The dynamics of law clerk matching: An experimental and computational investigation of proposals for reform of the market

Ernan Haruvy[a],[*], Alvin E. Roth[b], M. Utku Ünver[c]

[a]School of Management, University of Texas at Dallas, USA
[b]Harvard Business School and Department of Economics, Harvard University, USA
[c]Department of Economics, University of Pittsburgh, USA

Abstract

In September of 1998, the Judicial Conference of the United States abandoned as unsuccessful the attempt—the sixth since 1978—to regulate the dates at which law students are hired as clerks by Federal appellate judges. The market promptly resumed the unraveling of appointment dates that had been temporarily slowed by these efforts. In the academic year 1999–2000 many judges hired clerks in the fall of the second year of law school, almost two years before employment would begin, and before hardly any information about candidates other than first year grades was available. Hiring dates moved still earlier in the Fall of 2000 and 2001. The present paper explores proposed reforms of the market, experimentally in the laboratory, and computationally using genetic algorithms. Our results suggest that some of the special features of the judge/law-clerk market—in particular the feeling among many students and judges that students must accept offers when they are made–present obstacles to the success of the proposed reforms, including the latest reforms proposed by the judges, in 2002 and 2003. Unlike markets in which the inability to make binding contracts contributes to

[*]Corresponding author.
E-mail addresses: eharuvy@utdallas.edu (E. Haruvy), al_roth@harvard.edu (A.E. Roth), uunver@pitt.edu (M.U. Ünver).

market failure, in the law clerk market it is the ease with which binding contracts are forged that harms efficiency.

1. Introduction

Top law students compete fiercely for judicial clerkships, particularly at the appellate level, and federal judges vie with one another for prospective clerks. In the process, the time at which law students are hired, for jobs that they will begin only after completion of their third year of law school, has moved earlier in their law school career.

In September of 1998, the Judicial Conference of the United States abandoned as unsuccessful its attempt to prevent hiring of clerks before March 1 of their second year of law school.[1] Surveys of law students and judges conducted by Avery et al. (2001) reveal that by the 1999–2000 academic year hiring had moved much earlier, so that 63% of responding judges said that they had completed their clerkship hiring (for jobs beginning in 2002) by the end of January, 2000, in contrast to only 17% who had completed their hiring by January the previous year. That is, the timing of hires moved markedly earlier even in the 2 years immediately following the Judicial Conference's decision to stop trying to regulate the market.

The earlier clerks are hired in their law school career, the less information judges have available to help them distinguish one student from another. For example, when the hiring takes place in the fall semester of a student's second year of law school, only first year grades are available.[2] There is a strong sense among many market participants, captured in the Avery et al. surveys, that very early matching is inefficient. Nevertheless, the unraveling of appointment dates in this market is a problem of long standing, despite frequent attempts to modify the market in ways that would promote later appointments. From 1978 through 1998 there were six such attempts (see Roth and Xing, 1994, and Becker et al., 1994, for a discussion of the market's history through the early 1990s).

The unraveling of appointment dates is not a peculiarity of this legal labor market. Many markets, particularly entry level labor markets, have experienced similar problems. Roth and Xing (1994) discuss several dozen markets and submarkets that have experienced unraveling of appointment dates.[3] Several theoretical studies have shown how this unraveling can lead to ex post (and sometimes ex ante) inefficiency,

---

[1]More recent attempts at reform, in 2003 and 2004, will be discussed in the conclusion.

[2]Later hiring would not only make more grades available, but also Law Review editorial positions and articles, other written work, Moot Court competition results, and other information that, when it was available, used to play a role in law clerk selection. Later hiring would also give students more information about which judges might be interested in them.

[3]And see Avery et al. (2003) on the growth of early decision in college admissions.

because early matches, made before important information becomes available, can be mismatches.[4] Early matching can also reduce the scope of the market, causing it to break into many local markets.[5]

Some markets have addressed this problem of early matching by reorganizing themselves as centralized clearinghouses. Perhaps the best known is the market for new physicians in the United States, which established a centralized clearinghouse in the early 1950s, that was shown in Roth (1984) to produce matchings that are stable in the sense of Gale and Shapley (1962). Various attempted clearinghouse designs in England, Scotland, and Wales (Roth, 1991), together with experiments in the laboratory (Kagel and Roth, 2000; Ünver, 2005) helped confirm that this kind of stability is important to the success of such clearinghouses.[6] These clearinghouses have been subject to changes over the years, to keep them functioning in markets with increasingly complex demands (such as the desire of two-career couples to find work in the same city). Roth and Peranson (1999) report the latest redesign of the U.S. medical clearinghouse, and the Roth-Peranson design has since been adopted by other labor markets.[7]

Consequently, one of the reforms debated in the law literature has been the adoption of a centralized clearinghouse on the medical model. This has attracted both support and opposition (see the references in Roth and Xing, 1994; Avery et al., 2001). Avery et al. (2001) argue that some of the special features of the law clerk market, to be discussed below, may present serious obstacles to the successful implementation of a centralized clearinghouse in this market.

Another, more modest reform, already implemented in the 2001–2002 academic year, has been the creation of a web site on which federal judges may announce their hiring plans, specifically including the date on which they plan to start accepting applications.[8] These announcements, one might conjecture, would foster better cooperation among the judges.

In this paper, we report parallel experimental and computational studies to gain insight into the implications of centralized matching and announcements in the law clerk matching market. Experimentation in the laboratory allows us to get a look at how people respond when faced with the incentives induced by various forms of market organization that may not yet exist in practice. Computation, using genetic algorithms to model adaptive behavior, allows us to first reproduce the behavior observed in the laboratory, and then to explore how that behavior might have

---

[4]See the models in Roth and Xing (1994), Li and Rosen (1998), Sönmez (1999), Li and Suen (2000, 2001), Suen (2000), and Ostrovsky and Schwartz (2002).

[5]See Niederle and Roth (2003, 2004) and McKinney et al. (2005) for investigations of the consequences and causes of unraveling in the market for gastroenterologists.

[6]A number of other experiments have examined different aspects of matching. For example, see Chen and Sönmez (2002), and Pingle and Tesfatsion (2001).

[7]Other markets that have adopted it to date are, in the United States, Postdoctoral Dental Residencies, Osteopathic Internships, Osteopathic Orthopedic Surgery Residencies, Pharmacy Practice Residencies, and Clinical Psychology Internships, and, in Canada, Articling Positions with Law Firms in Ontario, Articling Positions with Law Firms in Alberta, and Medical Residencies. Centralized clearinghouses can also serve to facilitate market clearing in environments that have not suffered from unraveling: see e.g. Abdulkadiroğlu and Sönmez (2003), and Roth et al. (2004).

[8]https://lawclerks.ao.uscourts.gov.

further evolved had we been able to run longer experiments than feasible in the laboratory.[9] Earlier experimental results, particularly Kagel and Roth (2000) which studied the organization of medical markets in different regions of the British National Health Service, give us some reason to be confident that there is an important relation between behavior observed on a small scale in the laboratory and on a large scale in career-shaping labor markets. Similarly, the work of Ünver (2000, 2001) gives us reason to believe that computational simulations have some predictive power in these kinds of markets. In the conclusion we will further discuss the use of these tools, and their limitations, for forming conjectures, as in the present case, about market institutions that have not yet been tried in practice.

2. Background

2.1. Description of the law clerk market

Some important features of the law clerk market are described here. For a more detailed description see Avery et al. (2001), who found the following pattern of behavior in the contemporary market:

- interviews lead very quickly to offers;
- offers produce very quick responses;
- responses are generally acceptances (with there being a strong sense among many students and judges that offers must be accepted); and, in consequence,
- many students limit the judges to whom they apply, to avoid receiving an early offer from a less preferred judge.

The last two points—that many students feel obliged to quickly accept the first offer they receive, and that students (and judges) respond strategically to this when scheduling their interviews—are features that are much more pronounced in this market than in others we know of.[10] Avery et al. report that 73% of the students in their year 2000

---

[9]Time scale is among the hardest features of experiments, and of computations, to map onto field studies, both in duration (e.g. how long does an experiment last compared to the natural market that it models) and frequency (e.g. how frequently can subjects make decisions compared to participants in the natural market). So being able to get an idea of robustness across very different time scales is an important indicator of the potential generalizability of these results. That is, the computations will give us some assurance that our experimental results are not artifacts of slow learning in the laboratory, while the experiments will assure us that the behavior produced by the genetic algorithms is in fact similar to human behavior.

[10]The closest parallel in the literature seems to be the market for clinical psychologists (Roth and Xing, 1997) which, prior to being reorganized as a central clearinghouse, operated as a decentralized market in which applicants were often asked in advance if they would accept an offer if one were forthcoming. This also happens in the law clerk market, and in both markets the answers can apparently be relied on, in contrast to the much larger medical market to be discussed next. A similar phenomenon occurs in the contemporary market for admissions to elite American colleges, in which increasingly large parts of the market are accomplished through binding "early decision" programs (Avery et al., 2003).

sample accepted the first offer they received, including a majority of students whose first offer was not their first choice from among those for which they had interviews. This had to do largely with the perceived need to respond quickly, which precluded the possibility of waiting to see if a more preferred offer would arrive. Avery et al. report that 42% of the students in their sample responded to their first offer immediately, and 92% had responded within one week. In turn, some judges, fearing that they would not be the first choice of the most desirable students, insisted on conducting early interviews, and some students declined to accept such interviews rather than run the risk of accepting the interview, receiving the early offer, and then feeling compelled to take the position.

