

THE DYNAMICS OF REORGANIZATION IN MATCHING  
MARKETS: A LABORATORY EXPERIMENT MOTIVATED  
BY A NATURAL EXPERIMENT\*

JOHN H. KAGEL AND ALVIN E. ROTH

We create an environment in which congestion forces agents to match inefficiently early. We then introduce one of two centralized clearinghouse mechanisms. One of these has been successfully used to halt this kind of unraveling in a number of labor markets, while the other has failed. When it is costly for firms and workers to be mismatched compared with the costs of matching early, the experimental observations reproduce the field observations. Furthermore, the experiment permits us to observe the transition between a decentralized and a centralized market, both when the centralized market fails to control unraveling and when it succeeds, at a level of detail unavailable in field data.

Field studies of competitive entry-level professional labor markets reveal a common form of market failure involving the “unraveling” of hiring decisions, in which employment contracts from year to year become earlier and earlier in advance of employment, even when this becomes very costly (Roth and Xing 1994). Sometimes the cause is that the market is organized in such a way that firms and workers who wait until the efficient time to transact become caught in congestion and coordination failures (Roth and Xing 1997). In such markets, there is no way to avoid paying the costs of going early, without changing the rules of the market.<sup>1</sup> Some markets have successfully reorganized themselves by changing their rules to repair this kind of market failure. The best known of these is the American market for new physicians, the National Resident Matching Program (NRMP), in which hiring decisions unraveled to two years ahead of employment in the 1940s, and which reorganized itself around a centralized clearinghouse in the 1950s (Roth 1984). However, other

\* This research has been partially supported by grants from the National Science Foundation. We received valuable research assistance from Juergen Bracht, Hui-Jung Chang, Chi-Ren Dow, Juping Jin, and Nick Feltovitch. The experiments reported here were conducted at the University of Pittsburgh.

1. That is to say, there may be no equilibrium at which everyone goes late. There may even be no *feasible* strategies that yield efficient coordination with certainty. So the agents will be pressed to move early, despite the costs, and the task of the market designer will be to find rules that remove this necessity by allowing coordination to occur at the efficient time. Of course, hiring decisions must predate the start of employment, and not all early decisions are inefficient. See Li and Rosen (1998) and Suen (1999) for models in which early decisions may provide efficient insurance. In the experimental environment to be considered here, unraveling will be unambiguously inefficient.

© 2000 by the President and Fellows of Harvard College and the Massachusetts Institute of Technology.

*The Quarterly Journal of Economics*, February 2000

markets have experienced reorganizations that failed. Here we study these phenomena in controlled laboratory markets. We create a simple environment in which costly unraveling occurs, and then introduce one of two centralized clearinghouse mechanisms, one of which has been successfully used in a number of labor markets, while the other has been tried in several markets but failed to halt unraveling.

This experiment thus looks at matching mechanisms in the context in which they have most frequently arisen in naturally occurring markets, namely following market failure due to unraveling. This is perhaps the most striking difference between this and earlier experiments, which did not permit unraveling prior to the introduction of a centralized matching mechanism.<sup>2</sup>

In the experiment we find that, when it is relatively costly for the most desirable firms and workers to be mismatched compared with the costs of matching early, the experimental observations reproduce the field observations. This adds support to the contention that the differences observed in the field are due to the differences in the mechanisms, rather than to uncontrolled differences in the different markets to which these mechanisms have been applied. It also gives some indication of the robustness of these phenomena, since we now observe in a laboratory market with half a dozen firms and workers what has previously been observed in field studies of markets ranging in size from hundreds to tens of thousands of positions a year. Perhaps most importantly, this experiment gives us an opportunity to observe the transition between a decentralized market and a centralized market, both when this transition succeeds and when it fails.

The details of such transitions have so far been impossible to observe in field data. What we see in the experiment is somewhat surprising. In the high cost mismatch treatment, it is not the case that, when the centralized mechanisms are introduced they "roll back the clock" on appointment dates, which then stay fixed when the mechanism succeeds but start to unravel again when the mechanism fails. Instead what we observe is that, when the centralized mechanisms are introduced, there is only a small rollback of the unraveling that developed when the market was decentralized. But because of the congestion and competition in

2. See Sondak and Bazerman (1988, 1991) and Harrison and McCabe (1996). In fact, Harrison and McCabe speculate that the differences between their experimental results and the relevant field studies may be due to just this difference.

the market, some firms and workers who intend to make early matches find themselves unable to do so, and these participate in the centralized mechanism. In the market with the unsuccessful mechanism, their experience reinforces their desire to make early matches when they can, and the early transactions persist. But in the market with the successful mechanism, matching times move toward more efficient later matching. Thus, this experiment permits us to see not only the unraveling observed in field markets, but, during a successful transition to a centralized clearinghouse, also a gradual rollback of matching times in the reverse direction. This offers some new insight into transition behavior, a subject of considerable practical importance because of the number of markets in transition at any given time.

For example, in the 1990s there were major reorganizations of the American markets for osteopaths, radiation oncologists, and postseason college football bowls, and of some Canadian regional markets for new lawyers. A panel of senior judges called for a reorganization of the market for Federal court clerks (Becker, Breyer, and Calabresi 1994), following earlier unsuccessful attempts to halt unraveling.<sup>3</sup> The entry-level market for clinical psychologists has undergone several changes of rules (Roth and Xing 1997), culminating in a change to a centralized clearinghouse (APPIC 1998). And, after almost a half century of incrementally adapting its matching mechanism to changes in medical markets, in 1995 the NRMP commissioned the design of a new matching algorithm (Roth 1995, 1996) implemented in 1998 (Roth and Peranson 1997, 1999). (See Roth and Xing (1994) for discussion of all these markets through 1993.) Behavior during the transition can be very difficult to observe in field data, not least because some of the critical questions may concern behavior forbidden by market rules, and therefore concealed by those who engage in it. So experiments give us a window on the process that is otherwise unavailable.

This paper is organized as follows. Section I sets the stage for the present study, by describing the results of a natural experiment in different regions of the British National Health Service. Section II describes the experimental procedures. Section III

3. This attempt was also unsuccessful: in 1998 the Judicial Conference of the United States voted to rescind their policy of establishing benchmark dates of February 1, as the earliest date for receipt of application materials, and March 1 as the earliest date for interviews, in the face of widespread violations of these guidelines.

describes and analyses the results of the experiment. Section IV concludes by discussing the relationship between the experimental and the field observations.

### I. BACKGROUND: STABLE MATCHING VERSUS PRIORITY MATCHING

In the 1960s the various regional markets for new physicians in the United Kingdom unraveled to the extent that, by late in the decade, medical students were being hired for their first postgraduate ("preregistration") positions a year and a half prior to graduation. When a Royal Commission on Medical Education (1965–1968) issued a report on the problem, the clearinghouse organization of the American market presented an obvious alternative. Following the Commission's recommendation, many regional markets introduced centralized matching procedures. In this form of market organization, employers and employees contact each other (via applications, interviews, etc.) in a decentralized way, after which each employer submits a rank ordering of applicants to a central clearinghouse, and each applicant submits a rank ordering of positions. The clearinghouse uses these preference lists to produce a match according to some prespecified algorithm, and employers and employees are informed of the match. In the United Kingdom (and unlike in the U. S. market), the matches are compulsory at this stage, because positions do not belong to the individual hospitals, but are assigned by the health service.

But nothing was then known about the importance of the matching algorithm, and so different regions developed different algorithms to determine the match from the submitted preferences. Some of these succeeded in halting unraveling, and are still in use, while others failed and were quickly abandoned. In Roth {1991} (see also Roth {1990}) it was shown that all of those that failed produced matches that were *unstable*, in the sense that there could be a student and hospital program who would both prefer to be matched to each other than to accept the matching produced by the algorithm. In contrast, it was shown in Roth {1984, 1991} that (except for some problems with married couples and other linkages between two positions) the NRMP, as it was then, and the two largest of the successful British algorithms, produced *stable* matchings, at which no such mutually unhappy students and hospitals exist.

There are good theoretical reasons and considerable empirical evidence to support the notion that, to successfully eliminate

unraveling, a clearinghouse must either produce a stable matching, have power to compel compliance with all stages of the match, or have costs of going early that are high relative to the strategic advantages that going early provides (cf. Roth and Xing {1994}). However, since the empirical evidence comes from natural experiments, there are uncontrolled and unobserved differences among the markets that employed different algorithms. So there could possibly be other reasons (e.g., market demographics) why some algorithms failed, and it could be coincidence that these were unstable (particularly since Roth {1991} observed some unstable algorithms that had not failed, and since more recently some stable mechanisms have been observed to fail<sup>4</sup>).

The most common form of *unstable* matching algorithm, introduced in Edinburgh, Birmingham, and Newcastle, involved matching by priorities. In each region, the overall matching of students with hospitals was determined by determining a priority for each match, and making the individual matches of students with jobs in order of priority. That is, the first step of each of the three algorithms was to make all of the first priority matches. Then hospitals with unfilled positions and students still needing jobs were scanned to identify any second priority matches, and so on.

The algorithms introduced in Newcastle in 1967 {Leishman and Ryan 1970} and in Birmingham in 1966 {Alexander-Williams and Stephenson 1973} were almost identical. They each used the *product* of the student's ranking of the hospital and the hospital's ranking of the student as the basis for the priorities. If a hospital and student each ranked one another first (a "(1,1) match"), they had a priority of 1. If the hospital ranked the student first but the student ranked the hospital second (a "(1,2) match"), they had a priority of 2, as did a hospital who ranked a student second but was ranked first by the student (a "(2,1) match"). The two schemes differed in how they broke ties: in Birmingham ties were broken in the hospital's favor, so that a (1,2) match would have a higher priority than a (2,1) match. In Newcastle, ties were broken in the student's favor.<sup>5</sup>

Matches produced by such algorithms can be unstable, and

4. In 1995 Vancouver law firms discontinued their participation in a stable matching mechanism for articling positions (which continues to be used in Ontario and Alberta), and in 1997 the dental specialty of Periodontics pulled out of a stable mechanism (which continues to be used to match dental residencies for Advanced Education in General Dentistry, General Practice, Oral and Maxillofacial Surgery, Orthodontics, Pediatric Dentistry, and Prosthodontics).

5. At least initially (see Roth {1991} for details).

the centralized clearinghouses organized in this way all failed and were abandoned, having failed to halt unraveling, despite the fact that participation in the central clearinghouse was compulsory. For example, in Newcastle, by the time the scheme was abandoned in 1981, 80 percent of the preferences submitted by both students and positions indicated only a (mutual) first choice, indicating that these matches had in fact been arranged in advance. The manner in which priority algorithms are vulnerable to this kind of unraveling was analyzed in Roth {1991}.

The experience with stable matching algorithms was quite different. After abandoning their priority match algorithm, a matching algorithm that produces stable matches was adopted in Edinburgh in 1969, and in Cardiff around 1971. Like the stable algorithm adopted in the United States in the 1950s, these remain in use, and have largely ended the unraveling of appointment dates. These stable matching algorithms had as their basis Gale and Shapley's {1962} deferred-acceptance algorithm, adapted to local market conditions. The basic algorithm proceeds iteratively as follows:

Step 1: Each employer makes an offer to its highest ranked acceptable student.

