The collapse of a medical labor clearinghouse (and why such failures are rare)

The collapse of a medical labor clearinghouse (and why such failures are rare)

C. Nicholas McKinney, Muriel Niederle, and Alvin E. Roth[¥]

November 3, 2004

Abstract: The collapse of the clearinghouse for the entry-level gastroenterology labor market offers a unique opportunity to study the failures that sometimes afflict such markets.

To explore the reasons for the failure of the clearinghouse, we conduct an experimental investigation of demand shocks of the kind that occurred in the gastroenterology market. We find that a reduction in demand for positions leads to the collapse of the match only when it is detectable by firms before being detected by workers, as in the unexpected shock that took place in 1996, which could be seen by firms in their reduced applicant pools. Simple demand and supply imbalances do not seem to interfere with the operation of the centralized match.

Our results suggest that the market failed when applicants lost their ability to discern when they should reject an early, exploding offer. This suggests that the failure of the gastroenterology clearinghouse is related to the unraveling also seen in a number of decentralized markets.

[¥] C. Nicholas McKinney: Department of Economics and Business Administration, Rhodes College, mckinneyn@rhodes.edu; Muriel Niederle: Department of Economics, Stanford University, and NBER http://www.stanford.edu/~niederle, niederle@stanford.edu; and Alvin E. Roth: Department of Economics, Harvard University, and Harvard Business School, http://www.economics.harvard.edu/~aroth/alroth.html.
This work was partially supported by a grant from the National Science Foundation. We also thank David Brenner, M.D., Fred Gorelick, M.D., and Sean Sheehan, M.D. for agreeing to be interviewed about the demise of the GI match. Robert Beran, Elliott Peranson, and Cody Webb, M.D. helped orient us to the ongoing events in Gastroenterology. We thank David Cutler, Ernan Haruvy, Gerhard Orosel and an anonymous referee for helpful comments.

1. Introduction

Many entry-level labor markets (and quite a few other markets) have at some point in their history suffered from the "unraveling" of hiring decisions. Unraveling is typically a dynamic process, in which offers are made earlier from year to year, and can come to be written quite far before actual employment starts. In markets that experience unraveling, applicants typically receive "exploding" offers that must be accepted or rejected before other offers can be received and considered. The causes of this market failure are however still not well understood. [1]

In many markets there have been vigorous and sustained efforts to halt the unraveling of appointment dates. Efforts to simply impose uniform appointment dates, without any market structure to support them, have most often been unsuccessful. Some markets have successfully reorganized themselves around a centralized clearinghouse, which permits matching at a uniform, efficient time (Roth and Xing 1994).[2]

Not all centralized clearinghouses have been successful. There is a good deal of empirical evidence (see e.g. Roth 1984, 1991, Roth and Xing 1994) that a key element of the design of a successful clearinghouse is whether it produces matches that are *stable* in the sense that there exists no firm and worker who are not matched to one another but who would both prefer to be matched to one another rather than accepting the results of the centralized clearinghouse (cf. Gale and Shapley, 1962, Roth and Sotomayor, 1990).[3] However, some unstable matching mechanisms have been observed to persist for years (cf. Roth, 1991), and a very few stable matching mechanisms have been observed to fail.[4]

We know of about 100 markets and submarkets organized by a stable matching mechanism (often called simply a "match), but we know of only a handful in which a stable match was abandoned after operating successfully for several years. The gastroenterology market is the only one of these for which we have been able to gather evidence that suggests that the demise of the match is primarily related to events within the market itself. It therefore provides us with a unique opportunity to investigate why a stable centralized mechanism failed, while so many others continue to be successfully used.

---

[1] Roth and Xing, 1994, describe several dozen such markets and submarkets. Two markets that have recently been experiencing this kind of unraveling are the market for law clerks for Federal appellate judges (cf. Avery, Jolls, Posner and Roth 2001), in which offers have been made almost two years in advance of employment, and the market for college admissions (cf. Avery, Fairbanks and Zeckhauser 2003), in which elite colleges admit a high percentage of their entering classes "early decision.".

[2] For possible effects of a match on wages, see Niederle and Roth (2003a).

[3] This is a bit of an oversimplification. Many labor markets have special features that complicate the definition of stability: see e.g. Roth and Peranson, 1999.

[4] In theory, stability is neither a necessary nor a sufficient condition for the success of a match (see Roth and Xing, 1994, Li and Rosen, 1998, Li and Suen, 2000, and Suen, 2000).

This is an issue of considerable interest to market designers who are asked to consider how and when stable clearinghouses may improve market performance.[5] While it seems clear that centralized matches that produce *un*stable outcomes have been unsuccessful in organizing markets, we do not know when stable matches are successful. Understanding this will help us understand which other markets that currently do not have a centralized match may be able to successfully introduce a match, and which of the markets that now have a successful match may be prone to fail, and what could be done to help them. Finally, understanding the failure of a centralized match, and hence better understanding environments in which a centralized match succeeds, may also help us understand how *decentralized* markets of this kind clear, or fail to.

To address the reasons for the market failure in the gastroenterology market, after considering the history of the market prior to the collapse of the match, we will turn to laboratory experiments. The available field data, while suggesting hypotheses about the cause of the collapse (some of them put forward by market participants), do not allow these hypotheses to be distinguished, because they are all consistent with the history of the market. Experiments in the laboratory allow us to reproduce and vary on a small scale different conditions of supply and demand, and the kind of shocks experienced by the gastroenterology labor market just prior to the collapse of the match, in ways that the single observation of the history of the market cannot.

