Shapley discussed the algorithm in an abstract setting, it appears that, in various forms, equivalent algorithms have been developed in applied contexts both before and since, with the initial NRMP algorithm, dating from 1951, being the first we know of (Roth, 1984).

prefer her to any of the workers whose applications it held when the algorithm stopped. Furthermore, Gale and Shapley showed that, when preferences are strict, the particular stable matching produced by the worker-proposing version of the algorithm gives each worker her most preferred position among those she can get at any stable matching. Even more striking, the firm-proposing version of the algorithm gives every firm which fills q positions its q most preferred workers among those it can be matched to at any stable matching (Roth 1985, Roth and Sotomayor 1989). Much of the controversy about the organization of the NRMP focused on this difference between these two versions of the deferred acceptance algorithm.

But a deferred acceptance algorithm may fail to produce a stable matching in a market with some of the complexities of the NRMP, such as the presence of couples who submit rank order lists of pairs of positions. The key to the stability of the outcome in simple markets is that (in the worker-proposing version of the algorithm) no firm ever regrets having rejected a worker's application, since it only does so when it has an application it prefers, and it will be matched to this preferred application unless it receives applications it prefers even more. But, in a market containing couples, suppose that a firm f_1 receives an application from a worker w_1, and rejects an application from a less preferred worker $w\prime$ in order to hold w_1's application. Suppose further that w_1 is married to w_2, whose application is being held by firm f_2, because the pair (f_1, f_2) is high on the preference list submitted by the couple $\mathbf{c} = (w_1, w_2)$. Finally, suppose that firm f_2 now receives an application it prefers, and rejects the application of w_2. In order for the couple $\mathbf{c}$ to now apply to its next-choice pair of firms (f_3, f_4), w_1 must now be withdrawn from firm f_1. So firm f_1 now regrets having rejected worker $w\prime$, and there may be a potential instability involving f_1 and $w\prime$ (and if $w\prime$ is part of a couple this instability may involve another firm as well; see Appendix C).

The differences between simple and complex markets involve more than the failure of the deferred acceptance algorithm to produce stable matchings, but extend to the non-emptiness and structure of the set of stable matchings itself. Some of the important differences are summarized below, by noting theorems about simple matching markets which do not hold when the market contains couples or other linkages which create complementarities between positions or applicants. (See Roth and Sotomayor, 1990, for a comprehensive treatment and more detailed references to the literature.)

(i) In simple matching markets, firm and worker optimal stable matchings exist for all possible ROLs, and are produced by the firm and worker proposing variants of the deferred acceptance algorithm (Gale and Shapley, 1962, Roth, 1985).

(i') In markets with complementarities, no stable matching may exist, and even when stable matchings exist there may be no optimal stable matchings for either side of the market (Roth, 1984, Brian Aldershof and Olivia Carducci, 1996).

(ii) In simple markets, the same applicants are matched and the same positions are filled at every stable matching. (That is, any applicant who is unmatched at one stable matching is unmatched at every stable matching, and the positions which are unfilled are the same at every stable matching.) Furthermore, a firm which fills only some of its positions at a stable matching fills them with the same workers at every stable matching (Roth, 1986).

(ii') In markets with complementarities, different stable matchings may have different applicants matched and different positions filled (Aldershof and Carducci, 1996).

(iii) In simple markets, when the applicant proposing algorithm is used (but not when the program proposing algorithm is used) it is a dominant strategy for applicants to submit ROLs corresponding to their true preferences. No parallel assertion can be made about residency programs which have more than one position (Roth 1982, 1985).

(iii') In markets with complementarities, no algorithm exists that chooses a stable matching whenever one exists and makes it a dominant strategy for all agents to state their true preferences (Roth, 1985, Aljosa Feldin, in preparation).

So a major focus of this study was to assess the extent to which these theoretical possibilities play a role in the actual NRMP matches. In the course of this report it will become clear that, while it has always been possible to find stable matchings in the previous years' NRMP matches (a stable matching has been found in every match at least since the mid 1970's), it appears that no stable matching is precisely program-optimal or applicant-optimal in any of the years we have examined. However we will see that applicant-proposing and program-proposing algorithms continue to perform approximately as in the case of simple markets.

B. The Pre-existing NRMP Algorithm:

The pre-existing NRMP algorithm (the one in use in 1995 when this study began, and used through the 1997 matches) is the result of incremental modifications over a period of years. It is primarily, but not entirely, a program-proposing algorithm, and deals with match variations through a three-phase process. The first phase produces an initial match by ignoring most match variations, using the program-proposing deferred acceptance algorithm, modified to let couples hold on to many offers until a late stage in the algorithm. The match produced in this way will in general not be stable (because of the way it handles couples, and because the other match variations are ignored), so the second phase of the program identifies potential instabilities. The third phase of the program uses an algorithm to fix these instabilities one by one and produce a stable match. The processing in this third phase does not always have residency programs proposing. Instead, couples propose in part of the algorithm designed to fix instabilities due to couples, and applicants also propose in part of the algorithm which fixes instabilities related to supplemental (first year) matches. Thus the 1995 NRMP algorithm is a hybrid; program-proposing in its first phase (which performs the bulk of the matching), and applicant-proposing in some parts of its third phase.

The NRMP specialty matches like Thoracic surgery are run using an algorithm that is technically a little different than the original NRMP algorithm (it does not handle some of the NRMP match variations such as the use of supplemental lists to form multi-year matches, and it is organized in a single phase). But when no match variations are present, the specialty match algorithm and the 1995 NRMP algorithm are functionally equivalent to the program-proposing deferred acceptance algorithm in that they all produce the program-optimal stable matching.

II. The NRMP in the Years 1987 and 1993–6

Table 1 gives the descriptive statistics of the NRMP match in the five years we consider. Notice that in each year, a substantial number of the more than twenty thousand applicants who participate do so in ways which utilize the match variations which the NRMP allows–about 4 percent participate as couples, and 8-12 percent submit supplemental Rank Order Lists. In addition, in the 1990's about 7 percent of the three to four thousand programs which participate in each year have positions which could revert to other programs if they remain unfilled (accounting for almost 6 percent of the total quota of positions). So the match variations are a substantial part of the match. Before investigating how these match variations change the properties of stable matches and of strategic behavior, the first task is the design of an applicant proposing algorithm to produce stable matches that meet the match variation requirements of thousands of participants.

A. Specialty Matches: Thoracic Surgery in the Years 1991–94 and 1996

In contrast, the Thoracic surgery match is a simple match, with no match variations. Its basic descriptive statistics and match results are given in Table 1B.

III. Design of the Applicant-Proposing Algorithm

The process by which the applicant-proposing algorithm was designed is roughly as follows. First, a conceptual design was formulated and circulated for comment (Roth, 1996a). This was based on an algorithm for simple markets, modified to deal with the complexities of the NRMP. In order for this to be coded into a working algorithm, a number of choices had to be made concerning the sequence in which proposals would be made. The sequencing of proposals can be shown to have no effect on the outcome of simple matches, but could potentially effect the outcome when the NRMP match variations are present. Thus, like the overall architecture of the algorithm, the sequencing of proposals is a design question about which the existing theory gives some general guidance that

Table 1: Descriptive Statistics and Original Match Results for (A) the NRMP and (B) Thoracic Surgery

A. NRMP

Category	1987	1993	1994	1995	1996
Applicants (Active, ROLreturned)					
Primary ROLs	20,071	20,916	22,353	22,937	24,749
Applicants with supplemental ROLs	1,572	2,515	2,312	2,098	2,436
Results					
Primary matches	16,117	17,209	17,725	18,170	18,316
Supplemental matches	577	1,294	1,152	990	725
Couples					
Applicants who are coupled	694	854	892	998	1,008
Coupled applicants who matched	646	794	817	899	912
Programs					
Active programs	3,225	3,677	3,715	3,800	3,830
Active programs with ROL returned	3,170	3,622	3,662	3,745	3,758
Potential reversions of unfilled positions					
Programs specifying reversion	69	247	276	285	282
Positions to be reverted if unfilled	225	1,329	1,467	1,291	1,272
Programs requesting even/odd matching	4	2	6	7	8
Quotas:[a]					
Total quota before match	19,973	22,737	22,801	22,806	22,578
Changes during match processing					
Quota decreases					
Programs	22	120	143	124	130
Positions	45	357	357	327	336
Quota decreases					
Programs	23	127	142	128	138
Positions	46	338	338	303	326
Total quota after match (final quota)	19,972	22,756	22,820	22,830	22,588
Results					
Positions filled	16,694	18,503	18,877	19,160	19,041
Positions unfilled	3,278	4,253	3,943	3,670	3,547
Program filled	2,100	2,309	2,440	2,599	2,589

B. Thoracic Surgery

Category	1991	1992	1993	1994	1996
Applicant ROLs	127	183	200	197	176
Active programs	67	89	91	93	92
Program ROLs	62	86	90	93	92
Total quota	93	132	141	146	143
Positions filled	79	123	136	140	132

[a] Quotas include positions in active programs with no ROL returned. Changes during the match are caused primarily by reversions. In some cases, 1 position is reverted simultaneously to 2 programs, causing a net increase in the number of positions offered. In addition, a few positions may be dropped from the match during processing to accommodate requests for even/odd matching.

falls short of a complete engineering specification. Consequently, we performed computational experiments before making sequencing choices. In what follows, we first present the conceptual design (from Roth 1996a), and then discuss the sequencing experiments and implementation decisions.

A. The Conceptual Design

The algorithm described here is based on the instability-chaining algorithm in Roth and John H. Vande Vate (1990) (which finds stable matchings by resolving applicant-program instabilities one at a time) and on the general design of phase 3 of the pre-existing NRMP algorithm.

The object of the algorithm is to produce a stable matching, by assembling a set $\mathcal{A}(k)$ of residency programs and applicants and a tentative matching $M(k)$ with the property that there are no instabilities within the set $\mathcal{A}(k)$, and no applicant or program in $\mathcal{A}(k)$ is matched to anyone outside of $\mathcal{A}(k)$. When the set $\mathcal{A}(k)$ has grown to include all applicants and programs, the resulting match is stable, and the algorithm stops.

In the applicant-proposing algorithm, the initial set, $\mathcal{A}(0)$, consists of all positions offered in the match, and the initial tentative matching has all positions vacant. The algorithm begins by selecting an applicant $S(1)$ from the set of applicants in the match and adding $S(1)$ to $\mathcal{A}(0)$ to make the new set $\mathcal{A}(1)$.

At any step k of the algorithm, at which a new applicant $S(k)$ has just been added to form the set $\mathcal{A}(k)$, the new tentative matching $M(k)$ is formed as follows. First, applicant $S(k)[= S(k,1)]$ proposes down his Rank Order List (of programs which also rank $S(k)$), from the top, until the first program is reached which either has a vacant position or which prefers $S(k)$ to its least preferred current tentative match. In the latter case, this least preferred applicant, $S(k,2)$ is now rejected by the program in question, and this applicant now proposes down her ROL in a similar way, etc. Each applicant $S(k,n)$ displaced in this way similarly proposes down his/her ROL.