Step  $k$ : (i) Each worker rejects all but the highest-ranked of the acceptable offers she has received in steps 1 through  $k-1$ . (ii) Each employer who has had an offer rejected in part (i) of step  $k$ , makes an offer to its highest ranked acceptable worker who has not yet rejected its offer. (iii) The algorithm stops at any step  $k = T$  at which no rejections are issued, and the resulting matching places each worker with the employer (if any) whose offer she has not rejected. Any worker not holding an offer is left unmatched, as is any position not being held when the algorithm stops.

Since each worker holds at most one unrejected offer at any step in the algorithm and since no employer makes an offer twice to the same worker, the algorithm stops and produces a matching. The matching is stable, because if some employer would prefer a different worker than it is matched with, then it will have proposed to that worker at an earlier step in the algorithm and been rejected. This implies that these workers prefer the positions they have to this employer's offer, so the matching is stable.

Before describing the experimental design, a few words are in order about decentralized markets, and some of the problems they face that cause the unraveling that led to the adoption of centralized clearinghouses. One decentralized market that has

lent itself to study in the field is the entry level market for clinical psychologists in the United States (Roth and Xing 1994, 1997). Prior to its reorganization in 1999 as a centralized clearinghouse, that market operated as a one-day telephone market with rules very similar to those of the deferred-acceptance algorithm. Despite the very fast turnaround time of offers in that market, there was insufficient time for the market to clear, and a prime concern of employers was that, if they made offers to candidates who turned them down, they would miss the chance to make offers to candidates who would have accepted their offer if made in time. This concern caused them to pressure candidates for decisions prior to the start of the official market. In general, the combination of competition for good candidates and limitations on the number of offers it is feasible to make in a timely way after experiencing some rejections seems to lie at the heart of the unraveling of appointment dates in many markets that have experienced it. In the simple experimental environment described in the next section, the congestion arises from the fact that firms are allowed to make only one offer per period, so that if two firms make an offer to the same worker in a given period, at least one of them will fail to make a match in that period. Despite the fact that making early arrangements in this environment is costly, it will be unavoidable.

Although unraveling is often associated with instability, Roth and Xing (1994) showed that the instability of late matches is neither a necessary nor sufficient condition for unraveling to be an equilibrium behavior. This can also be influenced by the costs of unraveling.<sup>6</sup>

The costs to participants when a market unravels so that hires are made long in advance seem to come from a variety of factors, including uncertainty about the quality of the match and loss of planning flexibility. In the experiment described next there is no exogenous uncertainty; there is simply a fixed cost of contracting early. Thus, in these experimental markets (unlike some natural markets), there will be no ambiguity about whether early contracts are inefficient, since they will incur costs that can be avoided by contracting late.

6. While there is an incentive to transact early to resolve instabilities, the cost of early transactions may be too great to make them attractive.

## II. DESIGN OF THE LABORATORY EXPERIMENT

The plan of the experiment was to have subjects first gain experience with a decentralized matching market with sufficient competition and congestion to promote unraveling of appointments (Roth and Xing 1997). The subjects would have the opportunity to make early matches, but at a cost. Once subjects had time to adapt to this market, one of the two centralized matching mechanisms would be made available for those subjects who did not make early matches. Subjects in both conditions of the experiment would be given the same information about the centralized matching mechanism, so the only difference between the two conditions of the experiment would be that one employed a priority matching algorithm (of the kind used unsuccessfully in Newcastle) and the other a stable matching algorithm (of the kind used successfully in Edinburgh). The experimental market is only loosely based on the U. K. experience; the design is in no way intended to be a scale model of these markets. Rather, the experimental market abstracts some essential elements common to many markets, in a simple setting.

In each experimental market, there were twelve subjects, half of whom were randomly assigned the role of "firms" and half the role of "workers." Of the six firms, half were identified as "high productivity" and half as "low productivity," and similarly for the workers. A subject retained the same role (e.g., "high productivity firm") throughout the experiment. Each round of the experiment was a matching market. Subjects received payoffs in each market according to with whom they were matched. Unmatched firms and workers earned \$0 for that market.<sup>7</sup> Any (high or low productivity) firms matched to "high" workers earned \$15 plus or minus an individual payment of not more than \$1, as did any workers matched to high firms. Firms matched to low workers earned \$5 plus or minus an individual payment of not more than \$1, as did any workers matched to low firms. Thus, while the individual payments allowed different individuals to have different first, second, and third choices of partners with whom to match (and were chosen so preferences were strict), everyone on one side of the market would have the three high productivity individuals on the other side of the market as their three most

7. Being unmatched in this experimental market can be looked at as analogous to having to match in the secondary markets that follow the main match in field markets. These secondary markets (given names like "the scramble") are generally regarded as sources of less satisfactory matches.



profitable matches. The set of stable matchings will vary as the individual payments are varied, but all stable matchings will match the high productivity firms with the high productivity workers, and the low productivity firms with the low productivity workers. (So mismatches, of a high to a low, pay the high type on average \$10 less and the low type \$10 more than their expected payoff at a stable matching.) Average earnings for matching by type were common knowledge, but the individual payments were private information; i.e., firms knew who their first, second, third, . . . choices were, but did not know other firms' detailed preferences or the exact preferences of workers over firms.

The Appendix lists the payoff tables for five games; a total of 25 games were constructed from these five by permuting the identities of firms and workers. Since each player learned only his own payoff schedule, a player who plays, e.g., game 1 once with firm 1's payoff schedule and once with firm 2's cannot identify that the same game is being played. To further disguise the fact that the same game was repeated, subject numbers (as seen on each others' computer screens) were two-digit numbers that were scrambled randomly between matching markets (this also prevented subjects from identifying each other across matching markets).

Each matching market consisted of three periods, called (in order) periods -2, -1, and 0, to denote that the profit to a worker or firm from any match was reduced by \$2 if the match was made in period -2, by \$1 if the match was made in period -1, and by \$0 if the match was made in (the final) period 0 (with these costs subtracted from the match payoffs reported in the Appendix). Note from the Appendix that these costs still make it more desirable to match to a high productivity firm or worker in period -2 than to a low productivity firm or worker in period 0, but that it is more profitable to match with your *least* profitable high (low) productivity partner in period 0 than your *most* profitable high (low) productivity partner in period -2. In fact, the parameters chosen make the difference between the best and worst match within the high (or within the low) category smaller than the cost of going even one period early. So the cost of going early in these markets is high relative to the differential profitability of matches within a category, but low relative to the differential profitability of matches between categories.

In each matching market, employers could hire one worker, and likewise each worker could accept only one job. There were no

out-of-pocket costs to firms of making offers, but they were restricted to one offer in each period.<sup>8</sup> Workers could, and did, receive multiple offers in any given period, but could only accept their highest valued offer for that period, or reject all the offers. Once a match was made, firms were no longer permitted to make offers, and workers, although they could (and did) receive subsequent offers, had these offers automatically rejected.<sup>9</sup> Offers and rejections or acceptances were private information so that a firm could not tell which workers had received or accepted offers from other firms.<sup>10</sup> Similarly, workers could not tell which firms had made offers to other workers and whether these offers had been accepted or rejected. It can be readily verified that under these conditions it is not an equilibrium for all transactions to be conducted in the last period, even though that is the least-cost period.<sup>11</sup>

At the end of each matching market we reported all matches to all subjects, along with the period in which the match was made and the firms' and workers' productivity types. In this way subjects could tell at a glance when a match was made, if it involved a high or low productivity firm or worker, as well as who failed to be matched in that market. This easy visibility of instabilities (highs matched to lows and unmatched firms and workers) is characteristic of the U. K. regional matching markets where students in a region are all from the same medical school and the best employment is at the local teaching hospital {Roth 1991}.

Each experimental session began with ten consecutive decentralized matching markets. After the tenth decentralized match-

8. To insure timely operation of each session, firms had a limited period of time in which to make offers, as did workers in deciding whether or not to accept an offer. The default option had firms making no offer and workers rejecting all offers. Firms were encouraged to explicitly not make offers if that was their plan for the period, and workers were encouraged to explicitly reject offers if that was their choice for the period.

9. These rejections were timed to occur after a few seconds had passed so that they were indistinguishable from nonautomatic rejections by workers who were not yet matched.

10. This is modeled on the situation in the British markets, in which participation in the centralized system was mandatory, and hence early agreements were made informally and confidentially. The high incidence of these agreements (as in the Newcastle data referred to earlier) shows that in these small markets, such informal commitments were reliable. (In the much larger American market, informal commitments of this kind have not been reliable.)

11. In fact, this market with its one offer per period can be viewed as an extreme case of the telephone market studied in Roth and Xing {1997}, in which even the ability to make multiple offers in a fixed time period was insufficient for the market to clear without precommitments.

ing market it was announced that, in each subsequent market, periods -2 and -1 would proceed as before, but henceforth period 0 would employ a centralized matching algorithm. This sequencing of decentralized markets followed by centralized markets is designed to capture the fact that centralized matching markets only came into play in the United Kingdom (and elsewhere) following a history of decentralized matching subject to considerable unraveling.

For the centralized matching algorithm subjects were instructed that if they were not matched prior to round 0, they were to "submit a rank order list of their possible matches, and the centralized matching mechanism (which is a computer program) will determine the final matching, based on the submitted rank order lists." As in the field applications of these mechanisms, subjects were strongly encouraged to use the centralized matching mechanism. The (identical) instructions for each mechanism gave a description applicable to both algorithms:

This centralized matching mechanism is designed to produce the best match possible, taking everyone's preferences into account. For example, if there is a matching which gives everyone his or her first choice, this will be the outcome of the matching mechanism. Even if this is not possible, if your first choice also ranks you first, you can be sure of being matched to your first choice. And the higher you rank some firm (if you are a worker) or some worker (if you are a firm), the greater your chance of being matched to that firm or worker. Finally, if you submit a rank order list that lists all six of your possible matches, and if each of your choices includes you on their rank order list, then you can be sure of being matched. That is, the matching mechanism can only match a firm and a worker each of whom lists the other on his rank order list, and if everyone includes all of his six options, then the mechanism will leave no one unmatched.

It was further noted that the procedure was one that had been used in actual labor markets and that the procedure "makes it a good idea to rank your choices in terms of your payoffs." Subjects then proceeded through a trial run of the algorithm in which they were required to rank their choices in terms of their payoffs (and to rank all of their potential matches). In this trial run, preferences were such that there was an outcome that matched all firms and workers with their first choices, and this was produced by both algorithms, given the rank order lists submitted. The matches resulting from this trial run were reported to all subjects. Note that the instructions and introductory material were the same for both algorithms, so that the only difference between the

two conditions of the experiment was which algorithm was used.<sup>12</sup> As in the U. K. markets, there was no opportunity to decline to participate in a match once it was made.

In each experimental session we then conducted an additional fifteen matching markets, with periods -2 and -1 employing the same decentralized procedures described above, but with the centralized matching algorithm in place in period 0. Note, that as with the decentralized markets, subjects did not know who was or was not matched prior to submitting their preference lists in period 0. Matched firms and workers were prohibited from submitting lists. Further, unlike the trial run, subjects were free to submit incomplete lists and to rank firms or workers as they wished without regard to their rank order. Three experimental sessions employed the Newcastle matching algorithm and three sessions employed the deferred-acceptance algorithm.<sup>13</sup>

The five basic preference structures given in the Appendix were played in blocks of five, with each preference structure played once, in random order, within each block (but with the same randomized ordering in all sessions). The five games were designed to have different numbers of stable matchings (a low of 1 and a high of 9). Note that instabilities can occur under both algorithms if some firms and workers match early, even if all remaining firms and workers state their true preferences in period 0.

