NORTHWESTERN UNIVERSITY

Unravelling in Two-Sided Matching Markets and Similarity of Preferences

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Economics

By

Hanna Wiesława Hałaburda

EVANSTON, ILLINOIS

December 2007

ABSTRACT

Unravelling in Two-Sided Matching Markets
and Similarity of Preferences

Hanna Wiesława Hałaburda

This study investigates the causes and welfare consequences of unravelling in two-sided matching markets. "Unravelling" arises when agents contract with one another at an early stage, before much of the relevant information is available. Such early matches may lead to ex-post inefficiencies and are perceived as socially harmful. This study shows that similarity of preferences is an important factor driving unravelling. In particular, under the ex-post stable mechanism (which is the mechanism the literature focuses on), unravelling occurs more often for more similar preferences. This is because when preferences are very similar, many firms are likely to prefer the same workers. Some firms would rather contract early, under uncertainty, than wait for the information and face the competition.

This study also shows that unravelling leads to a loss of welfare. Thus, any Pareto-optimal mechanism must prevent unravelling. Moreover, the ex-post stable mechanism is Pareto-optimal if and only if it prevents unravelling. In markets where the ex-post stable

mechanism unravels, there exist ex-post *unstable* Pareto-optimal mechanisms that stop unravelling and are preferable to the ex-post stable one.

Acknowledgements

First, and foremost I am deeply grateful to my dissertation committee chairman, William Rogerson, for his support and guidance, and time he devoted to my development from a student to a researcher. I am also indebted to my dissertation committee members Kim-Sau Chung, Michael Whinston and Asher Wolinsky for their insightful comments and suggestions.

I also thank David Goldreich, Evgeny Lyanders, Marcin Peski and Balazs Szentes discussions and comments. I gratefully acknowledge the generous financial support from the Center for the Study of Industrial Organization and from Weinberg College of Arts and Sciences.

Finally, I would like to extend my deepest gratitude to my husband, Łukasz Pomorski, for his never-ending help, support, patience and advice.

Dedicated to my grandmother Helena Haƚaburda

Table of Contents

List of Figures

CHAPTER 1

Two-Sided Matching and Unravelling in the Literature

This study investigates causes and welfare consequences of unravelling in two-sided matching markets. A two-sided matching market is a market populated by two distinctive groups of agents, e.g. by firms on one side of the market and workers on the other. Agents on each side are heterogenous in characteristics and have individual preferences over agents on the other side of the market. The goal of every agent is to match with the best possible agent on the other side. Some of the most important social and economic transactions are conducted in matching markets: choosing a spouse, seeking admission to a college, finding employment, etc.

One of the most important problems encountered in many matching markets is "unravelling of appointment dates." This phenomenon occurs when contracts are signed long before the start of the job and before much of the relevant information is available. An example, taken from the market for judicial clerks, is discussed below to illustrate the scope and importance of the problem. Next, the literature on matching markets and in particular on unravelling is reviewed. The following chapters present a model of unravelling, analysis of the welfare losses due to this issue, and a discussion of ways of alleviating the problem.

A judicial clerkship is one of the most desired positions that a new lawyer can take just after he finishes law school. Law clerks are a very important part of a judge's office. Judges depend on their law clerks to aid them with their workload. The best candidates for clerks are students who have achieved good academic record, served on the law review board and performed well in moot court. However, the current system of employing clerks is such that binding agreements between judges and students are made before most of this important information is available.

Initially, positions of clerks were filled by students when they were in their last semester of law school. This provided plenty of information to the judges. However, judges competing for the best clerks were making offers and pressing students for acceptance earlier than the competition. Due to this process, appointments were made earlier from year to year, until students where appointed at the end of their first year of law school for a clerkship position that would not start until students graduated two years later. At the time of the appointment only students' first semester grades were available. With so little information, it is hardly surprising that by the start of the job many matches turned out to be mismatches.

What the market for court clerkships has experienced is a typical example of unravelling. The remainder of this chapter explores how the understanding of two-sided markets, and especially unravelling, has developed in the economics literature.

1.1. Gale and Shapley (1962)

Gale and Shapley (1962) is the first paper in the academic literature that discusses a two-sided matching market. The goal of Gale and Shapley (1962) was to investigate the matching process during college admissions. Every college can admit many students, but each student can attend only one college. Such a matching is called "many-to-one matching". Gale and Shapley (1962) first turned to a simpler case, where every college can admit only one student (a one-to-one matching). To give it a more natural interpretation, they called it a marriage market, with men and women looking for spouses. Since then, throughout the whole literature the one-to-one matching problem is called the "marriage problem", and the many-to-one matching problem is called the "college admissions problem", even if they describe, for instance, a labor market.

Gale and Shapley (1962) defined a *stable matching* to be a matching such that no man and woman would prefer to be matched with each other rather than with their currently assigned "spouses", and no man or woman would rather remain single than stay with their "spouse". This property has been considered desirable in a matching market.

The main result in Gale and Shapley (1962) is that in every market there exists a stable matching. This result is proven by construction, with the help of the *deferred acceptance algorithm*. The algorithm proceeds in several rounds. In the first round all men propose to their most preferred woman. A woman who receives more than one proposal rejects all except for the best one. In the next round all rejected men propose to their next-most-preferred woman. Again any woman holds on to only her most preferred offer and rejects all others. The process repeats in this way each round. The algorithm stops when no man is rejected.

The matching resulting from the deferred acceptance algorithm is stable. This is because every woman that a man would prefer to his match must have rejected him in the algorithm. And as for a woman – any man that she would prefer to her match must have not proposed to her, which means that every man that she would prefer is matched with his more preferred woman.

1.2. Unravelling in Two-Sided Markets

The subject of the present study is unravelling - a phenomenon that frequently arises in matching markets and that is often viewed as socially harmful by market participants. A market "unravels" when matches (contracts) in that market are made very early in time, before much of the relevant information is available. An extreme example of unravelling comes from the market for first-year law students, who in the 1980's were recruited for their summer internships before their first class in law school.[1] Naturally, at that stage the quality of the students can only be very roughly approximated and many of these early matches turn out to be inefficient by the time students finish their first year of classes.[2]

The phenomenon of unravelling was described for the first time in the academic literature – but not yet named so – by Roth (1984), who discussed the market of medical interns in the US as a case study in game theory. Roth (1984) studied a specific problem the market experienced as well as attempts, both unsuccessful and successful, to solve it.

A medical residency, or internship, is a position that every medical student needs to take after graduating from medical school in order to become an independent physician. Initially a student would secure his internship position only towards the end of his senior year in medical school. Since the establishment of this system at the beginning of the twentieth century, there have been more positions offered than students graduating from US medical schools. The competition among hospitals resulted in their attempts to make binding agreements with students a little earlier than their competitors. Thus, the date at which such appointments were made moved earlier from year to year. The phenomenon

[1] Roth and Xing (1994)
[2] A more detailed analysis of this problem follows bellow.

of these early agreements had already been noted as a problem by 1926. By 1944 the interns were appointed even as early as 2 full years before the internship would begin.

The first step to alleviate the problem was to institute a uniform date for making appointments. However, this effort did not prove fully successful. Around 1944, the Association of American Medical Colleges adopted a proposal that no information about medical students (transcripts or letters of reference) would be released by medical schools before the end of the junior year for students seeking internships starting in 1945. Thus, recruitment would be possible only a year in advance.

This method was effective in preventing appointments sooner than that. However, a new unexpected problem appeared. Students held on to multiple offers until the day they were required to accept one. The rejected hospitals would then make the vacated offer to the candidate next on their list, but often that candidate had by then already accepted another offer, sometimes inferior. Both the student and the hospital with an open offer were left unsatisfied. Reneging on an accepted offer is frowned upon, and if it does happen it leaves the abandoned hospital in a highly undesirable situation. The period between offers and students' decisions was shortened over the six years between 1945 and 1951 from 10 days to 12 hours. This, however, was not a satisfactory solution.

The major problem with unravelling was the poor quality of matches, i.e. a large number of mismatches. Even though under the uniform appointment dates the contracts took place much later, the quality of the matches did not improve as it was expected to. Neither side of the market was content with the new arrangements.

After this unsuccessful attempt to organize the market, a centralized clearinghouse was implemented. The clearinghouse operated on the basis of the rankings that both medical students and hospitals submitted. Given the rankings, the clearinghouse computed a stable matching and announced that matching to market participants. This centralized procedure was successful and stopped unravelling, although that did not happen immediately. The first algorithm, used in 1951, raised objections that a student who submitted a ranking according to his true preferences might receive a less preferable match than if he had submitted a different rank order. In response to these objections, a new algorithm was used in 1952. This algorithm was similar to Gale and Shapley's deferred acceptance algorithm. Interestingly, it was implemented 10 years before the paper describing the algorithm was published.

Roth (1984) undertakes an investigation into why the early attempts to stop unravelling were unsuccessful, but the stable clearinghouse did stop the problem. He suggests that the final centralized procedure was successful because it always produced a stable matching. The other methods produced unstable matchings, and the perspective of instability gave students and hospitals incentives to contract early. An algorithm guaranteeing the stable matching would take away this incentive, eliminate the reason to contract early, and thus stop unravelling.