Avery et al. conjecture that the ability of judges to elicit binding promises from applicants would also hamper the adoption of a centralized match on the medical model. Interestingly, although the medical matches attempt to prevent employers from asking for early commitments, or requesting information on how applicants will rank them in the match, medical matches are plagued with instances of requests for commitment and informal rankings. For example, Pearson and Innes (1999) reported that 15% of 1996 and 1997 graduates of the University of Virginia School of Medicine were asked for signals concerning what rank order list they intended to submit to the centralized match. Other surveys confirm this finding, and also that many medical students when confronted with such questions, answer deceptively (Anderson et al., 1999; Carek et al., 2000; Teichman et al., 2000). That is, in the large, national, first year medical market, in which students may not have future encounters with a residency director whose position they are not matched to, it appears to be accepted that if you are asked an unethical question, you may give a deceptive answer. In contrast, in the law clerk market, Avery et al. received virtually no reports of deceptive answers by law students to questions posed by Federal appellate judges. In fact, not only are deceptive answers rare or nonexistent, even negative answers seemed inappropriate to many students when asked to accept a position immediately rather than wait for (even) an already scheduled interview.

2.2. Prior experiments

Kagel and Roth (2000) and Ünver (2005) showed that a centralized clearinghouse that produces a stable matching is more effective at halting inefficiently early matching than are various unstable mechanisms. Those experiments compared clearinghouse designs found in different regions of the British National Health Service, and observed that the laboratory results obtained in a stylized simple setting corresponded to the outcomes observed in the British labor markets that employed those mechanisms. In those experiments, the inefficiency of early matching (which in medical markets involved loss of planning flexibility as well as information costs) was modeled in the experimental environment with a fixed cost of $1 for each period that a match took place before a final period.

In the law clerk market, however, the major potential source of inefficiency seems to arise from the lack of information in early periods about applicants' qualities.[11] In

---

[11]Judges lose little planning flexibility by hiring clerks early, since the number of positions they have is the same from year to year.

the experiments described below, we model the information about students as developing over time, as their grades become available.

3. Experimental design[12]

The information environment in the law clerk experiments is as follows. Judges have qualities 1 to 4, that are known from the beginning of the market. In each of three periods (or "years"[13]), grades of students are independently drawn from {0,1,2} with a discrete uniform density. The cumulative grade of a student in each year is the summation of the grades of the student in that year and previous years. Ties between students are broken arbitrarily only after year 3 grades become available (but before the application process starts in year 3). After all ties are broken, the applicant with the highest cumulative grade in year 3 has quality 4, the applicant with the second highest cumulative grade has quality 3, the applicant with the second to last grade has quality 2, and the applicant with the lowest grade has quality 1. The payoff to each subject is the product of his own quality and that of the subject to whom he is matched.[14]

In each condition, subjects experienced 20 markets, with the roles (judge or applicant) remaining the same throughout, judges' qualities remaining fixed, and applicants' grades determined randomly in each year of each market. So, offers made in year 1 can be contingent only on year 1 grades, while offers made in year 3 are made after all grades, and hence students' final qualities are known.

We model the institutional features of the law clerk market by having applicants decide, at the beginning of each year, to which judges, if any, to submit applications. Applicants can apply to as many available judges as they wish, and no judge may make an offer to an applicant who has not applied to him. However, when a judge makes an offer, the applicant must accept, unless a better judge made an offer simultaneously.[15]

---

[12]Subjects were recruited through the Computer Lab for Experimental Research at Harvard, through web based sign ups, newspaper ads and posters. The subject pool is fairly diverse, including many students from area universities, but also many area residents unaffiliated with a university. Experiments were programmed using the Z-Tree software of Fischbacher (1999). The complete instructions are available in the Appendix.

[13]Although we will often speak about the three periods as "years," they can be interpreted as any time period in which information becomes available. As remarked earlier, it is of limited usefulness to try to identify experimental periods with particular spans of time in related natural markets. And, indeed, we will see in Section 9 that the most recent plan for reforming the market envisions a three-period hiring process in which each period lasts a week. The computational results will support the view that the experimental results are robust to the length of the experiment and the frequency of the interaction.

[14]This environment is chosen to make the social efficiency issue clear, so that a socially efficient match is strictly assortative. That is, the socially efficient outcome matches the most productive judge to the most productive applicant, the second most productive judge and to the second most productive applicant, and so forth (cf. Becker, 1981). Later we will calculate welfare losses relative to this socially efficient outcome.

[15]We thus model students' "obligation to quickly accept an offer" more forgivingly than in the actual market, since we treat offers made in the same "year" as simultaneous, and allow the applicant to accept the best of however many offers are received in that period. However, an applicant who receives one or more offers cannot reject them all, e.g. an applicant who receives a single early offer must accept it.

To investigate the potential effectiveness of a centralized clearinghouse, we ran treatments with and without centralized matching in the final year. To investigate the effect of having judges announce when they will start filling their positions, we ran each condition with and without "announcements," which required that judges explicitly indicate, prior to year 1, the year at which they will begin accepting applications. These announcements (like the federal judges' web page) are seen by both students and other judges.

Finally, we explore the effect on a centralized match of the fact that students feel compelled to accept an offer if one is made. The fact that students must accept an offer from a judge if one is made makes them vulnerable to being hired when they are interviewed, even if they prefer to wait until the match. We ran markets with and without this kind of "coercion". In the "no coercion" treatments, we model applicants' ability to participate in the match if they choose by, in the experiment, allowing them to participate in the central match even if they have not previously applied to any judges. That is, in the no-coercion ("idealized") treatments of the experiment, an applicant may choose not to apply to any judge in years 1 and 2, but may nevertheless participate in the centralized match in year 3. The interpretation is that applicants apply to judges just prior to the match, and that no student who does so is coerced to accept a position before the match. However, as in the other treatments, students who apply *early* to judges are obliged to accept an early offer if one is made.

The table below summarizes the six different law clerk market treatments, and reports the number of cohorts (of 8 subjects—4 judges and 4 applicants) that were observed in each treatment. Each cohort participated in 20 markets. A total of 454 subjects participated in these six treatments, forming 57 different cohorts.[16]

	Decentralized	Centralized-idealized	Centralized-coerced
Without 8 cohorts	announcements	8 cohorts	10 cohorts
	5/21/2001 (2 cohorts)	5/7/2001 (3 cohorts)	5/26/2001 (4 cohorts)
	6/7/2001 (4 cohorts)	6/20/2001 (3 cohorts)	7/3/2001 (4 cohorts)
	7/18/2001 (2 cohorts)	7/2/2001 (4 cohorts)	
With announcement	10 cohorts	11 cohorts	10 cohorts
	5/21/2001 (3 cohorts)	5/30/2001 (4 cohorts)	7/10/2001 (7 cohorts)
	6/6/2001 (2 cohorts)	6/28/2001 (4 cohorts)	8/1/2001 (1 cohorts)
	7/18/2001 (2 cohorts)	7/31/2001 (1 cohorts)	8/2/2001 (2 cohorts)
	7/25/2001 (3 cohorts)	8/2/2001 (2 cohorts)	

In the three "announcement" treatments, prior to the start of each market, judges announce the year (1, 2, or 3) in which they become available to receive applications.

---

[16]The number of 454, rather than 456, reflects the fact that there were two subjects out of 454 who by mistake were able to participate in two different treatments.

In all six treatments, in each year (1–2 in the centralized treatments and 1–3 in the decentralized):

1. Applicants send applications to available judges.
2. Judges may choose any one applicant from the pool of applicants who have applied in a given year, and they are matched to this applicant unless a higher quality judge also chooses that applicant. If that applicant is unavailable, the judge is given the option to choose another applicant from the pool and so forth.

In the centralized-idealized treatment, year 3: Judges and applicants not matched by the end of year 2 are all matched by the computer in year 3, at the unique stable matching among those remaining in the market. That is, the unmatched judge with the highest quality is matched with the highest quality unemployed applicant. The second highest quality unmatched judge is then matched with the highest quality applicant from the remaining unemployed applicant pool, and so on. (Recall that even in the centralized-idealized treatment, when an applicant applies to a judge and receives an offer in years 1 and 2, she cannot decline this offer).

In the centralized-coerced treatment, year 3: Judges and applicants who were not matched by the end of the second year were placed via the central match in year 3. But to be eligible for matching to a particular judge following year 2, an applicant needs to have sent an application to that judge in either year 1 or year 2. The unmatched judge with the highest quality was matched with the highest quality unemployed applicant who had sent this judge an application in either year 1 or year 2 or both. The second highest quality unmatched judge was then matched with the highest quality applicant from the remaining unemployed applicant pool, who had sent this judge an application in either year 1 or year 2 or both, and so on.

Notice that in the centralized-coerced treatment, judges can still choose to participate in the centralized match, by not making offers to applicants in years 1 or 2. But students cannot choose to participate in the centralized match, because they need to apply to a judge in year 1 or 2 to become eligible to be matched, and may at that time be "coerced" into an early match.[17]

The treatments without announcement skip the announcement stage and all judges accept applications every year. For the treatments without centralized matching, year 3 is identical to year 2, except that qualities are known (ties are resolved), and following that year, the market ends.

In a single session, 24–32 subjects were divided into three to four groups of eight, such that no person knew who the other seven members of his group were. Each

---

[17]Note that our convention of referring to periods as "years" may be potentially misleading. Think of the periods as informationally distinct intervals (that is, new information must be revealed each period) prior to the match. Typically for law students, new grades are revealed at the end of each semester. But other information, such as election to Law Reviews and student government, prizes, results of competitions, etc. may be revealed sporadically. Since judges can choose when to conduct their interviews, judges who wish to make offers before some information is available have many opportunities to do so. Of course, judges who wished to do so could also simply conduct their interviews for the match long before it was scheduled to take place.

group of eight consisted of four judges and four applicants. The groups remained fixed throughout the experiment. Each subject saw on the screen his or her role (judge or applicant) and was notified that this role would remain fixed for the duration of the experiment, which would last 20 "markets."

In this experiment, judges and applicants may remain unmatched at the end of a market. In the context of the law clerk market, the interpretation is, for an applicant, that he does not become a clerk, but goes into the general labor market for new lawyers, and for a judge, that he hires a clerk who does not turn out to be one of the top ranked students.

4. Equilibrium analysis

The games considered in the experimental design have multiple Nash equilibria, which support both early and later matching.

