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] and M. Utku Ünver[c]

November, 2003

[a]University of Texas at Dallas

[b]Harvard Business School and Department of Economics, Harvard University

[c]Koç University

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 reform proposed by the judges, in March 2002, which is a one year moratorium on clerkship hiring. Unlike many markets in which the inability to make binding contracts contributes to market failure, in the law clerk market it is the ease with which binding contracts are forged that harms efficiency.

---

* We thank John Duffy and Muriel Niederle for helpful comments on drafts of this paper. Parts of this work were supported by grants from the National Science Foundation, and from Koç University. Instructions and data to accompany this work are available at www.utdallas.edu/~eharuvy/legalmatch.
In the experiment, we used z-Tree software (Urs Fischbacher, 1999)..

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 and 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, Jolls, Posner, and Roth (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 two 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, Breyer, and Calabresi, 1994, for a discussion of the market's history through the early 1990's).

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.

[1] In 2003 another reform has been attempted but the results are not yet known.

[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.

Roth and Xing (1994) discuss several dozen markets and submarkets that have experienced unraveling of appointment dates.[3] A number of theoretical studies have shown how this unraveling can lead to ex post (and sometimes ex ante) inefficiency, 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. The largest of these we know of is the market for new physicians in the United States, which established a centralized clearinghouse in the early 1950's, 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 2001b) helped confirm that this kind of stability is important to the success of such clearinghouses.[6] These clearinghouses have been subject to a number of 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 a number of 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, and Avery et al. 2001). Avery et al. (2001) argue that some of the special features of the law clerk market, to be

[3] And see Avery, Fairbanks and Zeckhauser (2003) on the growth of early decision in college admissions.
[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) and McKinney, Niederle and Roth (2003) 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.

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 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 and 2001a) 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.

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

[9] Time scale is among the hardest features of experiments, and of computations, to map onto field studies, 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.

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 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 largely had to do 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

[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, Fairbanks, and Zeckhauser, 2003).

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 (2001b) 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 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

---

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

[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 and data files are available at www.utdallas.edu/~eharuvy/legalmatch.

[13] Although we will often speak about periods as "years," they can be interpreted as any time period in which information becomes available.

[14] This environment is chosen to make the efficiency issues clear, so that an efficient match is strictly assortative, i.e. so that an efficient outcome matches the most productive firm to the most productive worker, and so forth (cf. Becker, 1981).

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]

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.

[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.

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 Announcements	8 cohorts 5/21/2001 (2 cohorts) 6/7/2001 (4 cohorts) 7/18/2001 (2 cohorts)	10 cohorts 5/7/2001 (3 cohorts) 6/20/2001 (3 cohorts) 7/2/2001 (4 cohorts)	8 cohorts 5/26/2001 (4 cohorts) 7/3/2001 (4 cohorts)
With Announcement	10 cohorts 5/21/2001 (3 cohorts) 6/6/2001 (2 cohorts) 7/18/2001 (2 cohorts) 7/25/2001 (3 cohorts)	11 cohorts 5/30/2001 (4 cohorts) 6/28/2001 (4 cohorts) 7/31/2001 (1 cohorts) 8/2/2001 (2 cohorts)	10 cohorts 7/10/2001 (7 cohorts) 8/1/2001 (1 cohorts) 8/2/2001 (2 cohorts)

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

<u>In all six treatments, in each year (1-2 in the centralized treatments and 1-3 in the decentralized):</u>

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.

<u>In the centralized-idealized treatment, year 3:</u>

[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.

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).

<u>In the centralized-coerced treatment, year 3:</u>

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 four groups of eight, such that no person knew who the other seven members of his group were. Each group of eight

[17] Here our convention of loosely referring to periods as "years" is potentially misleading. Think of the periods as short intervals prior to the match. Students who wish to participate in the match must apply to and be interviewed by judges before the match (at a time of the judges' choosing), and are vulnerable to

consisted of four judges and four applicants. 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 firms do not

being coerced during the interview (In contrast, in the centralized-idealized condition, subjects can in effect apply to all judges without risk of being forced to accept a position before the match).

make any offers. This strategy profile will result in a matching in which firm 4 is matched with the best applicant of year 1 in year 1, firm 3 will be matched to the second best applicant of year 1 in year 1, and so on. This strategy is an equilibrium since by making unilateral deviations from the strategy no agent can improve her payoff. The same strategy will be an 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 firms make offers in year 1 and year 2 even if firms receive applications. In year 3, all applicants apply to all firms. 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 an equilibrium, since by a unilateral deviation, i.e. by applying early, no applicant will be hired early since firms do not hire applicants in years 1 and 2, and a firm 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 firms, and firms 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 firms, at which time they can be hired. We conjecture that at any equilibrium in the centralized-coerced condition, at least one firm 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 firms find themselves in direct competition over an applicant who applied to both firms in year 2, the lower ranked of the two will 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 firm in period 1; firms hire no applicant in period 1, even if there are applications. In period 2, each applicant applies to all firms, and firm i extends an offer to the i^{th} 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 firm will not improve payoff, since firms 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 firm 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 firms in period 2, because applying to fewer firms does not have any effect in increasing their chances to be matched to better firms.

In the counterparts of these games in which firms announce their entry periods, the strategies above constitute equilibria whenever firms announce their availability starting in year 1. Moreover, 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 Sensitivity Analysis section.

A strategy for an agent in a genetic algorithm is represented by a string of integers. We use a bounded rational representation by only modeling some of the

---

[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 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 at 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... Such systems have recently come to be known as complex adaptive systems."

[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.

information sets. For instance, in the decentralized game with announcements, a firm strategy is encoded by 6 integers.

$$T\text{- } A^1\text{-}R^1\text{-}A^2\text{-}R^2\text{-}R^3$$

T is an integer in {1,2,3}. This decision variable is the year when the firm 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 firm may hire an applicant in year t. When A^t=0, the firm 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 firm is going to hire. The full strategy representations for applicants and firms 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 firm 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: firms 1 through 4, and workers. 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.

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

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 Appendix 2. 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.