The argument that a stable clearinghouse, i.e. a clearinghouse that always produces a stable matching, would stop or prevent unravelling became later known as the "stability hypothesis." The stability hypothesis quickly set off to be a major concern in papers analyzing two-sided matching markets and a number of studies investigated its validity. For example, Roth (1991) describes several regional UK markets for preregistration house-officers – a position similar to medical interns in US. There are 6 regional markets in UK: Newcastle, Birmingham, Edinburgh, Cardiff, Cambridge and The London Hospital Medical College. All of them experienced unravelling by the mid-twentieth century. Encouraged by the success of the clearinghouse in the US market, the markets in UK also substituted the decentralized market with centralized clearinghouses. At that time the particular algorithm that a clearinghouse used to produce matching was not considered important and different regional markets ended up employing different algorithms. Apart from the stable matching described above, some markets used "priority algorithms" or "linear programming algorithms."[3] These two algorithms do not guarantee that the outcome will be a stable matching.

The fact that different markets chose different algorithms to implement matching makes the natural experiment in the UK medical market an ideal testing ground for the stability hypothesis. Roth (1991) argues that while the clearinghouses that implemented stable matching were largely successful in eliminating unravelling, those that imposed priority algorithms were not, presumably because they tended to produce unstable matching. This interpretation is in line with the stability hypothesis proposed in Roth (1984). On the other hand, Roth (1991) also describes markets using the linear programming algorithm. Although this algorithm sometimes produces unstable matchings, clearinghouses that adopted it prevailed and succeeded in eliminating unravelling. In fact, even today

[3]Priority matching used the product of student's and hospital's ranking to determine priority of matching them as a pair. E.g., if the student and the hospital ranked each other first, it was (1,1) match, with priority 1. If the student ranked the hospital second instead, it was a (1,2) match and had priority 2. In linear programming matching, every position in a ranking is assigned a weight (more preferred partners have larger weights). Weight of every potential pair is based on the sum of the weights assigned to the respective ranks. The objective of the linear programming problem is to maximize the sum of match-weights. (Roth, 1991)

there is no unravelling and the algorithm is still in use in Cambridge and in The London Hospital Medical College. These examples go against the stability hypothesis. Roth (1991) holds that it is the small size of the market and thus high reputation and social costs that stopped the unstable mechanism from unravelling.

Mongell and Roth (1991) focus on unravelling in sororities and fraternities. These organizations initially admitted new members only in their senior year in college. However, by the second half of the nineteenth century, even freshmen were admitted. This phenomenon could be interpreted as unravelling: fraternities compete for new members and decide to offer them admission earlier and earlier, perhaps before important information about their qualities and potential can be learned. In some cases, even the students of the preparatory schools connected with the colleges were pledged and initiated. For sororities,[4] this situation lasted until 1928, when the National Panhellenic Conference adopted the "preferential bidding system" – a centralized algorithm producing matchings based on rank-orderings submitted by students and sororities. The algorithm is not a stable matching algorithm, and produces unstable matchings with respect to the reported preferences. However, the participants, aware of this property, misreport their preferences in such a way that the matching produced by the system is stable with respect to the *true* preferences. Thus, the stability hypothesis cannot be tested and rejected based on the behavior of students admitted into sororities.

The results in Roth (1984) and Roth (1991) were formulated based on field observations and natural experiments. Since the environment is not controlled in these, there

[4]Fraternities did not revert to a centralized matching mechanism. However, in 1930's the popularity of preparatory schools declined, so in a natural way fraternity recruiting was limited to the freshmen year in college.

is a suspicion that the relation between instability of matching and unravelling is only accidental. Kagel and Roth (2000) complement Roth's (1991) findings by conducting a laboratory experiment to provide more evidence on the stability hypothesis. The market in the experiment was initially decentralized, and agents preferred to contract early. Then one of two types of a clearinghouse was introduced. One of the clearinghouses was a clearinghouse producing stable matching, and the other was one using the priority algorithm from the UK medical market. In the experiment, unravelling persisted under the introduction of the priority algorithm, whereas unravelling stopped under the stable clearinghouse. Thus, the findings of Kagel and Roth (2000) lend support to the stability hypothesis. Unfortunately, Kagel and Roth (2000) only considered two of the three mechanisms featured in Roth (1991): the priority and stable algorithm. They did not consider the third possibility of the linear programming algorithm, even though that matching method successfully stopped unravelling in some of the UK medical markets. This gap was later filled by Unver (2001) and Unver (2005), which are discussed below.

In an overview paper, Roth and Xing (1994) summarize the history of several markets that experience, or have experienced, unravelling. These markets include a variety of medical markets, two legal markets (the market for Federal court clerkships and the market for new associate positions in large law firms), the market for football bowls, the market for new humanities and social-science graduates of elite Japanese universities, as well as fraternities and sororities. Roth and Xing (1994) identify four stages that markets experiencing unravelling go through. In stage 1, in a decentralized market contracts are signed earlier from year to year. In stage 2, a market organization (either an already existing one or a new one created for the purpose) attempts to enforce uniform appointment dates. Before such a date, offers must not be made, sometimes even interviews should not take place. Stage 3 is the introduction of a centralized matching mechanism. In stage 4 a centralized mechanism is in place, but a significant proportion of contracts is arranged before the operation of the mechanism.

The attempt to impose uniform appointment dates at stage 2 istypically the first response to unravelling. Some markets move directly from stage 1 to a centralized clearinghouse of stage 3. However, some central mechanisms are not successful in stopping unravelling. In such a case the market either moves to stage 4, or the mechanism is altogether abandoned and the market reverts to stage 1.

Roth and Xing (1994) carry out some additional analysis that sheds light on the stability hypothesis. They present theoretical examples of markets that unravel even though the clearinghouse imposes a stable matching, as well as examples of markets that do not unravel even if the matching in the second period is unstable. However, Roth and Xing (1994) do not identify any systematic factors that may drive their results and it has been unclear whether their examples constitute isolated anomalies or are perhaps a manifestation of deeper forces that cause unravelling.

An interesting complement to the theoretical analysis in Roth and Xing (1994) was provided by Unver (2001). Unver (2001) fills a gap in Kagel and Roth (2000) and investigates how successful the linear programming algorithm is in stopping unravelling. Two regional medical markets in the UK, Cambridge and The London Hospital Medical College, have used this algorithm for over 30 years. Despite the potential instability that the algorithm produces, these markets have not experienced unravelling since the introduction of the clearinghouse. That is not consistent with the stability hypothesis.

Roth (1991) sees the small size of the markets and social aspects as the factors preventing those markets from unravelling. Unver (2001) proposes that the reason for no unravelling may lay in the mechanism itself. He uses simulations of an evolutionary environment to show that the linear programming algorithm does not necessarily unravel. Unver (2005) complements this result with a laboratory experiment involving human decision makers. The design of the experiment is similar to Kagel and Roth (2000). Unver (2005) shows that agents adjust to the linear programming mechanism in the long run.[5] Moreover, agents behave strategically and manipulate their rank-order lists in such a way that they obtain almost stable outcomes.

The findings of Roth and Xing (1994) and Unver (2001) can be alternatively explained by the analysis that follows in Chapters 2 and 3. Chapter 2 shows that in markets where preferences are very similar, the stable clearinghouse does not necessary stop unravelling. Chapter 3 demonstrates that in such markets an unstable algorithm may be more successful in preventing unravelling than the stable one.

Sonmez (1999) studies incentives of agents to pre-arrange matches before the operation of a stable clearinghouse. The goal of the paper is to try to explain why many markets (e.g., the market for hospital interns in US or the market for new lawyers in Canada) still unravel in spite of clearinghouses that implement stable matching. In particular, he proposes that it is the stable matching itself that gives market participants an incentive to contract early and later manipulate their rankings submitted to the clearinghouse. For example, such a manipulation may arise when an employer has multiple positions to fill

[5]This is not true in the short-run. In Unver (2005) both simulations and the laboratory experiment indicate that in the short run linear programming performs as poorly as priority matching (and far worse than the Gale-Shapley mechanism) in terms of preventing unravelling.

and has very specific preferences over the set of workers it hires with possibly different skills (e.g., a hospital may seek to hire both an anesthesiologist and a surgeon). Since the stable matching is implemented in the worker-by-worker fashion, it is likely that such a firm will not get its preferred combination of workers. By pre-arranging a match with one worker (who may be made better-off than in the stable matching), the firm may end up with a better combination in the stable matching over the remaining candidates than without manipulation.

Papers reviewed above – except for Roth and Xing (1994) – examine unravelling only in markets where there are no side payments. Salaries in medical markets in the US and the UK are not subject to negotiation, and there are no side payments in sororities and fraternities. In contrast, Roth and Xing (1994) show that unravelling could occur also in markets where side payments are possible and cannot always be alleviated if employers offer different workers different wages. An example of a market with negotiable wages where unravelling occurs is the market for new lawyers. The competition of this market resulted in both unravelling of appointment dates and salary wars.[6]

Several markets described in Roth and Xing (1994), namely the market for federal court clerks, for gastroenterologists and the market of post season college football bowls were the subjects of further studies.