2. The Gastroenterology Market

The history of the market for GI fellowships (so called because of the older name of Gastro-Intestinal disease) is similar to that of many medical labor markets (Roth, 1984, Roth and Xing, 1994). Before 1985, it suffered from the unraveling of appointment dates, and a number of solutions were attempted prior to the adoption of a centralized match (cf. Gerson, 1999). In 1986 the Medical Specialties Matching Program (MSMP) was initiated to establish a uniform appointment date and permit applicants to complete at least two years of residency before deciding which sub-specialty to pursue. The fellowship clearinghouse was conducted a year in advance, i.e. after two years of internal medicine residency, and one year before employment

[5] Understanding under what conditions stable matches succeed is of lively interest in several contemporary design efforts. In New York City, the Department of Education initiated a stable match for the approximately 100,000 middle school graduates who entered high school in September 2004. In Boston, a study of the existing, unstable matching system is presently underway, with the goal of recommending design changes. (In this connection, see Abdulkadiroglu and Sonmez 2003.) And a recent anti-trust suit against the National Resident Matching Program prompted new legislation clarifying that such a match does not violate antitrust laws (Public Law 2004), and raised questions about how possible modifications in the match would change the performance of the market (cf Bulow and Levin, 2004, Crawford, 2004).

would begin (NRMP 1999). Applicants and fellowship programs submitted rank order lists of positions and applicants, respectively, that are processed to produce a stable match.

However, after 1996 the centralized match broke down and was abandoned. In 1996 around 300 positions participated in the match, by 1999 there were only 14, and in 2000 no centralized match for gastroenterologists was even attempted.

The demise of the match seems to have been set in motion in 1993-1994, when Gastroenterology subjected itself to a manpower analysis (Meyer et al 1996). Its main conclusions were that the US health care system and gastroenterologists would benefit from a reduction in gastroenterology fellowship programs. The Gastroenterology Leadership Council endorsed a goal of 25% to 50% reduction in the number of GI fellows over 5 years. Furthermore, an additional year of training was mandated: starting in the summer of 1996, three years of training were required to be board eligible, instead of two.

That is, in 1996 the *supply* of gastroenterology fellowships was sharply reduced, by administrative decision, and the time needed to become a gastroenterologist was increased by a year (although some three-year fellowship programs existed before 1996). However, this announced (and hence expected) reduction in supply triggered an *even larger reduction* in the number of residents who applied for GI fellowships. This seems to have been the start of the demise of the match. In 1996, for the first time, and despite the reduction in the number of positions offered, there were fewer applicants for GI fellowship positions than there were positions offered in the match. That is, despite the considerable reduction in the *supply* of positions, the market experienced a shortage in *demand* for positions, as residents stayed away from the market. This seems to have triggered a scramble among fellowship programs (see e.g. Gerson, 1999). Once the match broke down, and the commitment to uniform late appointment dates vanished, the market for gastroenterology fellows once again experienced unraveling and exploding offers (cf. Bauer et al 1999).

One hypothesis about what caused the demise of the gastroenterology match, put forward by market participants in the gastroenterology literature, is that a centralized match can only work when there is a surplus of applicants (cf. Gorelick, 1999, Toskes, 1999).

Of course, there may be different kinds of perceived shortages of applicants. There may simply be fewer applicants than positions, as in 1996. A more subtle kind of shortage is that there may not be enough "high quality" applicants to fill the high quality positions. Indeed, the perception among some GI fellowship directors is that, even though there are now once again more applicants than positions, there are not enough high quality applicants available. Of course, such perceived shortages can also arise in the eyes of applicants, regardless of the number of

positions available. In many markets there appears to be a perception among participants on both sides of the market that they are on the long side (i.e. there are never enough high quality opportunities on the other side of the market, cf. Toskes, 1998).

The matching algorithm used by the NRMP to conduct the medical matches generates a stable outcome regardless of any shortages of positions or applicants. However, the incentives for programs and applicants to participate in the centralized match, or strike a deal outside the match, may change in response to changes in supply and demand.

Before we present further hypotheses about the demise of the GI match, we first need to consider the difference in the strategic options facing firms (fellowship programs) and applicants. Two asymmetries seem particularly important in the present instance:

1. Asymmetry of actions: In both centralized and decentralized markets, applicants apply to firms to be considered for a position (e.g. apply for an interview). In decentralized markets, or when firms do not wait for a centralized market, firms then make offers, and applicants decide whether to accept or reject them.
2. Asymmetry of information following a shock: Receipt of applications gives firms an informational advantage; they know if they are getting many or few applications, and hence they have an early indication of shifts in the demand for positions. In contrast, information about the *supply* of positions is common to both firms and applicants, since available positions are announced well in advance. And in the absence of a shock, historical information will be a reliable guide to both firms and applicants.

This suggests three related hypotheses about why the shortage of applicants for gastroenterology positions in 1996 set off the collapse of the match. Each is a special case of the previous hypothesis.

1. The success of this kind of match depends on there being more applicants than positions: the centralized match fails when there are fewer applicants.
2. The match failed because there was a shock that reduced the demand for positions below the supply, but the match could have recovered from this shock if given the chance, once supply and demand stabilized.
3. The match failed because there was a shock that reduced the demand for positions below the supply, and because firms knew this (because they could see their reduced applicant pool) and applicants didn't. However, the match could have recovered from this shock if given the chance, once supply and demand stabilized, since then applicants would no longer be at an information disadvantage to firms.

While all three hypotheses take as their basis the fact that the gastroenterology match started to unravel the first year in which demand for positions was less than supply, they differ in their implications for a reorganization of the gastro match, for what we should expect to find in other markets with and without successful centralized clearinghouses, and for how such supply and demand shocks might be managed in the future.

3. Experiments

We create a simple matching environment, in which to study the effects of both stationary imbalances in supply and demand, and of shocks that create a shortage of applicants.

In each experimental condition, subjects first gain experience of unraveling by participating in fifteen decentralized markets. They then participate in fifteen markets in which centralized matching is available to subjects who choose to wait and use it. In the treatments that involve a shock, the shock occurs after the 30th market, and subjects participate in an additional 15 markets after the shock, with centralized matching available for those who wait to use it. (We use single shocks that reverse imbalance between demand and supply of positions, such as we observed in the market for gastroenterology fellowship positions.) Firms and applicants are both always fully informed about the number and types of firms in the market. We vary the information that applicants have about other applicants, and examine markets either with *full* information, in which firms and applicants have the same information about supply and demand, or *partial* information, in which only firms are directly informed about the number of applicants. (The partial information condition is intended to let us observe the market in conditions like that of the 1996 shock to the gastro market, which unexpectedly reduced the number of applicants.)