   Crossover Example:

   P_1 : a_1-a_2-a_3-a_4 $\quad\quad$ O_1: a_1-a_2-b_3-b_4

   $\rightarrow$ Crossover starting at 3rd digit $\rightarrow$

   P_2: b_1-b_2-b_3-b_4 $\quad\quad$ O_2: b_1-b_2-a_3-a_4

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:

   a_1-a_2-a_3-a_4 $\quad\quad$ $\rightarrow$ Mutate 2nd digit $\rightarrow$ $\quad\quad$ a_1-x_2-a_3-a_4

[22] Although mutation is a very important evolutionary idea (especially at the beginning of evolution), in genetic algorithms as soon as the fittest strategies start to emerge, this operator loses its importance. We exogenously decrease 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^m_{max}=0.07 to p^m_{min}=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.

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 Appendix 3 for the choice of genetic algorithm parameters. We show that the majority of the results are robust to choices of real parameters in a range within ± 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

Figure 1 shows that very few matches are in fact consummated by the centralized-coerced match. The top panel graphs the number of applicants going through centralized match in the experiment, over time as the subjects gain experience, and the bottom panel graphs the same relationship for the computations. Interestingly, as Figure 1 shows, the number of applicants entering the central match is increasing over time in the centralized-idealized treatments. This is the opposite of the trend in the centralized-coerced treatment, in which the number going through the match is decreasing.

Table 1 shows that period 3 matches are fewest in the centralized-coerced treatment. Pooling over announcement conditions, in the decentralized and centralized-idealized conditions 46.90% of the all matches are realized in period 3, taking an average over the last 5 markets. On the other hand the average percentage of the employed agents who use the centralized-coerced match is only 13% over the last 5 markets, with a 33.90% difference from the other treatments. For significance determination, each cohort’s average over the last five markets was treated as a single observation. With 39 cohorts of decentralized and centralized-idealized conditions and 18 cohorts of the centralized-coerced condition, our t-statistic is 5.50 with a two-tail p-value less than 0.0001 for 55 degrees of freedom. There is no significant difference between the percentage of applicants who use the centralized-idealized match and the percentage of applicants who are matched in period 3 in the decentralized market. (The prior is 1.91% larger, with a t-statistic 0.50 and p-value 0.6211 for 37 degrees of freedom). In the genetic algorithm simulations, 83.07% of the employed applicants are matched in period 3 in the centralized-idealized and decentralized treatments, and only 9.27% of the employed applicants use centralized-coerced match using averages over the last block of 50 markets.[24] Moreover, the percentages of employed agents who get matched in period 3 are similar for computational decentralized and centralized-idealized treatments, only the latter is slightly higher. Table 1 shows on firm level that in the centralized-coercive environment, matches are mostly made in periods 1 and 2.

The major effect of this observation can be seen in the difference in welfare between the centralized-coerced treatment and other treatments. Figure 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.

Pooling over announcement conditions, comparing the decentralized to the centralized-idealized conditions of the experiment, we find that the centralized-idealized market improves welfare by 0.37 on average over all 20 markets, and by 0.55 on average over the last five markets. With 18 cohorts of decentralized matching and 21 cohorts of centralized-idealized, our t-statistic of -1.96 has 37 degrees of freedom and is significant with a one-tail P-value of 0.029. The corresponding Wilcoxen Rank Sums Z-statistic is -1.84, with a one-tail P-value of 0.033. In the genetic algorithm simulation, the increase in welfare is even more prominent with an improvement by 1.09 on average over all 500 markets, 0.87 on average in the last 50 markets. This is in sharp contrast to the strong welfare decreasing effect of the centralized-coerced variation, as Figure 2 illustrates. Thus, in an environment in which students are subject to coercive offers when they interview, the results suggest that a centralized match will *not* have the welfare-

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

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 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.

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 firms 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 firms 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

[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.

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] 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. Figure 2 suggest that announcements may simply accelerate the underlying trend in either direction.

Figure 2 shows the changes in welfare for all six treatments in the experiment as well as for the genetic algorithm. Looking at the block of last five markets in each session, treating each cohort's average as a single independent observation, the addition of announcements resulted in no significant welfare changes in the decentralized and centralized-idealized conditions. In the decentralized setting, the one-sided T-test for the significance of announcement (16 d.f.) has a P-value of 0.364. The corresponding one-sided Wilcoxen Rank Sums P-value is 0.377. In the centralized-idealized case, the one-sided T-test for the significance of announcement (19 d.f.) gives a P-value of 0.157 and the corresponding one-sided Wilcoxen Rank Sums P-value is 0.128. The genetic algorithm had some small improvement in welfare due to announcement in these two conditions, but this improvement vanishes over time for the centralized-idealized treatment.

In the experimental data, announcements did not significantly delay hiring for any firm by either the T-test or Wilcoxen Rank Sums in either the decentralized or the centralized-idealized conditions. In the experimental data, what appears to hinder welfare increases following announcements is the success of firm 2 in hiring better quality

[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.

applicants, once firms 3 and 4 make later entry announcements. Table 3 demonstrates this assertion.

As in the decentralized and idealized treatments, where announcements appeared to not have much of an impact, in the centralized-coerced treatments, announcements had no significant effect on either average hiring periods by firms, excluding unmatched firms, or on welfare (one sided T-test, with 18 d.f., had a P-value of 0.184 and a corresponding Wilcoxen Rank Sums had a P-value of 0.238). Notwithstanding the lack of significance, welfare appears adversely affected by announcement in the centralized-coerced treatments, an observation confirmed by the genetic algorithm. Figure 3 suggests that this adverse effect is at least partly due to a higher number and quality of unmatched applicants. Table 3 suggests that in the experimental data another reason is the significant gains in quality of hired applicants by firm 1, the lowest quality firm, which receives more applications when its peers are not available to receive applications.

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 central 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. 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 firms go later in the centralized algorithm is a little different in the experiment than in the simulations. That is, which firms employ the centralized clearinghouse the most may not be robust. In Appendix 2, we present sensitivity analyses for both simulation and experimental design parameters.

8. 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]

9. 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

---

[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 How is "Fall" determined under the Plan?

