The Dynamics of Law Clerk Matching:
An Experimental and Computational Investigation of Proposals for Reform of the Market

Ernan Haruvya\textsuperscript{a} Alvin E. Roth\textsuperscript{b} and M. Utku Ünver\textsuperscript{c}

This draft: 9 November 2001

\textsuperscript{a}University of Texas at Dallas

\textsuperscript{b}Harvard Business School and Department of Economics, Harvard University

\textsuperscript{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. In an attempt to stop the further unraveling of appointment dates, a reform that has been implemented for the 2001 market is a web based public database of judges' hiring plans. Another reform that has been debated is a centralized clearinghouse modeled on the medical match. The present paper explores both these potential reforms, experimentally in the laboratory, and computationally using adaptive agents learning through genetic algorithms. 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 potential obstacles to the success of both of these reforms.

\textsuperscript{*} 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 authored by Urs Fischbacher at the University of Zurich, Institute for Empirical Research in Economics.
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. 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 of a student’s second year of law school, only first year grades are available. 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. Roth and Xing 1994 discuss several dozen markets and submarkets that have at one point

---

1 See Dorf (2000) for a law school dean’s perspective.
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.
in their history experienced unraveling of appointment dates. 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.

Some of these markets have addressed this problem of early matching by reorganizing themselves as centralized clearinghouses. The largest of these that 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. 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 medical clearinghouse, and the Roth-Peranson design has since been adopted by a number of other labor markets.

Consequently, one of the reforms that has been 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 discussed below, may present serious obstacles to the successful implementation of a centralized clearinghouse in this market.

Another, more modest reform, already implemented for 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

---

3 And see Avery and Zeckhauser 2001 for a discussion of 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), and Suen (2000).
5 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.
applications. These announcements, one might conjecture, would eliminate judges’ ability to blindside their peers and thereby foster better cooperation. Below is the page listing, as of August 15, 2001, the already announced clerk positions for the Ninth Circuit Court of Appeals to begin in September, 2003. Some of the early hiring dates already announced suggest that the unraveling of the market may not be slowed by public announcement of hiring dates.

---

6 https://lawclerks.ao.uscourts.gov
Ninth Circuit Court Announced Positions for Sept, 2003, as of August, 2001

<table>
<thead>
<tr>
<th>Status and Type of Position</th>
<th>Application Dates</th>
<th>Dates of Service</th>
<th>Ninth Circuit Court, Work Location, Judge Type, Judge Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Available Term</td>
<td>11/01/2001 until filled</td>
<td>08/07/2003 to 08/31/2004</td>
<td>United States Court of Appeals, San Francisco, CA Senior Circuit Judge: James R. Browning</td>
</tr>
<tr>
<td>Available Term</td>
<td>09/01/2001 until filled</td>
<td>09/01/2003 to 09/01/2004</td>
<td>United States Court of Appeals, Seattle, WA Senior Circuit Judge: Betty B. Fletcher</td>
</tr>
<tr>
<td>Available Term</td>
<td>09/01/2001 until filled</td>
<td>09/01/2003 to 09/01/2004</td>
<td>United States Court of Appeals, Seattle, WA Senior Circuit Judge: Betty B. Fletcher</td>
</tr>
<tr>
<td>Available Term</td>
<td>09/01/2001 until filled</td>
<td>09/01/2003 to 09/01/2004</td>
<td>United States Court of Appeals, Seattle, WA Senior Circuit Judge: Betty B. Fletcher</td>
</tr>
<tr>
<td>Available Term</td>
<td>11/01/2001 until filled</td>
<td>08/15/2003 to 08/14/2004</td>
<td>United States Court of Appeals, Pasadena, CA Senior Circuit Judge: Alfred T. Goodwin</td>
</tr>
<tr>
<td>Available Term</td>
<td>11/01/2001 until filled</td>
<td>08/15/2003 to 08/14/2004</td>
<td>United States Court of Appeals, Pasadena, CA Senior Circuit Judge: Alfred T. Goodwin</td>
</tr>
<tr>
<td>Available Term</td>
<td>open until filled</td>
<td>09/30/2003 to 08/30/2004</td>
<td>United States Court of Appeals, Seattle, WA Circuit Judge: Ronald M. Gould</td>
</tr>
<tr>
<td>Available Term</td>
<td>10/05/2001 until filled</td>
<td>08/30/2003 to 08/30/2004</td>
<td>United States Court of Appeals, Seattle, WA Circuit Judge: Ronald M. Gould</td>
</tr>
<tr>
<td>Available Term</td>
<td>10/05/2001 until filled</td>
<td>08/30/2003 to 08/30/2004</td>
<td>United States Court of Appeals, Seattle, WA Circuit Judge: Ronald M. Gould</td>
</tr>
<tr>
<td>Available Term</td>
<td>10/05/2001 until filled</td>
<td>08/30/2003 to 08/30/2004</td>
<td>United States Court of Appeals, Seattle, WA Circuit Judge: Ronald M. Gould</td>
</tr>
<tr>
<td>Available Term</td>
<td>open until filled</td>
<td>09/10/2003 to 08/30/2004</td>
<td>United States Court of Appeals, Pasadena, CA Senior Circuit Judge: Cynthia H. Hall</td>
</tr>
<tr>
<td>Available Term</td>
<td>10/01/2001 until filled</td>
<td>08/30/2004 to 08/30/2004</td>
<td>United States Court of Appeals, San Diego, CA Senior Circuit Judge: Margaret M. McKeehan</td>
</tr>
<tr>
<td>Available Term</td>
<td>12/01/2001 to 02/01/2002</td>
<td>09/01/2003 to 08/30/2004</td>
<td>United States Court of Appeals, Phoenix, AZ Circuit Judge: Michael Daly Hawkins</td>
</tr>
<tr>
<td>Available Term</td>
<td>05/01/2002 to 09/01/2002</td>
<td>09/01/2003 to 08/31/2004</td>
<td>United States Court of Appeals, Seattle, WA Circuit Judge: Richard C. Tallman</td>
</tr>
<tr>
<td>Available Term</td>
<td>09/01/2001 to 09/30/2003</td>
<td>09/01/2003 to 08/31/2004</td>
<td>United States Court of Appeals, Fairbanks, AK Circuit Judge: Andrew J. Kleinfield</td>
</tr>
<tr>
<td>Available Term</td>
<td>09/01/2001 to 09/30/2003</td>
<td>09/01/2003 to 08/31/2004</td>
<td>United States Court of Appeals, Fairbanks, AK Circuit Judge: Andrew J. Kleinfield</td>
</tr>
<tr>
<td>Available Term</td>
<td>09/01/2001 to 09/30/2003</td>
<td>09/01/2003 to 08/31/2004</td>
<td>United States Court of Appeals, Fairbanks, AK Circuit Judge: Andrew J. Kleinfield</td>
</tr>
</tbody>
</table>
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 do very much longer experiments than are in fact feasible in the laboratory. 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.