The Environment[6]

Participants in the experiment are assigned the role of either a firm or a worker (i.e. an applicant for work), of one of two types, High or Low. Each subject maintained the same role (e.g. as a High firm) throughout all the markets in an experimental session, and no subject participated in more than one session.[7] In each market, firms can match to at most one worker,

[6] Our environment is similar to the experimental environment in Kagel and Roth (2000), with some small procedural changes. So our results prior to the introduction of shocks serve to replicate their results and add confidence that they are not the artifact of some of the specific procedural choices. The imbalances in demand and supply and the shocks that are the subject of this experiment are unlike the treatments in prior experiments.

[7] In the gastroenterology market, like other entry level labor markets, only employers tend to participate in many markets. However, applicants have many opportunities to learn from and about the experience of

and each worker can match to at most one firm. For each participant, a match to a High type is worth 150 points plus or minus a private value between 0 and 10 points. A match to a Low type is worth 50 points plus or minus a private value between 0 and 10 points. Participants know other participants' types, but not their private values (although each participant knows his own private values for others).

Markets are divided into three periods, and in each period, firms that are not yet matched can make up to one offer. After all the firms make offers (or choose not to make an offer), workers accept or reject offers. An offer contains the type and an identification of the firm that made that offer. A worker who receives several offers only sees them one at the time, in a random order. That is, these are "exploding" offers, a worker must choose to accept or reject each offer without knowing whether he has any more offers coming in that period. Workers can only accept one offer per market. Accepted offers are announced to all participants in any subsequent periods, while unaccepted offers remain private. Matched firms are no longer allowed to make offers and unmatched firms cannot make offers to matched workers.

In the gastroenterology labor market, costs of unraveling consist of many components for both firms and workers, including uncertainty about the quality of the match and loss of planning flexibility. In our experiment, we model these various costs simply, by imposing a fixed cost of contracting early (as in Kagel and Roth, 2000). In particular, each market lasts for 3 periods, denoted periods -2, -1 and 0. To model costs of unraveling each participant who matches in period –2 incurs a cost of 20 points and each participant who matches in period –1 incurs a cost of 10 points. The penalties for matching early are deducted from the participants' earnings. Subjects who failed to match by the end of period 0 receive 0 points for that market. Participants receive $.008 for each point plus a $10 show-up fee.

There are two kinds of markets:

1. In a *decentralized* market, firms have 3 periods to match by making offers to workers, who can accept or reject the offers.
2. In a *centralized* market, the first two periods are as in the decentralized market. But in the last period, period 0, all firms and workers who have not already matched in periods -2 and -1 are matched at the firm-optimal stable matching.[8]

---

previous years, from colleagues, advisors and publications. For simplicity, the present experimental design gives both sides of the market experience in the same way.

[8] When the MSMP operated the gastroenterology match, a version of the firm-optimal deferred acceptance algorithm was used. For ongoing matches, the MSMP subsequently followed the NRMP in adopting the Roth-Peranson (1999) algorithm, built around an applicant-optimal deferred acceptance algorithm.

In each experimental treatment there are 3 low type firms and 3 low type workers. We vary the number of high type firms and workers to induce demand and supply imbalances and the kinds of demand shocks as experienced by the gastroenterology market.

Fixed Demand treatments:

To test for effects of demand and supply imbalances, we have 3 treatments, determined by the number of high type firms and workers. 3Firms – 3Workers is the treatment in which there are 3 high type firms and 3 high type workers (in addition to the 3 low type firms and workers). We compare the outcome of the 4 sessions of this balanced treatment to the 7 sessions in which there is excess supply of workers (3 Workers – 2 Firms) and the 7 sessions in which there is an excess supply of firms (3Firms – 2Workers). In all three treatments, both firms and workers are informed about demand and supply, and participate in 15 decentralized markets followed by 15 centralized markets.

The Shock Treatments:

The shock treatments started out with 3 High type firms and 4 High type workers. Then, after market 30, there was a single shock, eliminating 2 High type workers and hence resulting in a shortage of workers from markets 31 to 45.[9] We ran 10 sessions of each of those treatments.[10]

The shock conditions were conducted under one of two information conditions, Full or Partial information. (The conditions without a shock were all conducted under Full information.) In both information conditions, firms and workers are both fully informed about the number and quality of firms in the market. Firms also know the number and quality of workers. The information condition affected whether workers know about the number and quality of workers.

1. Full Information: Workers are fully informed about the number and types of workers in each market. So, when there is a shock that changes the number of workers, the workers know about the shock as soon as the firms do.
2. Partial Information: Workers are not informed about the number and types of workers in each market. So, when there is a shock that changes the number of workers, only the

[9] Under the Full Information design, the two workers who drop out after market 30 collect their payments and leave while the others finish the last 15 markets. In the partial information condition, these two workers are told to remain quietly at their workstations so as not to signal to the other participants that they have been removed from the market.

[10] All fixed market sessions and 4 of each of the shock sessions were run in the winter and spring of 2001-2002. 6 sessions of the shock treatments were run in the summer of 2003.

firms can see the shock at the start of the market. Workers only find out about other workers when early matches are made.[11]

Experimental Results:

As in Kagel and Roth (2000) the data are best explained by examining the High type participants, since Low types almost universally participate in the centralized match once it becomes available.[12] To filter out the High-Low or Low-High matches that occur from the imbalance of supply and demand, in each market we will look at the High types on the short side of the market. That is we are concerned whether the maximal number of possible High-High matches are achieved in each market.

Figure 1 graphs the percentage of High-to-High matches that are made late (i.e. in period 0). For the sessions that experienced shocks in the balance of supply and demand following market 30, the figure shows how participation in the centralized clearinghouse reacted to the shock, in both the short term (markets 31-35), and over the longer term (markets 31-45) under the new post-shock conditions of worker shortage.

To analyze the data, for each session we use averages over five markets (as in figure 1) as data points. To analyze differences within treatments, we use the Wilcoxon matched-pairs signed-rank test, while we use the Mann Whitney test for differences across treatments.