Avery, Jolls, Posner, and Roth (2001) give a detailed and extensive description of the market for federal judicial law clerks and the unravelling it experiences. They analyze data from surveys of law students and judges. Binding agreements that appoint the clerks (effective after they graduate from law school) are made two years in advance. The

[6]see Roth and Xing (1994)

interviews are conducted early, and it is customary for the judges to make the offer during the interview, and for the applicant to accept or reject the offer on the spot. Such early contracting, while prevalent,[7] is recognized as a significant problem by both academics and market participants. Avery et al. (2001) also discuss the possibility of a reform of the market for federal court clerks. They suggest that many of the solutions used in other markets affected by unravelling are unlikely to work in the judicial clerk market. The most promising method is perhaps the stable clearinghouse used in the US medical market. But since the medical-style matching process allows for informal but binding agreements at the interview stage, it may not work for the clerks market. This is so especially due to the current market culture where binding pre-arranged agreements are the common way to contract.

To alleviate the problem of informal agreements, Avery et. al. (2001), propose a modified matching procedure which imposes an additional constraint: those judges who wish their clerks to be eligible for United States Supreme Court clerkships (one of the most desired positions) would be required to participate in the centralized matching system.

Haruvy, Roth, and Unver (2006) further explore the reforms proposed in Avery et al. (2001). They conduct laboratory experiments and a computational study to analyze incentives and behavior of the agents under various proposed reforms of the clerks market. They find that, indeed, the most prominent obstacle to reform is the obligation perceived by many law students to accept an offer if one was made while interviewing for a position. The institution of centralized matching – so successful in the medical market – is unlikely

[7]In September of 1998, the Judicial Conference of the United States abandoned as unsuccessful the attempt to enforce uniform appointment dates for clerks to Federal appellate judges. That was the sixth such unsuccessful attempt to reform the market since 1978.

to reverse or stop unravelling and improve efficiency and welfare in the market, as long as this specific element of the market culture is prevailing.

Frechette, Roth, and Unver (2006) discuss successful reorganization of the market for post-season college football bowl games. Bowls strive to match teams that are ranked 1 and 2 nationally, as such championship games attract most viewers. The number of viewers is the measure of output in this industry. Before 1992 that market experienced severe unravelling. Binding agreements between teams and bowls were made so early that teams still had to complete four games remaining in the regular season after agreeing to the bowl. Within those four games the relative standing of a team could change substantially. Agreements were made so early that it was practically impossible to predict which teams would win college conferences, and thus almost impossible to get two highest ranked teams nationally to play opposite each other in a bowl. The National Collegiate Athletic Association (NCAA) attempted to prevent this unravelling for several seasons, until it gave up the effort altogether following the 1990-1991 football season.

A series of changes in the market have been made since 1992,[8] and have lead to a more centralized allocation process. Eventually, contracting was successfully postponed until the end of the season. By 1998 all the bowls organized themselves (without the interference of the NCAA) in the Bowl Championship Series. The Bowl Championship Series includes all conferences and all bowls. Teams nationally ranked 1 and 2 play a national championship game under this system. Other bowls are pre-set matches between champions of certain conferences. Since there are no arrangements with individual teams

[8]In years 1992-1994 the Bowl Coalition included all major conferences except for the Pacific Ten and Big Ten conferences, and all major bowls except for the Rose Bowl. During 1995-1997, the Bowl Coalition turn into the Bowl Alliance. The Bowl Alliance had the same members as the Bowl Coalition, but better organization.

before the final rankings and conference champions are known, the bowl matches have proven to be more efficient, in terms of the viewership.

Another market that has received substantial attention in the literature is the market for gastroenterology fellows. Medical students who want to specialize in gastroenterology need to complete a gastroenterology fellowship after their residency. The changes in the organization of the market for those fellowships provided a rare opportunity to study the failure of clearinghouses that produce stable matchings, especially the causes and consequences of such failure.

Prior to 1986 the market for gastroenterology fellowships – like the markets for fellowships in many other medical specialties – experienced unravelling. In 1986 gastroenterology joined other specialties in the Medical Specialties Matching Program, which uses the same clearinghouse as the match for medical interns. Up to 1995 the program was successful, until a planned reduction in number of fellowship position began in 1996. At that time the fellowship programs were surprised to notice that the other side of the market responded with an even higher reduction in the number of applicants. Residents – unaware that they were on the short side of the market – were willing to accept early offers. Once some of the matches were being made before the operation of the clearinghouse, the gastroenterology programs and applicants lost confidence in the effectiveness of the centralized mechanism and preferred to contract early. The clearinghouse for gastroenterology was suspended after 1999. Only 14 positions (from above 300 in the market) participated in the match that year. After 1999 the market organization was the same as before 1986.

A major concern in markets that unravel is the loss of efficiency caused by contracting before important information is available. However, inefficiency (and potential lack of stability) of matches may not be the only drawback of unravelling, as discussed in Niederle and Roth (2003). They show that unravelling reduces mobility in the market. To illustrate the idea they use the example of the gastroenterology market. Niederle and Roth (2003) find that before and after the clearinghouse was in use, residents were less mobile. That is, they stayed in the same hospital in which they completed their residency with a higher probability. In contrast, during the years of participation in the clearinghouse, the fellows were more likely to move to a different hospital, city or state when finishing their residency and starting the fellowship. One cause of the tendency to stay in the same area may be that unravelled markets seem to rely heavily on "networks" to hire candidates. Such networks form more closely with geographical proximity – denser networks within a hospital than between hospitals, or between hospitals in the same city than those in different states. These findings suggest that the clearinghouse increased the scope of the market, which likely improved efficiency in yet another way.

In a follow-up paper, McKinney, Niederle and Roth (2005) come back to the gastroenterology market. They investigate – through laboratory experiments – what factors might have contributed to the failure of the stable clearinghouse in the gastroenterology market. The central finding of this paper is that shocks to supply and demand can lead to collapse of the clearinghouse, but only in the presence of asymmetric information about the shock. In 1996 applicants in the gastroenterology market were on the short side of the market, but they were not aware of that. The candidates were expecting a higher probability of remaining unmatched after the operation of the clearinghouse. The experiments in McKinney, Niederle and Roth (2005) show that when shocks are common knowledge, they do not impair the effectiveness of the clearinghouse.

As shown by Niederle and Roth (2004), other features of the market that affect unravelling are commitment and the presence of exploding offers. Based on observations of real life examples as well as on a laboratory experiment, Niederle and Roth (2004) demonstrate that a market unravels only when firms can make exploding offers, and acceptance of such offers is binding. Without a deadline on the offers, or without commitment, early offers would not lead to inefficiencies. Without commitment, inefficient matches simply could not hold up. Moreover, without a deadline on the offer, workers would hold all their offers until uncertainty is resolved. Clearly, in such a case, it is not optimal for firms to make early offers. Thus, in markets without commitment and deadlines on the offers unravelling should not arise.

The previous papers approached unravelling by studying the stability of the matching that arises after the arrival of information. Another strand in the literature studies unravelling from the competitive equilibrium perspective. This approach leads to a different channel that may generate unravelling: risk-aversion and insurance obtained by contracting early. All papers described below consider models where side payments can be made between agents.

Li and Rosen (1998) consider insurance aspects of early contracting and characterize price competition and unravelling based on incomplete information about workers' future productivity. In their model, unravelling is socially desirable as it provides insurance for risk averse agents (workers) in the absence of complete insurance markets. The model introduces idiosyncratic uncertainty about the traits of a worker: he could be endowed with high or low productivity. All workers have an incentive to match with firms, but only firms employing high skill workers can obtain positive payoffs. Workers are initially uncertain about their own productivity and about how many productive workers exist in the marketplace. Because of that, they face the risk that their own productivity will turn out to be low or even if it is high, there will be a large number of other skilled workers who will compete away the salaries. These risks induce workers to contract early, before their characteristics are known. Unravelling therefore arises in equilibrium. On the one hand, it makes workers better off (they insure away some risks). On the other hand, it introduces inefficiencies in the firm-worker assignment. Such inefficiencies arise because parties contract before information about workers' productivity is available. Li and Rosen (1998) show that unravelling is more likely when the pool of potential workers is small, heterogenous, and skewed towards less-promising applicants. Note that the principal force driving unravelling in Li and Rosen (1998) is risk aversion. With risk-neutral workers the rationale described above would not be valid. Although risk aversion likely makes unravelling more prevalent, Chapter 2 shows that it is not necessary to generate the effect in the first place.

Li and Suen (2000) propose a similar model to investigate how aggregate and individual uncertainty affects early contracting in equilibrium. They also study the trade off between risk-sharing and inefficiency of matches that are made in the absence of full information about the workers' productivity. They find that more promising candidates are more likely to contract early and less promising agents are more likely to wait. This is because for the more promising candidates the insurance gains outweigh the sorting inefficiency. The difference becomes more significant as risk-aversion of workers increases.

Suen (2000) demonstrates that this type of unravelling is a result of risk aversion of workers, and that with risk-neutral workers (firms are assumed to be risk neutral in both cases) unravelling will not occur. This is because only with risk-aversion doe insurance gains outweigh the loss from inefficient matching resulting from early contracting.

As shown in Chapter 2, this result is not general. It holds for the special case of independent preferences of the firms. However, when firms' preferences are allowed to be similar (i.e., dependent on one another), unravelling can be obtained even with risk-neutral agents.

1.3. Other Issues in Two-Sided Matching Markets

Unravelling is not the only issue of interest in the literature on two-sided matching markets. Before the phenomenon of unravelling was noticed, the major concerns discussed in the literature included the properties of stable matching, extension of the results to the matching environment with side payments, as well as to many-to-one matching environment. Later, about the same time as the surge of attention on unravelling, came interest in strategic behavior of agents in two-sided matching markets, and strategy-proofness of matching mechanisms. Theses topics are shortly reviewed and discussed below.