Theorem. *In all of the experimental conditions, there are Nash equilibria at which all matches are made in period* 1. *In all but the centralized-coerced condition, there are also Nash equilibria at which all matches are made in period* 3.

(In the centralized-coerced condition, we conjecture that no such very late equilibrium exists, but there are equilibria at which all matches are made in period 2.) We prove the theorem by constructing some examples of the indicated kinds of equilibrium, starting with the games without announcements.

Early equilibria: In the decentralized game consider the following strategy profile: in the first year all applicants apply to all judges. Judges make an offer to the applicant with the same relative rank as the judge: if two applicants are tied, the higher ranked judge involved in the tie hires the subject with higher assigned subject number. In the second and third years no applications are made by unmatched applicants and unmatched judges do not make any offers. This strategy profile will result in a matching in which judge 4 is matched with the best applicant of year 1 in year 1, judge 3 will be matched to the second best applicant of year 1 in year 1, and so on. This strategy is in equilibrium since by making unilateral deviations from the strategy no agent can improve her payoff. The same strategy will be in equilibrium in the centralized-idealized game and in the centralized-coerced game.

Late equilibria: There are also third-year-hiring equilibria for the decentralized and centralized-idealized games. Consider the following strategy profile in the decentralized game. No applications are made in year 1 and year 2 by the applicants, and no judges make offers in year 1 and year 2 even if judges receive applications. In year 3, all applicants apply to all judges. Since ties are resolved before the application procedure, judge *i* tries to hire the applicant of the same quality in year 3. This strategy will result in the unique stable (assortative) matching in year 3. This is in equilibrium, since by a unilateral deviation, i.e. by applying early, no applicant will be hired early since judges do not hire applicants in years 1 and 2, and a judge will not be able to hire anybody early by deviating from its strategy since applicants do not submit early applications. Also in year 3, applicants cannot benefit by applying

to fewer than all available judges, and judges cannot benefit by making an offer to an applicant other than the one with the same rank.

Notice that this argument does not apply to the centralized-coerced condition, in which applicants who wish to participate in the centralized match must earlier apply to judges, at which time they can be hired. We conjecture that at any equilibrium in the centralized-coerced condition, at least one judge will hire by period 2. The intuition is that it is a strictly best response for all applicants to submit applications by period 2: the alternative is to remain unemployed. Moreover, if any two judges find themselves in direct competition over an applicant who applied to both judges in year 2, the lower ranked of the two can extend an offer to that applicant in year 2.

However, in the centralized-coerced game there is an all-second-period-hiring equilibrium. Consider the following strategy: Applicants do not apply to any judge in period 1: judges hire no applicant in period 1, even if there are applications. In period 2, each applicant applies to all judges, and judge i extends an offer to the ith ranked applicant of period 2. If there is a tie among applicants' scores, the offer is determined by the subject number. This is an equilibrium, since a unilateral deviation by an early application in period 1 on part of an applicant or early hiring by a judge will not improve payoff, since judges are not making offers in period 1 and applicants are not making any applications in period 1. Similarly, a unilateral deviation from period 2 to period 3 cannot improve the payoff for either judge or an applicant, since they would simply get the same match in period 3 that they forego in period 2. Also, applicants have no incentive to apply to less than all judges in period 2, because applying to fewer judges does not have any effect in increasing their chances to be matched to better judges.

In the counterparts of these games in which judges announce their entry periods, the strategies above constitute equilibria whenever judges announce their availability starting in year 1. There are many other equilibria. We conjecture that this multiplicity of equilibria would survive refinement, but have not proved this. We have here considered Nash equilibria which allow weakly dominated strategies, for simplicity, and because equilibrium refinements have shown little predictive ability in experiments. This multiplicity of Nash equilibria is one of the motivations for the experiment and corresponding computer simulations reported here.

5. Learning in the experiment: artificial adaptive agent simulations

Subjects in the laboratory, like judges and applicants in the market for law clerks after a regime change such as took place in 1998, are not expected to begin in equilibrium. Rather, as in most economic settings, economic agents will have to learn how to behave. Subjects' strategies—their mapping from available information (on the state of the world and on others' actions) to their own actions—will depend on their past experiences and "reinforcements" of past strategies. Simple reinforcement learning models have been useful for explaining and predicting behavior in a variety of simple games (see e.g., Roth and Erev, 1995; Erev and Roth, 1998; Feltovich,

2000). However, simple learning models are most useful when agents learn among a pre-specified and small set of strategies. In the relatively complex environment considered here, the sets of strategies available to the agents are quite large. Genetic algorithms incorporate the ability to generate and evaluate new strategies, and are thus a compact way of exploring learning in a large strategy space, while making increasing use of strategies that have worked well in the past. A genetic algorithm is motivated by biological evolution through variable selection from a heterogeneous population that maintains diversity through sexual reproduction and mutation (Holland, 1975).[18,19] If we can find similarities between the behavioral dynamics of human subjects and artificial agents, then we can use the computational techniques to test the robustness of behavior over longer time horizons.[20]

We apply computational artificial adaptive agent simulations to model the adaptive behavior of subjects in the experiments. Using these simulations, we can investigate short run and long run learning behavior. As will be shown later, short run results of the simulations are comparable to the behavior of human subjects in the experiment. Using the long run simulations, we will assess the impact of various market designs and demonstrate that the results of the experiment appear to be robust to time scale. Long run simulation results are given in the Appendix in the section on sensitivity analysis .

A strategy for an agent in a genetic algorithm is represented by a string of integers. We model only some of the information sets. For instance, in the decentralized game with announcements, a judge strategy is encoded by 6 integers.

$$T - A^1 - R^1 - A^2 - R^2 - R^3.$$

---

[18]Originally, Holland used genetic algorithms to solve complex optimization problems. Generally, this is the context researchers have used them in other disciplines. Goldberg (1989) and Michalewicz (1994) explain this approach in detail by giving an underlying theory for sufficient conditions in obtaining the maximum in optimization. However, economists and social scientists treat them as a possible model of the strategy updating or learning technique of interacting economic agents. See Holland and Miller (1991) and Dawid (1999) in this respect. Dawid (1999) has a theoretical attempt to find when a genetic algorithmic learning converges to an evolutionary steady state in multiple population games. However, our model is more complex than Dawid's.

[19]Genetic algorithms (and other adaptive learning techniques) have been widely used to capture the behavior of economic agents. This modeling approach is known as agent-based computational economics (ACE). Leigh Tesfatsion summarizes ACE on the web-pages http://www.econ.iastate.edu/tesfatsi/ace.htm and http://www.econ.iastate.edu/tesfatsi/caswork.htm as ''the computational study of economies modeled as evolving systems of autonomous interacting agents. ACE is thus a specialization to economics of the basic complex adaptive systems paradigm…''

[20]Pingle and Tesfatsion (2001) use adaptive artificial agent simulations and human subject experiments to study the impact of changing the level of a non-employment payoff on the evolution of cooperation between workers and employers participating in a sequential employment game with incomplete contracts. Ünver (2001a) uses adaptive artificial agents with genetic algorithms to examine an entry-level labor market game inspired by the hospital-medical intern markets in Britain. Ünver (2000, 2001b) compares the findings of this prior study with results of human subject experiments. Duffy (2001) conducts another benchmark study, this time to compare human and artificial subject behavior in a macroeconomic environment using individual learning methods. Also see Arifovic (1996) and Bullard and Duffy (1998) for the use of genetic algorithms to model macroeconomic adaptive learning.

T is an integer in $\{1,2,3\}$. This decision variable is the year when the judge is going to start accepting applications from applicants. This is automatically set to 1 in the treatments without announcements. A^t is in $\{0,1\}$. When $A^t = 1$, the judge may hire an applicant in year t. When $A^t = 0$, the judge will not hire an applicant in year t. ($A^3 = 1$ is automatically set at the beginning of the simulations, so it is not a decision variable.) R^t is in $\{1,2,3,4\}$. This decision variable is the threshold relative rank of the available applicants that the judge is going to hire, representing the lowest ranked candidate to whom he is prepared to make an offer in each period. The full strategy representations for applicants and judges are given in the Appendix.

In the genetic algorithm simulations, we initially generate a strategy pool for each type of agent. Then we let the artificial agents randomly choose strategies from their corresponding pool and play the matching market game. The "reinforcement" or "fitness" of a strategy is measured by its average payoff in a tournament of 1000 games (50 games per judge strategy on average). Strategies with higher reinforcements have higher probability of being selected for inclusion in the next pool of strategies and of being selected as "parents" to produce new strategies by combining their decision variables. (In this respect, the integer string representing a strategy plays a role like the strategy's "DNA" in biological evolution).

The tournament is repeated in the next market with the updated pools of strategies, and so on. We ran 20 sessions of simulations, each of which ran for 500 generations.[21] Note that there are 5 different types of agents in our experimental design: judges 1 through 4, and applicants. Such an artificial adaptive agent simulation is called "co-evolutionary" due to the fact that we observe distinct evolution of strategies of different types not only depending on their past behavior but also depending on the state and history of the behavior of others.

The genetic algorithm is formally defined by the details of how these strategies with different reinforcements are updated to produce the next generation strategies. These details, including the representation of strategies for each kind of agent, are given in the Appendix. But the basic component of a genetic algorithm is made up of three operators, which transform one population of strategies into the next generation:

1. Selection pressure ("survival of the fittest"): we directly copy the highest fitness strategies for inclusion in the next generation of agents.
2. Crossover ("sexual reproduction"): Four parent candidates are chosen from among the strategies in the current generation. Then we pair them, and select the fittest one from each pair. These two selected strategies are called parent strategies. We randomly determine a crossover point. Then we copy the decision variables of the first parent strategy before this crossover point, and the decision variables of the second parent strategy after the crossover point to form an offspring strategy.

---

[21]The longer simulations had 5000 generations. Their results are reported in the Appendix.