At some point in this process, an applicant $S(k,n)$ may be displaced who is a member of a couple, or who is displaced from a primary (second year) position for which she also holds a supplemental (first year) position. In either case, a second position now potentially becomes vacant, as the spouse of $S(k,n)$ is withdrawn from his tentative match, or as $S(k,n)$ is withdrawn from her supplemental match. In either case, the program whose position is left vacant, $P(k,n)$, is added to a "program stack" to be held for later processing. If $S(k,n)$ is a couple, then both couple members ($S(k,n,a)$ and $S(k,n,b)$) now propose down their joint ROL of pairs of programs, and they each may displace another applicant. Also, if any $S(k,n)$ has a supplemental ROL associated with her new tentative match, she proposes down it as well, which may also result in the displacement of another applicant. So both couples and supplemental matches may simultaneously displace more than one applicant. One displaced applicant is processed immediately and any others are added to an "applicant stack" for later processing.

Applicants propose down their ROL's in this way until the applicant stack is empty. (Applicants continue throughout to be able to propose to programs which may be on the program stack.) A residency program is then selected from the program stack, and all of the applicants in $\mathcal{A}(k)$ with whom it might form instabilities—*i.e.,* all of the applicants in $\mathcal{A}(k)$ who are preferred by the program to its least preferred current tentative match and who prefer this program to their current match—are added to the applicant stack, which is processed as before, with applicants proposing down their ROL, from the top.

When both the applicant and program stacks are empty, the tentative matching thus produced is $M(k)$: no instabilities for $M(k)$ are contained in the set $\mathcal{A}(k)$, and no applicant or program in $\mathcal{A}(k)$ is matched by $M(k)$ to anyone outside of $\mathcal{A}(k)$. The algorithm is now ready to pick a new applicant $S(k+1)$, and start the process again, for the set $\mathcal{A}(k+1)$.

When all applicants have been included in the set $\mathcal{A}(k)$, even-odd requests and program reversions are adjusted, which causes additions to the applicant and program stacks, which are handled as above. When these stacks are empty, the algorithm stops, and the last tentative match becomes final.

In a match with no match variations, the applicant and position stacks would always become empty, and the final match would be the applicant optimal stable matching. When the match variations are present, there is a possibility that at some stages of the algorithm the position stacks would never become empty–i.e. a cycle would occur, in which the same positions reappeared on the stack. So "loop detectors" need to be added to each stage k. Every loop must involve a position becoming unmatched and made vacant either because a couple or a supplemental assignment has been withdrawn from the position, or a position has been withdrawn from an applicant (e.g. in satisfying an even/odd constraint). So a loop detector can work by keeping a log of when positions become unmatched in these ways–*i.e.,* recording which applicant is unmatched from which position, during the processing of some step $\mathcal{A}(k)$. If the same pairs appear multiple times, a loop is in progress. How to proceed at this point may depend on the nature of the loop. (It is observed in Roth and Vande Vate (1990) that certain kinds of inessential loops can be rendered harmless by randomizing the order in which applicants and positions are processed from the stacks. Loops due to the non-existence of a stable matching would be more serious, but the prior experience of the NRMP suggests that these may be rare.)

Thus, the existing theory suggests the general architecture for an applicant proposing algorithm that can deal with instabilities one at a time as they are detected, and provides guidance on how the algorithm may possibly fail to find a stable matching. But to determine how often it might fail to produce a stable matching we need some computational experiments. The experiments reported next, which will help determine the details of the algorithm design, will also show that failures are rare: we will not observe even a single failure when we explore different versions of the algorithm on previous years' ROL data.

B. Sequencing Questions and Implementation Decisions

In a simple match, without the NRMP match variations, the applicant-proposing algorithm just described always produces the applicant-optimal stable match and the program-proposing algorithm always produces the program-optimal stable match, regardless of the order in which proposals are processed within the algorithm. One consequence of the fact that these optimal stable matches do not exist in general when the match variations are present is that the order in which applicants and programs are processed may have an effect on the match produced. Thus the sequence in which applicants and programs are processed at various points in the algorithm needs to be considered as part of the design of the applicant-proposing algorithm.

Two issues were considered in conducting and evaluating experiments related to the sequencing of operations in the algorithms.

(i) Do sequencing differences cause substantial or predictable changes in the match result (e.g. do applicants or programs selected first do better or worse than their counterparts selected later)?[6]

(ii) Does the sequence of processing affect the likelihood that an algorithm will produce a stable matching? (In connection with this latter point, recall that instability-chain algorithms can cycle –even when stable matchings exist and certainly when they do not. So one objective was to consider how sequencing decisions might influence the frequency of “loops” occurring in the algorithm.)

Experiments to test the effect of sequencing were conducted using data from three NRMP matches: 1993, 1994, and 1995.

1. Sequencing Experiments on the Pre-existing NRMP *Algorithm*

We investigated the effect of different sequencing of operations in variants of the pre-existing NRMP algorithm, in part to establish a baseline against which to compare the algorithm to be designed. In the pre-existing algorithm, programs are processed in ascending sequence by 6 digit program code number. To test the sensitivity of the results to this sequencing, computational experiments were run on the ROL data in which this sequencing was reversed, i.e. programs were processed in descending order by program code number. As expected, the results showed differences, but the differences were small: the largest difference was in 1994 when only 4 out of 3,662 programs which submitted ROL's received a different match under the alternative ordering, as did 4 out of 22,353 applicants. Not only are these differences very small, they do not appear to be systematic.[7]

[6] Even in a simple matching market, the order in which proposals are made can matter in versions of the Roth and Vande Vate (1990) instability chaining algorithm in which members of both sides of the market may be chosen to make the next proposal (in contrast to versions in which all proposals are made by one side of the market). Blum and Rothblum (forthcoming) show that, in such a version of the algorithm, late proposers have an advantage over early proposers

[7] We use the term ”very small” informally, but not merely to express an opinion of changes which affect on the order of .01 percent of applicants. These changes are also at least an order of magnitude smaller than the main effects we will find due to changes between program proposing and applicant proposing algorithms. And since the effects appear unsystematic,

(A fuller account of the results of these experiments appears in Appendix A.)

The pre-existing NRMP algorithm was also investigated for its sensitivity to the sequence in which reversions are processed. Rather than simply changing the order in which reversions occurred, the experiments involved setting the program quotas input to the match processing to be the final, post match quotas obtained from the original results produced by the pre-existing NRMP algorithm. All further reversion processing was then eliminated. These experiments then provided an indication of the differences caused, not only by changing the order of reversions, but also by altering when reversions enter into the match processing (i.e. all required reversions were assumed to take place simultaneously, at the beginning of match processing). No more than 2 programs or applicants were observed to be affected by such changes in any of the three years 1993-5 (see Appendix A).

Finally, it should be noted that no loops were detected in any of these experiments on the pre-existing NRMP algorithm. Consequently, despite the presence of match variations, sequencing does not appear to play a significant role in the operation of the pre-existing NRMP algorithm.

2. Sequencing Experiments on the Applicant-Proposing Algorithm

Computational experiments were conducted to measure the impact of

(i) The sequence in which applicants are admitted to the algorithm for processing;

(ii) The sequence in which couples are processed relative to other applicants; and

(iii) The sequence in which applicants ranked by a program are processed when attempting to fill a program that has been selected from the program stack.

To understand the results of the computational experiments (which are tabulated in detail in Appendix A) it is useful to compare the outcomes from each experiment to those from a fixed baseline. We chose as a baseline an applicant proposing algorithm in which applicants were processed in ascending order by their applicant codes, regardless of whether they were single or members of couples. (In all cases, when a member of a couple was processed, so was the other member. When applicants were processed in ascending code order a couple was selected for processing based on the code number of the spouse with the lower applicant code.) When a program was selected from the program stack, applicants were processed in ascending sequence by program rank number. All of these experiments were carried out on the ROL data from the NRMP matches in 1993, 1994, and 1995.

The experiments were conducted in a partial factorial design. The handling of couples had 3 treatments (couples intermixed with singles, couples first, couples last), the order of introducing applicants into the match had 2 treatments (ascending order by applicant code, or descending order), and the order of processing applicants when a program is pulled from the stack had 2 treatments

they do not appear to have any welfare implications on average.

(ascending order by program rank, or descending order). The results are that none of these sequencing decisions had a large or a systematic effect on the matching produced. In two thirds of the cases the match was the same as the baseline case. (In the majority of the remaining cases only 2 applicants received different matches, and the maximum number of applicants affected was 12 out of 22,937, which occurred when a couple received a worse match and initiated a chain of displacements. This happened in two of the eighteen cases and involved the same 12 applicants in both cases).

However there was an effect of sequencing on the internal processing of the algorithm. The number of loops encountered was fewest when couples were introduced to the match after single applicants. This is not too surprising in view of the fact that no loops would occur in the absence of match variations. The results indicate that loops are least likely to occur when the couples are introduced into the larger market with some tentative matches already assembled, as opposed to when couples enter first, so that the initial tentative matches involve only couples. Introducing couples last reduces the numbers of loops (and hence the potential that in some future match it would be difficult to find a stable matching) without changing the prospects of couples or single applicants in the match.

Finally, experiments related to the sequence in which reversions are processed were performed on an applicant proposing algorithm, similar to those performed on the pre-existing NRMP algorithm. Again, no substantial changes were induced by changing the order in which reversions were handled; no changes at all resulted in the 1993 match, and only 2 applicants and programs were affected in the 1994 and 1995 matches (see Appendix A). Thus for both the pre-existing NRMP algorithm and the applicant proposing algorithm, there is almost no difference between the results obtained with reversion processing and the results obtained by setting the quotas to the final quotas after reversions and eliminating further reversion processing. (This point simplifies the design of some of the experiments to compare the two algorithms, in connection with strategic behavior by residency programs, to be discussed later in this report.)

Based on the sequencing experiments described above, it was decided to sequence all proposals by couples after proposals by single applicants, since this was the order that produced the fewest internal loops.[8]

---

[8]The full details of the sequencing decisions are:

1. All single applicants are admitted to the algorithm for processing before any couples are admitted.

2. Single applicants are admitted for processing in ascending sequence by applicant code.

3. Couples are admitted for processing in ascending sequence by the lower of the two applicant codes of the couple.

4. When a program is selected from the program stack for processing, the applicants ranked by the program are processed in ascending order by program rank number.

5. The processing of programs requesting even numbers of matches and/or reversions of unfilled positions is deferred until all applicants have been admitted for processing.

6. Programs requesting even numbers of matches are processed in ascending sequence by program code. An applicant deleted from a program in order to leave an even number of matches in the program is placed on the applicant stack for processing.

Note that we did not at any point choose to randomize the processing order (randomization was shown in Roth and Vande Vate 1990 to allow the algorithm to escape from certain kinds of loops). The reason is that loops do not appear to be a problem with the processing sequences selected, and it was felt that a desirable feature of the match is that it should be precisely reproducible from the ROL data.

IV. Differences in the Matches Produced by the Two Algorithms

A. The NRMP

The pre-existing NRMP algorithm and the newly designed applicant proposing algorithm were compared by comparing the matches that they produce for the ROL's submitted in 1987 and 1993-1996. Table 2 gives the results of these comparisons. The first half of the table concentrates on the comparisons from the point of view of applicants, the second half from the point of view of programs.

Only about 0.1 percent of applicants are affected by the change in algorithms, and of these most prefer the match they receive under the applicant proposing algorithm. Note that in two of the five years the number of applicants matched changed by one (one fewer in 1987, one more in 1996). Recall that in a simple match a change from one stable matching to another would never change the number of applicants matched; so here is another case in which the match variations cause a difference, but a difference which turns out to be very small and unsystematic.