From each set of ten decentralized matching markets, one market was selected at random with subjects receiving the payoffs earned for that market. Similarly, for each set of fifteen centralized matching markets, one market was selected at random with subjects receiving the payoffs earned for that market. In addition, all firms and workers received an \$8 participation fee.

Five additional, low mismatch cost sessions were conducted prior to these sessions, with different subjects from the same subject pool. These games differed from those described in the Appendix only in that the payoff for a match between two high productivity types was \$5 less. All other payoffs and costs remained the same, so that the cost of being mismatched, relative to

12. Note also that the instructions and introductory material which accompany a market mechanism should be regarded as part of the mechanism, since different instructions might elicit different behavior. This was clear, for example, in the controversy preceding the recent redesign of the matching algorithm for the National Resident Matching Program, since one of the principal issues had to do with whether the material distributed to students was misleading (Roth and Peranson 1997, 1999).

13. Assignment of sessions to treatments was made in advance of each session.

the cost of early matches, was substantially smaller for high productivity types. The first of these sessions, using the Newcastle algorithm, was a pilot session, employing ten centralized matching markets. The remaining four sessions, two with Newcastle and two with the DA algorithm, used identical procedures to those already described, except for the lower return of matches (and hence a lower implicit cost of mismatches and nonmatches) for high productivity types. The effects of these parameter changes on the primary results reported will be discussed following the analysis of the high mismatch cost sessions.

All experimental sessions were conducted in the economics laboratory at the University of Pittsburgh. A broad cross section of undergraduate and graduate students (as well as an occasional staff member) from the University of Pittsburgh and Carnegie Mellon University participated. Subjects were recruited through notices posted throughout both campuses, newspaper advertisements, and postings on electronic bulletin boards.

To summarize, the decentralized experimental environment does not make it feasible for firms and workers to coordinate on an efficient match at period 0. *Both* centralized mechanisms do make such coordination feasible, since if all agents list all possible matches at period 0, no agent will be unmatched (and the agents are informed of this in their instructions). However, one mechanism produces stable matches, and the other does not. The experiment therefore allows us to investigate whether efficient coordination is achieved, as a function of both the stability of the match and the costs of early transactions.

### III. EXPERIMENTAL RESULTS

We begin by reporting aggregate results from the sessions with a high cost of mismatch. These data are comparable in form to what we observe in field data, and they show the manner in which the experimental results duplicate the field observations. Then we consider disaggregated results, which allow us to study the transition in more detail. Finally, we consider the results of the low cost mismatch sessions.

#### *III. 1 Aggregate Results in Markets with a High Cost of Mismatching*

Figure I reports aggregate data on early matches over time under the two different matching procedures. The natural metric

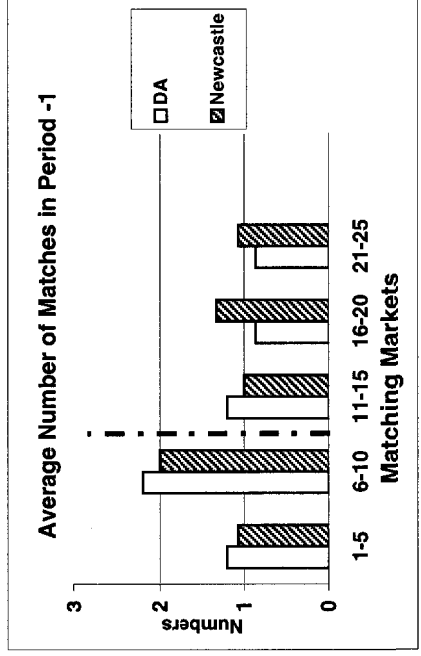
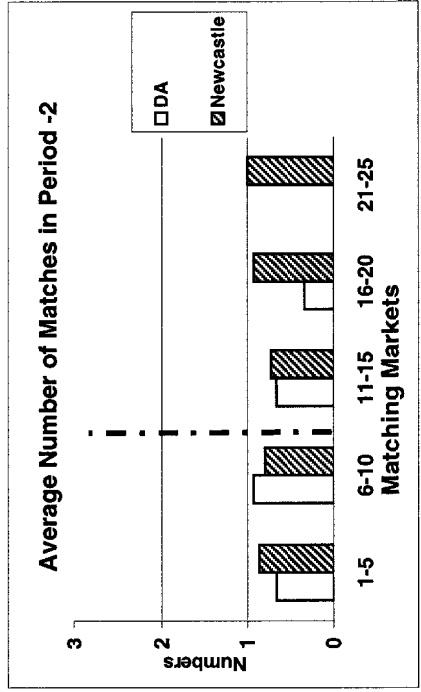
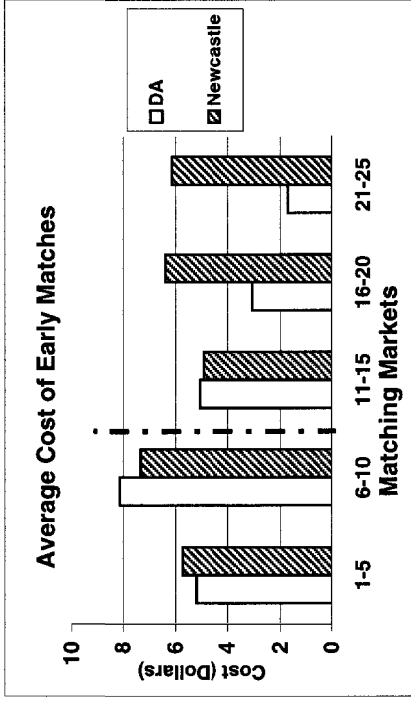


FIGURE I  
 Early Matches: Average Cost of Early Matches and Average Number of Matches in Periods -2 and -1

for pooling raw counts of early matches across periods -2 and -1 is the total cost of these matches, reported in the first panel of Figure I. Early matches by period are reported in the bottom two panels of Figure I. Data are reported in terms of averages for each consecutive set of five games (recall that each game was played once, in random order, within each set of five games).

In the decentralized matching markets (markets 1–10) we observe clear unraveling over time. The average cost of matching early went from \$5.46 per game in the first five games to \$7.74 per game in the next five games which is a statistically significant increase in the cost of matching early ( $t = -4.03$ , *d.f.* 5,  $p < .01$ , one-tailed  $t$ -test).<sup>14</sup> As the bottom panels of Figure I indicate, around 75 percent of the time there was one match in period -2, with this percentage roughly constant over time. In contrast, period -1 matches nearly doubled from an average of 1.1 to 2.1 between the first and last five decentralized games, accounting for nearly all of the increase in the cost of early matches. Finally, as expected, no significant differences in the first ten markets were observed between those sessions which would be switched to the Newcastle algorithm and those sessions switched to the deferred-acceptance (DA) matching algorithm starting in the eleventh market.

The introduction of a centralized matching algorithm led to an immediate reduction in the cost of early matches under both algorithms in all sessions. These cost reductions were statistically significant in both cases,<sup>15</sup> and essentially the same for both algorithms, as subjects had nothing but our claims regarding the beneficial properties of the matching algorithms to guide their decisions. Further, as the bottom panel of Figure I indicates, essentially all of these immediate cost reductions were accounted for by a reduction in period -1 matches.

However, the subsequent dynamics of the adjustment process differed substantially between the two algorithms. Under the DA algorithm, period -2 matches decreased steadily over time and were totally eliminated in the last five games. In contrast, period -2 matches increased modestly over time under the Newcastle algo-

14. The unit of observation is the average cost for each five games in an experimental session. Upper bounds to these costs are \$24 (everyone matched in round -2) and \$12 (everyone matched in round -1). The field data clearly suggest increasing costs of early matches (unraveling), hence the one-tailed  $t$ -test. Note that increasing costs were observed in all six sessions.

15.  $t = -2.59$ ,  $p < .10$  for sessions employing the DA algorithm;  $t = -10.0$ ,  $p < .01$  for the Newcastle algorithm (*d.f.* = 2 in both cases, one-tailed  $t$ -tests).

rithm. Period -1 matches decreased modestly over the last ten games under the DA algorithm, but increased slightly under Newcastle.

The net effect of these adjustments was that by the last five games (i) the cost of early matches had been reduced substantially under the DA algorithm compared with the last five games with decentralized matching (a reduction of \$6.40 per session;  $t = 5.25$ ,  $d.f. = 2$ ,  $p < .05$ , one-tailed  $t$ -test) and (ii) although there was a modest reduction in the average cost of early matches of \$1.20 per game under Newcastle, this difference was not significant at conventional levels ( $t = 2.60$ ,  $d.f. = 2$ ,  $p > .05$ , one-tailed  $t$ -test).<sup>16</sup> Thus, our results in the laboratory replicate those in the field: over the last five games the DA algorithm had reduced unraveling considerably compared with decentralized matching, and the Newcastle algorithm had failed to reduce the unraveling. Further, the dynamic adjustment process that emerges in the laboratory (which cannot be observed in field data) is a small, immediate improvement in performance under both mechanisms followed by a tendency to unravel over time under Newcastle versus steady reductions in early round matches under the DA algorithm.

### *III.2. Disaggregated Results in Markets with High Costs of Mismatching*

Figure II shows the cost of these early matches disaggregated by types, high firms to high workers (top panel) and low firms to low workers (bottom panel).<sup>17</sup> In the decentralized markets there were substantially more early matches by high than by low types, with an average cost of early matches for highs of \$4.50 per game in the first five games versus \$0.86 for lows. Costs increased slightly for highs in the last five decentralized games to \$5.24 per game, while increasing substantially more for lows to \$2.50 per game. Lows matched mostly in period -1 (79.1 percent of all low matches), with highs accounting for most of the period -2 matches (81.6 percent of all period -2 matches). The higher frequency of period -2 matches by high productivity types no doubt reflects the substantially higher cost to high productivity types of not being

16. Differences in the cost of going early between the two algorithms were statistically significant with Newcastle averaging \$6.14 per game versus \$1.74 with the DA algorithm ( $t = 3.77$ ,  $d.f. = 4$ ,  $p < .01$ , one-tailed  $t$ -test;  $p < .05$  using a nonparametric Mann-Whitney test as well). Note again that one-tailed tests are justified here on the basis of field data reported in Roth (1991).

17. Early matches, under both decentralized and centralized matching procedures, were almost exclusively between like types, with mismatches constituting less than 4.5 percent of all early matches.