A There is no fixed definition of Fall, nor is there any fixed starting date for the hiring season.

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.

[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."

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 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.

[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.

[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).

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 the distinctive feature 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.)

Another market modification, recently adopted, allows judges to announce the time they intend to begin receiving applications. We found that announcements had a very marginal contribution to welfare in idealized conditions and a substantial adverse effect in the more realistic matching setting.

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 proposed one year moratorium on hiring, 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.

References

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.

[32] For more on market culture, see Niederle and Roth (2003).

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

Avery, C., Jolls, C., Posner, R. A., and Roth, A. E. (2001), "The Market for Federal Judicial Law Clerks," *University of Chicago Law Review*, 68:793-902.

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

Becker, E. R., Breyer, S. G., and Calabresi, G. (1994), "The Federal Judicial Law Clerk Hiring Problem and the Modest March 1 Solution," *Yale Law Journal*, 104:207-225.

Bullard, J. and Duffy, J. (1998), "A Model of Learning and Emulation with Artificial Adaptive Agents," *Journal of Economic Dynamics and Control*, 22:179-207.

Chen Y. and Sönmez, T. (2002), "Improving Efficiency of On-Campus Housing: An Experimental Study,'' *American Economic Review*, 92: 1669-1686.

Carek, P. J., Anderson, K. D., Blue, A. V., and Mavis, B. E. (2000), "Recruitment Behavior and Program Directors: How Ethical Are Their Perspectives about the Match Process?" *Family Medicine*, 32:258-260.

Dawid, H. (1999), *Adaptive Learning by Genetic Algorithms: Analytical Results and Applications to Economic Models*, Second Edition. Springer-Verlag, Berlin.

Duffy, J. (2001), "Learning to Speculate: Experiments with Artificial and Real Agents," *Journal of Economic Dynamics And Control*, 25:295-319.

Erev, I. and 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. and Shapley, L. S. (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. and 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*, 335-350.

Kagel, J. H. and 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. and Rosen, S. (1998), "Unraveling in Matching Markets," *American Economic Review*, 88:371-387.

Li, H. and Suen, W. (2000), "Risk Sharing, Sorting, and Early Contracting," *Journal of Political Economy*, 108:1058-1091.

Li, H. and Suen, W. (2001), "Early Contracting Rush," working paper, University of Toronto and University of Hong Kong.

McKinney, C. N., Niederle, M., and Roth, A. E. (2002) "The Collapse of a Medical Labor Clearinghouse (and Why Such Failures are Rare)," December, 2002.

Michalewicz, Z. (1994), *Genetic Algorithms + Data Structures = Evolution Programs, Second Edition*. Springer-Verlag, Berlin.

Miller, J. H. (1998), "Active Nonlinear Tests (ANTs) of Complex Simulation Models," *Management Science* 44: 820-830.

Niederle, M. and 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*, forthcoming.

Niederle, M. and Roth, A. E. (2003), "Market Culture: How Norms Governing Exploding Offers Affect Market Performance," working paper.

Ostrovsky, M. and Schwartz, M. (2002), "Waiting Times and Coordination," working paper, Harvard University.

Pearson, R. D. and Innes, A. H. (1999), "Ensuring Compliance with NRMP Policy," *Academic Medicine*, 74:747-748.

Pingle, M. and 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 U.K.," *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. and 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. and 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. and 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. and 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. and 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.

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., and 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," PhD dissertation, University of Pittsburgh.

Ünver, M. U. (2001a), "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. (2001b), "On the Survival of Some Unstable Two-Sided Matching Mechanisms," working paper.

Table 1. Average Period of Hiring by Firm in the Legal Market[33]

	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
Firm 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
Firm 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
Firm 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
Firm 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

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 firm	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

Table 3. Average Quality of Hired Applicant by Firm 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
Firm 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
Firm 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
Firm 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
Firm 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

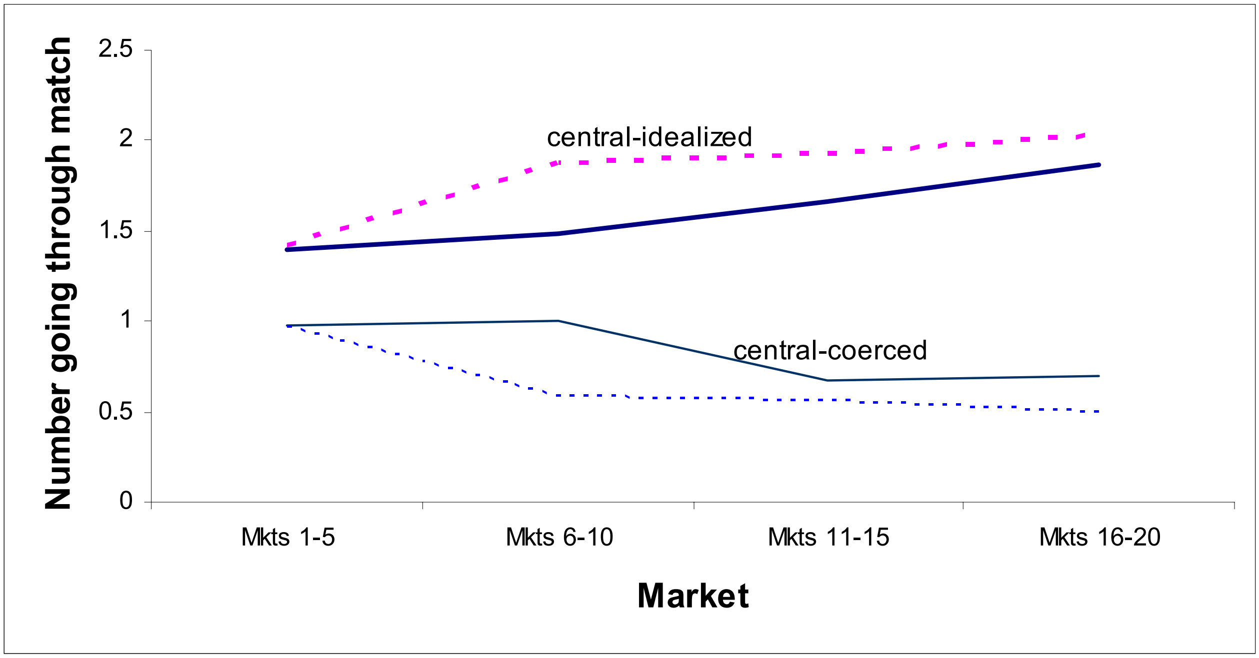
[33] In each box, we report the averages for the conditions with and without announcements. We report the without-announcement average before "/" and the corresponding announcement average after "/". The same notation is also used in Table 3.