The effects of a centralized Match under constant market conditions

For all 5 treatments, the decentralized market exhibits substantial early contracting. Furthermore, Markets 6-10 experience the same amount of late contracting as markets 11-15.

For all 5 treatments, the introduction of the centralized match increases the percentage of matches that are made efficiently late: markets 16-20 are significantly different from markets 11-15 in all treatments.[13] The success of the match is virtually the same across conditions.[14] [15]

[11] Accepted offers are made public, along with the type and identity of the firm and worker.

[12] Figures for the Low type participants, screen shots of the experimental conditions, and experimental instructions are found in McKinney, Niederle and Roth (2004) also available on the web at http://www.stanford.edu/~niederle/gastro.experiment.pdf .

[13] The Wilcoxon matched-pairs signed-rank test comparing the proportion of late matching in markets 11-15 with the proportion of late matching in markets 16-20 in the 10 pairs of each of the full and partial information treatments yields *p*-values of 0.04 and 0.01 respectively. Similar Wilcoxon matched-pairs signed-rank tests yield *p*-values of 0.07 for the 4 pairs in the 3 Firms – 3 Workers treatment, 0.02 for the 7 pairs in the 3 Firms – 2 Workers treatment and 0.03 for the 7 pairs in the 2 Firms and 3 Workers treatment.

[14] Comparing the percentage of late matches in markets 26-30 across all conditions using the Mann Whitney test we find p-values that are 0.4 or higher.

[15] For all the markets in which there are 3 possible matches between high type agents, the effectiveness of the centralized match increases as the participants gain experience with it: there is significantly more late matching in markets 26-30 than in markets 16-20. The Wilcoxon matched-pairs signed-rank tests yield *p*-

Both when there was a shortage of firms and when there was a shortage of workers, a centralized market was effective in increasing the percentage of late matches. This increase in late matches is due to a somewhat more efficient period 0 matching mechanism, and mostly to a reduction in early matchings.[16]

There is no suggestion in the data that centralized matches work well only when supply and demand are balanced, or only when firms are on the short side of the market. In each of the experimental sessions with no shocks (or before the shock), the centralized match, once introduced, achieved a steady rate of participation, suggesting that it is robust to varying conditions of supply and demand. Our results replicate and extend those of Kagel and Roth 2000 to the current environment.

The effect of the shock

We are now in a position to consider the effect of the shock that changes the market from one in which workers are in excess supply to one in which they are in short supply, starting in market 31. We have already seen that the centralized mechanism is widely adopted when it is available. In both information conditions its adoption improves efficiency, resulting in significantly higher total payoffs to firms and workers.[17] Prior to the shock there are no important differences between the two information treatments. [18]

---

values of less than 0.01 for both the full and partial information treatment, and yields p=0.07 for the 3Firms - 3Workers treatment. When there are only 2 high matches, the change is not significant, the p-values are 0.102 and 0.369 respectively.

[16] Indeed, comparing the proportion of early matches in markets 26-30 with the proportion in markets 11-15, a Wilcoxon matched-pairs signed-rank test for the 10 pairs of the full and partial information treatments yields p-values of less than 0.01, for the 4 pairs of the 3Firms - 3Workers treatment the p-value is 0.068, for the 7 pairs of each the 2 Firms – 3 Workers and the 3 Firms - 2 Workers treatment the p-values are 0.022 and 0.05 respectively.

[17] A Wilcoxon matched-pairs signed-rank test comparing earnings in markets 16-20 with markets 11-15 in each of the 10 pairs yields values of $p<0.01$ for both the full information treatment and the partial information treatment.

[18] Mann-Whitney tests comparing the average proportion of late matching between the 10 sessions in each treatment, yield p-values of 0.44 for markets 11-15, the last decentralized markets, 0.49 for markets 16-20, the first centralized markets, and 0.84 for markets 26-30 the last centralized markets before the shock. Similar results are obtained when looking at the proportion of early matches. Furthermore, there are no differences in the changes in the proportion in late matchings between the full and the partial information treatment, for the introduction of the centralized match, (markets 16-20 minus markets 11-15), the increase in adoption of the match (markets 26-30 minus markets 16-20), and the difference in late matching between the last centralized markets and the last decentralized markets (markets 26-30 minus markets 11-15). Similar results hold for the proportion of early matches.