Crossover Example :

$$P1 : a1 - a2 - a3 - a4 \qquad\qquad\qquad\qquad O1 : a1 - a2 - b3 - b4$$

$$\rightarrow \text{Crossover starting at 3rd digit} \rightarrow$$

$$P2 : b1 - b2 - b3 - b4 \qquad\qquad\qquad\qquad O2 : b1 - b2 - a3 - a4$$

3. Mutation: With a small probability each decision variable in each offspring strategy is mutated to form a new random decision variable.[22,23]

Mutation Example :

$$a1 - a2 - a3 - a4 \qquad \rightarrow \text{Mutate 2nd digit} \rightarrow \qquad a1 - x2 - a3 - a4$$

To summarize, "selection" of tournament winners, and using fitness to select "parent" strategies, allows the genetic algorithm to increase the probability that successful strategies will be played, as in reinforcement learning. Mutation and crossover create new strategies, and allow the algorithm to search the big strategy space defined by all possible realizations of the integer string.

A detailed sensitivity analysis is provided in the Appendix for the choice of genetic algorithm parameters using active non-linear tests (cf. Miller, 1998). We show that the majority of the results are robust to choices of real parameters in a range within $\pm 50\%$ deviation from the original parameters and different choices of remaining parameters.

6. Results

6.1. The effect of centralized matching: coerced versus idealized treatments

Fig. 1 shows that very few matches are in fact consummated by the centralized-coerced match. Interestingly, as Fig. 1 shows, the number of applicants entering the central match is increasing over time in the centralized-idealized treatments but

---

[22]Although mutation is important, in genetic algorithms as soon as the fittest strategies start to emerge, this operator loses its importance. We exogenously decrease the mutation rate over time, but the mutation rate never reaches zero. Mutation thus plays a role something like experimentation in reinforcement learning (see e.g. Roth and Erev, 1995), but also helps to generate new strategies, both of which stop the evolutionary process from getting locked in.

[23]The following are the genetic algorithm parameters used: We have 20 strategies for each agent type in the current pool. The 2 most successful strategies in each current pool are directly copied to the pool for the next market. We use a linear-crossover operator to create the remaining pool of the strategies of the next market based on success of strategies in the tournament. The crossover probability is $p = 0.8$ for crossing over the two parent strings starting from a randomly chosen decision variable. With probability $1 - p = 0.2$, these parents are directly copied to the next generation pool. The mutation probability for each decision variable is generally exogenously decreased over time from—about—$p_{\max}^{\mathrm{m}} = 0.07$ to $p_{\min}^{\mathrm{m}} = 0.01$. The parameter values are well within the range of parameter choices reported in the GA literature. For the initial pools of strategies we force applicants to make applications in every period, when they can. Otherwise the strategies are random. This resembles the initial behavior of the human applicants in the first market of the experiment. Note that we also tried random initial conditions for our simulations, which did not change the results we obtained as outlined in the Appendix on sensitivity analysis.
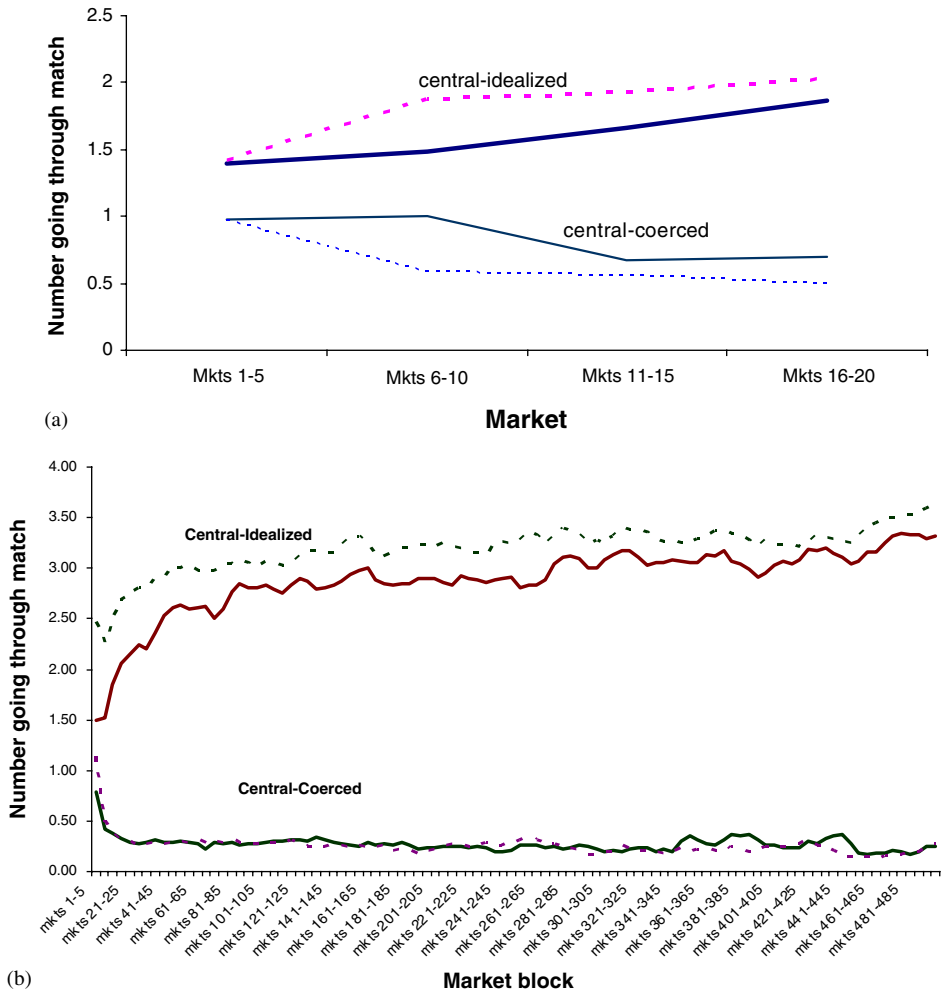
Fig. 1. Number of applicants matched through the centralized matching. Dashed lines represent announcement conditions. (a) Experimental findings. (b) Computational findings.

decreasing in the centralized-coerced treatment. This corresponds to our conjecture that coercion is partially responsible for the unraveling we observe in the legal market.

Looking at the proportion of period 3 matches in each set of treatments, we can reject (p-value < 0.001) the hypothesis that the proportion matched in the third period of the last five markets is the same for the centralized-coerced treatment's proportion (13%) and the other two treatments (46.90%). In the GA simulations, the corresponding proportions (using the last 50 markets) are 9.27% in the centralized-coerced match and 83.07% in the other treatments.[24] Again, this

----

[24]The differences in the genetic algorithm are almost always significant.

Table 1
Average period of hiring by judge in the legal market[a]

	Experimental—average of 20 mkts			Computational—average of 500 mkts		
	Decentral. NA/A	Central-idealized NA/A	Central-coerced NA/A	Decentral. NA/A	Central-idealized NA/A	Central-coerced NA/A
Judge 1	2.30/2.31	2.68/2.58	2.22/2.05	2.70/2.81	2.88/2.91	1.80/1.77
Judge 2	1.90/1.73	2.03/2.12	1.81/1.77	2.35/2.49	2.61/2.67	1.61/1.61
Judge 3	1.75/2.06	1.84/2.06	1.89/1.80	2.02/2.18	2.28/2.41	1.39/1.40
Judge 4	2.04/2.17	1.82/2.16	1.76/1.78	2.25/2.67	2.66/2.87	1.26/1.32

[a]In each box, we report the averages for the conditions with and without announcements. We report the without-announcement average before "/" and the corresponding announcement average after "/". The same notation is also used in Table 3.

confirms the conjecture regarding the detrimental effect of coercion. Table 1 shows average hiring periods and confirms that in general judges hire earlier in the centralized-coerced treatment than in the centralized-idealized treatment.

The effect of this observation can be seen in the difference in welfare between the centralized-coerced treatment and other treatments. Fig. 2 shows the welfare effects of the three institutional regimes—decentralized, centralized-coerced, and centralized-idealized—with and without announcements by judges of their earliest hiring dates.

Comparing the decentralized to the centralized-idealized conditions of the experiment, we find that the centralized-idealized market improves welfare by 0.55 on average over the last five markets. This welfare improvement is significant (t-statistic = −1.96, d.f. = 37, p-value = 0.029). In the genetic algorithm simulation, the increase in welfare is even more prominent with an improvement by of 0.87 in the last 50 markets. This is in sharp contrast to the strong welfare decreasing effect of the centralized-coerced variation, as Fig. 2 illustrates. Thus, in an environment in which applicants are subject to coercive offers when they interview, the results suggest that a centralized match will *not* have the welfare-enhancing effects that would result from the introduction of a centralized match in a non-coercive environment.[25]

The effect of centralized-idealized treatment in increasing welfare can be broken down into two components: welfare increase due to the decrease in *un*matched agents and welfare increase due to later hiring and reduction in number of *mis*matched agents. It is impossible to separate these two effects entirely, but we can make an approximation. As the below analysis suggests, the welfare increase in the centralized-idealized market is mostly attributable to the reduction in number of mismatches below the levels in the other treatments.