Matching markets are also studied in the macroeconomics literature. The modeling method differs considerably from the one introduced by Gale and Shapley (1962) and continued by the subsequent studies. This review of issues of interests in matching markets finishes with a short comment on this alternative approach to the problem.

1.3.1. Properties of Stable Matching

The main result in Gale and Shapley (1962) is that in every marriage problem there always exists a stable matching. They prove this result with help of the deferred acceptance algorithm, where men propose to women.[9] However, the procedure may be reversed with women proposing to men. Typically the two algorithms return different stable matchings, as in a typical marriage market there are multiple stable matchings. Gale and Shapley (1962) also show that the stable matching resulting from the men-proposed algorithm is the men-optimal stable matching, meaning that no other stable matching is preferred by any man. At the same time, it is also the worst possible stable matching for women. Of course, the stable matching resulting from the women-proposed algorithm is women-optimal: the best stable matching for women and the worst stable matching for men.

Gale and Sotomayor (1985b) establish that there is also no unstable matching that would make all men better off than the men-optimal matching. They also demonstrate that all agents matched under one stable matching are matched under any stable matching (but they may be matched to different partners), and that in general it is not possible to match all men and women in a stable matching, even when their numbers are equal. Without restrictions on preferences, it may happen that a stable matching leaves some agents unmatched, even when there are as many men as women.

Crawford (1991) conducts a comparative statics analysis for matching markets. He finds that if a new woman enters the market, then all men are weakly better off in the men-optimal matching, and all women are weakly worse off.

---

[9]This algorithm is described at the beginning of current chapter.

1.3.2. Two-Sided Matching with Side Payments

The algorithm in Gale and Shapley (1962) did not involve side payments between men and women. But in many markets side payments are important part of a contract, for example wages in labor market. Negotiable wages may change the properties of the matching market.

Crawford and Knoer (1981) extend the deferred acceptance theorem to matching markets with side payments. They call the modified algorithm the "salary-adjustment process." Using this new approach they demonstrate that there always exists a stable matching; i.e., assignment of workers and firms as well as wages such that no worker and firm finds it more preferable to match with each other (at some wage) than their assigned matches.

That was a new, constructive method of proving that result. Existence of a stable matching in such an environment – equivalent to equilibrium in cooperative game – was first proven by Shapley and Shubik (1971). Their proof did not involve the construction through the salary-adjustment process.

Kelso and Crawford (1982) improved the result of Crawford and Knoer (1981). They proved the validity of the salary-adjustment process and the existence result for markets where several of the assumptions, necessary in Crawford and Knoer (1981), were relaxed.

1.3.3. The College Admission Problem

The college admission problem differs from the marriage problem in that colleges have typically more than one place to fill (although students enroll in only one college), when each man and woman can match with only one spouse.

Gale and Shapley (1962) concluded that a stable matching in the college admission problem always exists, because the college admission problem is equivalent to the marriage problem. Every college has as many "incarnations" as places to fill and each "incarnation" has the same preferences over workers. The deferred acceptance algorithm applied to such a modified marriage market returns stable matching.

However, Roth (1985b) argues that the college admission and marriage problems are not equivalent. This is because colleges with more than one place to fill have preferences over groups of agents that cannot be easily translated into preferences over single agents. For instance, consider a college with two places to fill and four possible students a, b, c or d. Even though the strict preferences over the individuals are $a \succ b \succ c \succ d \succ \emptyset$, it may be $\{b, c\} \succ \{a, d\}$, or the other way around.

The result about existence of stable matching in any college admission problem still holds when the issue of preferences over groups of agents is taken into account. But, as Roth (1985) demonstrates, some of the results for marriage problem are no longer true for the college admissions problem. For example, Gale and Sotomayor (1985b) establish that in the marriage market there is no other matching (stable or unstable) that would make all men strictly better off than the men-optimal matching. However, as Roth (1985) shows, in the college assignment problem with some colleges having more than one place to fill, it is possible that there exists an unstable matching that makes all colleges strictly better off than under the college optimal stable matching.

Nonetheless, many of the results for the marriage market still hold for the college problem. The equivalent of another result in Gale and Sotomayor (1985b) states that all students matched in one stable matching will be matched in all stable matchings, and that every school will fill out exactly the same number of position under every stable matching. This result, along with others, was proven anew, with attention to the observations of Roth (1985), in Roth and Sotomayor (1989). That paper also discusses results for college market that could not be observed in the special case of the marriage market.

1.3.4. Strategic Behavior

Since the introduction of the clearinghouse in the US medical market, one of the major concerns is whether the algorithm used by a centralized matching mechanism provides incentives to misreport preferences. Because any clearinghouse relies on self-reported ranking lists, it may potentially be manipulated.

Dubins and Freeman (1981) demonstrate that in the college-proposed deferred acceptance algorithm, colleges cannot improve on their situation by misrepresenting their preferences.

Roth (1982) further explores game-theoretic aspects of matching problems. Inspired by the history of the US medical market, and the success of National Residents Matching Program, he studies strategy-proofness of possible centralized matching procedures. He concludes that it is not possible to find an algorithm that always yields a stable matching and gives all agents incentives to report their true preferences. At best it is possible that agents on one side of the market report their true rankings, and agents on the other side truthfully report their top-ranked position. That is, in a college-proposed deferred acceptance algorithm, colleges do best by reporting their full ranking-order truthfully. Each student truthfully reports the college of his first choice, but it may be optimal for a student to misreport rankings of other colleges. Similar results proven in an alternative way are presented in Gale and Sotomayor (1985a).

Both Dubins and Freeman (1981) and Roth (1982) consider environments where preferences are common knowledge to the market participants, but they are not known by the clearinghouse. For such environments Roth (1984) argues that even though there will be some misreporting of the preferences on the students' side, the resulting matching will be stable with respect to the true preferences, although it may not be college-optimal. This is because a misrepresentation by one student will trigger a response of other students who might in turn misrepresent their preferences. In equilibrium all students will report their preferences in such a way that there will be no blocking pair, and the outcome will be a stable matching. An example of a similar adjustment in the case of an unstable matching mechanism for sororities is presented in Roth (1991). Roth (1984) also suggests that in an environment where agents have little information about others' preferences, they may find it difficult or impossible to determine an "optimal" misrepresentation, and therefore they will not "play the system".

Other results on the properties of the stable matching in the marriage and college problems, as well as results on strategic behavior, are collected in the book by Roth and Sotomayor (1990). Gale (2001) also summarizes and comments on recent developments.

1.3.5. Search Models

Most of the markets that experience unravelling have a "seasonal" structure. That is, typically once a year new cohort of students graduates and is ready to take on new obligations: residency, internship, new job. Hospitals know that if they do not succeed in hiring an intern one spring, they need to wait a year to try again. However, there are many matching markets with more "continuous" structure, where firms and workers continuously enter and exit the market. The two types of matching markets require substantially different modeling. The former is a finite game (played once a year), where one looks for a form of a Nash equilibrium. The latter is an infinite game, where the goal is to find a steady state.

A typical example of the second modeling approach is an environment with firms and workers, where agents on opposites side of the market meet each other randomly. The quality of the agent on the opposite side is known only upon the encounter, and both agents decide whether to contract (and leave the market, at least for some time), or part and keep on searching. The decision depends on the agent's own quality, the other agent's quality and the expected payoff from further searching. As some agents contract and leave the market, new agents on both sides enter the market. Typically, when a firm and a worker match, they engage in some kind of production and share the surplus. For this reason, search models usually involve side payments, in form of wages.

These types of models study how the decision whether to enter a match or not changes with the environment. The environment variables include: for how long the matched pair will leave the market, how much information about the quality of the potential partner is revealed upon encounter, how often encounters happen, and what the rule for sharing the surplus is. In equilibrium, the matching decision affects the resulting matching outcome. One of the results is that when agent types (usually their qualities) are complements in the production, then agents on one side match with an agent of similar quality on the other side. This is so called positively assortative matching (e.g. Becker (1973), Shimmer and Smith (2000)). If the types are substitutes, the resulting matching is negatively assortative (e.g. Legros and Newman (2002)). That is, agents of high type match with agents of low type.

1.4. Structure of the Analysis

The analysis carried out in the present study adopts the finite-game modeling method, similar to Gale and Shapley (1962) and subsequent papers. The focus of this study is on causes and welfare consequences of unravelling.

Chapter 2 presents a model which allows to investigate unravelling. The main conclusion of that chapter is that unravelling is more likely in markets where preferences of agents are more similar.

Chapter 3 demonstrates that unravelling leads to a loss of welfare and to inefficiency in the Pareto sense. It also shows that it is possible to stop unravelling and improve welfare by carefully choosing the matching mechanism.

CHAPTER 2

Model of Unravelling

One of the most important determinants of a firm's success is its hiring policy. The hiring process involves collecting information to choose the best from among the candidates. However, it has been observed that in certain markets firms hire workers long before the needed information is available. For instance, in the market for hospital interns before 1945, appointments took place even as early as 2 years prior to students' graduation and the effective start of the job. A similar situation is still a concern in the market for federal court clerks.[1] Such behavior occurs in those markets because some employers see the best chance to hire the desired candidates in offering them the job before other employers do. In response, other employers rush with their own offers, and the hiring dates creep earlier and earlier. A more extensive review of such markets was done in Chapter 1.