Table 4. Average Entry Period in Announcement Treatments in the Legal Market

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

Figure 1. Number of Firms/Applicants Matched through the Centralized Matching in the Legal Market. Dashed lines represent announcement conditions.

a. Experimental Findings:



b. Computational Findings:

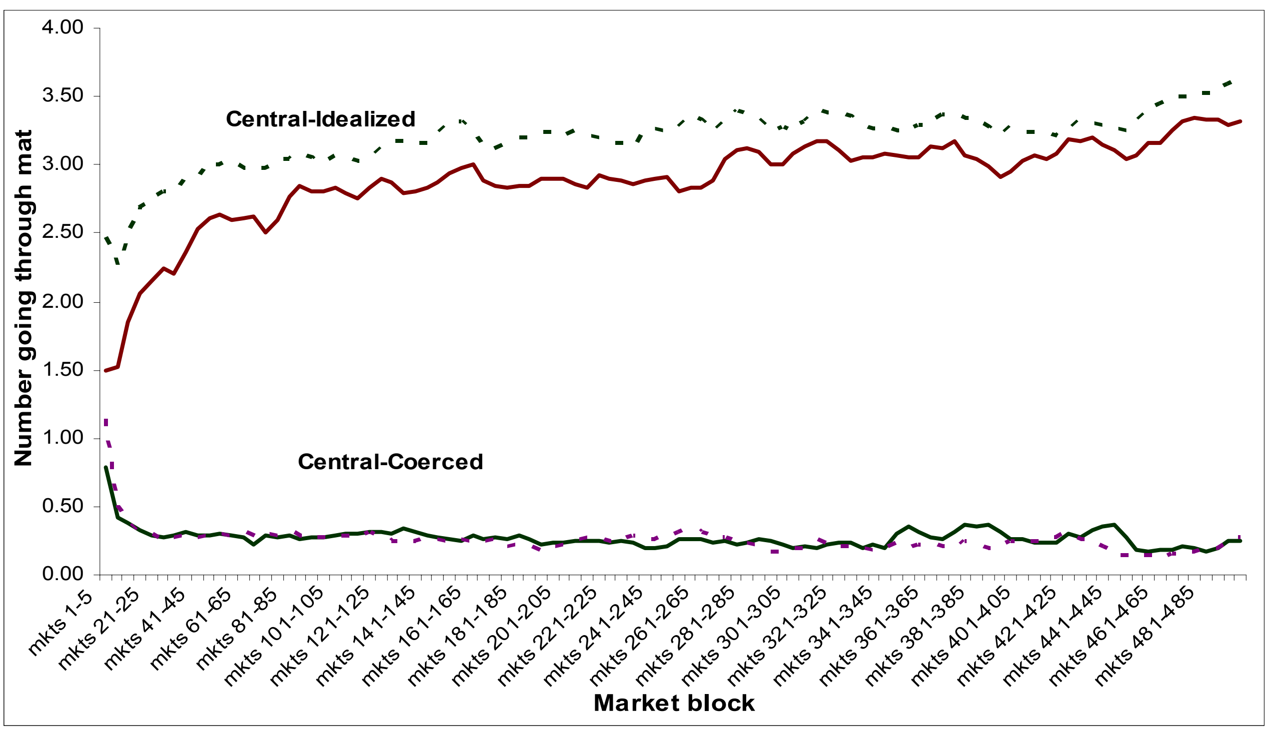
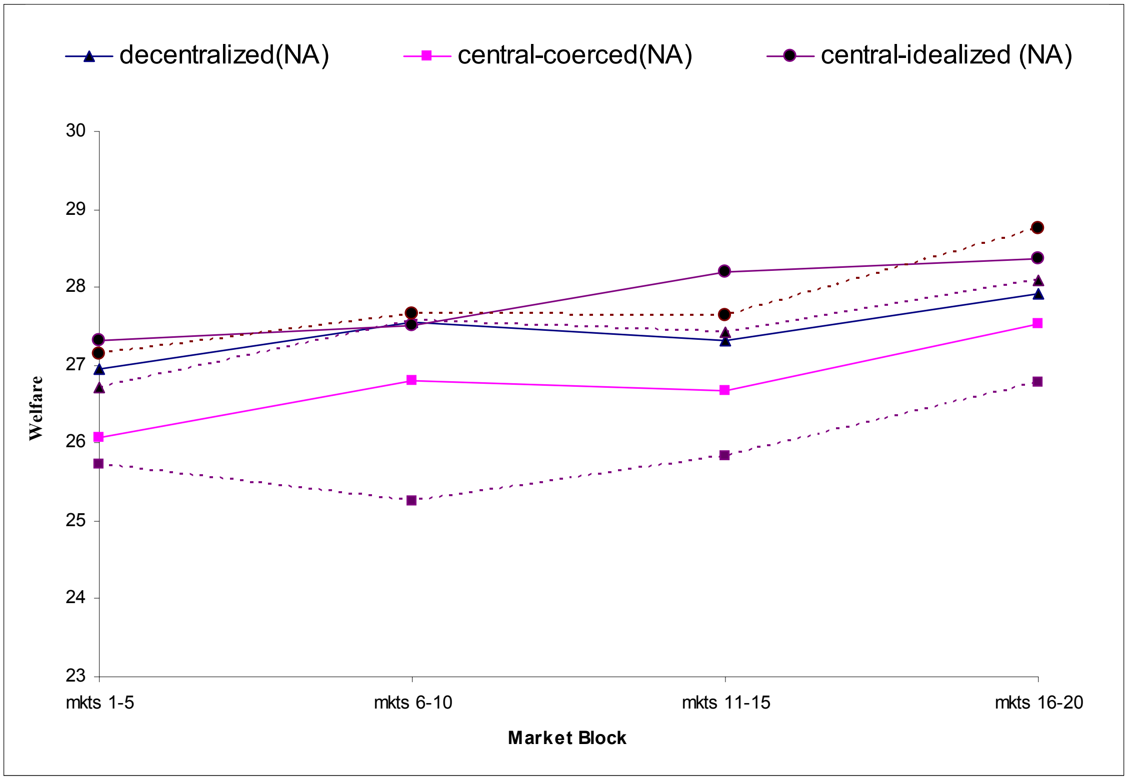