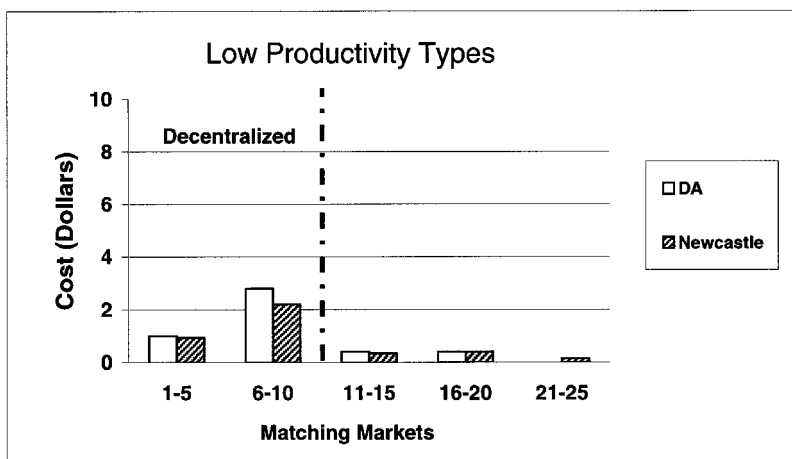
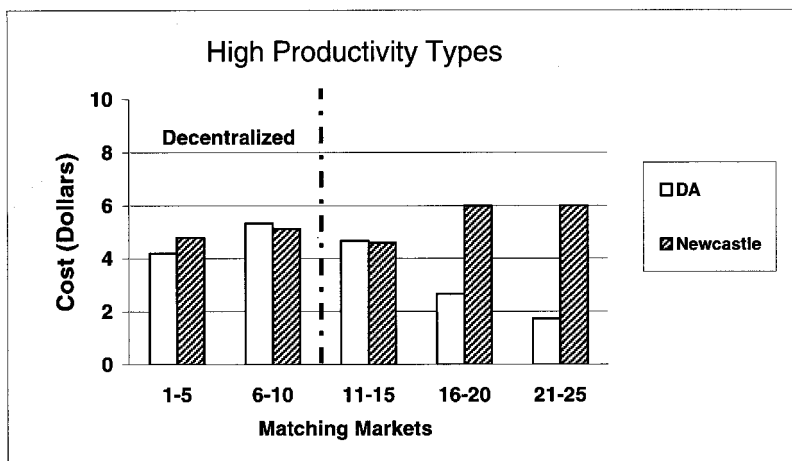


FIGURE II

Average Cost of Early Matches for High- and Low-Productivity Types

matched. Further, as the data in Figure II clarify, most of the unraveling that occurred over time in the decentralized markets resulted from low productivity types matching earlier.

The switch to the centralized markets had its largest immediate impact on low productivity types' early matches. These dropped from an average cost of \$2.50 per game in the last five decentralized games to \$0.38 per game in the first five games with

centralized matching (Figure II, bottom panel; reductions occurred in all six sessions). Costs of early matches for low productivity types continued to decline over time, with essentially no differences in behavior between the two algorithms.

The switch to centralized matching resulted in a modest reduction in the cost of early matches for high productivity types of \$0.64 per game in the first five games with centralized matching, with essentially no difference between the two algorithms. However, the dynamics of the adjustment process differed substantially as a function of the matching algorithm. Costs of early matches *decreased* steadily, and quite dramatically, for high types under the DA algorithm to \$1.74 per game in the last five games. In contrast, they increased some 30 percent to \$6.00 per game under the Newcastle algorithm. Thus, the two centralized matching algorithms had their greatest *differential* impact on early matches of high productivity types, with no differential impact on low productivity types.

These differential responses by highs and lows to the introduction of the centralized matching algorithms are perfectly consistent with the incentives inherent in the two mechanisms relative to the costs and benefits of early round matches.

With decentralized matching the primary potential cost to low productivity types of waiting to be matched in period 0 was that they might go unmatched. Both centralized matching algorithms insure a match to someone provided that all firms rank order all workers and all workers rank order all firms.<sup>18</sup> And, as Figure III shows, both matching algorithms sharply reduced the frequency of unmatched low productivity types: from 2.27 per game (37.8 percent) under decentralized matching to 0.62 per game (10.4 percent) with centralized matching, with essentially no differences between the two algorithms.<sup>19</sup>

In contrast, high productivity types have to worry about both mismatches and nonmatches. The Newcastle algorithm promotes mismatches (high firms matched to low workers), whereas the DA algorithm tends to eliminate them. This differential impact of the two algorithms resulted in an average mismatch rate of 0.84 pairs per game under Newcastle (14 percent of all high productivity firms and workers) compared with .20 pairs per game with the DA

18. Recall that this was explicitly pointed out to subjects in the instructions.

19. Unmatched high productivity types averaged 1.17 per game (19.5 percent) under decentralized matching and 0.15 per game (2.6 percent) with the centralized matching algorithms.

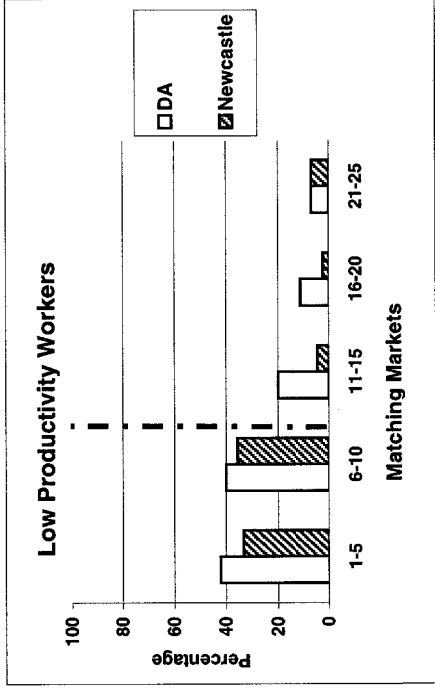
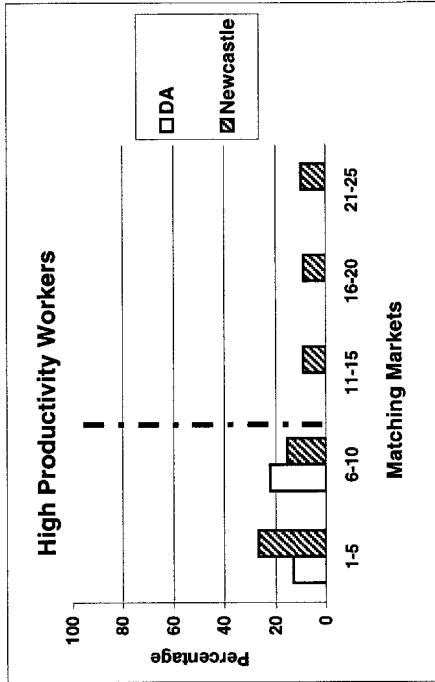
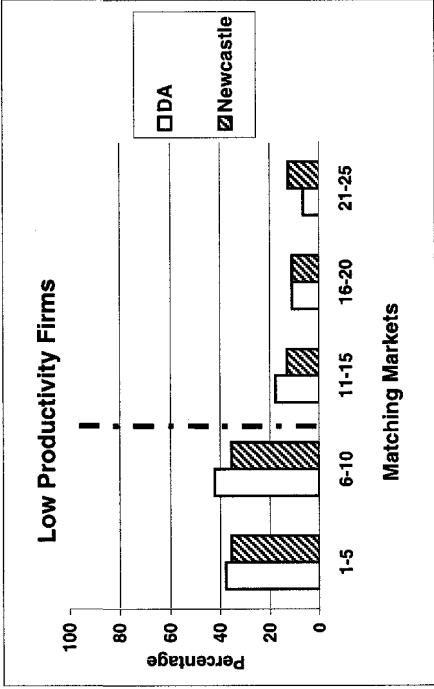
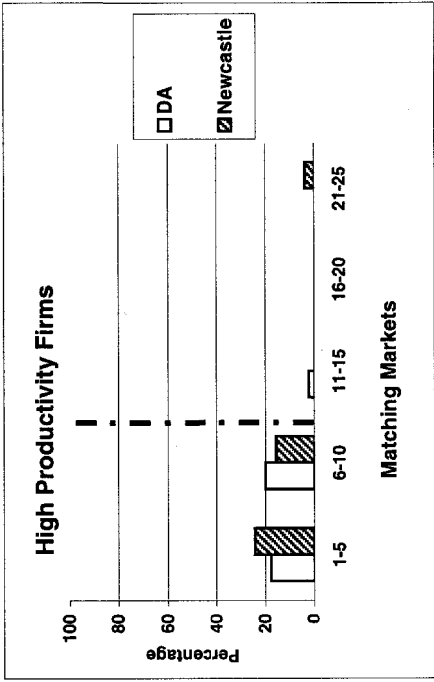


FIGURE III  
Unmatched Firms and Workers

algorithm (3.3 percent of all high productivity firms and workers;  $t = 2.03$ ,  $d.f. = 4$ ,  $p < .10$ , one-tailed  $t$ -test).<sup>20</sup> The cost of a mismatch for a high productivity type was, of course, much greater than the cost of an early match assuming, as was almost always the case, that early matches were between like productivity types. (In contrast, low productivity types stood to benefit from mismatches.) Thus, there was substantially more incentive for highs to match early under Newcastle to avoid mismatches than under the DA algorithm.<sup>21</sup>

The two matching algorithms could also be expected to produce differences in the number of instabilities (blocking pairs) in matches actually achieved, with Newcastle generating greater numbers of instabilities. This proved to be the case in the experiment as well: instabilities in final matches averaged 2.9 blocking pairs per game in the last five games with decentralized matching. The DA algorithm resulted in a sharp reduction in the number of instabilities, averaging one blocking pair in the last five games, a statistically significant reduction compared with the last five games with decentralized matching ( $t = 17.3$ ,  $d.f. = 2$ ,  $p < .01$ ).<sup>22</sup> In contrast, under the Newcastle algorithm the number of instabilities increased slightly from 2.9 to 3 a game. Under the DA algorithm there was an immediate drop in blocking pairs from 3 to 2.1 per game, with further reductions over time as a result of fewer early matches. In contrast, there was an immediate *increase* in the number of blocking pairs from 2.87 to 3.67 under Newcastle ( $t = 6.93$ ,  $d.f. = 2$ ,  $p < .05$ ), with the increase resulting from the increased number of mismatches that Newcastle produced. In this case the reduction in instabilities over time resulted from increased numbers of early matches by high productivity types serving to largely offset the instabilities inherent in the algorithm.

### *III.3. Strategic Behavior in Markets with High Costs of Mismatching*

So far we have looked at outcomes; now we turn to individual behavior. Figures IV and V report early offers and offers accepted

20. In the DA algorithm all the mismatches came from session S3 and were the result of one high firm not listing all workers and consistently ranking some low workers ahead of high workers. In contrast, the three Newcastle sessions all showed consistent mismatches throughout.

21. This incentive was even greater under the Newcastle algorithm than with decentralized matching as the latter averaged .30 mismatched pairs per game, averaged across all sessions.

22. Reductions occurred in all three sessions.

by period for firms and workers. We report early offers normalized as the percentage of eligible (unmatched) firms making offers. An offer value of 100 percent indicates that all unmatched firms made offers in a given round; an offer value of 50 percent indicates that half of the unmatched firms made offers in that round. Normalizing in terms of percentages adjusts the period -1 data for any differences in the *ability* of firms to make offers as a result of acceptances in period -2. Acceptance data for workers are also reported in percentage terms, as offers accepted divided by workers eligible to accept those offers. Further, we only report "serious" early offers. This means all offers of high productivity firms, but for low productivity firms only offers to low productivity workers are counted, as early offers from low productivity firms to high productivity workers were virtually never accepted.<sup>23</sup>

Data for high productivity firms and workers are reported in Figure IV. The most striking aspect of the data is that from game 20 on high productivity workers refused all period -2 offers under the DA algorithm. In contrast, under Newcastle there was a steady acceptance rate of at least 33 percent of all period -2 offers over the last five games. Differences in offer rates are not nearly as striking. Offers averaged around 50 percent per game over the last five games under Newcastle, which is not significantly different from the 35 percent offer rate over these last five games under the DA algorithm. Thus, the striking differences in period -2 early matches over the last five games between the two algorithms reported in Figure II can be largely attributed to differential acceptance rates of workers.