This situation, where contracting occurs long before the start of the job and before the relevant information is available, is called "unravelling". Such early matches often turn out to be inefficient when the job starts. This is because at the time of contracting, students often do not know what speciality they will want to pursue in two years, while employers do not know yet their needs, or the quality of the students. Unravelling has been recognized as a serious problem that affects many markets.[2] While measures designed to

[1]Haruvy, Roth and Unver (2005) report that "63% of responding judges said that they had completed their clerkship hiring [for jobs beginning in 2002] by the end of January, 2000, in contrast to only 17% who had completed their hiring by January the previous year."

[2]Examples include postseason college football bowls, entry-level law and medical markets, fraternity and sorority rushes. For a more extensive list, see Roth and Xing (1994).

preclude it (e.g. centralized clearing houses or enforcement of uniform hiring dates) have been introduced in these markets, they have not always been successful. Unravelling still occurs in spite of such measures, for example, in the market for gastroenterologists, federal court clerks in the US, and lawyers in Canada (Roth and Xing, 1994; Avery, Jolls, Posner and Roth, 2001; Haruvy, Roth and Unver, 2006). At the same time, some other markets for entry-level professionals have never seemed to experience unravelling, for instance, markets for new professors in finance, economics or biology.

Despite extensive research in the economics literature, the causes and welfare consequences of this phenomenon are not fully understood. In particular, we have only limited understanding of why unravelling occurs in some markets but not in others. The goal of this chapter is to investigate the causes of unravelling. It is shown that the similarity of preferences is an important factor contributing to unravelling. Chapter 3 focuses on issues of social welfare in markets where unravelling is possible.

To study unravelling, this chapter considers a two-sided matching market populated by firms on one side and workers on the other. Agents on each side are heterogenous and have preferences over agents on the other side of the market. Their aim is to match with the best possible agent on the other side. Workers' preferences over firms are identical. The similarity of firms' preferences over workers is a comparative statics parameter. The two extreme cases are independent and identical preferences, although intermediate levels of similarity are also explored. There are two periods. Firms and workers can contract in either period, but firms only learn their preferences in the second period. Firms and workers who contract in the first period exit the market. The agents who remain in the second period participate in a mechanism that produces a matching between them.

Unravelling in this model corresponds to contracting in the first period, before firms' preferences are known. Such early contracting takes place when a firm makes an offer in the first period, and this offer is accepted. This happens when contracting under uncertainty yields a higher expected payoff than the expected matching in the second period, for both the firm and the worker.

This chapter investigates unravelling when the mechanism in the second period is assumed to produce the ex-post stable matching. A matching is ex-post stable if everyone prefers the match to being unmatched, and if there is no blocking pair, i.e. a worker and a firm that both strictly prefer each other than their currently matched partners.

In particular, this chapter analyzes equilibria that arise for different levels of similarity in firms' preferences under the ex-post stable mechanism. The focus is on sequential equilibria in pure strategies. These equilibria depend crucially on the level of similarity of preferences. Specifically, unravelling only occurs in markets where this similarity is high enough. In markets where preferences are very similar, many firms are likely to prefer the same workers. Even before firms know their actual rankings, they are aware that once the information arrives, all firms will compete for the same workers. Some firms would not be able to hire their most desired candidates amid such competition. Those firms may have a better chance to hire the most preferred workers if they contract before the rankings are known. On the other hand, in any market with independent preferences the unique equilibrium outcome is for no unravelling to occur. As the similarity of preferences increases, equilibria involving unravelling become more likely, and it becomes less likely that "no unravelling" is an equilibrium.

A substantial part of the existing research, reviewed in Chapter 1, focuses on the issue of stability as the key to understanding unravelling. Roth (1991) and Kagel and Roth (2000) argue that an ex-post stable matching implemented in the market upon arrival of information should preclude early contracting under uncertainty. This so-called “stability hypothesis” (Roth, 1991) is based mainly on the observation that implementing an ex-post stable matching through a clearinghouse stopped unravelling in the US and UK medical markets. However, some clearinghouses with an ex-post stable algorithm have failed to stop unravelling. Examples include the gastroenterology market in the US, where the clearinghouse was abandoned in 1996 (Niederle and Roth, 2003), and the Canadian market for new lawyers, where despite the clearinghouse, a large number of firms contract with students a year before the graduation (Roth and Xing, 1994). Also, Roth and Xing (1994) show theoretical examples of unravelling even when the ex-post stable matching is expected upon arrival of information. However, there is no consensus on whether these examples are single anomalies, or if there is instead some systematic cause for the stability hypothesis to fail. This chapter shows that high similarity of preferences may lead to unravelling even under the ex-post stable mechanism.

The stability hypothesis is not the only explanation of unravelling in the literature. In Damiano, Li and Suen (2005) early contracting is the result of costly search. Li and Rosen (1998), Li and Suen (2000) and Suen (2000) point to workers’ risk aversion as the main cause of the phenomenon. Although risk aversion plays an important role and may be an additional cause of early contracting, it is not a necessary condition for the phenomenon. The model in this paper assumes risk-neutrality in order to separate incentives to unravel driven by similarity of preferences from incentives due to risk-aversion.

2.1. The Model

To investigate unravelling, a two-stage game is constructed between two types of agents: firms and workers. Firms and workers can contract in the first stage. If they do, they leave the market. In the second stage, the remaining agents are matched by a mechanism.

The market is populated by F firms, $f \in \{1, \ldots, F\}$, and W workers, $w \in \{1, \ldots, W\}$. There are more workers than firms, $W > F$. Each firm has exactly one position to fill, and each worker can take at most one job.

Workers have identical preferences over firms. All workers consider firm F to be the most desired one, firm $(F-1)$ – the second-best, and so on. The utility for a worker from being matched to firm f is u_f, and the utility from being unmatched is 0. The workers prefer to be hired by the worst firm than not to be hired at all, i.e. $0 < u_1 < u_2 < \ldots < u_F$. Let $\mathbf{u} \equiv [u_1,\, u_2, \ldots,\, u_F]$.

Firms may have different preferences over workers. Each firm's preferences are characterized by its ranking. Firm f's ranking over workers – denoted by $\mathcal{R}^f$ – is an ordered list of length W:

$$\mathcal{R}^f = (r_1^f, r_2^f, \ldots, r_W^f)$$

where r_1^f is the identity of the lowest ranked worker, and r_W^f – the identity of the highest ranked worker in firm f's ranking. Every worker has exactly one position in every firm's ranking. Let $\mathbf{R} = [\,\mathcal{R}^1,\, \ldots,\, \mathcal{R}^F]$ be the vector of all firms' rankings. For a subset of firms $\mathcal{F} \subseteq \{1, \ldots, F\}$, let $\mathbf{R}^{\mathcal{F}}$ be a similar vector for rankings of firms in $\mathcal{F}$.

The value to firm f of being matched to worker r_k^f is v_k. It is better to hire the worst worker than to retain a vacancy, i.e. $0 < v_1 < v_2 < \ldots < v_W$. Let $\mathbf{v} \equiv [v_1, v_2, \ldots, v_W]$. Matching value vectors, $\mathbf{u}$ and $\mathbf{v}$, are publicly known,[3] but rankings are each firms' private knowledge.

There are no transfers between firms and workers. When firm f is matched with worker r_k^f, the worker receives utility of exactly u_f and the firm receives a payoff of exactly v_k.

Let $\mathcal{W} \subseteq \{1, \ldots, W\}$ denote an arbitrary subset of workers. Similarly, $\mathcal{F} \subseteq \{1, \ldots, F\}$ denotes a subset of firms.

Definition 1 (matching)**.** *A* matching *between $\mathcal{F}$ and $\mathcal{W}$ is a function $\mu^{\mathcal{F},\mathcal{W}} : \mathcal{F} \to \mathcal{W} \cup \{\emptyset\}$ such that for any two firms f and f' in $\mathcal{F}$*

$$f \neq f' \qquad \Longrightarrow \qquad \mu^{\mathcal{F},\mathcal{W}}(f) \neq \mu^{\mathcal{F},\mathcal{W}}(f') \quad \textit{or} \quad \mu^{\mathcal{F},\mathcal{W}}(f) = \mu^{\mathcal{F},\mathcal{W}}(f') = \emptyset$$

Expression $\mu^{\mathcal{F},\mathcal{W}}(f) = \emptyset$ means that firm f is not matched with any worker in $\mu^{\mathcal{F},\mathcal{W}}$. When $\mu^{\mathcal{F},\mathcal{W}}(f) = w \in \mathcal{W}$, then firm f is matched with worker w in $\mu^{\mathcal{F},\mathcal{W}}$. In such a case, worker w is also matched with f. In general, any worker $w \in \mathcal{W}$ is matched in $\mu^{\mathcal{F},\mathcal{W}}$ if and only if there exists a firm $f \in \mathcal{F}$ such that $\mu^{\mathcal{F},\mathcal{W}}(f) = w$. Otherwise, a worker is unmatched in $\mu^{\mathcal{F},\mathcal{W}}$. Let $\boldsymbol{\mu}(\mathcal{F}, \mathcal{W})$ denote the set of all possible matchings between $\mathcal{F}$ and $\mathcal{W}$.