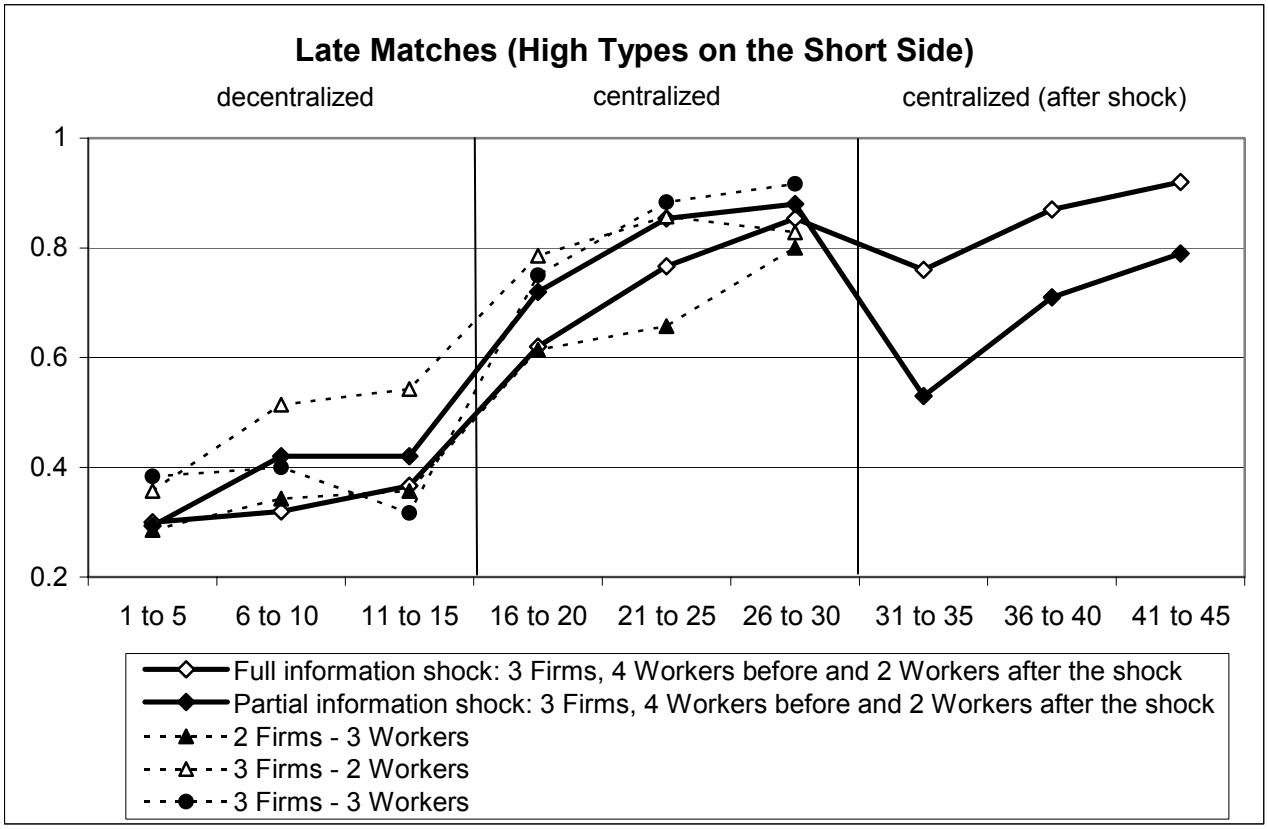


Figure 1: Percentage of High Matches that are made in period 0.

Following the shock that removes two High type workers, unraveling reoccurs in markets 31-35 in both information conditions, but only marginally in the Full Information condition, and significantly in the Partial Information condition. The Wilcoxon matched-pairs signed-rank test that compares the proportion of late matches in markets 25-30 with that in markets 31-35 in the 10 matched pairs of the partial information treatment yields a *p*-value of 0.01, and the test in the full information treatment yields a *p*-value of 0.26.[19]

The impact of the shock is significantly greater in the partial information treatment. Participation in the centralized match falls by 9.3 percent under full information, and 30 percent under partial information when comparing markets 31-35 with markets 26-30. A Mann Whitney test confirms that the lack of information leads to significantly more unraveling ($p<0.05$).

After the shock the amount of unraveling in the partial information treatment, 47% (in markets 31-35) is comparable to the one in decentralized market, 52% (in markets 11-15). A Wilcoxon matched-pairs signed-rank test delivers a *p*-value of 0.33. In contrast, under full information there is nowhere near as much unraveling after the shock (24%) as there was in the

decentralized markets (55%): a Wilcoxon matched-pairs signed-rank test yields a p-value of 0.007. Furthermore, there is a significant treatment effect, that is, the significant change of unraveling in the full information condition is significantly different from the (small) change in unraveling in the partial information condition. A Mann Whitney test on the differences in the proportion of early matches (n=20) delivers a p-value of 0.058, a significant effect.

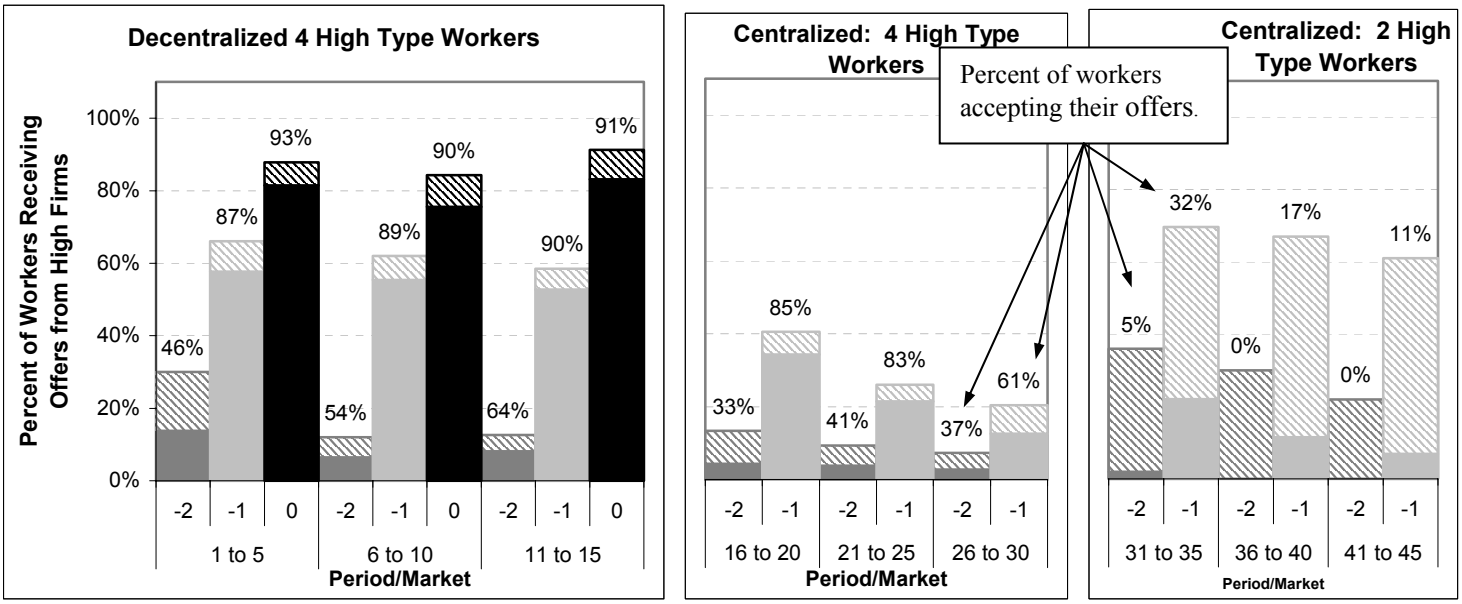
By the final 5 markets, markets 41-45, about 80% of the participants are again matching late (with in each treatment at least half the groups having 100% late matches). That is, the effect of the shock, minor in the full information condition and much larger in the partial information condition, is transient in both conditions. The effect of workers not having information about the number of other workers diminishes as they gain experience of markets with no subsequent shocks in the number of workers.