2. Background

2.1. Description of the Law Clerk Market (focusing on differences from the medical 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;

---

7 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.
• 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, in order to avoid receiving an early offer from a less preferred judge. That is, rather than apply widely to judges for whom they would be willing to work, many students avoid applying to desirable judges who are not their top choices, because the obligation to accept an offer if one is made would then preclude waiting and seeing if a more preferred offer would be forthcoming.

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 fact of the market when scheduling their interviews—are features that are much more pronounced in this market than in others that we know of. Avery et al. report that 73% of the students in their year 2000 sample accepted the first offer they received, and that a majority of students whose first offer was not their first choice from among those for which they had interviews nevertheless accepted their first offer. This largely had to do with the perceived need to respond quickly, rather than 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.

---

8 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
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. In particular, both of 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 in which those mechanisms had been employed. 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.

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, 2001).
It seems likely that in the law clerk market, however, the major potential source of inefficiency stems from the lack of information about applicants’ true qualities in early periods.\footnote{Judges lose little planning flexibility by hiring clerks early, since the number of positions they have is the...}

In the experiments to be described below, we model the information about students as developing over time, as their grades become available. It will therefore be useful to begin by seeing if the experimental results for the medical market can be replicated in the information/efficiency environment we use to model the law clerk market.

### 3. Experimental Design

Two sets of experiments were run. The first set, motivated by the medical environment in which applicants are completely free to turn down early offers, was run to examine unraveling in an experimental matching market with slowly revealed information on applicants’ qualities. The second set of experiments, with the same information structure, introduces some of the special institutional features of the law clerk market. To make the efficiency issues clear, productivity of applicants and workers was modeled so that an efficient match would be strictly assortative, i.e. so that at an efficient match the most productive firm would be matched to the most productive worker, and so forth (see e.g. Becker, 1981).

In all of these experiments, 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.

#### 3.1. The “Medical Model” Matching Market Experiment

The medical match experiment involved 64 subjects, from eight cohorts of eight subjects each. The experiment was conducted in three sessions, with one session of two cohorts on 4/5/2001 and two sessions of three cohorts each on 4/8/2001. The complete instructions and data files are available at [www.utdallas.edu/~eharuvy/legalmatch](http://www.utdallas.edu/~eharuvy/legalmatch).

A typical session involved 16-24 people divided into two or three groups of eight, such that no subject knew who the members of his or her group were. Each group of eight
people was arbitrarily divided into four “firms” and four “applicants.” Each subject saw on the computer screen his or her role and was notified that this role would remain fixed for the duration of the experiment, which would last 40 “markets.”

A subject’s payoff in each market was a subject’s “quality” multiplied by the “quality” of the subject with whom he or she matched. For example, if a firm of quality 3 hired an applicant of quality 4, both firm and applicant would receive a payoff of 12 tokens each. (Tokens were exchanged for money at the end of the experiment at the rate of $0.10 per token.) Each subject in the firm role was assigned a quality of 1, 2, 3, or 4 and told that this quality would remain fixed for the duration of the experiment. Each firm’s quality was known by both firms and applicants from the beginning of the market. Subjects in the applicant role were not told their qualities prior to the start of each market, rather, these were determined as follows.

Each market lasted three “years.” In each year, applicants received a “grade” of 0, 1, or 2, with an equal probability for each grade. Subjects were told that the qualities would be assigned at the end of year 3, with the highest quality of 4 going to the subject with the highest cumulative grade, the quality of 3 going to the subject with the second highest cumulative grade, and so on, with the applicant with the worst cumulative grade getting a quality of 1. Ties were arbitrarily broken. Thus the efficient matching, i.e. the one that resulted in the largest payment from the experimenters to the subjects, is the one that matches each firm to the worker whose quality is the same as the firm’s. But this matching cannot in general be identified until the third year, when the students’ rankings are determined from their cumulative grades.

Each year, all subjects in a group were informed about the grades of all applicants in the group, as well as which applicant had been hired by which firm and in which year. In the first stage of each year, each firm could extend an offer to one and only one applicant. In the second stage, an applicant would view all offers and would have the option of accepting or rejecting the highest offer he or she received in that year. If that offer was accepted, the proposing firm as well as the accepting applicant would not be

---

same from year to year.