Figure 2: Average Group Welfare in the Legal Market. For the six treatments in the legal market, dashed lines represent treatments with announcements.

a. Experimental Findings:



b. Computational findings:

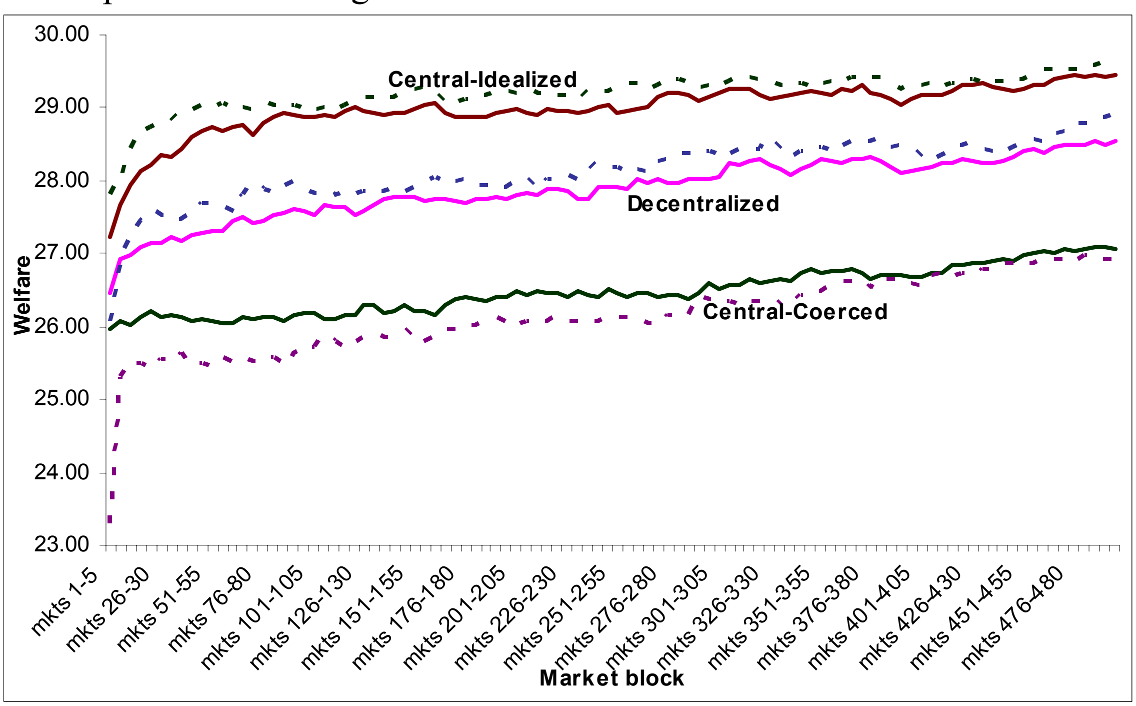
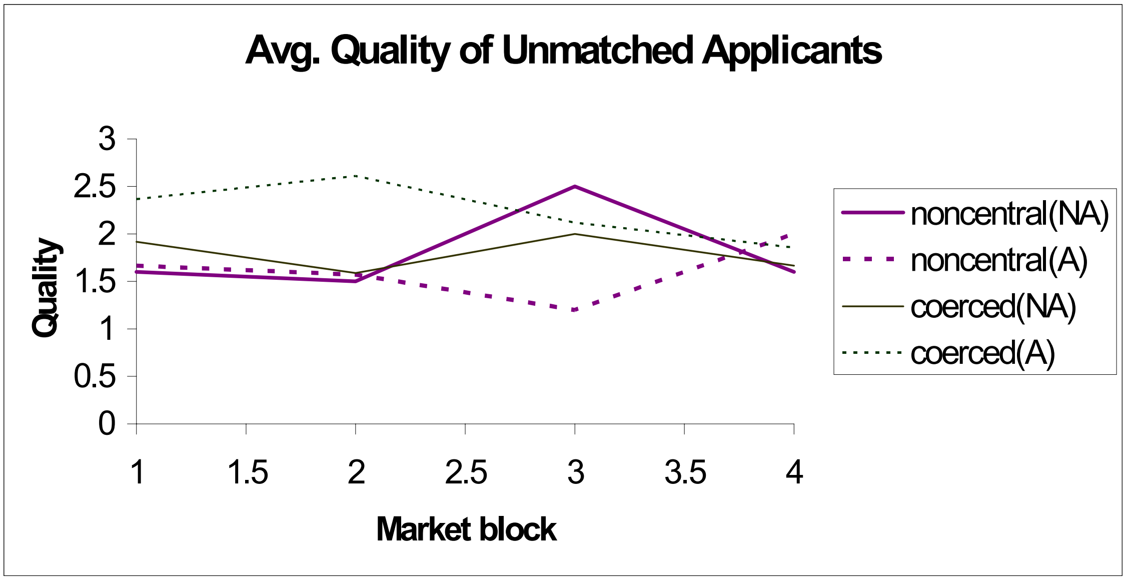
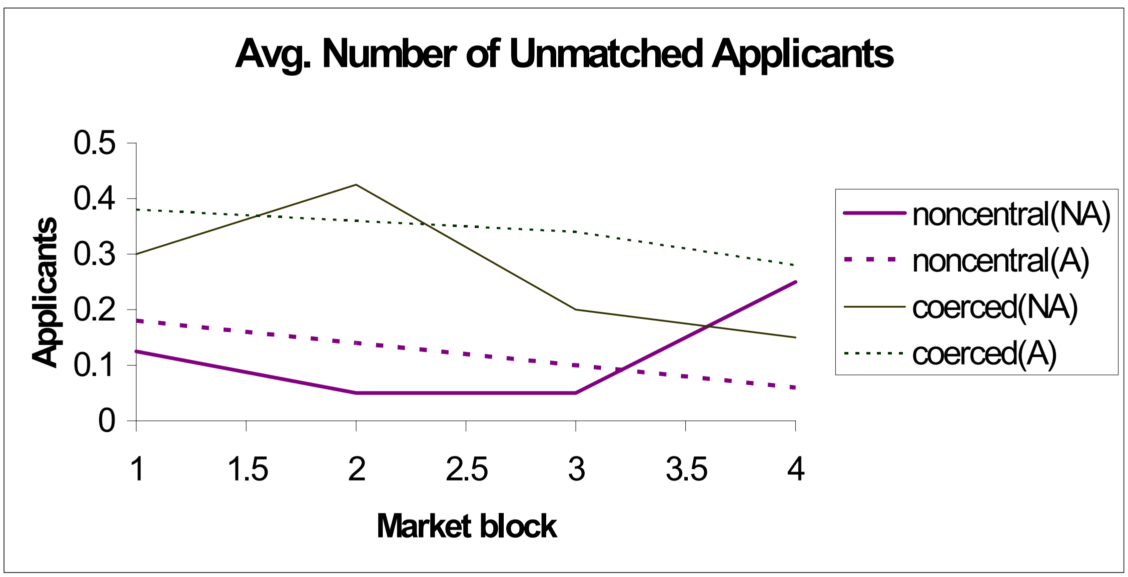
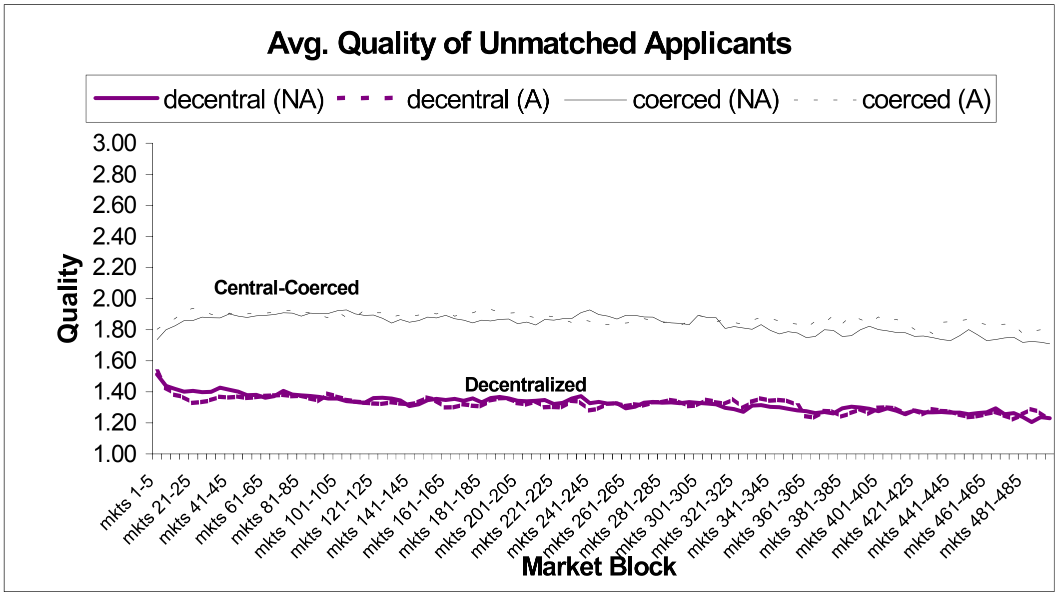