The existing literature emphasizes the importance of ex-post stability in the matching. Roth (1991) and Kagel and Roth (2000), for example, argue that the ex-post stable matching implemented after the arrival of information should preclude early contracting.

[3] v_k's, as well as u_f's, do not need to be the precise values of a match. For the analysis, it is enough if they are the expected values. The actual values may be realized, for example, only after the match is made.

The notion of ex-post stability[4] has been introduced by Gale and Shapley (1962). A matching is called *ex-post unstable* if there is a firm and a worker that would rather be matched to each other than remain in their current matches. A matching is called *ex-post stable* if it is not ex-post unstable.

For any $\mathcal{F}$ and $\mathcal{W}$, $\boldsymbol{\mu}(\mathcal{F}, \mathcal{W})$ is the set of all possible matchings. Which of those is ex-post stable depends on firms' preferences, $\mathbf{R}^{\mathcal{F}}$. Lemma 1 below establishes that for a given preference profile there is a unique ex-post stable matching between $\mathcal{F}$ and $\mathcal{W}$.[5] It can be characterized in the following way: The best firm in $\mathcal{F}$ is matched with its highest ranked worker in $\mathcal{W}$. Then, the next-best firm is matched with its highest ranked worker from among the remaining workers, etc. Every firm in $\mathcal{F}$ is matched to its highest ranked worker remaining in the pool after all the better firms in $\mathcal{F}$ have been matched.

Lemma 1 (ex-post stable matching)**.** *For any $\mathcal{W}$, $\mathcal{F}$ and $\mathbf{R}^{\mathcal{F}}$, there is a unique ex-post stable matching between $\mathcal{F}$ and $\mathcal{W}$.*

For any $f \in \mathcal{F}$, let $\mu_S^{\mathcal{F},\mathcal{W}}(f|\mathbf{R}^{\mathcal{F}})$ refer to the worker matched with f in the ex-post stable matching between $\mathcal{F}$ and $\mathcal{W}$ under firms' rankings $\mathbf{R}^{\mathcal{F}}$. Then

$$\mu_S^{\mathcal{F},\mathcal{W}}(f|\mathbf{R}^{\mathcal{F}}) \equiv \max_{k \in \mathcal{W}} \left\{ r_k^f \,|\, \forall\ i \in \mathcal{F} \quad \text{s.t.} \quad i > f \qquad \left(\mu_S^{\mathcal{F},\mathcal{W}}(i|\mathbf{R}^{\mathcal{F}}) \neq r_k^f \right) \right\}$$

Proof.

[4]In Gale and Shapley (1962) this property was called just "stability". In this it is called "ex-post stability" to emphasize the fact that a matching satisfying this property may nevertheless unravel, and thus in a sense be "ex-ante" unstable even though it is "ex-post" stable.

[5]With arbitrary workers' preferences the ex-post stable matching does not need to be unique (Gale and Shapley 1962).

The formula for the ex-post stable matching rule is obtained by the Gale-Shapley algorithm. The uniqueness follows from identical preferences of workers. The proof follows in two steps: (i) $\mu_S^{\mathcal{F},\mathcal{W}}$ is an ex-post stable matching over $\mathcal{F}$ and $\mathcal{W}$, (ii) there is no other ex-post stable matching.

(i) $\mu_S^{\mathcal{F},\mathcal{W}}$ is an ex-post stable matching over $\mathcal{F}$ and $\mathcal{W}$.

Proof by contradiction: Assume $\mu_S^{\mathcal{F},\mathcal{W}}$ is not an ex-post stable matching. Thus, there must be a blocking pair.Let $f \in \mathcal{F}$ and $w \in \mathcal{W}$ be the blocking pair. If f prefers w to $\mu_S^{\mathcal{F},\mathcal{W}}(f)$, then (by construction of the algorithm) w is matched with $f' > f$. If so, w prefers its current match to f. Thus, f and w are not a blocking pair.

(ii) there is no other ex-post stable matching.

Proof by contradiction: Assume there is another ex-post stable matching μ', different than $\mathfrak{o}_S$, i.e.

$$\exists\, f \in \mathcal{F} \text{ s.t. } \mu'(f) \not\equiv \mu_S^{\mathcal{F},\mathcal{W}}(f) \tag{2.1}$$

But it can not be true for the best firm in $\mathcal{F}$. If μ' does not match the best firm with its most desirable worker in $\mathcal{W}$, then this firm and this worker form a blocking pair. Also, (2.1) can not be true for the next best firm. If μ' does not match the next best firm with its most desirable worker available after the best firm got matched, this firm and this worker form a blocking pair. Similarly for any $f \in \mathcal{F}$. Thus, if μ' is a ex-post stable matching, it must be the same as $\mu_S^{\mathcal{F},\mathcal{W}}$, i.e. $\forall\, f \in \mathcal{F}\,, \mu'(f) = \mu_S^{\mathcal{F},\mathcal{W}}(f)$. □

A *matching outcome* refers to a matching between all firms, $\{1,\ldots,F\}$, and all workers, $\{1,\ldots,W\}$, realized at the end of the two stage game. The *ex-post stable outcome* – denoted by $\mathfrak{o}_S$ – is the ex-post stable matching between all workers, $\{1,\ldots,W\}$, and all firms, $\{1,\ldots,F\}$, in the market:

$$\mathfrak{o}_S(f|\mathbf{R}) \equiv \max_{k\in\{1,\ldots,W\}} \Big\{ r_k^f \mid \forall\, i \in \{f+1,\ldots,F\} \quad \Big(\mathfrak{o}_S(i|\mathbf{R}) \neq r_k^f\Big)\Big\}$$

Since the *ex-post stable outcome* is unique for every market, any other matching outcome is ex-post unstable.

Below, I drop $\mathbf{R}$ from the notation, keeping in mind that the ex-post stable matching depends on rankings. In the ex-post stable outcome, $\mathfrak{o}_S$, firm F is matched with its most preferred worker, r_W^F. Firm $(F-1)$ is matched with its most preferred worker excluding $w \equiv r_W^F$, who has been matched with firm F, etc. That is, any firm f is matched with its most preferred worker remaining in the pool after all firms better than f have been matched.

In some situations firms are asked to report their rankings and a matching is produced based on those reports. In these situations the matching is produced by a matching mechanism, also called a clearinghouse.

Definition 2 (matching mechanism)**.** *A* matching mechanism, $\mathcal{M}$, *is a function that maps* $\mathcal{F}$, $\mathcal{W}$, *and the reported rankings of firms,* $\widehat{\mathbf{R}}^{\mathcal{F}}$, *to a randomization over all matchings between* $\mathcal{F}$ *and* $\mathcal{W}$*:*

$$\mathcal{M} : (\mathcal{F},\mathcal{W},\widehat{\mathbf{R}}^{\mathcal{F}}) \mapsto Rand\Big(\boldsymbol{\mu}(\mathcal{F},\mathcal{W})\Big)$$

A matching mechanism is *incentive compatible* if no firm benefits from misreporting its preferences. All mechanisms considered in this paper are incentive compatible. A mechanism is called *ex-post stable* – and denoted $\mathcal{M}_S$ – if it applies the ex-post stable matching to the reported rankings with probability 1. It is easy to check that in this model the ex-post stable mechanism is incentive compatible. Therefore, the ex-post stable mechanism operating over $\mathcal{F}$ and $\mathcal{W}$ will produce the ex-post stable matching between $\mathcal{F}$ and $\mathcal{W}$.[6]

There are two periods in the model: $t = 1, 2$. Workers' preferences are commonly known in both periods. Firms learn their own preferences, as rankings, only at the beginning of period 2. Each firm's ranking is its private knowledge.

Denote by $\mathfrak{R}$ the set of all $W!$ possible rankings over workers. The rankings for all F firms, $(\mathcal{R}^1, \ldots, \mathcal{R}^F)$, are drawn from a joint distribution G over $\mathfrak{R}^F$. The model focuses on distributions where the marginal distributions of individual rankings are always uniform, allowing for different levels of similarity between the rankings.[7] Two special cases – of identical preferences and independent preferences – are defined below.

Let G_1 be the joint distribution where rankings of all firms are identical and the marginal distribution of any individual ranking is uniform on $\mathfrak{R}$. That is, any ranking

[6]Incentive compatibility means that there exists an equilibrium where all firms report their true preferences. In this model, the ex-post stable mechanism has a stronger property. For firm f only top workers $r^f_W, \ldots, r^f_{W-F+f}$ are relevant in producing the ex-post stable matching. Under the ex-post stable mechanism, misreporting this part of the ranking makes the firm strictly worse. Misreporting of the rest of the ranking is irrelevant for the equilibrium outcome. Therefore, under the ex-post stable mechanism, the unique equilibrium outcome is the ex-post stable matching between the agents that participate in the mechanism.

[7]The uniform prior is convenient for the presentation of the results. However, similar arguments can be made with other priors.

in $\mathfrak{R}$ may be drawn as $\mathcal{R}^f$ with equal probability of $\frac{1}{W!}$ and all firms will have the same ranking.

Definition 3 (G_1)**.** *For any F and $W > F$, let G_1 be the joint distribution over $\mathfrak{R}^F$ such that*

$$\forall\, \mathcal{R} \in \mathfrak{R} \quad Prob\big((\mathcal{R}^1, \ldots, \mathcal{R}^F) = (\mathcal{R}, \ldots, \mathcal{R})\big) = \frac{1}{W!}$$

and all the other probabilities are 0, i.e. $Prob(\exists\, f, i \text{ s.t. } \mathcal{R}^f \neq \mathcal{R}^i) = 0$.