10 This differs from the experiment of Kagel and Roth (2000), in which a firm could not tell which applicants had accepted offers from other firms. That experiment was motivated by the experience of the British medical markets, in which early matches were against the rules and hence were made privately.
able to take any more actions in the market, though both would still be able to view the progression of the market, the same as active subjects.

Following the play of 20 markets of three years each, the subjects were given another set of instructions for the next 20 markets. The next 20 markets would differ from the first 20 markets on one aspect: in the third year, subjects not yet matched would be matched by the computer, which would produce the unique stable matching among the as yet unmatched firms. That is, the highest quality available firm would be matched with the highest quality available applicant and so on.

3.2. The Law Clerk Matching Market

The information environment in the law clerk experiments is the same as described above: judges have known qualities 1 to 4, and applicants get grades over years 1 to 3 that determine their final qualities (also from 1 to 4), and the payoff to each subject is the product of his own quality and that of the subject he is matched to.

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 (i.e., in the same ‘year’).  

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 treatment 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.

---

11 We thus model students’ “obligation to quickly accept an offer” more forgivingly than in the actual market, since we treat all offers made in the same year as simultaneous, and allow the applicant to accept the best of however many offers are received in that year. However, unlike in the medical market experiment, an applicant who receives one or more offers cannot reject them all.
As a practical matter, having judges make announcements, and having students make applications, makes each experimental market take longer than the markets in the medical market experiment. So in each condition of this experiment subjects experienced only 20 markets, not 40. Therefore we examine each market institution separately, and do not mix decentralized and centralized markets as in the medical market experiment.

Finally, we explore the effect on a centralized match of the fact that students feel compelled to accept an offer if one is made. Since firms in general do not rank, in a central match, an applicant who has not applied for their position, the fact that students must accept an offer from a judge if one is made (see below for details of the experimental implementation) makes them vulnerable to being hired before the match 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 decline early offers 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 non-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 is coerced to accept a position before the match.

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

---

12 The number of 454, rather than 456, reflects the fact that there were two subjects out of 454 who were able to participate in two different treatments.
In the three “announcement” treatments, prior to the start of each market, firms announce the year (1, 2, or 3) in which they become available to receive applications.

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 firms.
2. Firms 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 firm also chooses that applicant.

In the centralized-idealized treatment, year 3:

Firms 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 firm with the highest quality is matched with the highest quality unemployed applicant. From the remaining unemployed applicant pool, the second highest quality unmatched firm is then matched with the highest quality applicant, and so on. (Notice that even in the centralized-idealized treatment when an applicant applies to a firm and receives an offer in years 1 and 2, she cannot decline this offer).
In the centralized-coerced treatment, year 3:

Firms 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 firm following year 2, an applicant needs to have sent an application to that firm in either year 1 or year 2. The unmatched firm with the highest quality was matched with the highest quality unemployed applicant that had sent this firm an application in either year 1 or year 2 or both. From the remaining unemployed applicant pool, the second highest quality unmatched firm was then matched with the highest quality applicant that had sent this firm 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 to become eligible to be matched, and may at that time be ‘coerced’ into an early match. Also notice that when an applicant applies to a firm and receives an offer in years 1 and 2, she cannot decline this offer.)

The treatments without announcement skip the announcement stage and all firms 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 consisted of four firms and four applicants. Each subject saw on the screen his or her role and was notified that this role would remain fixed for the duration of the experiment, which would last 20 “markets.”

In this experiment, firms 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. Artificial Adaptive Agent Simulations

We apply computational artificial adaptive agent simulations to model the adaptive behavior of subjects in the experiments. Using these simulations, we can simulate the short run and long run learning behavior of subjects. 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.

We use genetic algorithms for adaptive agents, instead of the simple learning models that have recently been shown to have some ability to track and predict human behavior in experiments with simple strategic environments. The reason is that, in this relatively complex environment, the sets of strategies available to the agents are quite large. Simple learning models have been used with some success to predict how agents will learn among a pre-specified (and small) set of strategies. Genetic algorithms, in contrast, incorporate the ability to generate and evaluate new strategies, and are thus a compact way of exploring learning in a large strategy space. They are shown to be efficient at the tradeoff between discovering new strategies and utilizing 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).

---

14 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.
15 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.”
We can compare the behavioral dynamics of human subjects and artificial agents in complex adaptive systems. If we can find similarities between the two then we can use the computational techniques to test the robustness of behavior over longer time horizons.\textsuperscript{16}

A strategy for an agent in a genetic algorithm is represented by a string of integers. For example, in the medical-matching market, a firm strategy is represented as a string of 5 decision variables:

\[ S^1 - R^1 - S^2 - R^2 - R^3 \]

\( S^t \) is in \{0,1\}. When \( S^t = 1 \), the firm makes an offer to an applicant in year \( t \). When \( S^t = 0 \), the firm does not make any offers in year \( t \). \( R^t \) is in \{1,2,3,4\}. This decision variable is the rank of the applicant (in year \( t \)) to whom the firm is going to make an offer, when \( S^t = 1 \). \( S^3 \) is automatically set to 1 at the beginning of the simulations, so it is not a decision variable, i.e. firms that are still unmatched in year 3 will always make an offer.\textsuperscript{17}

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 10,000 games. 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

\textsuperscript{16} Andreoni and Miller (1995) use genetic algorithms in auction environments and conclude that the adjustment process is similar for both humans and artificial agents. In the context of labor markets, 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 (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 Bullard and Duffy (1998) for a usage of genetic algorithms to model macroeconomic adaptive learning.

\textsuperscript{17} The full strategy representations for workers and firms in the medical and legal market are given in the appendix.
strategies by combination of their strategies. (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. 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.