To understand the effect of the shock, and why it affects the partial information condition so much more than the full information condition, we look at the pattern of early offers by high quality firms to high quality workers, and the high quality workers' acceptance or rejection of these offers following the shock. We will see that, after the shock, when there is a shortage of High type workers, firms increase the number of early offers they make. But the workers accept these offers much less frequently in the full information condition, in which they know they are now on the short side of the market, than in the partial information condition.

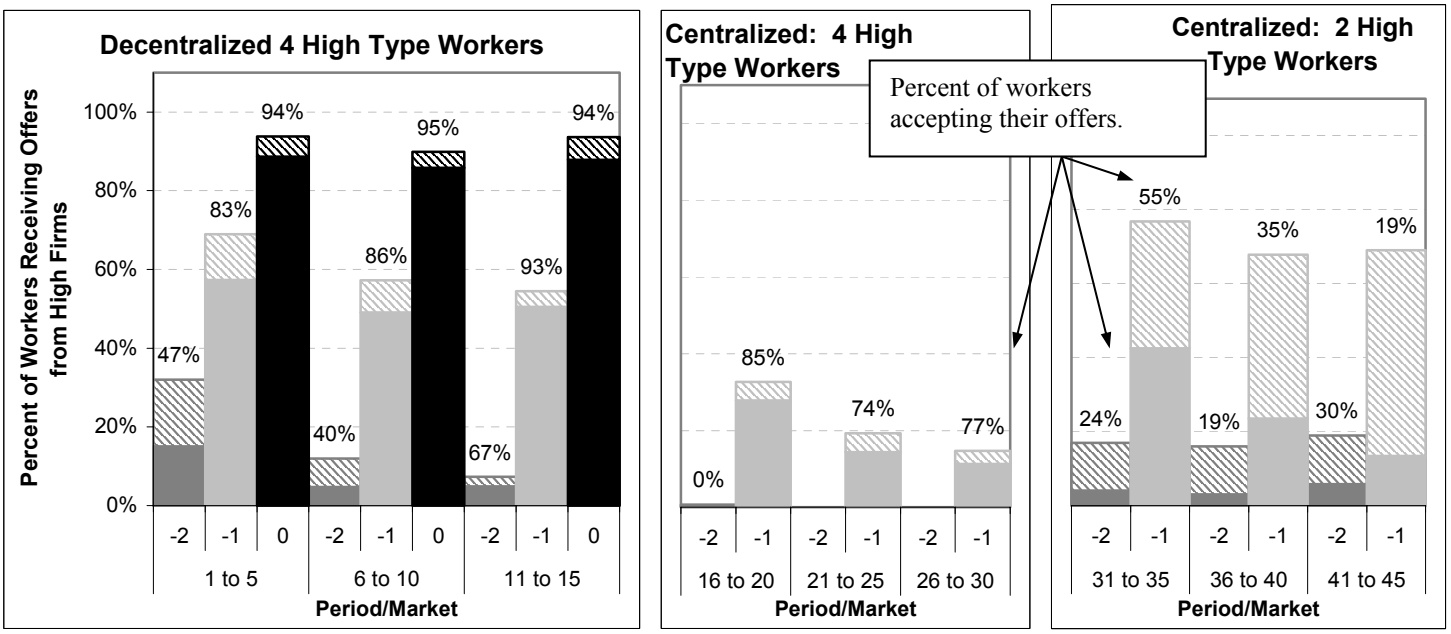
Figure 2 shows the normalized percentage of High type workers who receive an offer from a High type firm, and the solid part of the bar shows the proportion that accept those offers. The percentages are normalized once again by the number of High type participants on the short side of the market, and also by the number of High type participants remaining on the market. That is we report the number of high type workers that receive an offer from a high type firm divided by the number of high type workers who could have received an offer from a high type firm, had all the (remaining) high type firms made offers to different high type workers. (The numerical percentage on top of the bar, and the height of the solid part of the bar, indicate the percentage of these offers that were accepted.)

[19] Whenever a centralized match is in place, there are virtually no high matches that fail to form, that is, the reduction in participation in the centralized match almost equals the increase of early matchings.

Figure 2: Offers, Acceptances, and Rejections
Full information



Partial Information



The percent on the vertical axis is normalized by the number of High type participants on the short side of the market, that are still available. Thus, in markets 1-30, a bar of total height 100% in period –2 would imply that each of the three High firms made offers to a different High type worker. A bar of total height 100% in period –1 or 0 implies that all remaining High type workers minus one (as High type workers are on the long side of the market) receive an offer from a High firm. (The numerical percentage on top of the bar, and the height of the solid part of the bar, indicate the percentage of these offers that were accepted.) After the shock, in markets 31-45, a total height of 100% in period -2 implies that both of the High type workers received at least one offer from a High firm. In period –1 or 0, a total height of 100% implies that each remaining High type worker received an offer from a High firm.

Behavior in the decentralized markets is not surprising. In both treatments the firms are aware that they are in short supply, but to avoid congestion in period 0 a number of early offers are made and most are accepted. In markets 16-30, the centralized match solves the congestion problem and removes the incentives for high quality firms to make early offers to high quality applicants. The number of such offers falls significantly.[20] The proportion of early (period 1) offers continues to fall as market participants become experienced with the centralized match.[21]

After the shock firms are now on the long side of the market and have incentives to make early offers to secure a High type applicant. In both treatments, offers in period -1 increase significantly, by 49 and 62 percentage points, for the complete and partial information treatment respectively (with Wilcoxon matched-pairs signed rank tests comparing separately in each of the 10 pairs markets 26-30 with markets 31-35 yielding p-values below 0.01 in each case.)