The higher offer to acceptance ratio on the part of firms under both algorithms more than likely reflects the greater incentive high productivity firms had to match early to high productivity workers. Being the initiator in the matching process, firms making early offers to workers ranked these workers higher than the workers ranked them.<sup>24</sup> Early offers continued to occur even under the DA algorithm, in spite of the fact that the difference between the best and worst match to a high productivity worker

23. Only 0.9 percent of all low productivity firms' offers to high productivity workers were accepted in periods -2 and -1 (3 out of 353 offers).

24. In period -2 the average rank order difference between high productivity firms and the workers they made offers to was .70, compared with a maximum average difference of 1.3 (assuming that each high productivity firm made an offer to their highest ranked worker). In period -1 the average rank order difference was .53. These differences reflect the advantage in the matching process held by the proposer, an advantage consistently reflected in many matching environments and in the set of stable matchings (cf. Roth and Sotomayor {1990}).

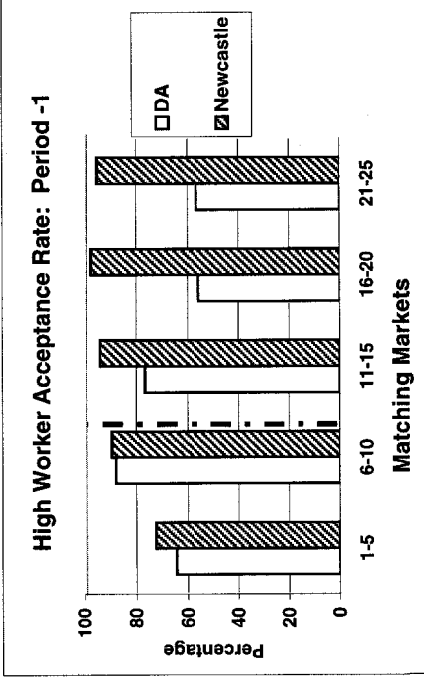
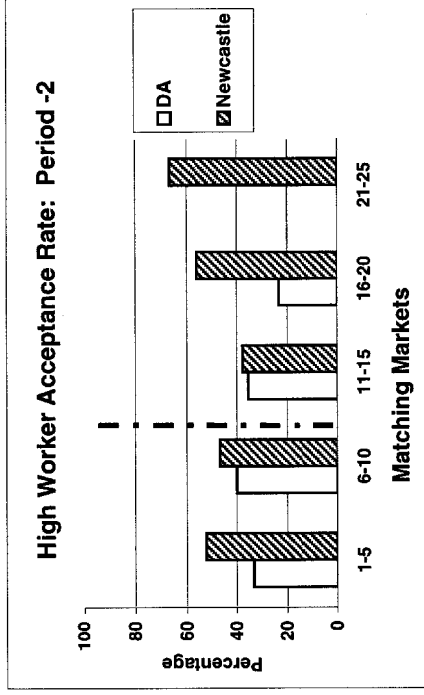
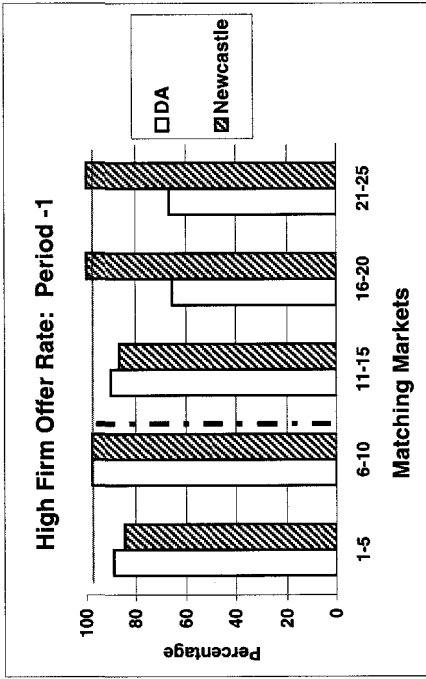
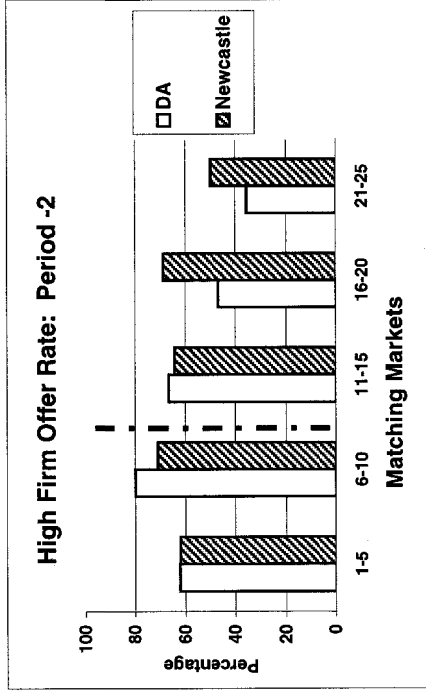


FIGURE IV  
High Productivity Firm Offer Rates and High-Productivity Worker Acceptance Rates

was less than the cost of matching in period -1. Since nonmatches and mismatches rarely occurred for high productivity types with the DA algorithm, the premium that high productivity firms were willing to pay for early matches probably reflects the relatively high frequency of nonmatches that occurred under decentralized matching.<sup>25</sup>

Under the Newcastle algorithm the percentage of high productivity firms making period -1 offers was 100 percent in the last ten games. In contrast, the offer ratio with the DA algorithm was just over 60 percent in the last ten games, which is significantly lower ( $t = 3.50$ ,  $d.f. = 4$ ,  $p < .025$ , one-tailed  $t$ -test). Further, as the bottom right panel of Figure IV indicates, there were substantial differences in period -1 acceptance rates, conditional on having received an offer, with the acceptance rate close to 100 percent over the last ten games under Newcastle compared with 50 percent under the DA algorithm ( $t = 2.60$ ,  $d.f. = 4$ ,  $p < .05$ , one-tailed  $t$ -test).<sup>26</sup>

Figure V reports offers of low productivity firms and acceptance rates of low productivity workers by round. Recall that most of the unraveling reported with decentralized matching resulted from period -1 matches between low productivity types. Figure V shows that this resulted from two factors: a modest increase in low productivity firm offer rates to low productivity workers and a sharp increase in low productivity workers' rate of accepting these offers.<sup>27</sup> Following the introduction of the centralized matching, for all intents and purposes low productivity workers' acceptance of early offers from low productivity firms immediately dropped to zero, and stayed there, for both algorithms: 1 out of a total of 33 period -2 offers and 6 out of a total of 50 period -1 offers were accepted.

The success of both centralized matching algorithms in eliminating unmatched workers in round 0 depends critically on the nature of the lists that unmatched firms and workers submit.

25. Only 1 of 270 (0.4 percent) high productivity types failed to be matched in games played with the DA algorithm compared with 18.3 percent prior to implementing the algorithm.

26. Period -1 offer and acceptance rates were uniformly lower under the stable matching algorithm; i.e., the highest offer (acceptance) rate under the DA algorithm was lower than the lowest offer (acceptance) rate under Newcastle.

27. The total period -1 offer rates by low productivity firms remained essentially constant between games 1-5 and 6-10 (a total of 74 and 73 offers, respectively), but the composition of offers changed to include more offers to low productivity firms in games 6-10 than in games 1-5 (54.8 percent to 40.5 percent;  $z = 1.73$ ,  $p < .05$ , one-tailed test).

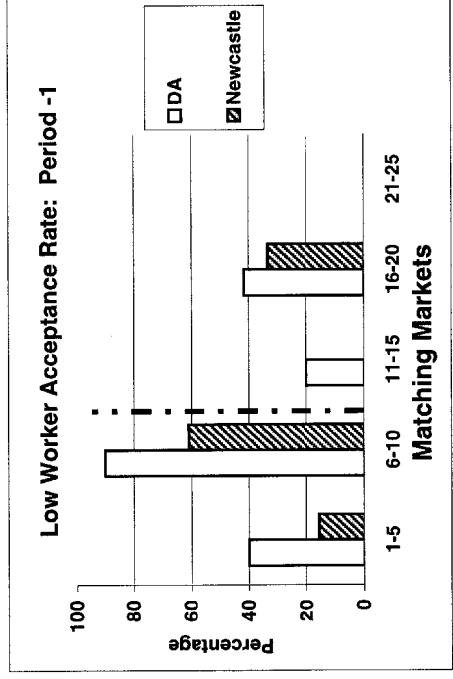
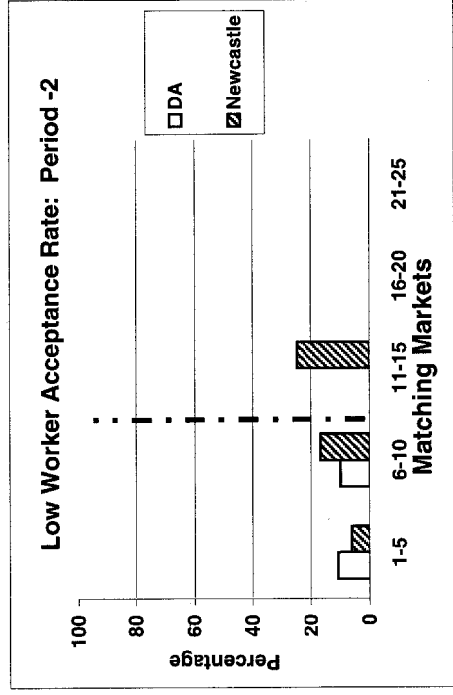
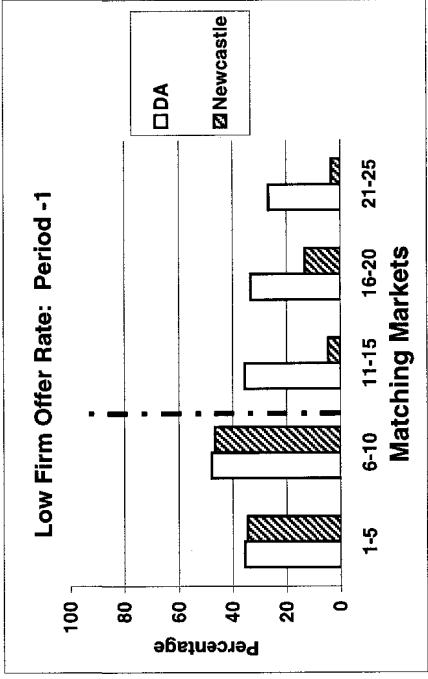
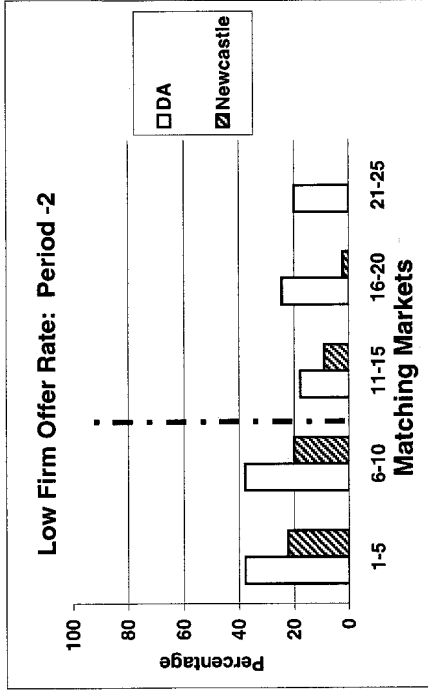


FIGURE V  
Low-Productivity Firm Offer Rates and Low-Productivity Worker Acceptance Rates



Unmatched high productivity firms and workers consistently failed to submit complete lists in round 0 under the DA algorithm, with list lengths averaging 5.15 for high firms and 5.40 for high workers (maximum list size is 6). Nevertheless, there were essentially no unmatched high productivity types (recall Figure III) since they consistently listed all highs as their first three choices (96.2 percent and 97.8 percent of first three choices for high firms and workers, respectively).<sup>28</sup> And under the DA algorithm (but not under Newcastle) high productivity types can be assured of matching to like types in round 0 if they list all highs as their first three choices, assuming no highs were matched to lows prior to round 0 (which virtually never happened). Under the Newcastle algorithm high productivity firms submitted slightly longer lists (average 5.79) with high productivity workers submitting somewhat shorter lists (average 4.52). So that even though high types consistently listed highs as their first three choices (97.3 percent and 100 percent for firms and workers, respectively), there were occasional, but consistent, numbers of unmatched high productivity workers under Newcastle (recall Figure III). The shorter lists of high productivity workers under Newcastle may represent an effort to avoid being matched to low productivity firms (recall that high productivity workers received numerous early offers from low productivity firms). But under the Newcastle algorithm these shorter lists could not achieve this outcome absent high (low) productivity firms listing only high (low) productivity workers, which did not happen.