Equally few programs are affected by the change of algorithms–and these constitute about 0.5 percent of all programs. Most but not all of the programs prefer the match they receive under the pre-existing NRMP algorithm, but in 1994 and 1996 slightly more programs would even have preferred the applicant proposing algorithm to the pre-existing NRMP algorithm. Most programs that receive a different match have only one applicant different between the matches produced by the two algorithms. The majority of differences have to do with filling a position with a different applicant; only a small number of positions

7. Programs requesting reversions of unfilled positions are processed in ascending sequence by the program code of the program "donating" the unfilled position(s). A program that "receives" a reverted position is placed on the program stack for processing.

8. After all reversions have been processed, the requests for reversions are reprocessed, in case any new reversions of unfilled positions are required as a result of changes made in the processing of reversions that have been processed since the last time this reversion request was considered.

9. When no further processing is required to satisfy all reversions, requests for even numbers of matches are reprocessed as in point 6 above, and if any changes are made, requests for reversions are reprocessed as in points 7 and 8 above. This iterative processing continues until no further changes are made by even processing or reversion processing. [The possible need for a reverted position to be "unreverted" is checked as part of the check for stability, by using original quotas for programs which have lost positions through reversions.]

move from being filled to unfilled or vice versa. Again, this is a consequence of the match variations; as already noted in the case of applicants, it turns out to be both very rare and unsystematic.

It may be helpful at this point to consider an example of precisely how the match variations can cause a deviation from the predictions of the theory for simple markets; for example how it can be that a few applicants do worse with the applicant proposing algorithm than with the program proposing algorithm. For example, if switching to the applicant proposing algorithm causes applicant A to improve his match from his 2nd to 1st choice, it may be that the 1st choice now requires a supplemental match that was not required before. If this new supplemental match displaces a previously matched but less preferred applicant in a program, that displaced applicant is forced to go farther down his/her list (i.e. does worse). Furthermore, matching that applicant may displace another applicant, who may displace another, and so on, causing a chain of applicants who do worse (even though, as expected of the applicant-proposing algorithm, this chain of events began with an applicant who did better than he would have had the program-proposing algorithm been used).

It is worth noting that when we refer to "only 0.1 percent" of applicants, we are talking about a change whose small size we will explain in what follows. But this does not necessarily imply that the associated change in welfare is small. Indeed, in the debate that led to this study, and after our report was circulated to the interested parties, a great deal of discussion stemmed from the view that the difference in welfare was likely to be large for the affected applicants, and likely to be small for the affected programs. This contributed to the decision to adopt the applicant proposing algorithm, a decision strongly lobbied for by the student organizations, and eventually unanimously adopted by the NRMP Board.[9]

B. Thoracic Surgery

Because there are no match variations in the Thoracic surgery matches for the years we consider, they are simple matches, and are well described by the existing theory. Consequently we know that the applicants will all do as well as possible at the stable match produced by the applicant proposing algorithm, and the programs will all do as poorly as possible at that stable matching. What the theory does not tell us is how large this effect will be; for that we need to look at the data. As discussed in the introduction, the effect turns out to be

[9]The argument about the size, and relative size, of the welfare effects for applicants and programs can be paraphrased in part roughly as follows. Both programs and applicants have some uncertainty in their rankings. There may not be that much difference between a program's seventh and seventeenth ranked candidates. Similarly, applicants may not be able to clearly judge whether they will get a better educational experience at their first or second choice programs. But applicants can clearly judge other factors in their preferences, such as whether they would prefer to live in Seattle or Miami, where these programs may be located. Therefore, a change of algorithms may have a big effect on the affected applicants, and only a small one on the affected programs.

Table 2: Comparison of Results Between Original NRMP Algorithm and Applicant Proposing Algorithm

Result	1987	1993	1994	1995	1996
Applicants:					
Number of Applicants	20	16	20	14	21
Applicant Proposing Result Preferred	12	16	11	14	12
Current NRMP Result Preferred	8	0	9	0	9
U.S. Applicants Affected	17	9	17	12	18
Independent Applicants Affected	3	7	3	2	3
Difference in Result by Rank Number					
1 rank	12	11	13	8	8
2 ranks	3	1	4	2	6
3 ranks	2	3	2	2	2
More than 3 ranks	2	1	1	2	3
	(max 9)	(max 4)	(max 5)	(max 6)	(max 6)
New Matched	0	0	0	0	1
New Unmatched	1	0	0	0	0
Programs:					
Number of Programs Affected	20	15	23	15	19
Applicant Proposing Result Preferred	8	0	12	1	10
Current NRMP Result Preferred	12	15	11	14	9
Difference in Result by Rank Number					
5 or fewer ranks	5	3	9	6	3
6 to 10 ranks	5	3	3	5	3
11 to 15 ranks	0	5	1	3	1
More than 15 ranks	9	4	6	0	11
	(max 178)	(max 36)	(max 31)		(max 191)
Programs with New Position(s) Filled	0	0	2	1	1
Programs with New Unfilled Position(s)	1	0	2	0	0

minimal: in the five years we studied, only four applicants and four programs would have been affected by a change in algorithms; in three of the five years, the applicant proposing algorithm would have produced the same match as the program proposing algorithm, indicating that this was the only stable matching in those years. (August Colenbrander, 1996, reports similarly small differences in the specialty matches he maintains.)

Table 3: Difference in Result when Algorithm Changed from Preexisting Specialty Match to Applicant-Proposing

Year	Difference
1991	none
1992	2 applicants improve, 2 program do worse
1993	2 applicants improve, 2 programs do worse
1994	none
1996	none

V. Differences in Sensitivity to Participant Behavior

The comparisons of match outcomes discussed in the previous section are all based on Rank Order Lists which were submitted for matches made by the pre-existing NRMP and specialty match algorithms. While the changes observed when the match was instead produced by the applicant proposing algorithm were small, a comparison of the algorithms also requires us to consider whether participants might have reason to submit different kinds of ROL's if the new algorithm were to be substituted for the pre-existing one. For this purpose, we consider whether participants could have favorably influenced the match, under either algorithm, by submitting different ROL's. The idea is both to assess how many participants could do so, and how the number is different for the two algorithms. This will also allow us to determine what kinds of advice can be given to participants about how to participate in the match, under either algorithm.

Once again, this is a subject about which the theory of simple matching markets tells us a great deal, for markets without the match variations found in the NRMP. To see how well the theory for simple markets approximates the NRMP matches, and also to assess the size of the effects to expect, again required computational experiments on the data. A quick review of the theory will help organize the discussion.

A. Strategic Behavior in Simple and Complex Matching Markets

In a simple matching market, without match variations, it has been shown (Roth 1982) that there do not exist any stable matching algorithms which completely remove the possibility that some applicant or program can get a better match by submitting an ROL that differs from his/her/its straightforward preferences. However, we already noted that:

> In simple markets, when the applicant proposing algorithm is used, but not when the program proposing algorithm is used, no applicant can possibly improve his match by submitting an ROL that is different from his true preferences. (No parallel assertion can be made about residency programs that have more than one position.)

So in simple markets, we would find strategic opportunities for applicants only when the program proposing algorithm is used, and the theory tells us what these might be. Specifically, consider the ROL of some applicant, and define a truncation of that ROL to be a shorter ROL which is the same as the original ROL for as many programs as it ranks. We can then say the that following:

> In simple markets when the program proposing algorithm is used, every applicant who can do better than to submit his true preferences as his ROL can do so by submitting a truncation of his true preferences. That is, if (holding all other ROL's constant) an applicant would be matched to his kth choice if he submitted his true preferences, and his j choice (with $j < k$) if he submitted some other ROL, then he can be matched to his j choice by submitting a truncation of his true preferences at the j choice. Furthermore, no part of his original ROL below the kth choice has any effect on the match. (Roth and Vande Vate 1991.)

It can also be shown that truncations are the kind of manipulation that can potentially be identified with the least information about others' preferences (Roth and Rothblum, 1999).

In simple markets, the reason that all successful manipulations can (also) be accomplished by truncations is that, in a simple market, a deferred acceptance algorithm never "backtracks:" no information in an agent's ROL is used beyond the point at which that agent is matched. Although we can't apply this result directly to the complex market, we can do computational experiments to assess how good an approximation is provided by concentrating only on truncations in the investigation of possible strategic manipulations in the NRMP. Specifically, if we find that information about agents' preferences among options below the point at which they are matched has little effect on the match, then we can be confident that investigating truncations will give us a comparably good approximation for the magnitude of possible strategic manipulations in the complex

NRMP market.[10]

For simple markets, the theory also tells us which applicants can potentially profit from manipulation, and how much:

> In simple markets, when the program proposing algorithm is used, the only applicants who can do better than to submit their true preferences are those who would have received a different match from the applicant proposing algorithm. Furthermore, the best such applicants can do is to obtain the applicant optimal match, and they can do this by submitting to the program proposing algorithm the truncation of their true preferences which stops at the match they would have gotten from the applicant proposing match. (see Gabrielle Demange, Gale and Sotomayor, 1987, and Roth and Sotomayor 1990.)

It is important to note that, even in the case of a simple match without match variations, in general an applicant would not have the information needed to submit such a truncation (and if he submitted a truncation that was one program too short he would become unmatched). But this result shows that, in a simple match, we can identify an upper bound on the number of applicants who could possibly profit from manipulating their rank order lists, by seeing how many applicants receive different matches at the two algorithms.

We cannot directly apply this upper bound to the NRMP, because it depends for its proof on the existence of optimal stable matchings for each side of the market, which we know (from the sequencing experiments) do not exist in the NRMP data. But the theory of simple matches allows us to use the computational results reported in Table 2 as a numerical benchmark against which to compare the computational estimates we will make of the scope for possible manipulation. That is, we can compare the estimates we get of how many applicants can potentially profit from strategically stating their ROLs with the numbers of applicants who were observed to get different matches from the two algorithms. If these numbers are close for the program-proposing algorithm (and close to zero for the applicant proposing algorithm), then the theory of simple matches provides a comparably close approximation for the situation in the complex NRMP market.

The case of programs that have more than one position is not so simple, even in the case of simple matches. Programs may, at least in theory, possibly profit both from truncating their ROL's, and from reducing the number of positions they submit to the match (either by making early arrangements with some applicants or by withholding positions to be filled by unmatched applicants after the match). The temptation for this latter kind of manipulation can be shown (Tayfun Sonmez 1996a,b) to be larger with the program proposing algorithm than with the applicant proposing algorithm. Thus, in addition to experiments

[10] This would free us from the computationally impossible task of investigating all possible manipulations by all participants.

with truncations of ROL's, we must also conduct computational experiments involving reductions in stated capacities.