Let G_0 be the joint distribution such that a ranking of any firm is drawn from a uniform distribution independently on any other firms' rankings.

Definition 4 (G_0)**.** *For any F and $W > F$, let G_0 be the joint distribution over $\mathfrak{R}^F$ such that*

$$\forall\, f \in \{1, \ldots, F\} \quad \forall\, \bar{\mathcal{R}}^f \in \mathfrak{R} \quad Prob\left((\mathcal{R}^1, \ldots, \mathcal{R}^F) = (\bar{\mathcal{R}}^1, \ldots, \bar{\mathcal{R}}^F)\right) = \left(\frac{1}{W!}\right)^F$$

Between the identical and the independent rankings, there is a continuum of cases of intermediate similarity, G_ϱ.

Definition 5 (G_ϱ)**.** *For $\varrho \in [0, 1]$,*

$$G_\varrho = \varrho G_1 + (1 - \varrho) G_0$$

Factor ϱ is a measure of preference similarity[8] and will be a comparative statics parameter in the analysis below. Preferences are said to be *more similar* under $G_{\varrho'}$ than under G_ϱ when $\varrho' > \varrho$. Since ϱ completely characterizes G_ϱ, the two are used interchangeably.

[8] Since preferences are determined by rankings, "rankings" and "preferences" are used interchangeably.

The marginal distributions are uniform under both G_1 and G_0, and also under G_ϱ. Therefore, the prior beliefs in $t = 1$ about firms' preferences are also uniform, for both workers and firms. That is, any worker may turn out to be the k-th worker ($k = 1, \dots, W$) of the given firm with equal probability.

A market in this model is characterized by the number of firms F, number of workers W, matching value vectors $\mathbf{u}$ and $\mathbf{v}$, similarity of preferences, ϱ and mechanism applied in the second period $\mathcal{M}$. Thus, a market is fully described by a tuple $(F, W, \mathbf{u}, \mathbf{v}, \varrho, \mathcal{M})$.

Figure 2.1 illustrates how the game unfolds. Market characteristics $(F, W, \mathbf{u}, \mathbf{v}, \varrho, \mathcal{M})$ and the preferences of the workers' are commonly known all the time. At the beginning of period 1 firms simultaneously decide whether or not to make an early offer, and if so, to which worker. Every firm can make at most one offer. After the early offers are released, every worker observes the offers he has obtained, if any. He does not see offers made to other workers. Based on his beliefs about other agents' strategies, every worker presented with an offer accepts or rejects it. He may accept at most one offer. If an offer is accepted, the matched firm and worker leave the market. Firms whose offers were rejected or who did not make an offer in $t = 1$, stay in the market for $t = 2$. In period 2, firms' rankings are realized and a matching mechanism $\mathcal{M}$ operates over the agents remaining in the market at this time. The first part of the paper – Section 2.2 – assumes the ex-post stable mechanism in the $t = 2$. The second part – Section 3 – considers other mechanisms. There is no discounting between the periods and making offers is costless.

Under an incentive compatible mechanism, firms truthfully report their rankings in $t = 2$. Therefore, both the firms and the workers make their strategic decisions only in

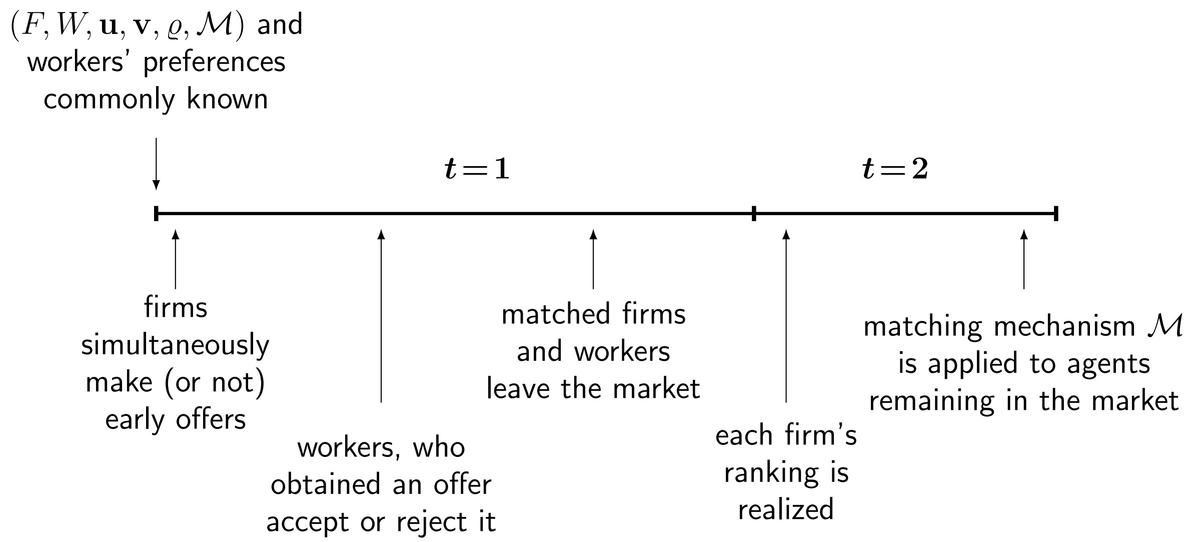


Figure 2.1. Timeline of the game

$t = 1$. First, every firm decides whether or not to make an offer and if so, to which worker. The analysis focuses on sequential equilibria in pure strategies, where a strategy of any firm f is $\sigma_f \in \{1, \ldots, W\} \cup \{\emptyset\}$. Since a worker can accept or reject an offer only if he has received it, a worker's strategy depends on the offers he has received. Let $\Omega_w \subset \{1, \ldots, F\}$ be the set of firms that have made an offer to worker w in $t = 1$. Then the worker's strategy, $\sigma_w(\Omega_w) \in \Omega_w \cup \{\emptyset\}$, is the offer that he accepts. Strategy $\sigma_w(\Omega_w) = \emptyset$ means that the worker rejects all offers. Let vector $\boldsymbol{\sigma}$ be the strategy profile for all firms and workers.

Let β_i denote agent's i beliefs. Agent i believes that agent j plays strategy σ_j with probability $\beta_i(\sigma_j)$. For any agent $j \not\equiv i$ and for any strategy σ_j, $\beta_i(\sigma_j) \in [0, 1]$. Let vector $\boldsymbol{\beta}$ denote a system of beliefs of all firms and workers.

Firms move first and simultaneously, so there is only one information set for each firm. When worker w makes a decision, his information set is characterized by the set of offers he has received, Ω_w.

Every firm's payoff depends on many variables: market characteristics $(F, W, \mathbf{u}, \mathbf{v}, \varrho, \mathcal{M})$, realized rankings of firms $\mathbf{R}$, and the strategies played by all agents in the market. That is, f's payoff is denoted by $\pi_f(F, W, \mathbf{u}, \mathbf{v}, \varrho, \mathcal{M}, \mathbf{R}, \boldsymbol{\sigma})$. The payoff expected by firm f at the beginning of the game depends on the market characteristics, f's strategy and its beliefs about other agents' strategies: $E\pi_f(F, W, \mathbf{u}, \mathbf{v}, \varrho, \mathcal{M}, \sigma_f, \beta_f)$. Similarly, any worker's utility and expected utility depend on the corresponding variables. For clarity, most of this notation is suppressed and only the variables essential for current analysis are indicated.

Let $E\pi_f(\sigma_f \,|\, \beta_f, \boldsymbol{\sigma}_{-f})$ be firm f's expected payoff from playing strategy σ_f given beliefs β_f and other agents' strategies $\boldsymbol{\sigma}_{-f}$. Similarly, let U_w be worker w's utility, and $EU_w(\sigma_w \,|\, \beta_w, \boldsymbol{\sigma}_{-w})$ – worker w's expected utility from playing strategy σ_w given beliefs β_w and other agents' strategies $\boldsymbol{\sigma}_{-w}$.

A sequential equilibrium in this model is a profile of strategies and a system of beliefs satisfying conditions stated in the definition stated below.