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 based on relative fitness (“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:

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

\[ \quad \rightarrow \text{Crossover starting at 3rd digit} \rightarrow \]

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

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

---

18 Each computer session is started with a different random number seed.
19 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
Mutation Example:
\[ a_1-a_2-a_3-a_4 \rightarrow \text{Mutate 2}^{\text{nd}} \text{ digit} \rightarrow \ a_1-x_2-a_3-a_4 \]

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.

5. Results of the Medical Market Experiment

We begin with the medical matching market to establish that the information structure studied here results in early hiring, which can be remedied by centralized matching, with welfare implications. Figure 1 shows hiring periods by firm quality for the two regimes of decentralized and centralized matching. The top panel shows the experimental data; the bottom panel shows the parallel computational results obtained with genetic algorithms. Whereas the bottom panel shows the reversal sharply, the top panel warrants a closer inspection.

In the top panel, representing hiring periods for the experimental sessions, the four curves correspond to the four firms of qualities 1 through 4. The dashed line separates the decentralized regime of markets 1-20 from the centralized regime of markets 21-40. The data are presented in blocks of 5 markets.

To see if the two regimes are different in hiring periods, we compare average hiring periods in the last block of each regime. Experimentally, average hiring period for 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.

20 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 \( p^m_{\text{max}}=0.07 \) to \( p^m_{\text{min}}=0.01 \). The initial pools of strategies for the simulations reported here were chosen to resemble the initial behavior of the human subjects in the first market of the experiment, so that the evolution of behavior could be compared to the experiment starting from the same initial conditions.
firm 1 in the last block of five markets in the decentralized regime is 2.62 as opposed to 2.95 in the last block of the centralized regime. Computationally, the long run results indicate that firm 1 hires on average in period 1.88 in the decentralized market and in period 2.78 in the centralized market. Experimentally, the pattern for firms 2 and 3 in the decentralized regime is somewhat flatter: Average hiring period for firm 2 in the last block of five markets in the decentralized regime is 2.29 as opposed to 2.48 in the last block of the centralized regime. Average hiring period for firm 3 in the last block of the decentralized regime is 2.20 as opposed to 2.50 in the last block of the centralized regime. Computationally, the pattern is even sharper for firms 2 and 3 in the long run: firm 2 hires in period 1.22 on average in the decentralized market and in period 2.59 in the centralized market. Firm 3 hires in the long run on average in period 1.18 in the decentralized market and in period 2.43 in the centralized market. Experimentally, for firm 4, the corresponding hiring periods are 1.53 in the decentralized regime and 1.90 in the centralized regime. Computationally, firm 4 hires on average in period 1.06 in the decentralized regime and in period 2.64 in the centralized regime in the long run.

In the experimental data, though the differences in hiring periods are always in the predicted direction, significance is only conclusive for firms 1 and 3. To determine significance, we compute the difference between the two regimes in the last block’s average hiring period for each cohort and each firm. We then test the hypothesis, separately for each firm, that this difference is zero. The one-sided t-test P-values—signed rank P-values in parentheses—for firms 1 through 4 are 0.020 (0.016), 0.193 (0.246), 0.005 (0.016), and 0.087 (0.125), respectively. The corresponding differences in the simulations are always significant. Therefore, we omit the statistical tests for the genetic algorithm.

In the decentralized regime of the experimental sessions, it is somewhat curious that experimental subjects in the roles of firm 4 appear to first move towards hiring later, before the unraveling begins toward hiring earlier. It appears that because subjects in this experiment are initially unfamiliar with this market structure, they have to first gain experience to understand that later hiring increases their information. Only then do they begin experiencing the incentives to hire earlier.
Figure 2 shows that the apparent later hiring results in a significant improvement in social welfare, as measured by the sum of earnings of applicants. Note that this sum of earnings is identical to the sum of earnings by firms. Hereafter, we refer to this sum as ‘welfare.’ Comparing the last block of five markets in each regime, we see average welfare at 26.63 in the decentralized regime and at 28.50 in the centralized regime. Treating each cohort as a single observation, the within-cohort-between-regime welfare difference (for the last block in each regime) is significantly different from zero by the paired comparison t-test (7 d.f.) with a t-statistic of 3.85 and a one-tail p-value of 0.0031. The value of the Wilcoxon Signed Rank Statistic is 17, yielding a one-sided p-value of 0.0053. The parallel artificial adaptive agent simulation findings give average welfare of 25.02 in the long run in the decentralized regime and 29.23 in the centralized regime.

However, note that not all firms benefit from this improvement, as the firms of quality 1 and 2 would often prefer the decentralized scheme. By Figure 3a, firm 1 and firm 2 hired, on average, lower quality applicants in the central scheme relative to the decentralized scheme in the experiment. As Figure 3b shows, this was also the case for both firm 1 and firm 2 under the genetic algorithm.

Since in the decentralized regime, not all applicants and firms were matched each market, we would like to know what proportion of the welfare gain from the decentralized to the centralized regime could be attributed to the elimination of non-matches. That is, if we were to match the unmatched to each other in the decentralized regime, what welfare gain would result? In the last block of five markets of the experiment, there were 12 failures to match (12 applicants and 12 firms) out of possible 160. The average quality of both unmatched applicants and unmatched firms in the last block of five markets was 1.5. Hence, for the last block comparison, for which the welfare difference was 1.875, roughly 0.17 of that difference was due to unmatched firms and applicants, while the rest is due to the improved efficiency of the matches when they are made later.