In the experiment and simulations, the welfare loss relative to the idealized setting due to unmatched and mismatched pairs in the decentralized and centralized coerced

_____

[25]Once again, the main result is that, when students can be coerced to take jobs before the match, few will succeed in remaining available to participate in the match. The magnitude of the welfare differences observed here, however, depends on details of our experimental implementation.
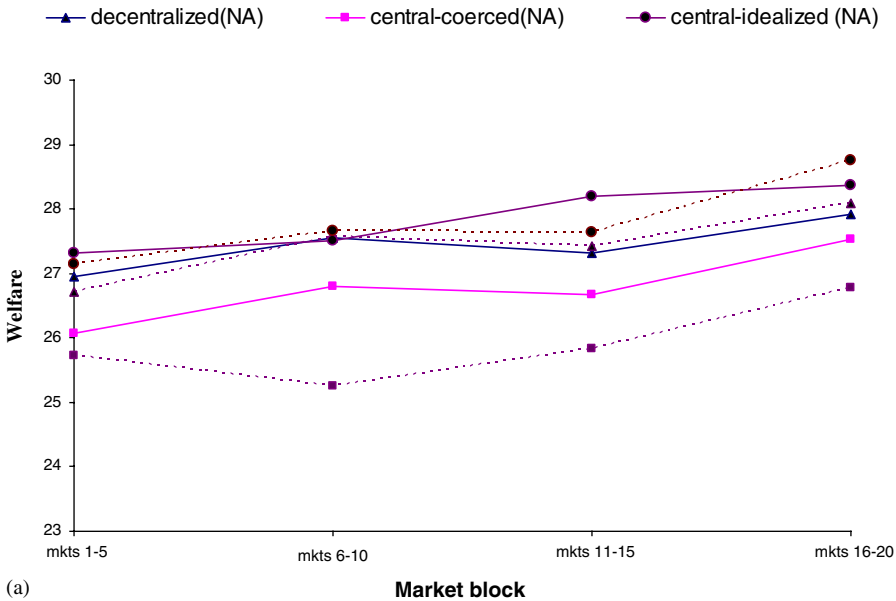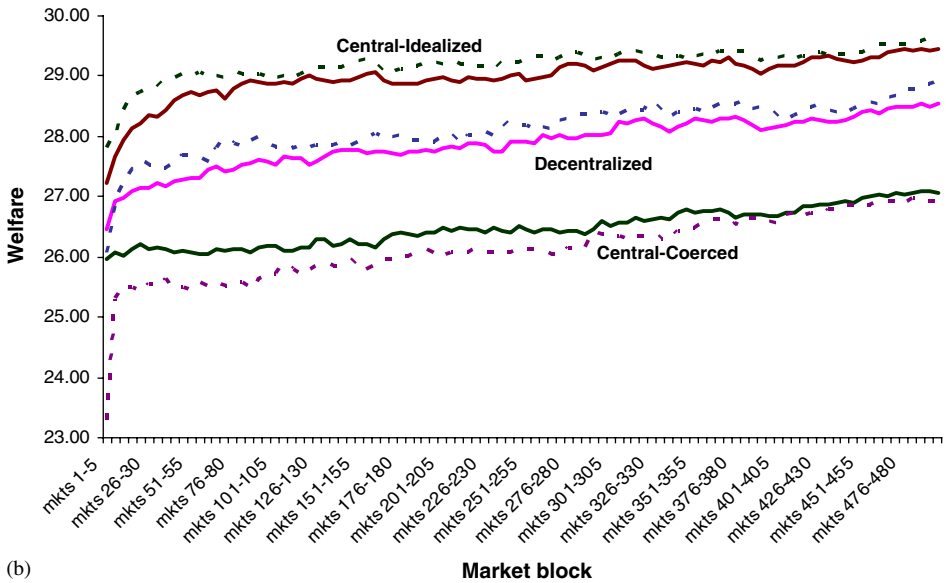
Fig. 2. Average group welfare in the legal market. For the six treatments in the legal market, dashed lines represent treatments with announcements. (a) Experimental findings. (b) Computational findings.

treatments (pooled over announcement conditions) can be summarized by Table 2. Statistics for the experiment correspond to the last block of 5 markets. For the simulations, the corresponding numbers are averages for the last 50 markets.

Table 2

Breakdown of welfare loss relative to the central-idealized treatment in the legal market

	Experimental—average of the last 5 markets		Computational—average of the last 50 mkts	
	Decentral.	Central-coerced	Decentral.	Central-coerced
Avg. number of unmatched pairs	0.14	0.22	0.24	0.33
Avg. quality of unmatched judge	1.15	1.20	1.16	1.13
Avg. quality of unmatched applicant	1.44	1.83	1.23	1.63
Approx. welfare loss due to unmatched	0.24	0.49	0.34	0.61
Approx. welfare loss due to mismatches	0.31	0.91	0.53	1.91

Fig. 3 shows the evolution of quality and quantity of unmatched applicants, whose averages in the final blocks are reported in Table 2. The average qualities of hired applicants are given in Table 3. This table shows that not all judges would prefer the centralized-idealized treatment to the centralized-coerced treatment despite the overall welfare gains under the centralized-idealized treatment. The lower quality judges (Judge 1 and Judge 2) would prefer the centralized-coerced treatment.

For the decentralized treatment, when the averages are taken in the last block (5 markets in the experiment; 50 in the simulation), there are on average 0.14 (0.24 in the simulation) unmatched pairs per market. Average quality of unmatched judges is 1.15 (1.16 in the simulation) and average quality of unmatched applicants is 1.44 (1.23). Therefore the approximate average welfare loss attributable to the unmatched pairs is 0.24 (0.34) in the decentralized markets. The difference in welfare between the centralized-idealized and decentralized markets is 0.55 (0.87) on average in the last block. Hence 0.31 (0.53) of the welfare increase in the centralized-idealized treatment is attributable to the decrease in the number of mismatches over the decentralized market.

For the centralized-coerced treatment, we see an average of 0.22 (0.33) pairs unmatched, 1.20 (1.13) as the average quality of unmatched judges and 1.83 (1.63) as the average quality of unmatched applicants. The approximate welfare loss due to unmatched agents is 0.49 (0.61). The welfare difference between the centralized-idealized and centralized-coerced treatment is 1.4 (2.52) on average in the last block. Therefore, we attribute a welfare raise of 0.91 (1.91) in the centralized-idealized treatment to the decrease in the number of mismatched pairs.

6.2. The effect of announcements

Public announcement by judges of when they will begin accepting applications has been proposed as a possible solution to unraveling.[26] Table 4 shows, for the

---

[26] As well as a way of ameliorating the information gathering problems facing students when judges accept applications at different times from one year to the next.
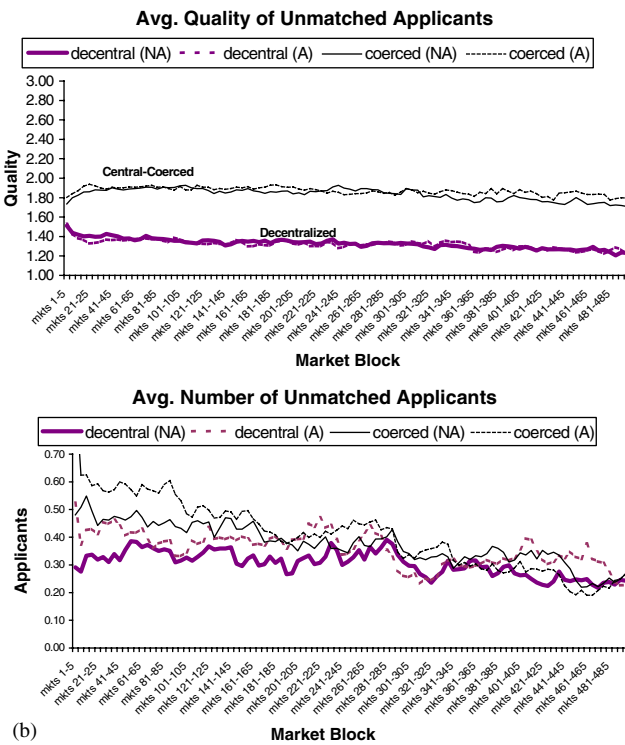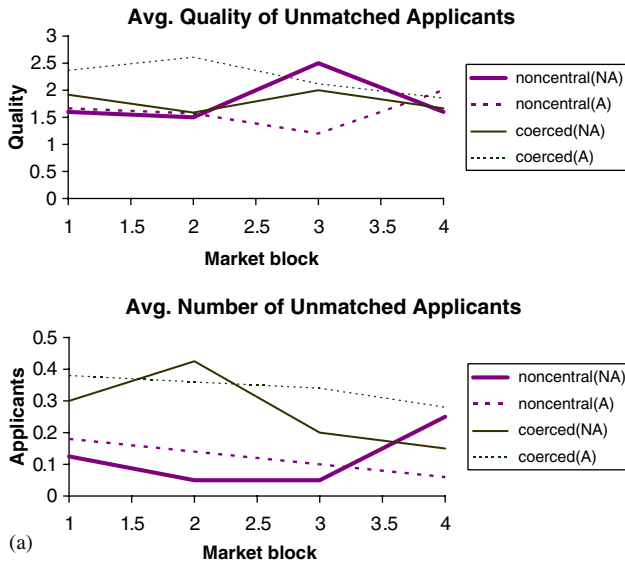
Fig. 3. Number and quality of applicants unmatched in the legal market. (a) Experimental findings. (b) Computational findings.

Table 3
Average quality of hired applicant by judge in the legal market

	Experimental—average of 20 mkts			Computational—average of 500 mkts		
	Decentral. NA/A	Central-idealized NA/A	Central-coerced NA/A	Decentral. NA/A	Central-idealized NA/A	Central-coerced NA/A
Judge 1	1.71/1.66	1.57/1.53	1.78/2.03	1.43/1.33	1.20/1.16	1.80/1.83
Judge 2	2.24/2.34	2.28/2.37	2.34/1.98	2.29/2.25	2.22/2.18	2.28/2.29
Judge 3	2.75/2.67	2.91/2.86	2.69/2.75	2.92/2.94	2.97/2.97	2.78/2.78
Judge 4	3.33/3.38	3.25/3.23	3.23/3.17	3.44/3.56	3.61/3.70	3.17/3.17

Table 4
Average entry period in announcement treatments in the legal market a judge's entry period is the first period in which a judge accepts applications from applicants

	Experimental—average of 20 mkts			Computational—average of 500 mkts		
	Decentral.	Central-idealized	Central-coerced	Decentral.	Central-idealized	Central-coerced
Judge 1	1.15	1.35	1.04	1.30	1.27	1.11
Judge 2	1.13	1.21	1.14	1.18	1.13	1.06
Judge 3	1.40	1.37	1.08	1.15	1.19	1.07
Judge 4	1.47	1.56	1.17	2.31	2.42	1.15

announcement treatments, average entry period by judge—the first period in which the judge would accept applications. We find that while such announcements may in fact result in later hiring in decentralized and idealized market settings (see Table 4), their welfare effects are small and gradually disappear over time. In the centralized-coerced matching setting, in which applicants must interview prior to the centralized match, and may be forced to accept an early offer, we find that announcements clearly do not improve welfare nor result in later hiring. Fig. 2 suggests that announcements may simply accelerate the underlying trend in either direction.