B. Experiments to Determine Upper Bounds for Profitable Strategic Behavior

1. Preliminary Experiments: Truncation of ROL*'s at the Match Point*

As noted above, in a simple market if an applicant is matched to his kth choice program, or if the lowest ranked applicant a program is matched to is its kth choice applicant, truncation of the ROL at the kth entry would have no influence on the match. This is because, in a simple match, the applicant or program proposing algorithms never have to "backtrack" on an ROL. But in the NRMP backtracking can occur, because of the match variations. So, before exploring what truncations, if any, could have a strategic effect on the match, it was first necessary to see whether truncations at the match point (i.e. deleting the $k + 1$ and higher choices for a participant who was matched to his kth choice) could influence the result of the match under either algorithm, and how much. The truncation of applicant ROLs and program ROLs were investigated separately, for each algorithm, for the 1993, 1994, and 1995 matches. In the majority of cases no change was produced when all ROLs were truncated at the match point; and in no case were more than 3 applicants affected by such truncations. (Over the more than 60,000 applicants involved in these experiments, only 4 were affected by truncations of applicants' ROLs; see Section B1 of Appendix B for the detailed results). So truncations at the match point, while not entirely without effect, do not play a substantial role; they are on the order of .01 percent of applicants, an order of magnitude smaller than the effects of changing algorithms.

Because we have now seen that information beyond the match point influences the outcome for only a tiny percentage of participants, concentrating on truncations will give us a comparably good approximation for the numbers of participants who could potentially profit from any kind of strategic manipulation of ROLs. The computational experiments which follow, therefore, will concentrate on identifying an upper bound on the number of participants who (if they had the necessary information) could potentially profit from strategic behavior involving truncations of their ROLs above the match point, and (for programs) reductions in the number of positions they offer in the match below the number of applicants to which they were matched.

2. Experiments to Determine Upper Bounds

As discussed above, the kinds of strategic manipulation to be considered involve truncation of ROL's by applicants or programs, and reductions in stated numbers of positions (quotas) by programs. Since we want to know how often a single agent can profitably manipulate the stated ROL, we could in principle conduct a separate experiment for each participant, but this would be computationally unfeasible. Consequently we need to design an efficient experiment that will let us tightly bound the number of individuals who can potentially profit from manipulating their ROLs.

The manipulations involving program quotas raise the question of how to handle reversions of positions when quotas are to be different from those in the data. Similar questions arise when truncating program ROLs, as this may increase unfilled positions. All of the experiments concerning strategic behavior of programs handle reversions by fixing quotas at the final quotas observed after the match with the original match data. None of the results are likely to be sensitive to this simplification, as shown by the results of the sequencing experiments discussed in Section III and detailed in Appendix A.

For each of the strategic manipulations whose potential magnitude is to be assessed, the chief difficulty in designing the experiments is that a change in a single rank order list or quota has two kinds of effects: it may potentially change the match of the applicant or residency program whose rank order list or quota is changed, but it may also potentially change the match of other applicants and residency programs. To see why, suppose that the rank order list of some applicant is truncated above his current match point, and the match under one of the algorithms is rerun after making (only) this change. Then the applicant whose rank order list was changed may do better (by being matched to a more preferred choice) or worse (by being unmatched instead of matched). At the same time other applicants may do better (and other residency programs may do worse) because of the availability of the position previously held by the applicant whose list was truncated.

This means that, if we truncate a group of applicant ROL's, for example, and see how many of the applicants in this group receive a better match as a consequence of this change, we will be looking at an overestimate of the number of applicants in the group who could have benefited from truncating their own ROL–many of them will have instead profited from someone else's truncation (even if that person himself became unmatched as a result of his own truncation). So the number obtained in this way would be an upper bound on the number that would have benefited by truncating their own ROL, while holding all others constant. But we do not have to settle for this upper estimate; we can refine it iteratively, by now continuing to truncate the ROL's only of those applicants whose match improved as a result of the previous (collective) truncations. This will allow us to further eliminate from the set of truncations those who profited from the truncations of applicants who were themselves harmed by their own truncation. Proceeding in this way, we can continue until no more reductions in the sample are achieved. This final number will still be an upper bound, of course, since even in a group of truncators who all do better when they all truncate their preferences, some may be profiting from the truncated ROL's of the others, not from their own truncation.

Experiments were conducted separately for applicants and for programs, and separately for each of the two algorithms. A computational experiment for applicants in a given year started by truncating all ROL's just above the (lowest) match point; i.e. every applicant's primary ROL was truncated just above the match he received when no ROL's were truncated using the algorithm in question. (For example, if an applicant originally matched to rank 3 on his primary ROL, the truncated ROL contained only his first 2 choices.) Of course, many applicants were left unmatched by this truncation, while others received preferred matches (these were the only two possibilities at this stage). Then at the next step, the ROL's of all those who had truncated their lists but did not improve were restored to their original length, and the process was repeated with the smaller number of truncations that remained. This process was repeated until it converged. Computational experiments for programs were structured similarly; starting with every program's ROL truncated just above the lowest ranked match it received. The full results for the NRMP for 1987 and 1993-96 are given in Appendix B. (Table B2 reports the results of the truncation of applicant ROL's for each algorithm, Table B3 the results for programs.)

The results can be summarized by looking at the final upper bounds of the number of applicants and the number of programs which could possibly benefit from truncating their Rank Order Lists.

The results are reported and analyzed below, first for the NRMP matches, and then for Thoracic surgery.

a. *Results for the* NRMP—The truncation experiments with applicants' ROLs yield the following upper bounds for the two algorithms in the years studied.

Table 4: Upper Limit of the Number of Applicants Who Could Benefit by Truncating Their Lists at One Above Their Original Match Point

	Upper Limit	
Year	Preexisting NRMP algorithm	Applicant-proposing algorithm
1987	12	0
1993	22	0
1994	13	2
1995	16	2
1996	11	9

As expected, more applicants can benefit from list truncation under the pre-existing NRMP algorithm than under the applicant proposing algorithm. Note that the number of applicants who could even potentially benefit from truncating their ROLs under the pre-existing NRMP algorithm is in each year almost exactly equal to the number of applicants who received a preferred match under the applicant proposing match (line 2 of Table 2). We will return to this point in a moment, but note that it suggests that this upper bound is very close to the precise number that would be predicted in the absence of match variations.

The truncation experiments with programs' ROLs yield the following upper bounds shown in Table 5.

As expected, some programs can benefit from list truncation under either algorithm. However, consistently more programs benefit from list truncation under the applicant-proposing algorithm than under the pre-existing NRMP al-

Table 5: Upper Limit of the Number of Programs That Could Benefit by Truncating Their Lists at One Above The Original Match Point

Year	Preexisting NRMP algorithm	Applicant-proposing algorithm
1987	15	27
1993	12	28
1994	15	27
1995	23	36
1996	14	18

gorithm. Note that although the numbers of programs in these upper bounds remain small, they are in many cases about twice as large as the number of programs that received a preferred match at the stable matching produced by the algorithm other than the one being manipulated. (That is, referring back to Table 2, we see for example that in 1995 only 14 programs preferred the matching produced by the pre-existing NRMP algorithm to the one produced by the applicant proposing algorithm, but we now find 36 programs in our upper bound of programs that could potentially profit from a manipulation of the applicant proposing algorithm.) It therefore seemed worthwhile to further examine these upper bounds, and see if they were overestimates.

For each algorithm, this was first done by taking a 50 percent sample of the programs contained in the upper bound for 1995, and restarting the truncation experiment with only these programs having truncated ROL's. The idea is that, if each of these programs can in fact benefit from its own truncation, the experiment would stop after the first iteration, with no further reductions in the upper bound. But if in fact the upper bound is an overestimate, and some of the programs in it are benefiting not from their own truncated ROL, but from the truncation of one of the other ROL's in the upper bound, then on average half of such "false positives" in our 50 percent sample would have been benefiting from the truncation by one of the programs in the other 50 percent, that are no longer truncated. In this case we would iterate until the number of truncators who improved their outcome again stabilized at a new, lower upper bound. This is in fact what happened, so the new estimates for 1995 (equal to twice the number obtained from the 50 percent sample) look as follows compared to the old ones.

These results confirm that the numbers that can benefit from the ROL truncations stated earlier are indeed overestimates.

A further analysis was undertaken for each of the five years, to compare the specific individual programs and applicants who appear in these upper bounds as potentially benefiting from ROL truncations with the programs and applicants whose results changed when the algorithm changed. This analysis indicated that those who could benefit from ROL truncations were, for the most part, those who did differently (generally worse) when the algorithm is changed from

Table 6: Refined Estimate of the Upper Limit of the Number of Programs That Could Improve Their Results by Truncating Their Own ROLs in 1995

Estimate	Preexisting NRMP algorithm	Applicant-proposing algorithm
Original results	23	36
Current estimate (still an upper limit)	12	22

their side proposing to the other side proposing (without ROL truncations). For example, the applicants who can benefit from ROL truncations when the program proposing algorithm is used are very largely the same as those who benefit when the algorithm is changed to an applicant proposing algorithm with no ROL truncations. Thus in this respect also, it appears that the theory for simple markets provides a good approximation of the situation in the NRMP match.

We next turn to the question of capacity manipulation by programs. Recall that in an actual match this could be considered by a program in the context of either an early agreement (for example with an independent applicant) or in anticipation that some positions would be filled post-match.

An initial experiment was run setting all program quotas to the number of positions filled with the algorithm in question with the original data. (This is analogous to the initial experiment involving truncations of the ROL's at the match point, rather than above it.) In a simple match without NRMP match variations, this would be expected to have no impact on the results. However, with NRMP match data some differences were observed, as noted below:

Table 7: Results with Input Quotas Set to Positions Filled, Compared to Original Results

	1993		1994		1995	
Result	Preexisting NRMP	Applicant-proposing	Preexisting NRMP	Applicant-proposing	Preexisting NRMP	Applicant-proposing
Programs						
Improve	12	2	9	none	25	none
Do Worse	none	none	3	2	9	2
Applicants						
Improve	none	none	3	2	6	2
Do Worse	12	2	9	none	27	none

With the applicant proposing algorithm, the differences are negligible. However, more differences were observed with the pre-existing NRMP algorithm, and the results obtained by setting the quotas to the original positions filled tended to produce better results for the programs.

In order to identify programs that could improve their remaining matches by further reducing their quotas, an iterative technique was employed similar to that used to investigate the effects of rank order list truncations. After several iterations revised downward the upper bounds obtained in this way, the resulting upper bounds on the number of programs that could potentially profit from stating lower quotas was as follows.

Table 8: Revised Estimate of the Upper Bound of The Number of Programs That Could Improve Their Remaining Matches by Reducing Quotas

Year	Preexisting NRMP algorithm	Applicant-proposing algorithm
1987	28	8
1993	16	24
1994	32	16
1995	8	16
1996	44	32

Again, these numbers are still estimates of the upper bound; further refinement would be possible. However, given the size of these numbers, it seems clear that only a very small number of programs (less than 1 percent) could improve their remaining matches by reducing their quotas. This does not appear to be an advisable strategy for programs to follow with either algorithm.

b. *Results for Thoracic Surgery*—Because the Thoracic surgery match does not have match variations, the theory tells us precisely which applicants and programs could improve their match by an optimal manipulation. As a check on our computational procedures, we confirmed these predictions by running the same computational experiments on ROL truncations as described for the NRMP matches. The results, summarized below, are as expected.

So, in Thoracic surgery as in the larger and more complex NRMP match, the opportunities for strategic manipulation are essentially non-existent under either algorithm. (Colenbrander, 1996, reaches essentially the same conclusions about the specialty matches he maintains.)

VI. Why the Differences are Small: Insights from the Theory of Simple Markets

All the results to this point can be characterized by noting that the theory of simple matches, without match variations, gives a good approximation for the direction of each of the comparisons, and, in addition, the size of all the changes has been very small. This section explores what insights we can get from simple markets to help explain why these differences are so small. The results in this section are based on computational comparisons similar to those discussed earlier, but now concerning hypothetical markets without any match variations.