In the genetic algorithm, under the decentralized regime, there are 0.43 applicants (and firms) unmatched on average in each market. It should be noted that this number decreases from 1.91 in the first five decentralized markets to 0.10 in the last five decentralized markets. The average applicant welfare is 25.95 in the last block of the
decentralized regime. There is a gain of 3.33 over the last block of the decentralized market in all centralized markets on average. In the last block of the decentralized regime, average quality of unmatched firms is 1.63 and the average quality of unmatched workers is 2.15. Therefore, the welfare gain that is attributable to the diminishing number of unmatched agents is only 0.35, while a welfare gain of 2.98 is related to better and later hiring in the centralized regime. This observation is supported by Figure 1b and Figure 3b.

Thus this experiment reproduces the results of Kagel and Roth (2000) in this environment, in which efficient matching is assortative by quality, and quality is revealed incrementally.

6. Results of the Law Clerk Market Experiment:

In the medical market experiments, applicants were not asked to send applications — the interpretation is that applicants automatically apply to all positions in which they are interested. Recall that, in the legal market, this assumption can no longer be made. The legal market is different from the medical market in that applicants may not be able to refuse an offer extended to them. Hence strategic interviewing and selective application submission are crucial and explicitly captured in the law clerk market experiment. We turn next to the effect of these institutional differences.

6.1. The Effect of Centralized Matching

Figure 4 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. For significance determination, each cohort’s average welfare over the last five markets was treated as a single observation. 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 Wilcoxon Rank Sums Z-statistic is -1.84, with a one-tail P-value of 0.033. Computationally, by the genetic algorithm, the increase in welfare is even more prominent with an improvement by 1.06 on average over all 500 markets, 0.69 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 4 illustrates. This decline in welfare from decentralized to centralized-coerced can be most easily understood by noticing, in Figure 5 that very few matches are in fact consummated by the centralized-coerced match. In the coercive environment, almost all matches are made before the centralized match. In our experimental environment, this means that the matches are mostly made in periods 1 and 2, as illustrated in Table 1. (The welfare effects of matching before the centralized match would be different in an experimental environment that allowed these early matches to happen later, when more information is available, so there is no strong reason to believe that centralized matching in a coercive environment always produces lower welfare than the decentralized match. The notable feature is that, in the coercive environment, centralized matching is not in fact used, but serves only to force most matching earlier than the centralized match.) Interestingly, as Figure 5 shows, the number of applicants entering the central match is increasing over time under the centralized-idealized treatments, improving welfare as a result. This is the opposite of the trend seen in the centralized-coerced treatment, in which the number going through the match is decreasing. The computational results using the genetic algorithm, also in Figure 5, show that whereas in the idealized centralized match, 3 out of 4 firms/applicants go through the match after 500 repetitions, virtually none go through the “coerced” match after 500 trials. 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-enhancing effects that would result from the introduction of a centralized match in a non-coercive environment.

---

21 The welfare drop also results at least partially from the increase in both the number of unmatched applicants and the average quality of unmatched applicants under the centralized-coerced treatment relative to the decentralized cases (although in all cases the rate of unmatched applicants is low). Recall that in the centralized-idealized condition, no possibility of unmatched applicants or firms exists.
6.2. The Effect of Announcements

Public announcements by judges of when they will begin accepting applications has been proposed as a possible solution to unraveling. Our findings are that while such announcements may in fact result in later hiring in decentralized and idealized market settings (see table 3), their welfare effects are small and gradually disappear over time. In the centralized-coerced matching setting, in which applicants are must interview prior to the centralized match, and may be coerced to accept an early offer, we find that announcements clearly do not improve welfare nor result in later hiring.

Figure 4 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 central-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 Wilcoxon Rank Sums P-value is 0.377. In the central-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 Wilcoxon 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 central-idealized treatment.

In the experimental data, announcements did not significantly delay hiring for any firm by either the T-test or Wilcoxon 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 applicants, once firms 3 and 4 make later entry announcements. Table 2 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

---

22 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.
corresponding Wilcoxon Rank Sums had a P-value of 0.238). Notwithstanding the lack of significance, welfare appears adversely affected by announcement in the central-coerced treatments, an observation confirmed by the genetic algorithm. Figure 6 suggests that this adverse effect is at least partly due to a higher number and quality of unmatched applicants. Table 2 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 in the law-clerk market experiment 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 19% to 60% 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 central-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, it is clear that:

- Decentralized matching leads to early hiring in the information/payoff environment explored here, both in the medical and in the legal environments.
- Centralized matching leads to later hiring and improvements of welfare in both the medical environment, and in the idealized, non-coercive legal environment.
- In the coercive legal 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.

8. Conclusions

Unraveling of hiring periods in the matching of clerks to judges has become a serious problem, resulting in aggravation to judges, applicants, and law professors. Such discontent is grounded in part in a feeling that efficiency is harmed. Although efficiency losses are hard to gauge in the actual law clerk market, they are straightforward to measure in a stylized experimental environment. This paper pursued this route.

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. 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
some grounds to speculate on the effects of proposed changes in the organization of the law clerk market.\textsuperscript{23}

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 feeling among many law students that they faced an obligation 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.