6.3. Strategic interviewing

Because applicants cannot refuse all offers extended to them in a given year, they have an incentive to interview strategically. In fact, many applicants in the experiment limit the judges to whom they apply to avoid being paired off early with a less preferred judge. This is consistent with the findings of Avery et al. (2001). A question on their student survey asked "Did you limit the number of judges to whom you applied based on a concern that some of your less-preferred judges would offer you interviews or positions before you had heard from your more preferred judges?" Roughly 55% of the respondents answered "yes" in the year 2000 survey.

Experimentally and computationally, we find that applicants apply selectively to judges ("strategic interviewing"), as follows:

1. Experimentally, all treatments are characterized by strategic interviewing as demonstrated by about 14% of applicants interviewing with judge 1 in year 1 versus 61–74% of applicants interviewing with judge 4 in year 1. Strategic interviewing is apparent computationally as demonstrated by a 17% to 57% gap between the proportion of applicants applying to judge 4 and judge 1 in year 1.
2. Experimentally, though in year 2 strategic interviewing is still prevalent, it is most prominent in the centralized-idealized (a 60% gap between first and last judge) and least prominent in the centralized-coerced (a 26% gap) treatments. (Recall that in the centralized-coerced treatment, failure to interview with a judge means loss of eligibility to match to that judge in the centralized match.) Computationally, we also observe strategic interviewing except in the centralized-coerced treatments. In the centralized-coerced treatment, since judge 4 usually hires an applicant earlier than the other judges, available applicants mostly apply to judges 1–3 in year 2. (Recall that applying to all remaining judges is not as dangerous as it would be if the rules did not allow the applicant to accept the best offer received.)
3. Experimentally and computationally, strategic interviewing in year 2 is diminished relative to year 1.
4. Both experimentally and computationally, the proportion of applicants interviewing with any judge in the first year is lowest in the centralized-idealized treatment with announcements.
5. Experimentally, the proportion of applicants interviewing with the low quality judges (1, 2) in the first year is highest in the (pooled across announcement conditions) decentralized treatment. Computationally, this proportion is highest in the (pooled across announcement conditions) centralized-coerced treatments.

7. Recovering strategies in the simulations

Although it is not feasible to extract subjects' strategies from observations of their actions in the laboratory experiments (since a strategy is the function that determines a player's action even in unreached contingencies), it is possible to recover the corresponding strategies in the GA simulations. The compositions of surviving strategies in the last 10 generations of the simulations are summarized in two tables (Table 5 for the judges and Table 6 for the applicants). In Table 5, judges' strategies are classified for judges of different quality. We observe that better quality judges learn to play less diverse strategies than lower quality judges in general, and the diversity of strategies is highest in the central-coerced treatment and lowest in the centralized-idealized treatment (except the decentralized treatment with announcements). This shows that learning or adaptation is most effective for the better judges who generally hire better quality applicants and in the centralized-idealized

Table 5
Judges' strategy composition in the last 10 markets

Judge	Central-ideal w/o ann		Central-ideal w ann		Central-coerced w/o ann		Central-coerced w ann		Decentral w/o ann		Decentral w ann	
	Strategies	Perc. (%)	Strategies	Perc. (%)	Strategies	Perc. (%)	Strategies	Perc. (%)	Strategies	Perc. (%)	Strategies	Perc. (%)
1	1111	22.73	1111	16.15	1111	8.70	1111	9.98	1111	13.58	1111	4.95
	1121	9.00	1121	22.20	1121	6.35	1121	17.28	1112	5.08	1121	4.48
	1151	6.28	1151	7.60	1131	8.68	1131	7.73	1123	3.73	1122	7.85
	1211	9.98	1211	3.50	1141	5.90	1151	2.63	1152	5.25	1222	2.05
	1221	5.60	1221	3.95	1211	11.98	1221	18.58	1153	7.15	1223	3.60
	1251	3.98	1421	4.48	1221	16.35	1231	7.13	1211	4.83	1421	3.85
	1411	2.95	1451	2.20	1231	4.35	1251	2.35	1251	9.03	1452	3.88
	1421	8.03	1511	5.78	1241	2.20	1311	3.00	1312	5.25	1453	2.63
	1511	5.78	1541	3.90	1321	5.35	1321	3.70	1511	6.03	1512	3.73
	1521	14.55	1551	5.05	1421	4.30	1331	4.95	1512	4.08	1522	4.18
	1531	3.03	2511	5.45	1431	5.60	1521	5.55	1513	3.80	1553	4.90
	1551	2.95	2521	2.43	1511	6.05	1531	4.88	1551	4.68	2512	13.18
	Other	5.18	3551	9.20	1521	4.45	2521	5.20	1552	3.25	2522	3.38
			Other	8.13	Other	9.75	Other	7.08	1553	3.33	2523	7.28
									Other	20.98	2553	3.43
											Other	26.68
2	1111	14.35	1111	6.28	1141	4.65	1111	3.30	1124	21.08	1114	4.05
	1121	6.18	1121	16.15	1211	5.85	1131	4.25	1134	16.65	1123	2.78
	1131	12.93	1131	2.98	1231	2.95	1151	4.78	1144	9.45	1124	4.75
	1141	3.00	1151	3.20	1241	5.38	1221	2.83	1154	4.30	1132	2.23
	1151	8.88	1221	11.73	1251	4.53	1241	8.40	1214	2.20	1134	2.55
	1211	5.43	1231	7.43	1311	4.40	1251	2.93	1224	4.28	1144	5.38
	1221	2.05	1251	16.95	1341	19.13	1341	2.83	1234	4.48	1154	3.18
	1231	18.60	1311	3.20	1351	4.38	1421	4.35	1254	2.50	1224	14.33
	1241	7.73	1351	2.58	1441	16.23	1431	6.80	1324	6.35	1234	5.13
	1251	2.35	1411	3.15	1451	6.40	1441	13.60	1334	4.23	1253	2.98
	1321	6.53	2511	2.55	1531	2.53	1451	14.38	1344	6.28	1254	2.03
	1331	4.58	2521	3.88	1541	11.83	1541	7.20	1413	2.95	1324	4.40

Table 5 (*continued*)

Judge	Central-ideal w/o ann		Central-ideal w ann		Central-coerced w/o ann		Central-coerced w ann		Decentral w/o ann		Decentral w ann	
	Strategies	Perc. (%)	Strategies	Perc. (%)	Strategies	Perc. (%)	Strategies	Perc. (%)	Strategies	Perc. (%)	Strategies	Perc. (%)
	Other	7.43	2531	7.55	1551	5.50	1551	6.75	1434	4.45	1334	6.38
			Other	12.40	Other	6.28	2541	12.58	Other	10.83	1424	2.45
							Other	5.05			1524	6.30
											1554	2.20
											2554	3.73
											3554	9.65
											Other	15.55
3	1321	3.18	1331	26.00	1141	2.80	1241	3.15	1234	5.35	1334	31.45
	1331	17.50	1351	2.20	1241	2.33	1341	17.85	1334	13.38	1434	28.78
	1341	5.05	1431	32.83	1341	12.70	1411	4.80	1344	7.85	1444	2.88
	1351	5.65	1441	17.25	1411	7.75	1441	18.30	1424	3.48	1454	5.50
	1431	30.40	1531	3.00	1441	28.75	1451	29.70	1434	30.20	1534	7.33
	1441	22.63	2531	5.98	1451	22.28	1541	10.45	1442	3.23	1554	2.35
	1451	10.78	2551	5.05	1541	15.50	2541	9.18	1454	4.98	2534	13.93
	Other	4.83	Other	7.70	1551	5.43	Other	6.58	1534	15.53	Other	7.80
					Other	2.48			1544	5.10		
									1554	2.23		
									Other	8.70		
4	1541	9.58	2551	19.80	1431	12.78	1411	2.70	1451	3.23	1554	6.68
	1551	88.95	3551	77.00	1441	15.93	1431	13.10	1544	4.78	2554	18.08
	Other	1.48	Other	3.20	1451	24.25	1441	19.90	1554	87.50	3554	72.40
					1541	32.95	1451	20.45	0	4.50	Other	2.85
					1551	11.43	1541	28.38	1	100.00		
					Other	2.68	1551	2.70				
							2541	10.28				
							Other	2.50				

Explanation of strategy $TR_1R_2R_3$: $T = 1,2,3$ entry period, $R_t = 1,2,\ldots5$ cutoff hiring decision in period t ($R_t = 5$ means no hiring). R_3 is set to 1 in centralized treatments and T is set to 1 in no announcement treatments.