However, the reaction of applicants differs sharply between treatments. While firms have an incentive to make early offers to secure a high type applicant applicants now would do better not to accept such early offers. However, in the partial information treatment, applicants are unaware of this change: 55% of the workers accept offers made in period -1 in markets 31-35, compared to 32% under full information (the difference between the two information conditions is significant, the Mann Whitney test (n=20) yields a p-value = 0.0373.

Overall, our experimental results show that a shock to the number of applicants that reverses the imbalance of supply and demand and makes applicants the short side of the market, like the one that the gastroenterology market experienced in 1996, can affect participation rates in the match. The experiments show that this is particularly so when workers do not realize that they are, unexpectedly, on the short side of the market. The results thus tend to support our third hypothesis, that the match failed because there was a shock that reduced the demand for positions below the supply, and because firms knew this and applicants did not. (Recall how the low *demand* for positions in 1996 was an unexpected shock.)

[20] The proportion of high quality applicants who receive an offer from a high quality firm in period -1 falls by 18 percent after the introduction of the centralized match (markets 16-20) in the complete information treatment, and by 22 in the partial information treatment compared to the average proportion over the last five markets before the match (markets 11-15). Both reductions are significant, a Wilcoxon matched-pairs signed-rank test on each of the 10 pairs yields *p*-values of 0.07 and 0.01 respectively. Comparing the 10 pairs of markets 6-10 with 11-15, yields no significant effect ($p>0.4$) and only changes of 3 or 4 percentage points.

[21] From markets 16-20 to 26-30 (from the first five to the last five markets in the match before the shock), the proportion of early offers falls by 20 and 18 percent for the complete and partial information treatment respectively, a significant difference ($p<0.01$ with a Wilcoxon matched-pairs signed rank test).

4. Connections to other markets

A large part of economics is devoted to investigating what markets accomplish at equilibrium. But we know much less about the mechanics of *how* markets clear, or fail to. This gap in what we know is particularly acute for markets, like labor markets, in which workers and jobs are heterogeneous, and so an offer is not just a price offered to the whole market, as in a financial or commodity market. Market failures offer a window through which to investigate this process, and the gastroenterology market offered a particularly clean transition from a late, efficient market organization to one with early, dispersed offers.

In the laboratory, we can see that when matchings become early, they become inefficient, in that they don't maximize the sum of participants' welfare. And in the field data, Niederle and Roth (2003b) found that when gastroenterology fellowship positions were offered through a centralized clearinghouse, the market became significantly more national than it was before the match, or since its demise. The mobility of GI fellows, as measured by whether their GI fellowship is in the same hospital, city or state as their former residency, significantly increased with the use of a centralized match. [22]

For many markets (and marketplaces), the ability to achieve efficient gains from trade has to do with some aspect of the thickness of the market, of how many possible transactions can simultaneously be considered. Dispersion of offers in time is one of the things that can diminish a market's thickness, since agents on both sides of the market are faced with sequences of choices among very few alternatives.

Our results suggest that markets unravel, and thickness is lost, when applicants lose their ability to reject early exploding offers. In the gastroenterology market, this seems to have resulted from a situation that temporarily created incorrect beliefs about supply and demand. More generally, unraveling requires conditions in which there are both firms that want to make early offers and applicants who want to accept them. To create thickness, a market must be attractive enough so that participants want to wait to transact in it, and safe enough for them to do so.

This generalizes to other contemporary market failures as well. The market in which federal appellate judges in the United States hire new graduates of top law schools as law clerks has also been afflicted with unraveling of appointment dates In that market, applicants are strikingly reluctant to reject offers by federal judges who have interviewed them, and indeed it is sometimes an explicit condition for receiving an interview that an offer will be accepted if made

[22] Frechette, Roth, and Unver (2004) show using Nielsen ratings of college football bowl games that unraveling caused inefficiencies in the early 1990s, that have been reduced by later matching.

(Avery, et al. 2001, Becker Breyer and Calabresi 1994.) Much of the market for admission to elite undergraduate colleges has moved early in a similar way, with some colleges explicitly using "binding early decision programs" that favor applicants who apply early to only one college, on the understanding that they will attend if accepted (Avery et al. 2003). In both of these markets, like the gastroenterology market, applicants are at least sometimes faced with early options that they may not hold and compare with other potential offers.

The results of the present experiment suggest the hypothesis that this ability to compare (and reject) offers may be one of the critical but fragile determinants of how well such markets clear.

5. Conclusions

The results of the experiment suggest that the collapse of the gastroenterology match was related to the peculiar situation in which gastroenterology found itself in the late 1990's. After a reduction in the number of gastroenterology positions, the market nevertheless suffered such a shortfall in applicants that applicants were in shorter supply than positions. Not only did this give fellowship programs incentives to fill their positions early, it made applicants (who knew of the reduction in positions, but not of the shortage of well qualified applicants) eager to accept offers whenever they received them.

The results of the experiment also suggest why failures of stable matches are rare. Anticipated imbalances in supply and demand, visible to both sides of the market, did not cause declines in match participation of anywhere near the magnitude caused by the unanticipated shock that created a shortage of workers (that workers were initially unaware of). And while supply and demand shocks themselves may not be rare in these labor markets, shocks that change which is the short side of the market appear to be quite rare. None of the other internal medicine subspecialty matches (Cardiovascular Disease, Pulmonary Disease and Infectious Disease) experienced such a shock.[23]

Since the loss of the match, the number of positions has again stabilized, raising the question of whether this market might be able to successfully employ a match again. Our results suggest that a centralized match would once again be successful. And reintroducing a centralized

[23] From 1990 to 1998 the ratio of applicants to positions offered in the Cardiovascular match varied from a high of 1.6 to a low of 1.3. For Pulmonary disease those ratios varied from a high of 1.5 to a low of 1.1, and for Infectious disease (from 1994 to 1998) those ratios vary from a low of .68 to a high of .92 (Niederle and Roth 2004). Thus, unlike in the Gastroenterology market, the short side of these markets did not change, although in Infectious diseases the applicants were in short supply, and in the other matches the positions were in short supply. Our experiments provide an additional data point that markets in which

match would increase the scope of the market and the mobility of gastroenterologists, as well as allowing matches to be made after more information has become available.