List lengths for low productivity types averaged 5.01 and 5.37 for firms and workers under the DA treatment and 5.92 and 5.21 for firms and workers in the Newcastle sessions. Average numbers of high productivity types listed as the first three choices were substantially smaller than for high productivity firms and workers, averaging 57.4 percent and 69.3 percent for firms and workers under the DA treatment and 47.4 percent and 78.2 percent for firms and workers in the Newcastle sessions. The failure of all low productivity types to make complete lists accounts for the nonnegligible numbers of low types that went unmatched under both algorithms.<sup>29</sup> In terms of maximizing potential earnings, these short lists were a clear mistake for low productivity types.

28. In cases where a firm or a worker listed fewer than three choices, the base underlying our calculation of percentage of high types listed remained three.

29. Under the DA algorithm this could have been compensated for by exclusively listing low productivity types first. But this did not happen either.

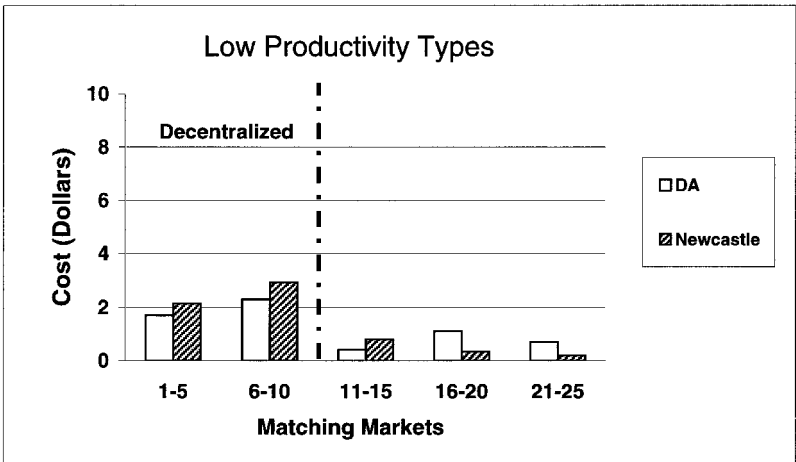
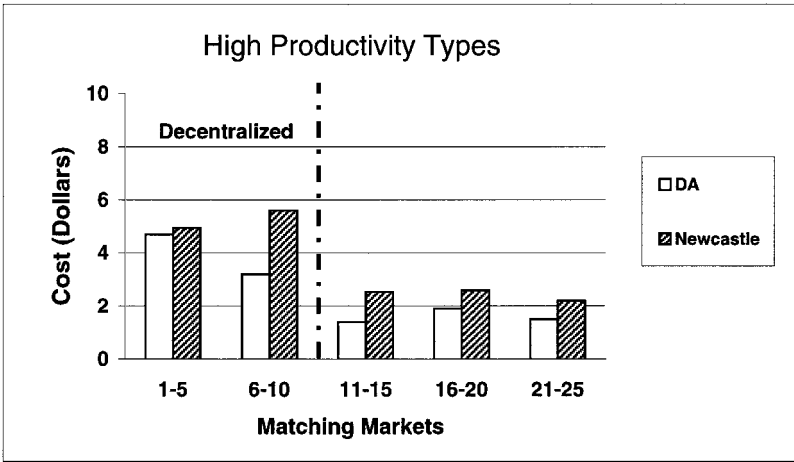


FIGURE VI

Average Cost of Early Matches for High and Low-Productivity Types for Low Mismatch Cost Sessions

III.4. Behavior in Markets with Lower Costs of Mismatching

Comparing results from the low mismatch sessions with those just reported indicates that two of our main results are sensitive to the cost of being mismatched, relative to the costs of matching early. Figure VI shows costs of early matches, disaggregated by types, for these low mismatch cost sessions, high firms to

high workers (top panel) and low firms to low workers (bottom panel).<sup>30</sup> As with the high mismatch sessions, the switch to centralized matching had its largest immediate impact on low productivity types' early matches, virtually eliminating them regardless of treatment condition. The impact of centralized matching on high productivity types was, however, different from the high mismatch cost sessions in two important ways. First, the introduction of centralized matching resulted in a large, immediate, reduction in the cost of early matches for high productivity types under both centralized matching algorithms; an average cost reduction of \$2.60 (54 percent) in the first five markets following the introduction of centralized matching versus the last five matching markets prior to its introduction. This compares with an average cost reduction of \$0.64 (13.1 percent) per market in the high mismatch cost markets. These differences in average cost reductions are significant at the 5 percent level ( $t = 2.28$ ,  $d.f. = 9$ ,  $p < .05$ , two-tailed test). Second, starting from this substantially lower base rate of early matching, there was no significant unraveling over time under Newcastle and no significant movement toward more efficient later matching under the DA algorithm.

Recall that the only difference between the high and low mismatch cost sessions is that the cost of being mismatched for high productivity types is substantially lower, with no corresponding reduction in the cost of matching early.<sup>31</sup> Under both high and low mismatch cost treatments, the Newcastle algorithm produced about 1 mismatched pair in each market (1.15 pairs and 0.84 pairs in low and high mismatch cost treatments, respectively).<sup>32</sup> Using an expected value of one mismatch per market, this produces an expected cost to highs of being mismatched under Newcastle of \$0.83 per market in the low mismatch cost sessions versus \$1.67 in the high mismatch cost sessions (in addition, there are increased costs to highs of being unmatched in the high mismatch cost sessions).<sup>33</sup> High productivity types paid an average of \$1.22 per market per person to match early in the low mismatch cost

30. Data for all five sessions are included since the Newcastle session with only ten centralized matching markets exhibited essentially the same behavior as the other two Newcastle sessions.

31. And in one Newcastle session there were only ten centralized markets.

32. If firms and workers had waited to be matched in round 0 and submitted complete and straightforward rank orderings, the average mismatch rate under Newcastle would have been one pair in each market.

33. A one-sixth chance of being mismatched in both cases and a cost of being mismatched of \$5 and \$10 between low and high cost treatments. All calculations

sessions compared with \$2.77 per market per person in the high mismatch cost sessions. In short, under the Newcastle algorithm, going from low to high mismatch cost sessions the expected cost of being mismatched doubled. This was accompanied by a roughly doubling of the premium high productivity types were willing to pay to avoid these mismatches by matching early (early matches almost never produced mismatches).

Taken together with the results of the sessions in which mismatches were more costly, these results confirm that unraveling is mediated both by the stability and instability of matches, and by the cost of mismatches (cf. Theorem 1 in Roth and Xing {1994}).<sup>34</sup> The differences between the two algorithms are much more important when the cost of mismatching is high relative to the cost of going early. However, what the theoretical results and the field data did not reveal were the details of the transition between decentralized and centralized matching, much of which was unanticipated.

#### IV. DISCUSSION

In the last two decades laboratory experiments have come to be an increasingly well-accepted part of economic research, but the present experiment is unusual in being motivated by a natural experiment. It may therefore be useful to consider not only the results of the experiment, but what they add to the field observations, and what the field observations add to them. In this connection, of course, we need to consider not only the field observations from Britain that motivated the particular comparisons made in this experiment, but the several dozen markets and submarkets in which unraveling of appointment dates is well documented, many of which have experienced some form of market reorganization as a result {Roth and Xing 1994}.

Perhaps the clearest lessons have to do with robustness. In a laboratory market with six positions, we reproduced behavior that had previously been observed in field markets ranging in size from tens of thousands of positions (in the United States) to hundreds of positions (in Britain). So the robustness of these results now

---

here are on an individual agent basis compared with the figures, which report costs for pairs of agents.

34. Roth and Xing {1994} examined a variety of dynamic models, and established that instability of the outcome at the final period is neither a necessary nor a sufficient condition for unraveling to occur at a perfect equilibrium.

spans four orders of magnitude in market size, and many variations on market structure, rules, costs, and supply and demand for positions. This robustness conveys information in two directions—the experimental results show that the field results are even more robust (particularly regarding the size of the market) than could have been anticipated from field data, and the field results show that laboratory markets using relatively small payoffs can reproduce behavior observed in markets with very large (career-shaping) incentives, and with rich institutional structures and histories.<sup>35</sup>

In addition to adding robustness, the results of this experiment also add an element of confidence to the interpretation of the natural experiment observed in Britain (Roth 1991). Edinburgh and Cardiff have other differences from Newcastle and Birmingham than simply the matching algorithm by which they organized (or attempted to organize) their regional medical markets. It is always possible that the differences among these four regions (in market size and concentration, interpenetration with neighboring markets, local culture, etc.) could have contributed to the market outcomes in a way that was incorrectly attributed to the matching mechanism employed. The control available in the laboratory, which allows us to look at markets that differ *only* in the matching mechanism employed, confirms that this difference by itself can have the predicted consequences.

The most interesting and potentially important lessons from the experiment, however, seem to us not to be those that simply reinforce or confirm what has been learned in the field, or demonstrate the power of experimental methods, but those that allow us to make observations that are so far unavailable in field data. These concern the transitions from the decentralized markets to the centralized ones. While there can of course be no

35. Because most laboratory experiments are designed to investigate phenomena that cannot be studied in the field, it is not often that an experiment gives us the opportunity to make this comparison directly, and so a relatively common cause of skepticism about experimental results takes the form of questioning whether results observed in the lab can tap into the phenomena of natural markets involving very high incentives. One way to try to address such questions is with experiments in which the size of the incentives are varied (see, e.g., Slonim and Roth (1998) or Cooper et al. (1999)); another is to try to find natural environments with large incentives which resemble laboratory environments which have produced results of interest (see, e.g., Telser (1995), Gertner (1993), Metrick (1995), and Berk, Hughson, and Vandezande (1996)). The field studies of unraveling in labor markets, and the natural experiment observed in the United Kingdom, presented us with the opportunity of reversing the direction of such investigations, and reproducing in the laboratory a phenomenon already robustly observed in the field.

guarantee that the unobserved transitions in the field markets look like the observed transitions in the lab, the fact that the success and failure observed in the field is reproduced in the high mismatch cost sessions lends some plausibility to the hypothesis that the transitions may also share many common elements. The experimental observations therefore suggest an agenda for future research in the field, and in the theory of two-sided matching, to try to observe and explain the experimentally observed patterns, and investigate their robustness and significance.