The small differences between algorithms we have been seeing reflects that, in each of the years studied, the set of stable matchings has been small, as measured by the number of participants who receive different matches from the program-proposing and applicant-proposing algorithms.[11] It is therefore of interest to consider how the set of stable matchings looks for comparably large markets when we concentrate on simple matches. For this purpose, we consider the very simple matching markets with n firms (each with one position) and n applicants, as n approaches the size of the markets we are studying, namely the specialty markets like Thoracic surgery, and the general NRMP match.

One factor that strongly influences the size of the set of stable matchings (which coincides with the core in this simple model) is the correlation of preferences among programs and among applicants. When preferences are highly correlated–i.e. when similar programs tend to agree which are the most desirable applicants, and applicants tend to agree which are the most desirable programs–the set of stable matchings is small. (When preferences are perfectly correlated then there is a unique stable matching, so both algorithms would produce the same matching.) However as the correlation of preferences goes down, the size of the set of stable matchings grows, and more and more participants would be matched differently by the two algorithms. This is true independently of the size of the market.

It turns out, however, that the size of the market also plays a critical role, in an interesting way. Consider the case in which preferences are uncorrelated (so the set of stable matchings is large). If every applicant could somehow interview and be interviewed for all of the positions, then the set of stable matchings would grow larger and larger (even as a percentage of the number of applicants who could get different stable matchings) as the number of applicants and positions grew. Figure 1 shows that this percentage grows to over 90 percent by the time n reaches 1,000.

But of course in a real market there is a limit to how many interviews an applicant can go on, or a program can conduct. And when we take this into account, we see that the set of stable matchings quickly becomes very small as the market becomes large.

Specifically, let k equal the number of interviews a candidate can go on, and let n equal the number of applicants and positions in the market. Then even when preferences are completely uncorrelated, as k/n becomes small, the

[11] This is the natural measure for the size of the set of stable matchings in the present context, since the concern is with how many market participants will be affected by a change in algorithms. Note however that it is different from the more common measure of the size of the set of stable matchings, the number of distinct stable matchings. If 20 applicants receive different assignments at different stable matchings, there could be as many as 210 = 1024 different stable matchings, in case the 20 applicants can be resolved into 10 independent pairwise interchanges of positions, or there could be as few as 2 stable matchings, if all 20 applicants are involved in a single irreducible cycle. In either case, if there are 20,000 jobs being filled, we have been focusing on the approximately 20 applicants who receive different assignments when we conclude that the set of stable matchings is small.

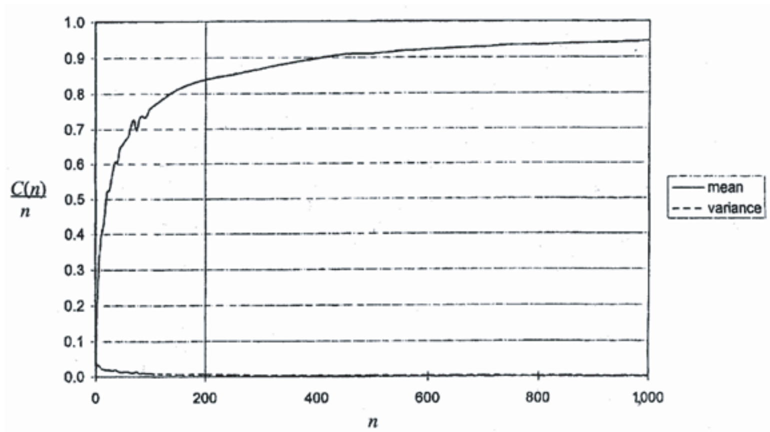

Figure 1: Size of the Set of Stable Matchings as a Fraction of n, When $k = n$ (Uncorrelated Preferences). *Note:* $C(n)$ is the number of applicants who get different stable matches when the market size is n.

set of stable matchings becomes small. So for example, if $k = 15$ (not an unreasonable approximation for the NRMP) and $n > 10,000$, fewer than 0.1 percent of applicants would receive a different match from the two algorithms.[12] That is, even with completely uncorrelated preferences, we see in this simple market the same one-in-a-thousand order of magnitude that we see in the NRMP. And for simple markets the size of the specialty matches like Thoracic surgery, with n on the order of 100 positions, if we suppose that applicants interview at no more than $k = 10$ programs we find only about 2 percent of applicants receiving different matches from the two algorithms. Figure 2 graphs the curves for fixed k , as n goes from 10 to 10,000.

Especially in view of the fact that preferences are not uncorrelated in the medical matches, this means that the orders of magnitude of the effects studied in the actual matches are very comparable to what we should expect of simple matches with similar k and n. Thus (once we look at both k and n) these simple markets turn out to provide a good approximation not only for the direction of the effects we are seeing, but also for their size.

The reason this is important for the present study of the NRMP and specialty matches is that, in the theoretical study of simple markets, we can look at what would happen when we know agents' true preferences, not just the ROLs which they submit to the match; whereas in the study of real matches we have been using as data the submitted ROLs. So one theoretical possibility might have been that the reason we find such small potential for strategic manipulation is that our data has been collected after such manipulation has already taken

[12]And the variance (based on 1,000 randomly generated simple markets) is well under .001 percent.

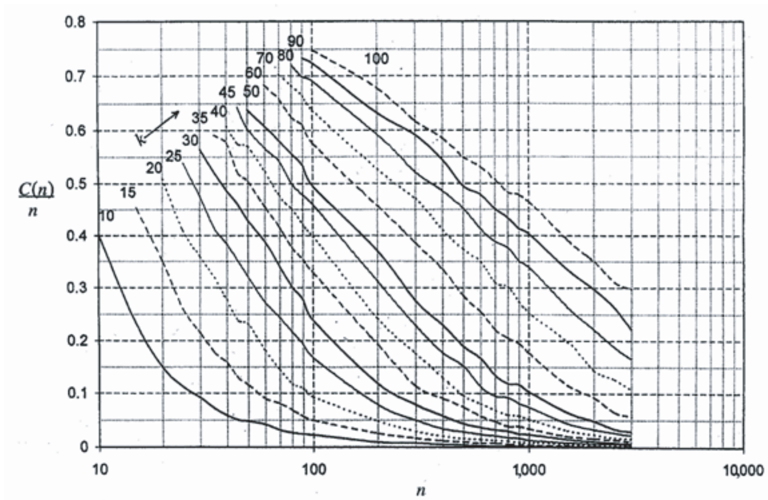

Figure 2: Size of the Set of Stable Matchings as a Fraction of n for Different Values of k (Uncorrelated Preferences). *Note:* $C(n)$ is the number of applicants who get different stable matches when the market size is n; k is number of programs on an applicant's ROL.

place. That is, one counter-hypothesis might explain our results by positing that, perhaps there are substantial opportunities for strategic manipulation, but these have been exhausted by the time we look at the ROLs submitted to the match, because the participants have already behaved strategically in an optimal way. Another counter-hypothesis could be that the hybrid nature of the pre-existing NRMP algorithm in fact produces matches that are far from the worst possible stable matching for applicants, and that the set of stable matchings is therefore substantially larger than we detect. The results discussed in this section show that these hypotheses are implausible, because when we looked at similarly sized artificial matches, in which we can examine the hypothetical participants' true preferences, we find that the set of stable matchings is close to the size we have computed from the ROL data. Thus the study of simple markets provides an explanation of not only the direction of the effects we have been examining, but also their small size.[13]

[13]It remains an open problem to develop analytical results which explain why the core of this simple market shrinks as the market grows when the number of interviews an applicant can go on remains constant. The fact that every worker who does get a different job at different stable matchings is involved in certain sorts of preference cycles may provide an avenue for obtaining such results.

VII. Theory and Computation in Economic Design: Some Methodological Reflections

Perhaps the first rule of any design effort is that "details matter." The details determine what outcomes are even feasible, and so they matter in the most basic aspects of design, and they have implications for all of the market's properties, so they matter for the subtlest aspects of the design's consequences. So every design effort will be different. But if we are to develop a body of knowledge about design practice in economics, we need to think about the methodological issues that may be common to many design efforts. This section is an attempt to put the methodological issues encountered in the NRMP design and evaluation into a context that may be useful for other design efforts. Specifically, this design effort involved the continual interplay among various aspects of simple theory, computational experiments, and theoretical computation. The simple theory guided the design of computational experiments on the complex system, which provided unpredicted results that were then explained by theoretical computation.

The reason there are gaps between theory and design is that, just as design is detailed, theoretical models must often be sparse, to be useful for organizing and directing work in a variety of applications whose connections may become apparent only with the benefit of theory. Much of this paper has therefore been concerned with filling the gaps between simple abstract markets and complex real ones. But before we discuss the filling of gaps, it is useful to recall the essential role played by the theory of simple matching markets. This role ranged from suggesting the basic design of the clearinghouse algorithm and the comparisons of the algorithms, to directing attention to aspects of the market in which problems might be anticipated, and to offering insights into how these might be overcome.

It was the existing simple theory, and the empirical studies it permitted to be conducted on field data, that pointed to the importance of stable matchings. And, although counterexamples showed that stable matchings might not exist in the complex American medical market (Roth, 1984), the theory of simple markets suggested a general architecture for an algorithm to find stable matchings. Furthermore, it showed that algorithms in which proposals were issued by applicants could be expected to produce stable matchings as favorable as possible to applicants. In short, the body of theory which existed prior to the start of this design, e.g. as summarized in Roth and Sotomayor (1990), already constituted a rough road map for the mechanism design and evaluation reported here.

At the same time, the existing body of theory, through counterexamples designed to explore its limits (inspired by empirical studies of existing markets), pointed to questions that needed to be answered. These included the role of sequencing in design of the algorithm, the frequency with which the algorithm might fail to find a stable matching, and the frequency with which opportunities for strategic manipulation might arise. These all required *estimations of mag-*

nitudes, which in turn required computational experiments on the data. Some of these computational experiments were straightforward to conduct. But for estimating how often strategic opportunities might arise, the theory played an essential role in the design of the computational experiments.

Specifically, although the main conclusions about strategic behavior do not carry over from the simple to the complex market, the theory of the simple case gives us not only final conclusions, but also insight into the way that strategic behavior works. In the case of misrepresentation of ROLs, the way an applicant might gain an advantage, in either the simple or complex markets, is to state an ROL that causes him, at some point in the algorithm, to make a rejection that would not have been made if he had submitted his true preferences.[14] This rejection causes a residency program to have a vacancy and hence make an offer to another applicant, who in turn may make a different rejection than he would have if the original applicant had stated his true preferences. It is the propagation of this "vacancy chain" through the market that raises the possibility that an applicant could do better than to state his true preferences. The fact that the potential advantage comes from what rejections are made implies in the simple model that the possibility of profitable strategic misrepresentations of ROLs can be investigated by looking at only the small subset of misrepresentations that consist of truncations. To see if this was approximately true for the complex market required a computational experiment, and (when this proved to be the case) it became computationally feasible to investigate the strategic properties of the complex market, through an experiment concentrating on truncations. So the theory allowed us to see what computational experiments would give us the answer to a question that the theory alone could not answer.