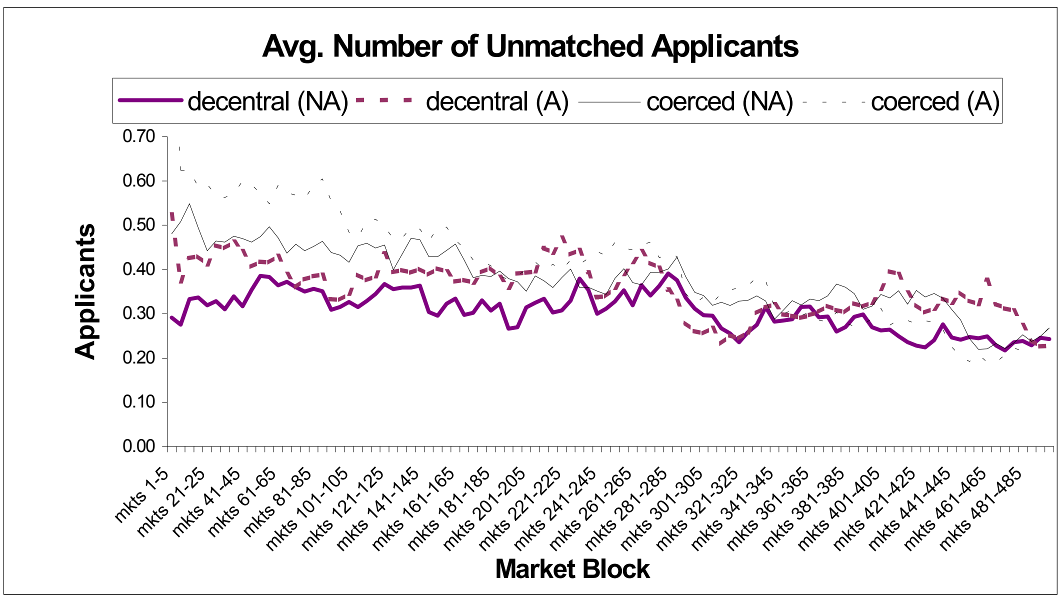
Figure 3. Number and Quality of Applicants Unmatched in the Legal Market.
a. Experimental Findings:





b. Computational Findings:





Appendix: Genetic Algorithm and Representation of Strategy Strings

The outline of one simulation session can be stated as follows:

Algorithm: Generate the initial population of strategies for each pool (that has a population of s strategies at a time) that will cause an outcome in matching behavior similar to the experimental subjects in market 1.