The form of the transition we observed in the high mismatch cost sessions is at least a little surprising. It would have been plausible to surmise after reading Roth {1991}—as one of us surmised after writing it—that the unobserved behavior during the transitions from decentralized to centralized matching in the United Kingdom was very different than the transitions we observe experimentally in the high mismatch cost sessions. It might have been supposed that, once a centralized mechanism was adopted, most market participants would be eager to avoid the cost of going early by “giving it a try,” and the differential success of the two algorithms would be due to the different experience that willing participants would have with the two algorithms. In fact, the transitions we witnessed in the laboratory tell a different story. When the costs of being mismatched are high, we see that the unraveling learned in the decentralized markets is not abandoned when the centralized market is introduced. Instead, it goes on (in round 11) pretty much as in round 10, with firms trying to make early matches, and workers accepting good early offers. But, in both markets, congestion forces some firms and workers to match late, even when they try to go early. And then, accidentally, they experience the centralized mechanisms, and have either bad experiences that confirm their desire to go early, or good experiences that increasingly cause (workers especially) to wait for good matches.<sup>36</sup>

36. A reviewer has suggested that this transition behavior might be an artifact of the lack of information about the algorithm in the experimental environment, and that transition behavior in the actual markets employing centralized clearinghouses might be quite different. This is of course one of many empirical questions raised by the fact that the experimental environment is not at all designed as a scale model of any one of the labor markets that have been studied. However, we are skeptical that information about the algorithms is sufficient in any of the actual markets to allow participants to easily optimize their behavior. Even in markets in which the algorithm is described to participants, the description tends to be summary rather than complete, because the handling of special cases can make the algorithms complex. And the process of discovering the markets' strategic properties (let alone disseminating them) can be a lengthy

The fact that workers and firms behaved differently during the transition in these high mismatch cost sessions is also intriguing. In the experiment this clearly must be due to the asymmetry introduced by the fact that firms can choose to whom to make offers, while workers can only accept or reject offers they receive. In the experiment (although not in the field) this is the only asymmetry between firms and workers. But in the field as well as in the experiment, unraveling depends on a willingness on both sides of the market to enter into early contracts. That workers learned to refuse early offers when the stable mechanism was employed, before firms learned not to make them, probably reflects that the cost of unraveling fell differentially on the two sides of the market. In many markets in which centralized clearinghouses have been organized, a preponderance of the organizational effort has been spent on the firms. While in many cases this is dictated by the institutional structure of the market, our experimental results suggest that increased attention to the workers' side of the market might have an important role to play in the probability of a successful transition.

In the experiment we also observed that the deepest and most persistent unraveling involved the high productivity firms and workers. Drawing inferences from this about the behavior of firms and workers in field markets is quite speculative, especially since the way we created more and less desirable firms in our laboratory market is not the result of careful study of some field market. But there is anecdotal evidence from field markets that suggests that this too may be a robust result. (See particularly the account of the market for law clerks in the U. S. federal appeals courts, in Roth and Xing {1994}.)

In closing, we note that two-sided matching markets have been studied with unusual methodological breadth. In addition to field and experimental studies, the subject enjoys a comprehensive body of theoretical literature,<sup>37</sup> and a growing body of engineering-oriented design practice. The usefulness of each of these approaches is enhanced by the insights gained from the others.

---

one—see, e.g., Roth {1984} and Roth and Peranson {1997, 1999}, for attempts in this direction for the American medical clearinghouse that began in the 1950s.

37. Motivated not only by labor markets (see, e.g., Roth and Sotomayor {1990}, Crawford {1991}, Sonmez {1997}, and Blum, Roth, and Rothblum {1997}), but also by marriage (see, e.g., Becker {1981} and Pollak {1994}).

APPENDIX: PAYOFF TABLES  
GAME 1 PAYOFF TABLE

High productivity firms			Low productivity firms			
$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$f_6$	
$w_1 = 15.80$	$w_1 = 15.50$	$w_1 = 15.20$	$w_2 = 15.60$	$w_1 = 15.70$	$w_1 = 15.40$	
$w_2 = 15.10$	$w_2 = 14.90$	$w_2 = 14.40$	$w_1 = 15.00$	$w_3 = 15.20$	$w_2 = 14.70$	
$w_3 = 14.90$	$w_3 = 14.60$	$w_3 = 14.30$	$w_3 = 14.80$	$w_2 = 14.90$	$w_3 = 14.6$	
$w_4 = 5.60$	$w_6 = 5.70$	$w_6 = 5.40$	$w_6 = 5.80$	$w_6 = 5.50$	$w_6 = 5.20$	
$w_5 = 5.00$	$w_4 = 5.20$	$w_5 = 4.70$	$w_5 = 5.10$	$w_5 = 4.90$	$w_5 = 4.40$	
$w_6 = 4.80$	$w_5 = 4.90$	$w_4 = 4.60$	$w_4 = 4.90$	$w_4 = 4.60$	$w_4 = 4.30$	
High productivity workers			Low productivity workers			
$w_1$	$w_2$	$w_3$	$w_4$	$w_5$	$w_6$	
$f_1 = 15.80$	$f_1 = 15.50$	$f_1 = 15.20$	$f_3 = 15.80$	$f_3 = 15.50$	$f_2 = 15.20$	
$f_2 = 15.10$	$f_2 = 14.90$	$f_2 = 14.40$	$f_2 = 15.10$	$f_2 = 14.90$	$f_3 = 14.40$	
$f_3 = 14.90$	$f_3 = 14.60$	$f_3 = 14.30$	$f_1 = 14.90$	$f_1 = 14.60$	$f_1 = 14.30$	
$f_4 = 5.60$	$f_6 = 5.70$	$f_6 = 5.40$	$f_6 = 5.60$	$f_6 = 5.70$	$f_6 = 5.40$	
$f_5 = 5.00$	$f_4 = 5.20$	$f_5 = 4.70$	$f_4 = 5.00$	$f_4 = 5.20$	$f_4 = 4.70$	
$f_6 = 4.80$	$f_5 = 4.90$	$f_4 = 4.60$	$f_5 = 4.80$	$f_5 = 4.90$	$f_5 = 4.60$	
Final matches under Newcastle (when above preferences are submitted):						
Firm:	1	2	3	4	5	6
Worker:	1	6	5	2	3	4
Final matches under deferred acceptance (when above preferences are submitted):						
Firm:	1	2	3	4	5	6
Worker:	1	2	3	5	4	6

GAME 2 PAYOFF TABLE

High productivity firms			Low productivity firms			
$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$f_6$	
$w_1 = 15.30$	$w_2 = 15.40$	$w_3 = 15.10$	$w_1 = 15.20$	$w_2 = 15.60$	$w_3 = 15.70$	
$w_3 = 15.00$	$w_1 = 14.80$	$w_2 = 14.70$	$w_3 = 14.90$	$w_1 = 15.40$	$w_2 = 15.30$	
$w_2 = 14.90$	$w_3 = 14.60$	$w_1 = 14.40$	$w_2 = 14.50$	$w_3 = 14.80$	$w_1 = 15.00$	
$w_4 = 5.20$	$w_5 = 5.60$	$w_6 = 5.70$	$w_4 = 5.30$	$w_5 = 5.40$	$w_6 = 5.10$	
$w_5 = 4.90$	$w_6 = 5.40$	$w_4 = 5.30$	$w_6 = 5.00$	$w_4 = 4.80$	$w_5 = 4.70$	
$w_6 = 4.50$	$w_4 = 4.80$	$w_5 = 5.00$	$w_5 = 4.90$	$w_6 = 4.60$	$w_4 = 4.40$	
High productivity workers			Low productivity workers			
$w_1$	$w_2$	$w_3$	$w_4$	$w_5$	$w_6$	
$f_3 = 15.30$	$f_1 = 15.40$	$f_2 = 15.10$	$f_1 = 15.20$	$f_2 = 15.60$	$f_3 = 15.70$	
$f_2 = 15.00$	$f_3 = 14.80$	$f_1 = 14.70$	$f_3 = 14.90$	$f_1 = 15.40$	$f_2 = 15.30$	
$f_1 = 14.90$	$f_2 = 14.60$	$f_3 = 14.40$	$f_2 = 14.50$	$f_3 = 14.80$	$f_1 = 15.00$	
$f_4 = 5.20$	$f_5 = 5.60$	$f_6 = 5.70$	$f_6 = 5.30$	$f_4 = 5.40$	$f_5 = 5.10$	
$f_5 = 4.90$	$f_6 = 5.40$	$f_4 = 5.30$	$f_5 = 5.00$	$f_6 = 4.80$	$f_4 = 4.70$	
$f_6 = 4.50$	$f_4 = 4.80$	$f_5 = 5.00$	$f_4 = 4.90$	$f_5 = 4.60$	$f_6 = 4.40$	
Final matches under Newcastle (when above preferences are submitted):						
Firm:	1	2	3	4	5	6
Worker:	2	3	1	5	6	4
Final matches under deferred acceptance (when above preferences are submitted):						
Firm:	1	2	3	4	5	6
Worker:	1	2	3	4	5	6



**GAME 3 PAYOFF TABLE**

High productivity firms			Low productivity firms		
$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$f_6$
$w_3 = 15.20$	$w_2 = 15.40$	$w_3 = 15.50$	$w_1 = 15.10$	$w_2 = 15.40$	$w_3 = 15.10$
$w_2 = 14.90$	$w_3 = 15.10$	$w_1 = 15.10$	$w_3 = 14.50$	$w_1 = 14.70$	$w_2 = 14.80$
$w_1 = 14.50$	$w_1 = 14.70$	$w_2 = 15.00$	$w_2 = 14.30$	$w_3 = 14.50$	$w_1 = 14.40$
$w_4 = 5.10$	$w_6 = 5.40$	$w_5 = 5.10$	$w_6 = 5.20$	$w_5 = 5.40$	$w_6 = 5.50$
$w_6 = 4.50$	$w_4 = 4.70$	$w_4 = 4.80$	$w_5 = 4.90$	$w_6 = 5.10$	$w_4 = 5.10$
$w_5 = 4.30$	$w_5 = 4.50$	$w_6 = 4.40$	$w_4 = 4.50$	$w_4 = 4.70$	$w_5 = 5.00$

High productivity workers			Low productivity workers		
$w_1$	$w_2$	$w_3$	$w_4$	$w_5$	$w_6$
$f_2 = 15.20$	$f_3 = 15.40$	$f_2 = 15.50$	$f_1 = 15.10$	$f_3 = 15.40$	$f_1 = 15.10$
$f_1 = 14.90$	$f_2 = 15.10$	$f_3 = 15.10$	$f_3 = 14.50$	$f_1 = 14.70$	$f_3 = 14.80$
$f_3 = 14.50$	$f_1 = 14.70$	$f_1 = 15.00$	$f_2 = 14.30$	$f_2 = 14.50$	$f_2 = 14.40$
$f_4 = 5.10$	$f_5 = 5.40$	$f_6 = 5.10$	$f_5 = 5.20$	$f_6 = 5.40$	$f_5 = 5.50$
$f_6 = 4.50$	$f_4 = 4.70$	$f_5 = 4.80$	$f_4 = 4.90$	$f_5 = 5.10$	$f_6 = 5.10$
$f_5 = 4.30$	$f_6 = 4.50$	$f_4 = 4.40$	$f_6 = 4.50$	$f_4 = 4.70$	$f_4 = 5.00$