As a benchmark, we ran experimental sessions of the medical market setting and found that in our medical market experimental environment, centralized matching was effective in reversing the unraveling and in improving welfare. We then found that to a lesser extent it was also effective in an idealized legal market environment, in which applicants could avoid early offers with the attendant obligation to accept. However, once we added more realistic features—specifically, judges would not rank applicants whose applications had not been received, but that applicants might not be able to resist accepting an immediate offer from a judge to whom they had applied—centralized matching was not successful, and almost all matches were arranged before the centralized match. (So, depending on the timing of the match, this could in fact 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 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, are likely succeed in reversing the unraveling of this market. 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

\textsuperscript{23} 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
may be necessary to change the culture of the market in which law students feel compelled to accept the first offer they receive.24
References


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

<table>
<thead>
<tr>
<th></th>
<th>Experimental – average of 20 mkts</th>
<th>Computational—average of 500 mkts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Decentral. NA/A</td>
<td>Central-idealized NA/A</td>
</tr>
<tr>
<td>Firm 1</td>
<td>2.30 / 2.31</td>
<td>2.68 / 2.58</td>
</tr>
<tr>
<td>Firm 2</td>
<td>1.90 / 1.73</td>
<td>2.03 / 2.12</td>
</tr>
<tr>
<td>Firm 3</td>
<td>1.75 / 2.06</td>
<td>1.84 / 2.06</td>
</tr>
<tr>
<td>Firm 4</td>
<td>2.04 / 2.17</td>
<td>1.82 / 2.16</td>
</tr>
</tbody>
</table>

Table 2. Average Quality of Hired Applicant by Firm in the Legal Market

<table>
<thead>
<tr>
<th></th>
<th>Experimental – average of 20 mkts</th>
<th>Computational—average of 500 mkts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Decentral. NA/A</td>
<td>Central-idealized NA/A</td>
</tr>
<tr>
<td>Firm 1</td>
<td>1.71 / 1.66</td>
<td>1.57 / 1.53</td>
</tr>
<tr>
<td>Firm 2</td>
<td>2.24 / 2.34</td>
<td>2.28 / 2.37</td>
</tr>
<tr>
<td>Firm 3</td>
<td>2.75 / 2.67</td>
<td>2.91 / 2.86</td>
</tr>
</tbody>
</table>

Table 3. Average Entry Period in Announcement Treatments in the Legal Market

<table>
<thead>
<tr>
<th></th>
<th>Experimental – average of 20 mkts</th>
<th>Computational—average of 500 mkts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Decentral.</td>
<td>Central-idealized</td>
</tr>
<tr>
<td>Firm 1</td>
<td>1.15</td>
<td>1.35</td>
</tr>
<tr>
<td>Firm 2</td>
<td>1.13</td>
<td>1.21</td>
</tr>
<tr>
<td>Firm 3</td>
<td>1.40</td>
<td>1.37</td>
</tr>
<tr>
<td>Firm 4</td>
<td>1.47</td>
<td>1.56</td>
</tr>
</tbody>
</table>
Figure 1. Average Hiring Period by Firm for Decentralized and Centralized Matching in the Medical Market. For experiments markets 1-20 are in a decentralized framework; markets 21-40 have centralized matching in year 3. For computations markets 1-100 are in a decentralized framework, markets 101-200 are in a centralized framework.

a. Experimental Findings:

b. Computational Findings:
Figure 2. Average Group Welfare as Measured by the Sum of Quality Products of Matched Pairs in the medical market. Average is taken over groups and markets in a block of five markets. Markets 1-20 are in a decentralized framework; markets 21-40 have centralized matching in year 3. In the bottom graph, the computational results, the regime change occurs after 100 markets.

a. Experimental Findings:

b. Computational Findings:
Figure 3. Average Quality of Applicants Hired by Firms in the Decentralized and Centralized Medical Matching Market. For experiments, markets 1-20 are in a decentralized framework; markets 21-40 have centralized matching in year 3. For computations markets 1-100 are in a decentralized framework, markets 101-200 are in a centralized framework.

a. Experimental Findings:

![Quality of Hired Applicants](image)

b. Computational Findings:

![Quality of Hired Applicants](image)
Figure 4: 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 5. 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 6. Number and Quality of Applicants Unmatched in the Legal Market.
a. Experimental Findings:
b. Computational Findings:

![Graph of Avg. Quality of Unmatched Applicants]

![Graph of Avg. Number of Unmatched Applicants]
Appendix: Genetic Algorithm and Representation of Strategy Strings

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

1. 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.

2. For $g=1\ldots G$, the total number of generations, run the algorithm for the existing set of strategies for each pool.

2.1. Make a **tournament** of $T$ matching games by randomly choosing strategies $i=1\ldots s$ from each pool $k=1\ldots 5$. 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.

2.2. For $i=1\ldots h$, **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.

2.3. For $i=1,\ldots,(s-h)/2$, **crossover** 2 parents for the $(2i-1)$'th and $(2i)$'th spots in the offspring generation for each type $k=1\ldots 5$ using the following technique:

   2.3.1. Use **tournament selection** to determine two parents $P_{2i-1}^k$ and $P_{2i}^k$:
   
   2.3.1.1. Choose four parent candidates $C_1$, $C_2$, $C_3$, $C_4$ for type $k$ randomly using the discrete uniform density.
   
   2.3.1.2. The higher fitness strategy of $C_1$, $C_2$ and $C_3$, $C_4$ become the two parents $P_{2i-1}^k$, $P_{2i}^k$ for type $k$.

   2.3.2. With probability $p$, crossover the parents, with probability $1-p$ directly copy the parents as the offspring using single point **linear crossover**.

   2.3.2.1. If crossover is adopted, randomly draw a crossover digit, $c$ in $\{1,2,\ldots,l^k-1\}$, in the strategy string of the size $l^k$. Otherwise set $c=0$.

   2.3.2.2. Copy the digits $1,\ldots,c$ of $P_{2i-1}^k$ and $c+1,\ldots,l^k$ digits of $P_{2i}^k$ to form the child $O_{2i-1}^k$.

   copy the digits $1,\ldots,c$ of $P_{2i}^k$ and $c+1,\ldots,l^k$ digits of $P_{2i-1}^k$ to form the child $O_{2i}^k$.