For g=1…G, the total number of generations, run the following algorithm for the existing set of strategies for each pool.

- Make a **tournament** of T matching games by randomly choosing strategies i=1,…,s from each pool k=1,…,5. The existing 5 types are firm 1, firm 2, firm 3, firm 4 and workers. Determine the reinforcement (or fitness) of each strategy as the average payoff that it brought to the players who adopted it in the tournament.
- For i=1,..,h, **select** the i'th **highest fitness** strategy of each type k to the next generation offspring. Return these strategies to the population pool for crossover.
- For i=1,…,(s-h)/2, **crossover** 2 parents for the (2i-1)'th and (2i)'th spots in the offspring generation for each type k=1,…,5 using the following technique:
  - Use **tournament selection** to determine two parents P^k_{2i-1} and P^k_{2i}: Choose four parent candidates C_1, C_2, C_3, C_4 for type k randomly using the discrete uniform density. The higher fitness strategy of C_1, C_2 and C_3, C_4 become the two parents P^k_{2i-1}, P^k_{2i} for type k.
  - With probability p, crossover the parents, with probability 1-p directly copy the parents as the offspring using single point **linear crossover**.
    - If crossover is adopted, randomly draw a crossover digit, c in $\{1,2,..,l^k-1\}$, in the strategy string of the size l^k.
    - Otherwise set c=0. Copy the digits 1,..,c of P^k_{2i-1} and c + 1 ,.., l^k digits of P^k_{2i} to form the child O^k_{2i-1}, copy the digits 1,..,c of P^k_{2i} and c+1,..,l^k digits of P^k_{2i-1} to form the child O^k_{2i}.
- For i=1,..,s, **mutate** each decision variable d=1,…,l^k in the offspring strategy O^k_i of each type t with probability $q = (1 - g / G)\, p^m_{max} + (g / G)\, p^m_{min}$ where g is the current generation number. Let $O^k_i(d)$ be the current decision variable.

- If mutation is adopted, randomly draw an integer *x* in $\{r_1,\ldots,r_2\}$, the range of the current decision variable $O^k_i(d)$, and replace it with *x*.
- If mutation is not adopted, directly copy the existing digit.

- The strategies for generation g+1 are the offspring of generation g. ◊

The artificial adaptive agents are constructed to choose among strategies represented by strings of decision variables. The strategies are conditioned on the rank of players as well as the current information available in each year. The applicants are ex-ante identical, so they use the same pool of strategies. The firms have different ex-ante qualities; therefore firms of different types consider different pools of strategies. Therefore there are 5 pools of strategies. The strategies are coded using **integer coding**.[34] We use a bounded rational representation for the strategies.

In the law clerk market simulations, a firm strategy is represented as a string of 6 decision variables:

$$T\text{- }A^1\text{-}R^1\text{-}A^2\text{-}R^2\text{-}R^3$$

Variable T is an integer in {1,2,3}. This decision variable is the year when the firm is going to start accepting applications from applicants. This is automatically set to 1 in the treatments without announcements in all generations. In the treatments with announcements, T is chosen from its full domain and evolves over time. A^t is in {0,1}. When A^t=1, the firm may hire an applicant in year t. When A^t=0, the firm will not hire an applicant in year t. A firm may announce admitting applications in a period, but this firm does not have to hire an applicant in the same period. This is the reason why we use two variables T and A^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 rank of the applicant that the firm is going to hire,[35] in the case A^t=1. When the applicants have lower ranks than R^t, simply the firm does not hire anybody in that period. Otherwise, it

[34] Each decision variable is represented by an integer.
[35] R^t is the rank of least acceptable applicant among the available ones.

hires the best applicant. In treatments with announcements, depending on the values of T some values of A^t and R^t may not be used.

An applicant strategy is a string of 20 decision variables:

$$S^1_1\text{-}N^1_1\text{-}S^1_2\text{-}N^1_2\text{-}S^1_3\text{-}N^1_3\text{-}S^1_4\text{-}N^1_4 \text{ - } S^2_1\text{-}N^2_1\text{-}S^2_2\text{-}N^2_2\text{-}S^2_3\text{-}N^2_3\text{-}S^2_4\text{-}N^2_4\text{- } N^3_1\text{-}N^3_2\text{-}N^3_3\text{-}N^3_4$$

Variable S^t_r is in {0,1}. When applicant is ranked r^{th} among the others, if S^t_r=1, she sends at least 1 application in year t; otherwise if S^t_r=0 she does not send any applications in year t. (S^3_r=1 is automatically set, so it is not a decision variable.) N^t_r is in {1,2,3,4} and denotes the number of firms that she will send an application in year t when she is ranked r at year t[36] and S^t_r=1. If none of these firms are available, she sends an application to the best available firm.[37]

The bounded rationality feature of the strategy representations comes from one source. We do not model all information sets using these representations. For example, subjects observe who already got matched and left the market before each period and they also observe actual grades of applicants not only their current rankings. As we find in our results, the current representations model subject learning pretty well in the experiment even though they are bounded rational.

Appendix: Sensitivity Analysis

Simulation Parameters

In this section we report results of two sets of sensitivity tests for the artificial adaptive agent simulations. These tests aim to see how much the results achieved through simulations depend on the choice of genetic algorithm parameters.

In the first set of tests, we conduct comparative static exercises by changing one parameter at a time. In three tests of this first set, we change number of simulation markets from 500 to 5000, we change number of simulations from 20 to 100 and we

[36] If none of these best N^t_r firms are available, she only sends an application to the best available firm.
[37] To keep the information sets simple, in the computational simulations ties are broken arbitrarily in every period, so there are never two students with the same rank.

change number of tournament games from 1000 to 10000 one by one. In each of the tests we measure mean welfare of applicants in all treatments, welfare difference between centralized-idealized and decentralized treatments, welfare difference between centralized-idealized and centralized-coerced treatments, welfare difference between decentralized and centralized-coerced treatments, and welfare difference between announcement and no-announcement treatments in each of the three market designs. We take the average of these in last 50 markets and report in Table 5 as well as the results for the original simulations reported earlier. We observe that average welfare across all treatments is almost the same in every exercise.