Table 6
Applicant action composition in the last 10 generations

Period	Central-ideal w/o ann		Central-ideal w ann		Central-coerced w/o ann		Central-coerced w ann		Decentral w/o ann		Decentral w ann	
	Actions	Perc. (%)	Actions	Perc. (%)	Actions	Perc. (%)	Actions	Perc. (%)	Actions	Perc. (%)	Actions	Perc. (%)
1	1515	2.13	1511	2.40	1135	4.53	1324	5.90	1521	2.45	1151	2.33
	1555	2.70	1515	5.40	1222	5.15	2155	3.58	1525	3.73	1515	2.08
	2155	2.20	1551	4.38	1322	4.50	2545	4.03	1551	4.33	1555	7.43
	2515	6.28	1552	2.68	2321	4.68	3233	8.23	1552	3.80	2255	5.40
	2552	2.40	1555	2.50	2324	2.80	3332	3.03	1555	4.78	2551	3.50
	2555	4.98	2515	3.50	2342	4.35	3515	2.55	2515	3.95	2555	7.55
	5251	3.53	2555	4.58	2344	2.73	4224	2.15	2551	2.63	3555	4.43
	5255	2.48	5155	5.18	3223	3.75	4244	2.23	2555	6.48	5115	11.53
	5515	6.35	5225	3.53	3231	3.53	4341	2.58	4331	2.10	5255	5.85
	5525	5.78	5255	2.23	3321	2.18	4432	3.20	4341	2.65	5515	5.43
	5551	9.18	5515	14.65	3332	4.30	4433	3.48	5151	2.80	5551	4.88
	5555	39.25	5525	3.18	3343	3.30	5511	3.00	5155	5.83	5555	21.85
	Other	12.78	5551	5.38	4424	3.63	5515	2.88	5255	5.08	Other	17.78
			5555	21.68	4444	4.65	5531	4.48	5515	9.15		
			Other	18.78	5515	4.40	5555	9.05	5551	4.35		
					5531	2.03	Other	39.68	5555	16.48		
					5551	2.08			Other	19.45		
					5555	2.08						
					Other	35.38						
2	1515	7.23	1515	7.68	1332	3.55	1332	2.20	1515	4.20	1155	4.65
	1521	2.30	1551	7.35	2522	2.45	1521	3.90	1551	2.65	1515	6.18
	1555	4.23	1555	3.55	2524	3.18	2343	2.63	1555	6.83	1551	4.25
	2555	8.85	5155	9.30	3334	3.48	2525	2.13	2155	2.65	1555	9.18
	3555	2.38	5255	4.45	3341	4.33	4212	4.03	2555	4.50	5155	11.53
	5155	4.33	5511	2.83	3343	2.88	4241	2.73	4343	3.83	5511	4.48
	5511	5.20	5515	9.08	3521	3.60	4324	3.93	5151	3.75	5515	10.78
	5515	5.03	5522	2.05	4244	4.03	4333	4.10	5155	5.65	5551	10.03
	5535	2.70	5525	2.23	4332	2.95	4341	2.80	5255	11.25	5555	17.55

Table 6 (*continued*)

Period	Central-ideal w/o ann		Central-ideal w ann		Central-coerced w/o ann		Central-coerced w ann		Decentral w/o ann		Decentral w ann	
	Actions	Perc. (%)	Actions	Perc. (%)	Actions	Perc. (%)	Actions	Perc. (%)	Actions	Perc. (%)	Actions	Perc. (%)
	5551	10.63	5551	17.53	4334	4.35	4343	5.35	5511	3.65	Other	21.40
	5555	31.73	5555	23.18	4344	8.73	4344	3.75	5515	7.40		
	Other	15.43	Other	10.80	4444	2.68	4431	2.58	5531	2.05		
					4455	2.25	4444	7.60	5541	2.75		
					5511	2.78	4524	2.63	5551	6.33		
					5521	2.80	5212	3.33	5555	10.98		
					5551	4.43	Other	46.35	Other	21.55		
					Other	41.58						
3	N/A								3323	2.73	3323	4.78
									3434	6.53	3344	2.80
									3443	5.08	4243	3.80
									3444	4.53	4324	4.68
									4241	3.60	4333	2.30
									4323	2.30	4334	4.90
									4333	4.38	4342	4.25
									4334	8.15	4344	14.35
									4344	9.00	4424	5.40
									4423	5.65	4433	13.03
									4424	4.60	4434	8.50
									4433	9.70	4443	5.28
									4434	15.80	4444	11.33
									4443	5.78	Other	14.63
									Other	12.20		

Explanation of actions $N_1 N_2 N_3 N_4$ for period t: N_r refers to number of applications sent to top available judges in period t in case the applicant is ranked rth in period t ($N_r = 5$ refers to not sending any applications).

treatment where welfare is the highest. We observe that strategies that delay hiring decisions to later periods survive in high percentages for higher quality judges and in the centralized-idealized treatments. In Table 6, applicant strategies are classified according to different actions in different periods. In general, the highest diversity is observed in the centralized-coerced treatments and lowest diversity is observed in the centralized-idealized treatments (although the strategy space for the decentralized treatment is far bigger). Strategies that delay application decisions to later periods are highest in percentages in the centralized-idealized and then in the decentralized treatments.

8. Comparing the computational and experimental results

We use computational tools to explore the robustness of our experimental results. When the computational and experimental results are the same, the very different ways in which they are obtained suggest that they possess a good deal of robustness. When they are different, they point to outcomes that may be much more situation specific.

Both in the experiments with human subjects and in the simulations with adaptive computational agents:

- Decentralized matching leads to early hiring.
- Centralized matching leads to later hiring and improvements of welfare in the idealized, non-coercive environment.
- In the coercive environment, few participants are matched by the centralized clearinghouse when it is available, and it does not improve welfare.
- Announcements of when judges will begin accepting applications have little effect.

On the other hand, there is less reason to believe that all of the details of how the observed welfare gains were realized are highly robust. For example, comparing the decentralized and centralized-idealized conditions in Table 1 shows that which judges go later in the centralized algorithm is a little different in the experiment than in the simulations. That is, which judges employ the centralized clearinghouse the most may not be robust. In the Appendix, we present sensitivity analyses for both simulation and experimental design parameters.

Note how the laboratory and computational experiments complement each other. The purpose of the experimental design, which systematically varies some of the conditions (announcements and matching technology), is to determine if these variables produce reliable differences. The computational and laboratory experiments both produce the same effects for the experimental variables. At the same time, small differences between the laboratory and computational experiments suggest that some of the incidental results of the two kinds of experiments, concerning parameters of the experiments that were not experimental variables (since judges' qualities were the same in all conditions), may not be reliably generalized.

9. Recent developments in the market

Most law clerks for terms beginning in 2003 were hired in the very early Fall of 2001, many in September. In March of 2002 (after these experiments were conducted), a large majority of Federal appellate judges voted to approve a proposal stating that ''…the hiring of law clerks in the Fall after the first year of law school is an unacceptable practice,'' and that they therefore endorsed ''a moratorium on law clerk hiring during the Fall of 2002, save for those judges who have yet to hire law clerks for the 2003–2004 term.''[27] The proposal calls for no structural changes in the market, remarking that

> The one good thing about the current situation is that it affords an easy solution to the existing problem. In order to switch from the current system, which focuses on second year law students, to a hiring system that focuses on third year law students and law graduates, judges need only abstain from clerk hiring in the Fall of 2002 and resume again in the Fall of 2003.

The results of the present experiment suggest that the adoption of this plan can be at most a short term solution, as the features of the market that promote unraveling remain in place.[28] And the initial experience of the plan led to revisions intended to prevent offers from being made too early in the Fall, reflecting some of the difficulties encountered in 2003.

Starting in 2004, the judges' plan divides the Fall into three periods. Applications and letters of reference for law clerk positions are not supposed to be submitted until the day after Labor Day. Judges are not supposed to schedule interviews until the first Monday after Labor Day, and no interviews are supposed to be conducted, or offers made, until the second Monday after Labor Day. So, under the revised plan, there is a week in which applications are submitted and read, a week in which interviews are scheduled but not yet conducted (during which applications continue to be read), and, starting the third week after Labor Day, a period in which interviews may be conducted and offers made (with no guidelines about when offers can be made, how long they must be left open, etc.).

---

[27]Letter to law school deans, March 11, 2002, signed by Chief Judge Edward R. Becker, and Judge Harry T. Edwards.

[28]A FAQ accompanying the letter by Judges Becker and Edwards makes clear that the proposal does not make effective changes to the elements of the market identified here as causing difficulties. For example, it includes the following questions and answers.

Q   Does the Plan endorse Summer interviewing?

A   No. …However, the Plan does not forbid a law student who, say, is from Virginia and working in Tulsa during the Summer from talking with a judge who is otherwise available to chat…

Q   Are judges forbidden from making ''exploding offers,'' i.e., offers that require an applicant to respond promptly to an offer?

A   The Plan does not purport to address how an offer is given by a judge. This is for each judge to determine. However, no applicant is obliged to act on an offer if the terms are unacceptable, nor is an applicant obliged to accept the first offer that he or she receives.

Notice that within this 3 week period, information will evolve differently than in the longer period over which this market has traditionally played out. However a student who is called early in the morning on the first Monday after Labor Day, and invited to an interview the morning of the second Monday after Labor Day by, for example, his third choice judge, would find it valuable to know which other judges might be interested in him (since he may be faced with an offer he can't turn down if he accepts that interview.) The results of the present experiment and computations, together with the fact that nothing in the revised plan reduces the coercive power of judges with respect to students, or judges' ability to make exploding offers, therefore strongly suggests that in years to come, the strategic scheduling of interviews (on the part of both students and judges) will persist.

10. Conclusions

Unraveling of hiring of clerks by Federal appellate judges has become recognized as a serious problem.[29] The present paper uses experimental and computational methods, in combination, to explore some of the proposed reforms to this market, and the obstacles that they face.

Because the experimental environment is a vast simplification of the conditions in the actual law clerk market, and because the experimental subjects are not judges and law students engaged in some of their most important labor market decisions, caution is obviously required in speculating about how the experimental results might generalize to the field. However, there are also limits to what can be accomplished in a field study. One unavoidable limit is that a field study (such as Avery et al., 2001) can only study the market as it is, and not as it might be if major changes in market design were introduced.

Experience with medical labor markets gives us some confidence that experiments on simple markets have some predictive power for complex field markets, and that large differences due to market design can be detected. In the present study, we have, in addition to the experiments with human subjects, also observed the same main effects in a computational simulation using genetic algorithms as adaptive agents. This suggests, at the very least, that the results we observe experimentally are due to the incentives present in the strategic environment, and not due to some undetected properties of the subject pool used for this experiment.[30] The genetic algorithms also

---

[29]The report of the ad hoc committee chaired by Judges Becker and Edwards, included with their March 11, 2002 letter, states that ''Many judges, law school deans and faculty members, and law students now agree that the existing law clerk hiring practices are irrational, unfair, and entirely indefensible.''