2.4. For $i=1\ldots s$, **mutate** each decision variable $d=1\ldots l^k$ in the offspring strategy $O_i^k$ of each type $t$ with probability $q = (1 - g / G) p^m_{\text{max}} + (g / G) p^m_{\text{min}}$ where $g$ is the current generation number. Let $O_i^k(d)$ be the current decision variable.
2.4.1. If mutation is adopted, randomly draw an integer \( x \) in \( \{r_1, \ldots, r_2\} \), the range of the current decision variable \( O_i^k(d) \), and replace it with \( x \).

2.4.2. If mutation is not adopted, directly copy the existing digit.

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.

In the medical-matching market, a firm strategy is represented as a string of 5 decision variables:

\[
S^1 - R^1 - S^2 - R^2 - R^3
\]

\( S^i \) is in \{0,1\}. When \( S^i = 1 \), the firm makes an offer to an applicant in year \( t \). When \( S^i = 0 \), the firm does not make any offers in year \( t \). \( R^i \) is in \{1,2,3,4\}. This decision variable is the rank of the applicant (in year \( t \)) to whom the firm is going to make an offer, when \( S^t = 1 \). Note that \( R^i \) is a type of relative ranking, since there is public information about the applicant’s availability in year \( t \). If the applicant ranked at \( R^i \) is not available, the firm sends an offer to an applicant ranked in a neighborhood of \( R^i \) (\( S^3 \) is automatically set to 1 at the beginning of the simulations, so it is not a decision variable, i.e. firms that are still unmatched in year 3 will always make an offer).

An applicant’s strategy is represented by a string of 20 decision variables:

\[
A^1_1 - F^1_1 - A^1_2 - F^1_2 - A^1_3 - F^1_3 - A^1_4 - F^1_4 - A^2_1 - F^2_1 - A^2_2 - F^2_2 - A^2_3 - F^2_3 - A^2_4 - F^2_4 - F^3_1 - F^3_2 - F^3_3 - F^3_4
\]

\[25\] Each decision variable is represented by an integer.

\[26\] When \( R^i \) ranked applicant is not available, the firm first checks the applicant ranked at \( R^i - 1 \), if that is not available it checks the applicant ranked at \( R^i + 1 \), until one available applicant is found.
At r is in {0, 1}. Suppose that an applicant is ranked in r’th place in year t. When At r=1, the applicant may accept an offer in year t. When At r=0, the applicant will not accept any offers in year t. (A3 r=1 is automatically set at the beginning of the simulations, so it is not a decision variable, i.e. unmatched workers are always willing to take a position in year 3). Ft r is in {1, 2, 3, 4}. This decision variable is the threshold rank of the firm, in the case A1 r=1. When the offers to the applicant have lower ranks than Ft r, simply the applicant does not accept any offers in that period. Otherwise, she accepts the best offer.

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

T- A1-R1-A2-R2-R3

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. A1 is in {0, 1}. When A1=1, the firm may hire an applicant in year t. When A1=0, the firm will not hire an applicant in year t. (A3=1 is automatically set at the beginning of the simulations, so it is not a decision variable.) R1 r is in {1, 2, 3, 4}. This decision variable is the threshold rank of the applicant that the firm is going to hire, in the case A1 r=1. When the applicants have lower ranks than R1 r, simply the firm does not hire anybody in that period. Otherwise, it hires the best applicant.

An applicant strategy is a string of 20 decision variables:

S1 1-N1 1-S1 2-N1 2-S1 3-N1 3-S1 4-N1 4 - S2 1-N2 1-S2 2-N2 2-S2 3-N2 3-S2 4-N2 4 - N3 1-N3 2-N3 3-N3 4

S1 r is in {0, 1}. When applicant is ranked r’th among the others, if S1 r=1, she sends at least 1 application in year t; otherwise if S1 r=0 she does not send any applications in year t.

---

27 If any of the firms that is as good as Ft r is not available, then the threshold firm is determined as the best available firm.
28 If any of the applicants as good as ranked at R1 r is not available, then the threshold rank is determined as the rank of the best available applicant.
(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 and S^t_r=1. If none of these firms are available, she sends an application to the best available firm.  

---

29 If none of these best N^t_r firms are available, she only sends an application to the best available firm.  
30 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. To keep the actions in every information set simple, the space of strategies represented are smaller than the actual strategy space of the market games. We tried different limitations and present here the limitations (in terms of co-existence of actions corresponding to different information sets in the same strategy), which produce results that are most similar to the experimental sessions in the short run. Note that different plausible limitations do not change the results extensively.
Appendix: Instructions for the Centralized-Coerced Treatment with Announcements

WELCOME

This is an experiment about economic decision making. It is important that during the experiment you remain SILENT. If you have any questions, or need assistance of any kind, RAISE YOUR HAND but DO NOT SPEAK. We expect and appreciate your cooperation.

The decisions made in this experiment are hiring decisions. Accordingly, your role will be either “firm” or “applicant.” If you look at the screen in front of you right now, you will see your role. Your role, firm or applicant, will stay the same throughout the experiment.

The experiment will have 20 “markets,” which will last three “years” each.

To get a positive payoff in any given market, a firm will need to hire one, and only one, applicant in that market. An applicant will need to be hired by one, and only one, firm in that market.

In each group, there are four firms and four applicants. The firms are numbered 1 through 4, and the applicants are numbered 5-8.

The firms and applicants are assigned “qualities.” Your payoff as a firm is your quality multiplied by the quality of the applicant you have hired. Similarly, your payoff as an applicant is the product of your quality and your employing firm’s quality. For example, if a firm of quality 3 hires an applicant of quality 4, both firm and applicant will receive a payoff of 12 tokens each.

Firms’ qualities are simply their assigned participant number. In other words, if you are firm 3, your quality is 3. If you are firm 4, your quality is 4.
Applicants’ “qualities,” in contrast, have nothing to do with their assigned number and depend solely on the applicant’s “grades.”