We observe that welfare in decentralized markets is catching up with the welfare in centralized-idealized markets in the 5000 generation treatment, although the latter is always higher. Moreover decentralized markets continue to raise more welfare than the centralized-coerced markets. The differences between announcement and no-announcement sessions are usually robust. In all cases announcements increase welfare slightly except the centralized-coerced treatment. In the original experiments for this treatment, announcement causes less welfare but this is not significant.

In the second set of tests, we conduct active nonlinear tests (ANTs) to obtain a multivariate sensitivity analysis. This is a technique found by Miller (1998). An ANT is a hill climbing optimization procedure, which tries to maximize some objective in the simulation by randomly searching in the parameter space. We use 100 iterations for the optimization procedure. Using this technique, the worst case scenarios for simulation models can be easily determined. The results show how much the results obtained in the original simulations depend on the choice of parameters and how much at worst results will be distorted if different parameters are chosen. We form the search space of deviations from the original parameters as {-50%, -40%, -30%. -20%, -10%, 0%, +10%, +20%, +30%, +40%, +50%} for the parameters, which have real number values. These parameters are the ratio of selected best strategies for the next market under selection pressure, crossover probability, initial mutation probability, final mutation probability, strategy population. We define the search space of other parameters as selection pressure and no-selection pressure operator; random generation of initial strategies and forcing initial worker generation to submit applications in each period (i.e. using strategies similar to initial human subject strategies). We try to maximize and minimize the welfare measures used in the comparative static exercises. The results are displayed in Table 6. We observe that the welfare differences between market types are very robust and the choice of parameters does not affect the fact that highest welfare is raised under centralized-idealized markets, followed by decentralized markets, followed by centralized-coerced markets. Mean welfare across all treatments is close to maximum under the original sessions. However the lack of selection under pressure accompanied by a combination of other parameters can decrease mean welfare substantially. This is due to substantial decrease in welfare for the centralized-coerced treatment with announcements.

Sensitivity Analysis on Experimental Design Parameters

In this section, we conduct robustness analysis by changing experimental design parameters. With different experimental designs, we run additional simulations. We impose two types of changes on payoff structure and on grade generation process for applicants. We choose new payoffs so that the maximum possible welfare of applicants is the same as the original experiments and they decrease the marginal utility of match quality for type j: in the original design this marginal utility is j, in subsequent comparative static exercises we have 3j/5 and 3j/7.[38] For different grade generation processes, we uniformly draw the grades of students from {0,1,…,5} and {0,1,…,10} instead of {0,1,2}. These decrease the probability of ties. The welfare measures discussed in the previous section are also calculated for these new exercises. The results are displayed in Table 7 for the last 50 markets of the simulations. We observe that with decreasing marginal utility of match quality, mean welfare of treatments increases, and particularly the welfare differences between treatments decrease. With decreasing probability of ties among student grades, the results are not substantially affected across different market designs. However announcements become slightly more effective and quicker in raising welfare.

[38] In the original design, utility of quality j agent from a quality i match is ij, in the subsequent exercises it is 3(ij+5)/5 and 3(ij+10)/7 respectively.

Table 5. Robustness Analysis for Simulations: Comparative Statics on Simulation Length, Simulation Number, and Tournament Length

Comparative Statics	Original Simulations	With 5000 generations	With 100 simulations	With 10000 games per tournament
Mean welfare of applicants in all treatments	**28.35 (0.0628)**[39]	28.77	28.38	28.56
Welfare diff. between central-idealized and decentral.	**0.87 (0.0495)**	0.26	0.78	0.69
Welfare diff. between central-idealized and central-coerced	**2.48 (0.0646)**	2.25	2.45	2.48
Welfare diff. between decentral. and central-coerced	**1.61 (0.0912)**	1.99	1.67	1.79
Welfare diff. between central-idealized A and NA	**0.16 (0.0541)**	0.07	0.17	0.11
Welfare diff. between decentral. A and NA	**0.25 (0.0969)**	0.14	0.27	0.28
Welfare diff. between central-coerced A and NA	**-0.12 (0.0603)**	0.17	0.01	0.07

Table 6. Robustness Analysis for Simulations: Multivariate Active Nonlinear Tests (ANTs) on other GA parameters

ANTs	Max.	Min.
Mean welfare of applicants in all treatments	28.81	23.23
Welfare diff. between central-idealized and decentral.	2.77	0.25
Welfare diff. between central-idealized and central-coerced	10.70	1.99
Welfare diff. between decentral. and central-coerced	7.93	0.98
Welfare diff. between central-idealized A and NA	0.37	-0.17
Welfare diff. between decentral. A and NA	0.25	-0.03
Welfare diff. between central-coerced A and NA	0.42	-6.39

39 The numbers in parentheses are standard errors of the benchmark simulations.

Table 7. Robustness Analysis for Experimental Design Through Simulations: Comparative Statics on Marginal Utility of Match Quality and Grade Generation of Students

Comparative Statics	Marginal Utility of Productivity for type j=3/5j	Marginal Utility of Productivity for type j=3/7j	Grades of applicants are uniformly drawn from {0,1,…,5}	Grades of applicants are uniformly drawn from {0,1,…,10}
Mean welfare of applicants in all treatments	29.01	29.29	28.41	28.40
Welfare diff. between central-idealized and decentral.	0.52	0.37	0.86	0.75
Welfare diff. between central-idealized and central-coerced	1.49	1.06	2.52	2.31
Welfare diff. between decentral. and central-coerced	0.97	0.69	1.66	1.55
Welfare diff. between central-idealized A and NA	0.09	0.07	0.12	0.17
Welfare diff. between decentral. A and NA	0.15	0.11	0.17	0.45
Welfare diff. between central-coerced A and NA	-0.07	-0.05	-0.07	0.09