[30]One way to think of the genetic algorithm results is like those of an experiment conducted on Mars. In general, we wouldn't find Martian evidence to be compelling evidence about terrestrial markets, but if we observe the same thing in a terrestrial market (in the lab) and in a Martian market (in the behavior of the genetic algorithms), that gives some support to the notion that we are seeing a property of the market and not of the subject population (students or genetic algorithms), and that therefore the results are likely to carry over also to the population of interest (participants in the clerkship market). That is, observing the same results both in the lab and in the genetic algorithms makes it less likely that what we are seeing is an artifact of the particular characteristics of either the experimental subjects or the adaptive agents.

give us a chance to look at the evolution of behavior in these markets over a wide range of time scales, which gives us a further robustness check. Thus the results of the present study give us grounds to speculate on the effects of proposed changes in the organization of the law clerk market, and the obstacles they face.[31]

Specifically, we set out to investigate whether the market design that has been effective in the medical market could succeed in the institutional environment of the law clerk market as reported by Avery et al. (2001). They speculated that the obligation perceived by many law students to accept an offer if one was made would be among the most important obstacles to reform in the law clerk market, and our experiments provide a test of this hypothesis.

We found that centralized matching was effective in an idealized legal market environment, in which applicants could avoid early offers with the attendant obligation to accept. However, in the presence of one of the most distinctive features of this market—that applicants could be obliged to accept an offer while interviewing for a position—centralized matching was not successful, and almost all matches were arranged before the centralized match. (So, depending on the timing of the match (and of interviews for the match), this could even reduce welfare compared to decentralized matching, as it did in our experimental environment.)

We pessimistically conclude that neither the centralized matching clearinghouse that is so effective in medical markets, nor the announcements currently implemented by federal judges, nor the latest, revised, "three-period" Fall hiring plan, are likely succeed in reversing the unraveling of this market as long as the distinctive elements of market culture that we have discussed remain in place.[32] To the extent that the experiment and genetic algorithm capture the important differences between medical and legal markets, this suggests that, before this market failure can be resolved, it may be necessary to change the culture of the market in which law students feel compelled to accept the first offer they receive. Unlike many markets in which the inability to make binding contracts contributes to inefficiency, in the law clerk market, ironically, it is the promiscuous ability to make binding contracts that harms efficiency.

Acknowledgements

---

[31]In general, this paper is part of a stream of work on matching markets that seeks to employ field observation, theory, experiments, and computation as aids to market design. (We have not emphasized in this paper the theoretical background, which involves stable matchings in the manner of Gale and Shapley, 1962: see e.g. Roth and Sotomayor, 1990). For discussion of other design work involving these combinations of tools, see Roth (2002).

[32]For more on market culture, see Niederle and Roth (2005).

http://kuznets.fas.harvard.edu/~aroth/papers/judgesexperiment.jedc.appendices.pdf.
In the experiment, we used z-Tree software (Urs Fischbacher, 1999).

Appendix A. Online Appendices

The online version of this article contains additional supplementary data. Please visit doi:10.1016/j.jedc.2005.02.002.

References

Abdulkadiroğlu, A., Sönmez, T., 2003. School choice: a mechanism design approach. American Economic Review 93, 729–747.

Anderson, K.D., Jacobs, D.M., Blue, A.V., 1999. Is Match Ethics an Oxymoron? American Journal of Surgery 177, 237–239.

Arifovic, J., 1996. The behavior of the exchange rate in the genetic algorithm and experimental economies. Journal of Political Economy 104, 510–541.

Avery, C., Jolls, C., Posner, R.A., Roth, A.E., 2001. The market for federal judicial law clerks. University of Chicago Law Review 68, 793–902.

Avery, C., Fairbanks, A., Zeckhauser, R., 2003. The Early Admissions Game: Joining the Elite. Harvard University Press, Cambridge, MA.

Becker, G.S., 1981. A Treatise on the Family. Harvard University Press, Cambridge, MA.

Becker, E.R., Breyer, S.G., Calabresi, G., 1994. The federal judicial law clerk hiring problem and the modest March 1 solution. Yale Law Journal 104, 207–225.

Bullard, J., Duffy, J., 1998. A model of learning and emulation with artificial adaptive agents. Journal of Economic Dynamics and Control 22, 179–207.

Carek, P.J., Anderson, K.D., Blue, A.V., Mavis, B.E., 2000. Recruitment behavior and program directors: how ethical are their perspectives about the match process? Family Medicine 32, 258–260.

Chen, Y., Sönmez, T., 2002. Improving efficiency of on-campus housing: an experimental study. American Economic Review 92, 1669–1686.

Dawid, H., 1999. Adaptive Learning by Genetic Algorithms: Analytical Results and Applications to Economic Models, second ed. Springer, Berlin.

Duffy, J., 2001. Learning to speculate: experiments with artificial and real agents. Journal of Economic Dynamics And Control 25, 295–319.

Erev, I., Roth, A.E., 1998. Predicting how people play games: reinforcement learning in experimental games with unique, mixed strategy equilibria. American Economic Review 88, 848–881.

Feltovich, N., 2000. Reinforcement-based vs. belief-based learning models in experimental asymmetric-information games. Econometrica 68, 605–641.

Fischbacher, U., 1999. z-Tree—Zurich Toolbox for Readymade Economic Experiments—Experimenter's Manual, mimeo.

Gale, D., Shapley, L., 1962. College admissions and the stability of marriage. American Mathematical Monthly 69, 9–15.

Goldberg, D.E., 1989. Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley, Reading, MA.

Holland, J.H., 1975. Adaptation in Natural and Artificial Systems. The University of Michigan Press, Ann-Arbor, MI.

Holland, J.H., Miller, J.H., 1991. Artificial Adaptive Agents in Economic Theory. American Economic Review: Papers and Proceedings of the 103rd Annual Meeting of the American Economic Association, pp. 335–350.

Kagel, J.H., Roth, A.E., 2000. The dynamics of reorganization in matching markets: a laboratory experiment motivated by a natural experiment. The Quarterly Journal of Economics 115, 201–235.

Li, H., Rosen, S., 1998. Unraveling in matching markets. American Economic Review 88, 371–387.

Li, H., Suen, W., 2000. Risk Sharing, Sorting, and Early Contracting. Journal of Political Economy 108, 1058–1091.

Li, H., Suen, W., 2001. Early contracting rush, working paper, University of Toronto and University of Hong Kong.

McKinney, C.N., Niederle, M., Roth, A.E., 2005. The collapse of a medical labor clearinghouse (and why such failures are rare). American Economic Review (forthcoming).

Michalewicz, Z., 1994. Genetic Algorithms + Data Structures = Evolution Programs, second ed. Springer, Berlin.

Miller, J.H., 1998. Active nonlinear tests (ANTs) of complex simulation models. Management Science 44, 820–830.

Niederle, M., Roth, A.E., 2003. Unraveling reduces the scope of an entry level labor market: gastroenterology with and without a centralized match. Journal of Political Economy 111, 1342–1352.

Niederle, M., Roth, A.E., 2005. Market Culture: How Norms Governing Exploding Offers Affect Market Performance, working paper.

Niederle, M., Roth, A.E., 2004. The gastroenterology fellowship match: how it failed, and why it could succeed once again. Gastroenterology 127, 658–666.

Ostrovsky, M., Schwartz, M., 2002. Waiting Times and Coordination, working paper.

Pearson, R.D., Innes, A.H., 1999. Ensuring compliance with NRMP policy. Academic Medicine 74, 747–748.

Pingle, M., Tesfatsion, L., 2001. Non-employment benefits and the evolution of worker-employer cooperation: experiments with real and computational agents. Economic Report No. 55, Iowa State University.

Roth, A.E., 1984. The evolution of the labor market for medical interns and residents: a case study in game theory. Journal of Political Economy 92, 991–1016.

Roth, A.E., 1991. A natural experiment in the organization of entry level labor markets: regional markets for new physicians and surgeons in the UK. American Economic Review 81, 415–440.

Roth, A.E., 2002. The economist as engineer: game theory, experimental economics and computation as tools of design economics. Econometrica 70, 1341–1378.

Roth, A.E., Erev, I., 1995. Learning in extensive-form games: experimental data and simple dynamic models in the intermediate term. Games and Economic Behavior 8, 164–212.

Roth, A.E., Peranson, E., 1999. The redesign of matching markets for American physicians: some engineering aspects of economic design. American Economic Review 89, 748–780.

Roth, A.E., Sotomayor, M., 1990. Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis. Econometric Society Monograph Series. Cambridge University Press, Cambridge.

Roth, A.E., Xing, X., 1994. Jumping the gun: imperfections and institutions related to the timing of market transactions. American Economic Review 84, 992–1044.

Roth, A.E., Xing, X., 1997. Turnaround time and bottlenecks in market clearing: decentralized matching in the market for clinical psychologists. Journal of Political Economy 105, 284–329.

Roth, A.E., Sönmez, T., Ünver, M.U., 2004. Kidney exchange. Quarterly Journal of Economics 119, 457–488.

Sönmez, T., 1999. Can pre-arranged matches be avoided in two-sided matching markets? Journal of Economic Theory 86, 148–156.

Suen, W., 2000. A competitive theory of equilibrium and disequilibrium unravelling in two-sided matching. RAND Journal of Economics 31, 101–120.

Teichman, J.M., Anderson, K.D., Dorough, M.M., Stein, C.R., Optenberg, S.A., Thompson, I.M., 2000. The urology residency matching program in practice. Journal of Urology 163, 1878–1887.

Ünver, M.U., 2000. Computational and Experimental Analyses of Two-Sided Matching Markets, Ph.D. dissertation, University of Pittsburgh.

Ünver, M.U., 2001. Backward unraveling over time: the evolution of strategic behavior in the entry-level british medical labor markets. Journal of Economic Dynamics and Control 25, 1035–1080.

Ünver, M.U., 2005. On the survival of some unstable two-sided matching mechanisms. International Journal of Game Theory (forthcoming).