**EXACTLY HOW ARE APPLICANTS’ QUALITIES DETERMINED?**

Think of the applicants as attending school for three years. Following each year of school, each of the applicants gets a grade of 0, 1, or 2, with 2 being the best possible grade and 0 being the worst possible grade. The computer generates these grades randomly, with each of 0, 1, and 2 having an equal chance of occurrence.

Each year, the grades are summed up, and the applicant is given his or her cumulative grade. In the third year, applicants are assigned qualities as follows: The applicant with the highest cumulative grade in the third year gets a quality of 4, the applicant with the second highest cumulative grade gets a quality of 3, the applicant with the third highest cumulative grade gets a quality of 2, and the applicant with the lowest cumulative grade gets a quality of 1.

For example, let’s say that applicant 5 got grades of 1, 0, and 2 in the three years. His cumulative grade would be the sum of the three: \(1 + 0 + 2 = 3\). That cumulative grade is **NOT** his quality, however. In fact, applicant 6 had a cumulative grade of 5, applicant 7 had a cumulative grade of 4 and applicant 8 had a cumulative grade of 2. Since applicant 5 had the third highest cumulative grade, he would be third ranked, resulting in a quality of 2. Applicant 6, who is best ranked, gets a quality of 4.

If there are no ties, the applicant qualities will be 1 through 4, with the highest quality of 4 going to the best ranked applicant; that is, to the applicant with the highest total grade. The quality of 3 will go to the second ranked applicant in total grades, and so on. The worst ranked applicant will get a quality of 1.
In case of ties, applicants having the same cumulative grade get ranked arbitrarily relative to each other. Then we assign qualities 1, 2, 3, and 4 to the four applicants according to their ranks, as before, with best ranked getting a quality of 4 and worst ranked getting a quality of 1.

**IF YOU HAVE ANY QUESTIONS, PLEASE RAISE YOUR HAND.**

**THE FINAL YEAR**

In the final year of each market, one additional piece of information will be revealed – the final quality of each applicant. Whereas in each of the first two years, participants see only the applicants’ cumulative grades, in the final year, the applicants’ final qualities will be shown as well.
PAYMENT

Since the highest possible quality for a firm or applicant is 4, the maximum payoff anyone can make per market is $16. At the end of the experiment, we will determine your dollar payoffs by dividing your total token amount by the number of markets and multiplying by 2. Hence you will get DOUBLE your average token earnings in dollars. This is in addition to your show up fee.
THE STAGES OF A MARKET

Prior to year 1:
Firms announce which year they will begin receiving applications.

In each year:
9. Applicants send applications to firms.
10. Firms may hire any one applicant from the pool of applicants who had applied in a given year.

At the end of year 2:
Firms and applicants that were not matched by the end of the second year are matched by the computer. Applicants can only be matched to firms which had received their applications in either year 1 or year 2 of that market.
**Announcement Stage**: Precedes the start of the market. In that stage, participants in the FIRM role declare the year in which they wish to be available to APPLICANTS to send them resumes. That choice is 1, 2, or 3, corresponding to first, second, and third year, respectively. Prior to the year indicated, applicants cannot send resumes to the firm and hence the firm cannot hire applicants.
Each year: APPLICANTS enter in the box in the top right hand corner the id number of a firm to which they wish to send a resume (apply for a job). Clicking on the “send resume” button underneath the text box completes the application process. Applicants can apply to more than one position, with the limitation that the firm receiving the application must be on the market. Remember to click on Finish when done sending applications. The top table shows which firms will receive your applications. The next table shows which firms are available and which have hired or not yet entered the market. The last table shows applicants’ cumulative grades.
Each year: Following the applicants’ decisions, each firm makes a hiring decision, by entering the number of an applicant in the box in the top right hand corner and pressing the “Make an offer” button. The applicants available to the firm are displayed in the top table under the title “Applicants who had sent you a resume.” Notice that the information tables on the firm’s screen closely resemble those on the applicants’ screen.
At the end of year 2: If you are not matched by the end of year 2 of any market, the computer will match you in year 3 as follows: Unmatched firms and applicants will be sorted by qualities. The firm with the highest quality will be matched with the highest quality unemployed applicant that had sent this firm an application in either year 1 or year 2 or both. From the remaining unemployed applicant pool, the second highest quality firm will be matched with the highest quality applicant that had sent this firm an application in either year 1 or year 2 or both, and so on.

Hence, to be eligible for hiring by a firm following year 2, an applicant needs to have sent an application to that firm in either year 1 or year 2.

Note that as a firm, each year you are allowed to make ONLY one hiring decision. If no higher quality firm made an offer to the same applicant, an offer will result in a hiring of that applicant. If a higher quality firm made an offer to the same applicant you made an offer to, the applicant will be hired by the other (higher quality) firm, and the computer will prompt you to make another hiring decision for that year. Once you made an offer which was accepted, you are “married” to that hired applicant for the duration of the market, and in future years in that market you will not get a decision but rather a message informing you that you had already hired a particular applicant.

If you have any questions please raise your hand.

Caution: The above example was selected arbitrarily and in way intends to suggest the actual qualities of players.
We now ask that you answer the quiz in front of you. Once you are done with all the
questions, raise your hand and one of the experimenters will come to you.

**Quiz**

<table>
<thead>
<tr>
<th>Applicant</th>
<th>YEAR 1</th>
<th>YEAR 2</th>
<th>YEAR 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

1. Imagine that the above table reflects the final grades after year 3 for the four
   applicants. You are firm 4. You hired applicant 7. Your payoff for this market is
   ___________________