While computational experiments on the data allow us to get answers that may not be available from simple theory, they do not necessarily let us understand why the answers are what they are. In addition, results obtained from exploring a large and complex data set with a large and new piece of software (the new algorithm) need to be checked in some way, to make sure that the results are not due to some unanticipated artifact of the way the algorithm deals with the complexities of the data.[15] That is, although properly constructed computational experiments on the data offer us answers to questions we cannot answer with theory alone, we need both to check and to understand these answers before we can have the confidence in them that we would like to have before recommending that the new algorithm be considered for use in the market.

We addressed these issues in two ways, by computational experiments on

[14] The propagation of vacancy chains as such in simple markets is a topic that was explored in the course of this design effort, and is reported in Blum Roth and Rothblum (1997).

[15] Of course it was necessary to check directly that the program worked, and in fact it was easy to confirm that the matchings it produced were stable as well as feasible with regard to all the match variations. So the question we are referring to here is not whether the program does what it was designed to do, but rather whether the apparent small size of the set of stable matchings might have to do with some aspect of how it handles the match variations.

the data from the Thoracic Surgery matches, and by theoretical computation to determine how the size of the set of stable matches behaved in large simple markets. The first of these allowed us to exercise the software on a medical market free of the match variations present in the general medical market. The theory therefore permitted us to interpret the comparisons between the two algorithms as unambiguously measuring the size of the set of stable matchings. The small size of this set therefore had no possibility of resulting from some aspect of how the algorithm deals with match variations.

The computations on thousands of randomly generated simple markets with fixed length of ROLs and varying numbers of participants allowed us to see how the size of the set of stable matchings shrinks as the market grows, which establishes a new kind of core convergence result. This shows that the match variations in the medical market do not substantially contribute to the size of the set of stable matchings, since the results on the market data are entirely consistent with the results for similarly sized simple markets. Given the theoretical results on strategic misrepresentation, this core convergence result also shows that it is always a best reply for all but a tiny percentage of participants in large simple markets to state their true preferences.

Note that we distinguish between what we call the "computational experiments" on the actual NRMP data and the "theoretical computation" on the randomly generated simple markets. This has to do with our view that "theory" resides in the simplicity of the model and systematic nature of the conclusions, rather than the body of mathematical technique traditionally associated with theory. The theoretical computations tell us how the difference between the applicant and firm optimal stable matches varies with the size of a simple market. This new, computational result, combined with existing theory, allows us to interpret this as precisely measuring the size of the core of the market, and to determine the implications this has for the possibility of profitable strategic manipulation. The theorems explaining why the core *must* converge as it does will surely follow (see Feldin (1998) for some progress in this direction).

In summary, the design process discussed here involved interplay among various aspects of simple theory, computational experiments, and theoretical computation.[16] We suspect that, as we build a body of engineering practice in economics, this will prove to be a general pattern.

VIII. Concluding Remarks

The crisis of confidence that threatened to undermine participation in the NRMP was serious precisely because the kind of market failure which the NRMP was initially developed to correct arose when residency programs and applicants lost confidence in the existing market. But by the time of this modern crisis, the historical market failure and how it was corrected by the NRMP were understood.

[16]Laboratory experiments also have a role to play, although not one we will discuss here; but see Kagel and Roth (1999).

So was the fact that similar market failures in British medical markets had occurred and been corrected with stable matching mechanisms, while unstable mechanisms had failed (Roth 1990, 91). In addition, the general class of market failures due to unraveling of appointment dates had been identified in many markets (Roth and Xing 1994). So, although physicians who had participated in the unraveling of the American medical market and in the formation of the NRMP were no longer active, it was not difficult to communicate to the participants in the modern market why it was desirable to focus on changes in the market which would not re-ignite the unraveling of appointment dates (Roth 1996b).[17] Thus, although what we knew about two sided matching markets did not provide an immediate solution to the design of a new market for physicians, it provided clear guidelines and suggested clear approaches.

It was nevertheless troubling to us at the outset of this design effort that not only did none of the standard theorems about simple matching markets apply directly to the medical market, but counterexamples to the conclusions of many of them were known when the complications of the actual market are present. These counterexamples had the potential to be of great importance, as in the possibility (which does not arise in simple markets) that different stable matchings might yield different levels of employment. And, indeed, our results show that in this market this possibility is real, and so cannot be ruled out with better theory. But (see Table 2), of the more than 100,000 applicants in the years we studied in detail, only two applicants (one in 1987 and one in 1996) would have changed from employed to unemployed or vice versa at the different stable matchings we consider. Because this difference was both tiny and unsystematic, it did not play a role in the market design.

This, and the related results about the small number of applicants who receive different matches at the different stable matchings, point to a need to develop theory in ways which will tell us not only about the possibility of different effects, but also about their probability and likely magnitudes. It seems to us that questions about magnitudes of the sort we encountered in the course of this design will often arise in efforts to employ economic theory in the design of institutions for complex markets. Theoretical computation can be a big help in this effort, as it was in the present case in clarifying the unexpected consequences of the simple fact that applicants can interview at only a small fraction of the available positions.

---

[17]Indeed the initial study proposal (Roth 1995) quoted Hippocrates' famous dictum that when preparing to treat a disease

> "The physician must be able to tell the antecedents, know the present, and foretell the future- must mediate these things, and have two special objects in view with regard to disease, namely, to do good or to do no harm."

In this connection it is worth mentioning that, particularly because confidence in the market was the key issue, the study was conducted in an unusually public way, with progress reports posted regularly on the internet (see http://www.economics.harvard.edu/~aroth/nrmp.html) and widely distributed to interested organizations of physicians and medical students. A final report, briefly summarizing the overall results as in Tables 1 and 2 was presented to the medical community at large in Roth and Peranson (1997).

More generally, just as there is a chemical engineering literature (and not just literature about theoretical and laboratory chemistry), and a medical literature (and not just a biology literature), economists need to develop a scientific literature concerned with practical problems of design. An engineering-oriented design literature, and the theory that supports it, will be different from the basic science on which it depends, both in emphasis and in method. And if we do not develop such a literature, the practical problems of design will be relegated to the arena of "just consulting," and we will fail to benefit from the accumulation of knowledge which is so evident in other kinds of engineering.

Appendix A: Results of the Computational Experiments Concerned with Sequencing

Table A1 presents results from the sequencing experiments on the preexisting NRMP algorithm. Table A2 summarizes results of the experiments related to sequencing in the applicant-proposing algorithm. Table A3 compares the results when input quotas are set to final quotas and reversion processing is eliminated to the initial results.

Table A1 — Effects of Sequence in Which Programs Are Processed

A. *Results with Programs Processed in Descending Code Order Compared to Original Results with Preexisting* NRMP *Algorithm*

Result	1993	1994	1995
Programs			
Improve	none	2	2
Do Worse	2	2	none
Applicants			
Improve	2	2	none
Do Worse	none	2	2

B. *Sequencing of Reversions: Results with Input Quotas Set to Final Quotas and Reversion Processing Eliminated, Compared to Original Results with Preexisting* NRMP *Algorithm*

Result	1993	1994	1995[a]
Programs			
Improve	2	2	none
Do Worse	none	2	2
Applicants			
Improve	none	2	2
Do Worse	2	none	none

Notes: In 1994, when some programs and applicants did better while others did worse, there was no correlation between the change in result and the code numbers of the applicants and programs.

[a]Subsquently, the years 1987 and 1996 were also examined with similar results: no applicants were affected in 1987, two were affected in 1996.

Table A2 — Summary of Results of Experiments Related to Sequencing in the Applicant-Proposing Algorithm

Sequence of processing	1993	1994	1995
A. *Baseline Results (when Program Selected from Stack, Applicants Processed in Ascending Program Rank Number Sequence)*			
Applicants ascending; singles and couples intermixed			
Match result[a]	—	—	—
Loops detected	3	6	4
B. *Applicant and Couples Processing Sequence (when Program Selected from Stack, Applicants Processed in Ascending Program Rank Number Sequence)*			
Applicants ascending; couples last			
Match result	same	same	same
Loops detected	0	0	0
Applicants ascending; couples last			
Match result	same	same	same
Loops detected	2	0	0
Applicants descending; couples first			
Match results	2 applicants worse	2 applicants worse	same
Loops detected	3	6	1
Applicants ascending; couples first			
Match result	2 applicants worse	2 applicants worse	same
Loops detected	1	59	3
C. *Sequence of Processing Applicants Ranked by Program Selected from Program Stack (When Program Selected from Stack, Applicants Processed in Descending Program Rank Number Sequence)*			
Applicants ascending; singles and couples intermixed			
Match result	same	9 applicants improved, 3 applicants worse[b]	same
Loops detected	17	148	62
Applicants ascending; couples last			
Match result		9 applicants improved, 3 applicants worse[b]	same
Loops detected	2	2	0

[a] This is the base result to which others are compared.
[b] In part C, the results for the two experiments for 1994 (couples intermixed and couples last) were the same. In both cases, the differences in the results in part C as compared to the baseline results in part A were caused by chains resulting from two applicants doing worse in part C when compared with part A.

Table A3 — Results with Input Quotas Set to Final Quotas and Reversion Processing Eliminated, Compared to Initial Results with Applicant Proposing Algorithm

Result	1993	1994	1995[a]
Programs			
Improve	none	none	none
Do worse	none	2	2
Applicants			
Improve	none	2	2
Do worse	none	none	none

[a] Subsequently, 1987 and 1996 were also examined, with no applicants affected in 1987 and a single chain of nine affected in 1996.

Appendix B: Results of the Computational Experiments Concerned with Truncation of ROLs and Capacity Reductions

The results of truncation at the match point are reported in Table B1. Table B2 shows the results for iterative truncations of Applicant ROLs, while Table B3 shows the corresponding results for iterative truncations of program ROLs.