Final matches under Newcastle (when above preferences are submitted):

Firm:                    1            2            3            4            5            6

Worker:                4            3            2            1            6            5

Final matches under deferred acceptance (when above preferences are submitted):

Firm:                    1            2            3            4            5            6

Worker:                1            2            3            4            5            6

**GAME 4 PAYOFF TABLE**

High productivity firms			Low productivity firms		
$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$f_6$
$w_2 = 14.90$	$w_3 = 15.20$	$w_3 = 15.70$	$w_3 = 15.90$	$w_1 = 15.60$	$w_2 = 15.30$
$w_3 = 14.80$	$w_1 = 15.10$	$w_2 = 15.20$	$w_1 = 15.40$	$w_2 = 15.50$	$w_1 = 14.80$
$w_1 = 14.60$	$w_2 = 14.50$	$w_1 = 14.90$	$w_2 = 15.30$	$w_3 = 14.80$	$w_3 = 14.70$
$w_6 = 5.90$	$w_4 = 5.60$	$w_5 = 5.30$	$w_6 = 4.90$	$w_4 = 5.20$	$w_5 = 5.70$
$w_5 = 5.40$	$w_5 = 5.50$	$w_4 = 4.80$	$w_4 = 4.80$	$w_5 = 5.10$	$w_6 = 5.20$
$w_4 = 5.30$	$w_6 = 4.80$	$w_6 = 4.70$	$w_5 = 4.60$	$w_6 = 4.50$	$w_4 = 4.90$

High productivity workers			Low productivity workers		
$w_1$	$w_2$	$w_3$	$w_4$	$w_5$	$w_6$
$f_3 = 14.90$	$f_2 = 15.20$	$f_2 = 15.70$	$f_3 = 14.90$	$f_3 = 15.20$	$f_1 = 15.70$
$f_2 = 14.80$	$f_3 = 15.10$	$f_1 = 15.20$	$f_2 = 14.80$	$f_1 = 15.10$	$f_2 = 15.20$
$f_1 = 14.60$	$f_1 = 14.50$	$f_3 = 14.90$	$f_1 = 14.60$	$f_2 = 14.50$	$f_3 = 14.90$
$f_4 = 5.90$	$f_5 = 5.60$	$f_5 = 5.30$	$f_6 = 5.90$	$f_4 = 5.60$	$f_5 = 5.30$
$f_6 = 5.40$	$f_4 = 5.50$	$f_6 = 4.80$	$f_4 = 5.40$	$f_5 = 5.50$	$f_6 = 4.80$
$f_5 = 5.30$	$f_6 = 4.80$	$f_4 = 4.70$	$f_5 = 5.30$	$f_6 = 4.80$	$f_4 = 4.70$

Final matches under Newcastle (when above preferences are submitted):

Firm:                    1            2            3            4            5            6

Worker:                2            3            1            5            6            4

Final matches under deferred acceptance (when above preferences are submitted):

Firm:                    1            2            3            4            5            6

Worker:                1            3            2            6            4            5

GAME 5 PAYOFF TABLE

High productivity firms			Low productivity firms		
$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$f_6$
$w_2 = 14.90$	$w_2 = 15.50$	$w_2 = 15.80$	$w_2 = 15.90$	$w_1 = 15.30$	$w_3 = 15.30$
$w_3 = 14.80$	$w_3 = 14.70$	$w_1 = 15.50$	$w_3 = 15.60$	$w_2 = 15.00$	$w_1 = 14.90$
$w_1 = 14.20$	$w_1 = 14.60$	$w_3 = 15.10$	$w_1 = 15.30$	$w_3 = 14.60$	$w_2 = 14.50$
$w_5 = 5.90$	$w_5 = 9.30$	$w_6 = 5.30$	$w_6 = 4.90$	$w_5 = 5.50$	$w_4 = 5.80$
$w_6 = 5.60$	$w_4 = 5.00$	$w_4 = 4.90$	$w_4 = 4.80$	$w_6 = 4.70$	$w_6 = 5.50$
$w_4 = 5.30$	$w_6 = 4.60$	$w_5 = 4.50$	$w_5 = 4.20$	$w_4 = 4.60$	$w_5 = 5.10$

High productivity workers			Low productivity workers		
$w_1$	$w_2$	$w_3$	$w_4$	$w_5$	$w_6$
$f_2 = 14.90$	$f_2 = 15.50$	$f_2 = 15.80$	$f_3 = 14.90$	$f_1 = 15.50$	$f_3 = 15.80$
$f_1 = 14.80$	$f_3 = 14.70$	$f_3 = 15.50$	$f_1 = 14.80$	$f_3 = 14.70$	$f_1 = 15.50$
$f_3 = 14.20$	$f_1 = 14.60$	$f_1 = 15.10$	$f_2 = 14.20$	$f_2 = 14.60$	$f_2 = 15.10$
$f_4 = 5.90$	$f_4 = 5.30$	$f_5 = 5.30$	$f_6 = 5.90$	$f_5 = 5.30$	$f_4 = 5.30$
$f_5 = 5.60$	$f_5 = 5.00$	$f_4 = 4.90$	$f_5 = 5.60$	$f_4 = 5.00$	$f_5 = 4.90$
$f_6 = 5.30$	$f_6 = 4.60$	$f_6 = 4.50$	$f_4 = 5.30$	$f_6 = 4.60$	$f_6 = 4.50$

Final matches under Newcastle (when above preferences are submitted):						
Firm:	1	2	3	4	5	6
Worker:	5	2	6	4	1	3
Final matches under deferred acceptance (when above preferences are submitted):						
Firm:	1	2	3	4	5	6
Worker:	3	2	1	6	5	4

OHIO STATE UNIVERSITY  
HARVARD UNIVERSITY

## REFERENCES

- Alexander-Williams, J., and Ivor G. Stephenson, "Appointment of Preregistration House Officers," *British Medical Journal*, IX (June 1973), 605–606.
- APPIC—Association of Psychology Postdoctoral and Internship Centers {1998}, "APPIC Internship Match—Q & A" <http://www.appic.org/t06matchqa.html> (accessed February 16, 1998).
- Becker, Edward R., Stephen G. Breyer, and Guido Calabresi, "The Federal Judicial Law Clerk Hiring Problem and the Modest March 1 Solution," *Yale Law Journal*, CIV (1994), 207–225.
- Becker, Gary S., *A Treatise on the Family* (Cambridge, MA: Harvard University Press, 1981).
- Berk, Jonathon, Eric Hughson, and Kirk Vandezande, "The Price Is Right, But Are the Bids? An Investigation of Rational Decision Theory," *American Economic Review*, LXXXVI (1996), 954–970.
- Blum, Yosef, Alvin E. Roth, and Uriel G. Rothblum, "Vacancy Chains and Equilibration in Senior-Level Labor Markets," *Journal of Economic Theory*, LXXVI (October 1997), 362–411.
- Crawford, Vincent P., "Comparative Statics in Matching Markets," *Journal of Economic Theory*, LIV (1991), 389–400.

- Cooper, David J., John H. Kagel, Wei Lo, and Qing Liang Gu, "Gaming against Managers in Incentive Systems: Experimental Results with Chinese Students and Chinese Managers," *American Economic Review*, LXXXIX (1999), 781–804.
- Gale, David, and Lloyd Shapley, "College Admissions and the Stability of Marriage," *American Mathematical Monthly*, LXIV (1962), 9–15.
- Gertner, Robert, "Game Shows and Economic Behavior: Risk-Taking on 'Card Sharks'," *Quarterly Journal of Economics*, CVIII (1993), 507–522.
- Harrison, Glenn W., and Kevin A. McCabe, "Stability and Preference Distortion in Resource Matching: An Experimental Study of the Marriage Market," in R. M. Isaac, ed., *Research in Experimental Economics, Volume 8* (Greenwich, CT: JAI Press, 1996).
- Leishman, A. G., and R. P. Ryan, "Appointment of Provisionally Registered House-Officers by Computer Match," *The Lancet* (August 29, 1970), 459–461.
- Li, Hao, and Sherwin Rosen, "Unraveling in Matching Markets," *American Economic Review*, LXXXVIII (1998), 371–387.
- Metrick, Andrew, "A Natural Experiment in Jeopardy," *American Economic Review*, LXXXV (1995), 240–253.
- Pollak, Robert A., "For Better or Worse: The Roles of Power in Models of Distribution within Marriage," *American Economic Review, Papers and Proceedings*, LXXXIV (May 1994), 148–152.
- Roth, Alvin E., "The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory," *Journal of Political Economy*, XCII (1984), 991–1016.
- , "New Physicians: A Natural Experiment in Market Organization," *Science*, CCL (1990), 1524–1528.
- , "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*, LXXXI (June 1991), 415–440.
- , "Proposed research program: Evaluation of changes to be considered in the NRMP algorithm" (1995), Consultant's report, mimeo and <http://www.economics.harvard.edu/~aroth/nrmp.html>.
- , Report on the design and testing of an applicant proposing matching algorithm, and comparison with the existing NRMP algorithm (1996), Consultant's report, mimeo and <http://www.economics.harvard.edu/~aroth/phase1.html>.
- Roth, Alvin E., and Elliott Peranson, "The Effects of a Change in the NRMP Matching Algorithm," *Journal of the American Medical Association* (September 3, 1997), 729–732.
- Roth, Alvin E., and Elliott Peranson, "The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design," *American Economic Review*, LXXXIX (1999), 748–780.
- Roth, Alvin E., and Marilda Sotomayor, *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*, Econometric Society Monograph Series (Cambridge, UK: Cambridge University Press, 1990).
- Roth, Alvin E., and X. Xing, "Jumping the Gun: Imperfections and Institutions Related to the Timing of Market Transactions," *American Economic Review*, LXXXIV (September 1994), 992–1044.
- Roth, Alvin E., and X. Xing, "Turnaround Time and Bottlenecks in Market Clearing: Decentralized Matching in the Market for Clinical Psychologists," *Journal of Political Economy*, CV (April 1997), 284–329.
- Slonim, R., and Alvin E. Roth, "Learning in High Stakes Ultimatum Games: An Experiment in the Slovak Republic," *Econometrica*, CXVI (May 1998), 569–596.
- Sondak, Harris, and Max H. Bazerman, "Matching and Negotiation Processes in Quasi-Markets," *Organizational Behavior and Human Decision Processes*, XCIV (1989), 261–280.
- Sondak, Harris, and Max H. Bazerman, "Power Balance and the Rationality of Outcomes in Matching Markets," *Organizational Behavior and Human Decision Processes*, C (1991), 1–23.
- Sonmez, Tayfun, "Manipulation via Capacities in Two-Sided Matching Markets," *Journal of Economic Theory*, LXXVII (1997), 197–204.
- Suen, Wing, "A Competitive Theory of Equilibrium and Disequilibrium Unraveling in Two-Sided Matching" *Rand Journal*, forthcoming.
- Telser, Lester G., "The Ultimatum Game and the Law of Demand," *Economic Journal*, CX (November 1995), 1519–1523.