In summary, this paper exploits a rare event—the failure of the stable matching mechanism used to organize the market for gastroenterologists—to explore how such clearinghouses fail, and why failures are rare. The fact that the failure of the gastroenterology match seems to have been the result of an unusual, one-time shock to the market, helps us to advise gastroenterologists that it appears that a reintroduction of the match would once again be successful. And the fact that simpler shocks do not disrupt the operation of the match explains in part why this kind of clearinghouse has operated so successfully, over many years, in markets that have certainly experienced shocks in supply and demand.

More generally, while theories of equilibrium do a good job of explaining what markets should do when they clear, we have few descriptions of how markets (especially those that are not commodity or financial markets) do clear. The results of the experiment reported here underline the importance of those elements of the market environment that gave applicants the confidence to decline early offers.

applicants are in short supply may be equally as stable as markets in which applicants are abundant, and that the success of the market for infectious disease fellowships is not an anomaly.

References

Abdulkadiroğlu, Atila, and Tayfun Sönmez "School Choice: A Mechanism Design Approach" *American Economic Review* 93-3: 729-747, June 2003.

Avery, Christopher, Andrew Fairbanks and Richard Zeckhauser, *The Early Admissions Game: Joining the Elite*, Harvard University Press, Cambridge, MA 2003.

Avery, Christopher, Christine Jolls, Richard A. Posner, and Alvin E. Roth, "The Market for Federal Judicial Law Clerks", *University of Chicago Law Review*, 68, 3, Summer, 2001, 793-902.

William T. Bauer, William Fackler, Kavita Kongara, Christie Matteoni, Bo Shen and Michael Vaezi, *Comment to It's Time to Bring the Best and Brightest Back to Gastroenterology*, Gastroenterology vol. 116, No.4, 1999, 1014.

Becker, Edward R., Stephen G. Breyer and Guido Calabresi, The Federal Judicial Law Clerk Hiring Problem and the Modest March 1 Solution, 104 Yale L.J. 207 (1994)

Bulow, Jeremy and Jonathan Levin, "Matching and Price Competition," working paper, December 2003.

Crawford, Vincent P. "The Flexible-Salary Match: A Proposal to Increase the Salary Flexibility of the National Resident Matching Program", working paper, 2004.

Frechette, Guillaume, Alvin E. Roth, and Utku Unver, "Unraveling Yields Inefficient Matchings: Evidence from Post-Season College Football Bowls" , working paper, 2004.

Gale, David and Lloyd Shapley, "College Admissions and the Stability of Marriage," *American Mathematical Monthly*, 69, 1962, 9-15.

Gerson L. To match or not to match in gastroenterology: an interview with David Brenner, MD, Chair of AGA's Manpower and Training Committee [on-line]. AGA Trainee & Young GI news 1999;5. Available at: http://www.gastro.org/trainee/trainee6.html.

Gorelick, Fred S. M.D., "Striking Up the Match", Comment From the Editors Gastroenterology, Vol 117, 1999, 295.

Kagel, John H. and A.E. Roth, "The dynamics of reorganization in matching markets: A laboratory experiment motivated by a natural experiment," *Quarterly Journal of Economics*, February, 2000, 201-235.

Li, Hao and Sherwin Rosen, "Unraveling in Matching Markets," *American Economic Review*, Vol. 88, 1998, 371-387.

Li, Hao and Wing Suen, "Risk Sharing, Sorting, and Early Contracting", *Journal of Political Economy*, Vol. 108, 2000, 1058-1091.

Meyer, Gregg S., Itzhak Jacoby, Henry Krakauer, Don W. Powell, Jeanette Aurand and Peggy McCardle "Gastorenterology Workforce Modeling", *Journal of the American Medical Association*, September 4, 1996, Vol 276, No.9, 689-694.

National Resident Matching Program, Specialties Matching Services*, MSMP Data, 1999 Appointments*, 1999.

Niederle, Muriel, and Alvin E. Roth, "Relationship Between Wages and Presence of a Match in Medical Fellowships," *JAMA*, *Journal of the American Medical Association,* vol. 290, No. 9, September 3, 2003a, 1153-1154

Niederle, Muriel and Alvin E. Roth, "Unraveling reduces mobility in a labor market: Gastroenterology with and without a centralized match," *Journal of Political Economy*, vol 111, no.6, December, 2003b, 1342-1352.

Niederle, Muriel, and Alvin E. Roth, "The Gastroenterology fellowship match: how it failed, and why it could succeed once again", *Gastroenterology*, 127, 2, August 2004, 658-666..

Pension Funding Equity Act of 2004, Public Law No: 108-218, http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=108_cong_public_laws&docid=f:publ218.108.pdf

Roth, A.E. "The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory", *Journal of Political Economy*, Vol. 92, 1984, 991-1016.

Roth, A.E. "A Natural Experiment in the Organization of Entry Level Labor Markets: Regional Markets for New Physicians and Surgeons in the U.K.*", American Economic Review*, vol. 81, June 1991, 415-440.

Roth, A. E. and Elliott Peranson, "The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design," *American Economic Review*, 89, 4, September, 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 University Press, 1990.

Roth, A.E. and X. Xing "Jumping the Gun: Imperfections and Institutions Related to the Timing of Market Transactions," *American Economic Review*, 84, September, 1994, 992-1044.

Suen, Wing "A competitive theory of equilibrium and disequilibrium in two-sided matching", *RAND Journal of Economics,* 31, Spring 2000, 101-120.

Toskes, Phillip P "It's Time to Bring Back the Best and Brightest Back to Gastroenterology!," *Gastroenterology*, Vol 115, 1998, 1002-1005.

Toskes, Phillip P , *Reply to It's Time to Bring the Best and Brightest Back to Gastroenterology*, Gastroenterology, Vol. 116, 1999, 1014.