Table B1 — Truncations at the Match Point

1993	1994	1995
Difference in Result for Both the Preexisting NRMP Algorithm and the Applicant-Proposing Algorithm When Applicant ROLs are Truncated at the Match Point:		
none	2 applicants improve, same positions filled	2 applicants improve, same positions filled
Difference in Result for the Preexisting NRMP Algorithm When Program ROLs are Truncated at the Match Point:		
none	none	2 applicants do worse, same positions filled
Difference in Result for the Applicant-Processing Algorithm When Program ROLs are Truncated at the Match Point:		
none	3 applicants do worse, same number of positions filled, but not same positions (3 programs filled 1 less position; 1 program filled 1 more position; 1 program filled 2 more positions; 1 additional position was reverted from one program to another)	none

Table B2 — Results for Iterative Truncations of Applicant ROLs

	1987				1993				1994				1995				1996			
	Original NRMP algorithm		Applicant-proposing		Original NRMP algorithm		Applicant-proposing		Original NRMP algorithm		Applicant-proposing		Original NRMP algorighm		Applicant-proposing		Original NRMP algorithm		Applicant-propsing	
Run	T	T&I	T	T&1	T	T&1	T	T&1	T	T&1	T	T&1	T	T&1	T	T&1	T	T&1	T	T&1
1	16,117	4,324	16,116	4,317	17,209	4,546	17,209	4,536	17,725	4,935	17,725	4,934	18,170	5,763	18,170	5,758	18,316	5,805	18,316	5,806
2	4,324	1,894	4,317	1,887	4,546	2,093	4,536	2,082	4,935	2,361	4,934	2,359	5,763	2,907	5,758	2,899	5,805	2,915	5,806	2,917
3	1,894	898	1,887	891	2,093	1,036	2,082	1,032	2,361	1,185	2,359	1,183	2,907	1,572	2,899	1,559	2,915	1,569	2,917	1,571
4	898	437	891	429	1,036	514	1,023	498	1,185	602	1,183	598	1,572	857	1,559	844	1,569	861	1,571	864
5	437	203	429	194	514	258	498	241	602	292	598	287	857	473	844	460	861	481	864	482
6	203	93	194	84	258	135	241	116	292	151	287	143	473	251	460	238	481	271	482	271
7	93	41	84	31	135	73	116	52	151	75	143	66	251	136	238	124	271	157	271	155
8	41	24	31	13	73	48	52	25	75	40	66	31	136	79	124	67	157	89	155	87
9	24	18	13	6	48	34	25	12	40	27	31	17	79	45	67	31	89	57	87	55
10	18	14	6	2	34	27	12	5	27	18	17	7	45	31	31	17	57	36	55	33
11	14	12	2	0	27	24	5	2	18	14	7	3	31	22	17	8	36	24	33	21
12	12	12	-	-	24	22	2	0	14	13	3	2	22	18	8	4	24	19	21	15
13	–	–	–	–	22	22	–	–	13	13	2	2	18	16	4	2	19	15	15	13
14	–	–	–	–	–	–	–	–	–	–	–	–	16	16	2	2	15	14	13	12
15	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	14	13	12	11
16	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	13	12	11	10
17	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	12	11	10	9
18	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	11	11	9	9

Notes: Columns labeled “T” report the number of matches involving truncated ROLs. Columns labeled “T & 1” report the number of matches involving truncated ROLs and “improved” matches.

Table B3 — Results for Iterative Truncations of Program ROLs

	1987				1993				1994				1995				1996			
	Original NRMP algorithm		Applicant-proposing		Original NRMP algorithm		Applicant-proposing		Original NRMP algorithm		Applicant-proposing		Original NRMP algorithm		Applicant-proposing		Original NRMP algorithm		Applicant-proposing	
Run	T	T&I	T	T&1	T	T&1	T	T&1	T	T&1	T	T&1	T	T&1	T	T&1	T	T&1	T	T&1
1	2,967	1,345	2,967	1,349	3,342	1,457	3,342	1,462	3,369	1,514	3,369	1,517	3,444	1,538	3,444	1,541	3,410	1,445	3,410	1,444
2	1,345	670	1,349	675	1,457	740	1,462	748	1,514	809	1,517	813	1,538	783	1,541	790	1,445	727	1,444	725
3	670	347	675	353	740	382	748	394	809	441	813	444	783	420	790	431	727	384	725	384
4	347	186	353	194	382	201	394	216	441	249	444	255	420	237	431	248	384	213	384	212
5	186	100	194	110	201	107	216	122	229	138	255	145	237	130	248	141	213	114	212	115
6	100	55	110	66	107	64	122	79	138	79	145	86	130	77	141	89	114	71	115	72
7	55	33	66	44	64	37	79	52	79	44	86	52	77	50	89	62	71	50	72	52
8	33	21	44	33	37	22	52	37	44	31	52	39	50	35	62	47	50	35	52	39
9	21	17	33	29	22	15	37	30	31	23	39	32	35	29	47	41	35	26	39	30
10	17	15	29	27	15	13	30	28	23	20	32	30	29	26	41	38	26	21	30	25
11	15	15	27	27	13	12	28	28	20	18	30	29	26	24	38	36	21	19	25	23
12	–	–	–	–	12	12	–	–	18	16	29	28	24	23	36	36	19	18	23	22
13	–	–	–	–	–	–	–	–	16	15	28	27	23	23	–	–	18	17	22	21
14	–	–	–	–	–	–	–	–	15	15	27	27	–	–	–	–	17	16	21	20
15	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	16	15	20	19
16	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	15	14	19	18
17	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	14	14	18	18

Notes: Columns labeled "T" report the number of matches involving truncated ROLs. Columns labeled "T & 1" report the number of matches involving truncated ROLs and "improved" matches.

Appendix C: Formal Definitions of Stability

Simple matching markets

For markets without linkages between positions we use the "college admissions" model as reformulated in Roth (1985) and Roth and Sotomayor (1990, Chapter 5). There are two finite and disjoint sets, $\mathcal{F} = \{f_1, ..., f_n\}$ and $\mathcal{W} = \{w_1, ..., w_m\}$, of firms and workers. For each firm f in $\mathcal{F}$, there is a positive integer q_f, which indicates the number of (identical) positions f has to offer.

An outcome is a matching of workers to firms, such that each worker is matched to at most one firm, and each firm is matched to at most its quota of workers. It will be convenient to denote a firm that has some number of unfilled positions as matched to itself in each of those positions, and similarly an unmatched worker will be matched to herself. To give a formal definition, define for any set $\mathcal{X}$ *an unordered family of elements* of $\mathcal{X}$ to be a collection of elements, not necessarily distinct, in which the order is immaterial.

A *matching* μ is a function from the set $\mathcal{F} \cup \mathcal{W}$ into the set of unordered families of elements of $\mathcal{F} \cup \mathcal{W}$ such that:

(i) $|\mu(w)| = 1$ for every worker w and $\mu(w) = w$ if $\mu(w) \notin \mathcal{F}$;

(ii) $|\mu(f)| = q_f$ for every firm f, and if the number of workers in $\mu(f)$, say r, is less than q_f, then $\mu(f)$ contains $q_f - r$ copies of f;

(iii) $\mu(w) = f$ if and only if w is in $\mu(f)$.

Each worker has preferences over the firms (and the possibility of remaining unmatched in the market), and each firm has preferences over the workers (and the possibility of leaving a position unfilled). All preferences are transitive, and strict (recall that in the markets we consider participants are obliged to submit rank orders which are necessarily strict). We will write $f_i >_w f_j$ to indicate that worker w prefers f_i to f_j. Similarly, $w_i >_f w_j$ represents firm f's preferences $P(f)$ over individual workers. Firm f is acceptable to worker w if $f >_w w$, and worker w is acceptable to firm f if $w >_f f$, *i.e.,* an acceptable firm is one which is preferable to being unmatched, and an acceptable worker is one which the firm prefers to leaving a position unfilled.

Each worker's preferences over alternative matchings correspond exactly to her preferences over her own assignments at the two matchings. Things are not quite so simple for firms, because even though we have described firms' preferences over workers, each firm with a quota greater than 1 must be able to compare groups of workers in order to compare alternative matchings. It will be sufficient for our purposes to assume merely that a firm's preferences over groups of employees it could be matched with (*i.e.,* over groups of not more than q_f workers) are such that, for any two assignments that differ in only one worker, it prefers the assignment containing the more preferred worker (and is indifferent between them if it is indifferent between the workers). Any preferences of this sort are called *responsive* to the firm's preferences over individual workers (Roth, 1985).

A matching μ is individually irrational if $\mu(w) = f$ for some worker w and firm f such that either the worker is unacceptable to the firm or the firm is unacceptable to the worker. Such a matching will also be said to be *blocked* by the unhappy agent. A firm f and worker w will be said together to block a matching μ if they are not matched to one another at μ, but would both prefer to be matched to one another than to (one of) their present assignments. That is, μ is *blocked by the firm-worker pair* (f, w) if $\mu(w) \neq f$ and if $f >_w \mu(w)$ and $w >_f \sigma$ for some σ in $\mu(f)$. (Note that σ may equal either some worker w' in $\mu(f)$, or, if one of firm f's positions is unfilled at $\mu(f)$, σ may equal f). Matchings blocked in this way by an individual or by a pair of agents are unstable in the sense that there are agents with both the incentive (because preferences are responsive) and the power (under rules which allow any firm and worker to conclude an agreement with each other) to disrupt such matchings. So we can now define a matching μ to be *stable* if it is not blocked by any individual or any firm-worker pair.[18]

Complex matches

In the medical markets served by the NRMP, the employers are residency programs and the workers are physicians applying to those programs. The simple model of the previous section does not allow for the variety of matching requirements observed in the medical market, for which purpose we will have to distinguish between different kinds of applicants, and different kinds of residency programs.

Let the set of applicants be $\mathcal{A} = \mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{C}$, where $\mathcal{A}_1$ is the set of (single) applicants who seek no more than one position, $\mathcal{A}_2$ is the set of applicants who may want two jobs, and who submit supplemental lists of first year jobs in connection with any second year position on their ROL which requires a complementary first year position (and does not come with one automatically), and $\mathcal{C}$ is the set of couples, who submit a single ROL listing pairs of positions. A member of $\mathcal{C}$ is a couple $\{a_i, a_j\}$ such that a_i is in the set $\mathcal{A}_3$ (of husbands) and a_j is in the set $\mathcal{A}_4$, and the sets $\mathcal{A}_1$, $\mathcal{A}_2$, $\mathcal{A}_3$, and $\mathcal{A}_4$ are sets of applicants, who together make up the entire population of individual applicants, which will be denoted $\mathcal{A}' = \mathcal{A}_1 \cup \mathcal{A}_2 \cup \mathcal{A}_3 \cup \mathcal{A}_4$. (The $\mathcal{A}_i$ may not be disjoint, since members of a couple may also submit supplemental lists.) The reason for denoting the set of applicants both as $\mathcal{A}$ and as $\mathcal{A}'$ is that from the point of view of a potential employer, the members of a couple $\mathcal{C} = \{a_i, a_j\}$ are two distinct applicants who seek distinct positions (typically in different residency programs), while from the point of view of the couple they are one agent with preferences over pairs of positions.

[18] This definition of stability appears to account only for coalitions of size one or two, but in fact accounts for coalitions of any size (i.e., stable matchings are in the core; see Roth and Sotomayor, [1990]).

The set of residency programs is $\mathcal{R} = \{r_1, ..., r_n\}$ and associated with each program r is a positive integer q_r indicating how many positions it seeks to fill. However, for some programs r, q_r may not be a constant at every point in the matching process. There are two reasons why q_r may change. A residency program r may have an agreement with another residency program r' (typically within the same hospital) that if r can only fill $k < q_r$ positions, the remaining $q_r - k$ positions will be added to the capacity of r'. In such a situation, the algorithm will change q_r to k and q'_r to $q'_r + (q_r - k)$. (It can also happen that the $q_i - k$ unfilled positions revert to more than one other residency program, and so the total number of positions need not remain constant, and different positions from a given program may revert to different programs.) The other reason why quotas may vary is that some residency programs wish to have an even number of residents, so a residency program r with quota q may have its quota reduced to $q' = q - 2$ in the event that it can only be matched to a maximum of $q - 1$ residents. (These quota adjustments take place after an initial attempt to make a stable match, and cause the matching algorithm to continue from the current match; in what follows, discussion of stability will refer to the current quota of a program r at any point in the algorithm, except as indicated.)

Applicants in the set $\mathcal{A}_1$ submit ROL's over residency programs, and hence have preferences just like the workers in the simple model discussed earlier. Applicants in the set $\mathcal{A}_2$ have on their ROL's at least one second year program which requires (but does not supply) first year training as well, and these applicants submit a supplemental ROL for each such position, indicating their preferences for first year positions, conditional on being matched to a given second year position. Each couple $c = \{a_i, a_j\}$ in the set $\mathcal{C}$ submits, as a single ROL, a ranked list of ordered pairs of positions, i.e. an ordered list of elements of $\mathcal{R} \times \mathcal{R}$ whose first element is some (r_i, r_j) which is the couples' first choice pair of positions for a_i and a_j respectively, and so forth. Each residency program submits as their ROL an ordered list of members of $\mathcal{A}'$, i.e. of individual applicants (whether or not they are members of a couple).

Having thus defined the form in which different kinds of agents state their preferences, we can now define stable matchings. A matching μ with range $\mathcal{R} \cup \mathcal{A}'$ is defined as in the simple market, except that for an applicant a in the set $\mathcal{A}_2$ it may be that $|\mu(a)| = 1$ or 2 if $\mu(a)$ matches a to a program for which it has submitted a supplemental ROL. In case $|\mu(a)| = 2$ we will write $\mu(a) = (r_1, r_2)$, where r_1 is the (second year) residency program on a's primary ROL, and r_2 is the (first year) residency program on a's supplemental ROL when a is matched with r_1. (When $|\mu(a)| = 1$ it must be that $r = \mu(a)$ is on a's primary ROL.)

As in the case of the simple market considered earlier, we will say a matching is *stable* if it is not blocked by any individual agent or by a pair of agents consisting of an individual and a residency program, or by a couple together with one or two residency programs.

A matching μ is blocked by an individual applicant (in the set $\mathcal{A}_1$ or $\mathcal{A}_2$), or by a residency program, if μ matches that agent to some individual or residency program not on its ROL, precisely as in the simple model. A matching is blocked by an individual couple $\{a_i, a_j\}$ if they are matched to a pair (r_i, r_j) not on their ROL. Of course no individual or couple blocks a matching at which he or it is unmatched.

A residency program r and an applicant a in the set $\mathcal{A}_1$ together block a matching μ precisely as in the simple market, if they are not matched to one another and would both prefer to be. A residency program r and an applicant a in the set $\mathcal{A}_2$ together block a matching μ if r prefers a to one of its matches under μ (i.e. $a_j >_r \sigma$ for some σ in $\mu(r)$), and if either $r >_a r_1 \in \mu(a)$ where the preferences $>_a$ correspond to a's primary ROL, or $r >_a r_2 \in \mu(a)$ and $>_a$ corresponds to a's supplemental ROL for the position $r_1 \in \mu(a)$.

A couple $c = \{a_1, a_2\}$ and residency programs r and r' block a matching μ if $(r, r') >_c \mu(c)$ and if either

(i) $a_1 \notin \mu(r)$, $a_1 >_r \sigma$ for some $\sigma \in \mu(r)$ and either $a_2 \in \mu(r')$ or $a_2 >_{r'} \sigma'$ for some $\sigma \in \mu(r')$ or

(ii) $a_2 \notin \mu(r')$, $a_2 >_r \sigma'$ for some $\sigma' \in \mu(r')$ and either $a_1 \in \mu(r')$ or $a_1 >_{r'} \sigma$ for some $\sigma \in \mu(r)$.

References

[1] **Aldershof, Brian and Carducci, Olivia M.**, "Stable Matchings with Couples." *Discrete Applied Mathematics*, August 1996, *68*(1-2), pp. 203-07.

[2] **AMA-MSS.** Resolutions for the 1995 Interim Meeting, http://www.bcm.tmc.edu/ama-mss/i95res.htm#11.

[3] **American Medical Students Association and Public Citizen Health Research Group**. "Report on Hospital Bias in the NRMP," http://pubweb.acns.nwu.edu/~alan/nrmp2.html, 1995 .

[4] **Ausubel, Lawrence M.; Cramton, Peter; McAfee, R. Preston; and McMillan, John**. "Synergies in Wireless Telephony: Evidence from the Broadband PCS Auctions." *Journal of Economics & Management Strategy*, Special Issue: Market Design and the Spectrum Auctions, Fall, 1997, 6(3), pp. 497-527

[5] **Blum, Yossi; Roth, Alvin E.; and Rothblum, Uriel G**. "Vacancy Chains and Equilibration in Senior-Level Labor Markets," *Journal of Economic Theory*, October 1997, 76(2), pp. 362-411.

[6] **Blum, Yosef and Rothblum, Uriel G**. "'Timing is Everything' and Marital Bliss." *Journal of Economic Theory*, forthcoming in 1999.

[7] **Colenbrander, August**. "Match Algorithms Revisited." *Academic Medicine*, May 1996, 71(5), pp. 414-415.

[8] **Cramton, Peter**. "The FCC Spectrum Auctions: An Early Assessment." J*ournal of Economics & Management Strategy*, Special Issue: Market Design and the Spectrum Auctions. Fall 1997, 6(3), pp. 431-495.

[9] **Demange, Gabrielle; Gale, David; and Sotomayor, Marilda**. "A Further Note on the Stable Matching Problem." *Discrete Applied Mathematics*, 1987, 16, pp. 217-222.

[10] **Feldin, Aljosa**. "Core Convergence in Two-Sided Matching Markets: Some Theoretical Considerations." Mimeo, 1998, Department of Economics, University of Pittsburgh.

[11] Gale, David and Shapley, Lloyd**. "College Admissions and the Stability of Marriage." American Mathematical M**onthly, January 1962, 69(1), pp. 9-15.

[12] **Kagel, John H. and Roth, Alvin E**. "The Dynamics of Reorganization in Matching Markets: A Laboratory Experiment Motivated by a Natural Experiment," mimeo, (revised) 1999.

[13] **Ledyard, John O., Porter, David; and Rangel, Antonio**. "Experiments Testing Multiobject Allocation Mechanisms." *Journal of Economics & Management Strategy*, Special Issue: Market Design and the Spectrum Auctions. Fall, 1997, 6(3), pp. 639-675.

[14] **McAffee, R. Preston and McMillian, John**. "Analyzing the Airwaves Auction," *Journal of Economic Perspectives*, Winter, 1996, 10(1), pp. 159-175.

[15] **McMillian,John**. "Selling Spectrum Rights," *Journal of Economic Perspectives*, Summer 1994, 8(3), pp. 145-162.

[16] **McMillan, John**. "Why Auction the Spectrum?" *Telecommunications Policy*, April 1995, 19, pp. 191-199.

[17] **Milgrom, Paul**. "Auction Theory in Practice: The Simultaneous Ascending Auction," mimeo, 1997, Stanford University.

[18] **Peranson, E. and Randlett, R.R.** "The NRMP Matching Algorithm Revisited: Theory Versus {ractice," *Academic Medicine*, June 1995, 70(6), pp. 477-484.

[19] **Peranson, E. and Randlett, R.R**. "'Comments on Williams' 'A Reexamination of the NRMP Matching Algorithm,' " *Academic Medicine*, June 1995, 70(6), pp. 490-494.

[20] **Plott, Charles R.** "Laboratory Experimental Testbeds: Application to the PCS Auction." Journal of Economics & Management Strategy, Special Issue: Market Design and the Spectrum Auctions. Fall, 1997, 6(3), pp. 605-638

[21] **Roth, Alvin E.** "The Economics of Matching: Stability and Incentives." Mathematics of Operations Research. 1982, 7, pp. 617-628.

[22] **Roth, Alvin E**. "he Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory." Journal of Political Economy. December 1984, 92(6), pp. 991-1016.

[23] **Roth, Alvin E.** "The College Admissions Problem is not Equivalent to the Marriage Problem." Journal of Economic Theory. 1985, 36, pp. 277-288.

[24] **Roth, Alvin E.** "On the Allocation of Residents to Rural Hospitals: A General Property of Two-Sided Matching Markets." Econometrica. 1986, 54, pp. 425-427.

[25] **Roth, Alvin E.** "New Physicians: A Natural Experiment in Market Organization," Science, 14 December 1990, 250, pp. 1524-1528.

[26] **Roth, Alvin E.** "A Natural Experiment in the Organization of Entry Level Labor Markets: Regional Markets for New Physicians and Surgeons in the U.K." American Economic Review. June 1991, 81(3), pp. 415-440.

[27] **Roth, Alvin E**. "Proposed research program: Evaluation of changes to be considered in the NRMP algorithm." Consultant's report, 1995, http://www.pitt.edu/˜alroth/nrmp.html.

[28] **Roth, Alvin E.** "Interim Report #1: Evaluation of the current NRMP algorithm, and preliminary design of an applicant-proposing algorithm." Consultant's report, March 1996, http://www.pitt.edu/˜alroth/nrmp.html.

[29] **Roth, Alvin E**. "The NRMP as a Labor Market." Journal of the American Medical Association, April 3, 1996b, 275(13), pp. 1054-1056.

[30] **Roth, Alvin E. and Peranson, Elliott.** "The effects of the change in the NRMP matching algorithm." Journal of the American Medical Association. September 3, 1997, pp. 729-732.

[31] Roth, A.E. and Rothblum, U.G., **"Truncation Strategies in Matching Markets: In Search of Practical Advice for P**articipants." Econometrica, January, 1999, 67(1), pp. 21-43.

[32] **Roth, Alvin E. and Sotomayor, Marilda.** "The College Admissions Problem Revisited." Econometrica. May 1989, 57(3), pp. 559-570.

[33] **Roth, Alvin E. and Sotomayor, Marilda**. "Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis." Econometric Society Monograph Series, Cambridge University Press, 1990.

[34] **Roth, Alvin E. and Vande Vate, John H**. "Random Paths to Stability in Two-Sided Matching." Econometrica. November 1990, 58(6), pp. 1475-1480.

[35] **Roth, A.E. and Vande Vate, J.H.** "Incentives in Two-Sided Matching with Random Stable Mechanisms." Economic Theory. 1991, 1(1), pp. 31-44.

[36] **Roth, Alvin E. and Xing, Xiaolin**. "Jumping the Gun: Imperfections and Institutions Related to the Timing of Market Transactions." American Economic Review. September 1994, 84(4), pp. 992-1044.

[37] **Salant, David J.** "Up in the Air: GTE's Experience in the MTA Auction for Personal Communication Services Licenses." Journal of Economics & Management Strategy, Special Issue: Market Design and the Spectrum Auctions. Fall 1997, 6(3), pp. 549-572.

[38] **Shiller, Robert J.** "Macro Markets: Creating Institutions for Managing Society's Largest Economic Risks." Clarendon Lectures in Economics. Oxford: Oxford University Press, 1993.

[39] **Sonmez, Tayfun**. "Manipulation via Capacities in Two-Sided Matching Markets." discussion paper, Department of Economics, University of Michigan, 1996.

[40] **Sonmez, Tayfun.** "Can Pre-Arranged Matches be Avoided in Two-Sided Matching Markets?" discussion paper, Department of Economics, University of Michigan, 1996.

[41] **United States District Court for the Western District of Missouri, Western Division (1996).** "United States of America v. Association of Family Practice Residency Directors, Final Judgment." May 28, 1996.

[42] **Williams, K.J.** "A Reexamination of the NRMP Matching Algorithm." Academic Medicine. June 1995, 70(6), pp. 470-476.

[43] **Williams, K.J.** "Comments on Peranson and Randlett's 'The NRMP Matching Algorithm Revisited: Theory versus Practice'," Academic Medicine. June 1995, 70(6), pp. 485-489.

[44] **Wilson, Robert B**. *Nonlinear Pricing*. Oxford: Oxford University Press, 1993.