Definition 6 (equilibrium)**.** *In the game with market* $(F, W, \mathbf{u}, \mathbf{v}, \varrho, \mathcal{M})$*, a profile of strategies and system of beliefs* $(\boldsymbol{\sigma}^*, \boldsymbol{\beta}^*)$ *constitute a sequential equilibrium when*

(1) *strategies are sequentially rational given the beliefs, i.e.*

  (f) *in its only information set, given the beliefs and the strategies of other firms and of workers, every firm* $f \in \{1, \ldots, F\}$ *chooses* σ^*_f *that maximizes its* *expected payoff, i.e.*

$$E\pi_f(\sigma_f^* \,|\, \beta_f^*, \boldsymbol{\sigma}_{-f}^*) \geq E\pi_f(\sigma_f \,|\, \beta_f^*, \boldsymbol{\sigma}_{-f}^*) \qquad \forall\, \sigma_f \in \{1, \ldots, W\} \cup \{\emptyset\}$$

(w) *in each information set Ω_w, given the beliefs and the strategies of firms and other workers, each worker $w \in \{1, \ldots, W\}$ chooses his strategy, conditionally on the set of received offers, $\sigma_w^*(\Omega_w)$ such as to maximize his expected utility, i.e.*

$$EU_w(\sigma_w^* \,|\, \Omega_w,\, \beta_w^*, \boldsymbol{\sigma}_{-w}^*) \geq EU_w(\sigma_w \,|\, \Omega_w,\, \beta_w^*, \boldsymbol{\sigma}_{-w}^*) \qquad \forall\, \sigma_w \in \{\Omega_w\} \cup \{\emptyset\}$$

(2) *beliefs are consistent with the strategies played, in particular*

(f) *for any firm $f \in \{1, \ldots, F\}$, its beliefs β_f^* are*

$$\forall\, w \in \{1, \ldots, W\} \quad \beta_f^*(\sigma_w) = \begin{cases} 1 & for \quad \sigma_w = \sigma_w^* \\ 0 & otherwise \end{cases}$$

$$\forall\, i \in \{1, \ldots, F\} \setminus \{f\} \quad \beta_f^*(\sigma_i) = \begin{cases} 1 & for \quad \sigma_i = \sigma_i^* \\ 0 & otherwise \end{cases}$$

(w) *for any worker* $w \in \{1, \dots, W\}$ *given the set of offers* Ω_w*, his beliefs* β^*_w *are*

$$\forall\, f \in \Omega_w \quad \beta^*_w(\sigma_f|\Omega_w) = \begin{cases} 1 & for \quad \sigma_f = w \\ 0 & otherwise \end{cases}$$

$$\forall\, f \in \{1, \dots, F\} \setminus \Omega_w \quad \beta^*_w(\sigma_f|\Omega_w) = \begin{cases} 1 & for \quad \sigma_f = \sigma^*_f \\ 0 & otherwise \end{cases}$$

$$\forall\, k \in \{1, \dots, W\} \setminus \{w\} \quad \beta^*_w(\sigma_k|\Omega_w) = \begin{cases} 1 & for \quad \sigma_k = \sigma^*_k \\ 0 & otherwise \end{cases}$$

The beliefs are consistent with the strategies played on the equilibrium path. Firms make their decisions simultaneously at the beginning of the game. They can not observe anything off the equilibrium path. Workers observe only the set of their own offers when making a decision to accept or reject. The only relevant possibility for an off-equilibrium path is when a worker receives an offer he did not expect to receive. A property of sequential equilibrium allows for determining in a unique way the sequentially rational beliefs and strategies even on the nodes not reached on the equilibrium path. When a worker has received an offer he had not expected, the worker updates his beliefs only about the firm that has made him the off-equilibrium offer. Now he believes that the firm has made him an offer, instead of making it to some other worker or not making it at all. But it does not change the worker's beliefs about any other firm.

Offers made and accepted in period 1 constitute unravelling.

Definition 7 (unravelling)**.** Unravelling *is a situation in which some firms and workers contract in* $t = 1$*, before firms know their own preferences.*

2.2. Unravelling under Ex-post Stable Mechanism

The ex-post stable matching is considered desirable in the existing literature. It has been proposed that ex-post stable mechanism prevents unravelling (Roth 1991, Kagel and Roth 2000). It also has been argued that the ex-post stable outcome maximizes social welfare (Bulow and Levin 2003). Moreover, the ex-post stable mechanism is often adopted by the clearinghouses introduced to prevent unravelling. The mechanism was chosen for the clearinghouses either independently, as in the market for medical residents in 1952, or by recommendation of economists, as for Boston public schools in 2005.[9] It can also be argued that the ex-post stable matching is one of the equilibria in a decentralized market (without a clearinghouse), after the information about preferences arrives.

Given that the literature focuses on ex-post stable mechanisms, this section investigates unravelling under the ex-post stable matching mechanism. Subsection 2.2.1 focuses on equilibria without unravelling, while Subsection 2.2.2 describes equilibria when unravelling occurs.

The mechanism applied to the reported rankings in $t = 2$ is assumed in this section to be the ex-post stable one, $\mathcal{M}_S$. Mechanism $\mathcal{M}_S$ is not only incentive compatible, but in all equilibria it produces the ex-post stable matching among the agents remaining in $t = 2$. Unless unravelling occurs in $t = 1$, it produces the ex-post stable outcome, $\mathfrak{o}_S$.

[9]See Kimberly Atkins, "Commitee OKs new school assignment plan", Boston Herald, Jul 21, 2005.

For both firms and workers, the decision whether to contract early presents a trade-off. A worker who receives an offer from firm f in $t = 1$ chooses between u_f – a sure payoff from accepting the offer – and a lottery in $t = 2$, where he possibly can be matched to a better firm, or a worse firm, or even remain unmatched. A firm decides between contracting early – which with uniform prior yields expected payoff of the average value of workers – and the ex-post stable matching in $t = 2$, where better firms may be matched with firm f's most preferred workers.

When firms expect the ex-post stable outcome in $t = 2$, their expected payoffs depend on their own position and the similarity of preferences in the market. The ex-post stable matching has two properties that are of particular interest here. One is that lower ranked firms receive lower expected payoff in the ex-post stable matching, and the other is that firms' expected payoffs decrease as preferences become more similar.

In a given market, a lower ranked firm expects a lower expected payoff from the ex-post stable outcome than a higher ranked firm expects. In $t = 2$ firm f gets its best worker remaining in the pool after all better firms $i > f$ have been matched. As there are fewer workers left for worse firms, it is more likely that such firms' best workers are already gone. Because of this property, worse firms are more likely to prefer early contracting under $\mathcal{M}_S$ than better firms.

Therefore, to unravel, firms need to be good enough to be accepted in $t = 1$ and bad enough to want to contract early. This leads to the "unravelling in the middle" – a property that in a typical market it is not the best or the worst firms, but firms "in the middle" that unravel. In special cases, also firms at the extremes of the spectrum contract early. It is possible to find equilibria in which any firm – except for the best one – unravels.

Moreover, firms' expected payoffs decrease as preferences become more similar. While the best firm, F, is always matched with its most preferred worker, for all other firms the expected value of $\mathfrak{o}_S$ strictly decreases as ϱ increases. Higher similarity of preferences increases the probability that other firms prefer the same workers as firm f does. Better firms are more likely to be matched with top workers of firm f in the ex-post stable outcome, and firm f will be matched with its lower-ranked workers with higher probability. Because of this property, as preference similarity increases, more firms prefer to contract early.

Let $E\pi_f(\mathfrak{o}_S|\varrho)$ denote firm f's expected payoff in the ex-post stable outcome in a given market. For the special cases of $\varrho = 0$ and $\varrho = 1$, G_0 and G_1 are used instead. Then the following lemma summarizes the properties of $\mathfrak{o}_S$.

Lemma 2. (properties of $\mathfrak{o}_S$)

(1) In any market $(F, W, \mathbf{u}, \mathbf{v}, \varrho, \mathcal{M}_S)$, for any $f > 1$, $E\pi_{f-1}(\mathfrak{o}_S|\varrho) < E\pi_f(\mathfrak{o}_S|\varrho)$.

(2) Holding other market parameters constant, for any $f < F$,

$$\varrho < \varrho' \quad \Longrightarrow \quad E\pi_f(\mathfrak{o}_S|\varrho) > E\pi_f(\mathfrak{o}_S|\varrho')$$

Proof. The proof proceeds separately for each point:

(1) ***Proof.***

Probability, that firm $f-1$ gets its worker $k > W-F+f$ is

$$(1-\varrho)\cdot P(W,f-1,k) =$$

$$= (1-\varrho)\frac{F-f+1!}{(F-W-f+1+k)!}\frac{(k-1)!}{W!}(W-F+f-1) =$$

$$= (1-\varrho)\cdot P(W,f,k)\cdot\frac{F-f+1}{F-W-f+1+k}\frac{W-F+f-1}{W-F+f}$$

Since, f and W are fixed, the ratio decreases with increasing k. It is more probable for the better firm to be matched with its better workers. Formally, the inequality in expected payoffs of firms f and $f-1$ follows from FOSD. □

(2) ***Proof.***

Notice that

$$E\pi_f(\mathfrak{o}_S|\varrho) = \sum_{k=1}^{W} v_k \cdot Prob(\mathfrak{o}_S(f) = r_k^f|\varrho) = \varrho\cdot E\pi_f(\mathfrak{o}_S|G_1) + (1-\varrho)E\pi_f(\mathfrak{o}_S|G_0)$$

because

$$Prob(\mathfrak{o}_S(f) = r_k^f|\varrho) = \varrho\cdot Prob(\mathfrak{o}_S(f) = r_k^f|G_1) + (1-\varrho)Prob(\mathfrak{o}_S(f) = r_k^f|G_0)$$

Moreover,

$$E\pi_f(\mathfrak{o}_S|G_0) = \sum_{k=W-F+f}^{W} v_k\cdot P(W,f,k) >$$

$$> v_{W-F+f}\sum_{k=W-F+f}^{W}\cdot P(W,f,k) = v_{W-F+f} = E\pi_f(\mathfrak{o}_S|G_1)$$

Let $\varrho' > \varrho$, then

$$E\pi_f(\mathfrak{o}_S|\varrho') =$$

$$= \varrho \cdot E\pi_f(\mathfrak{o}_S|G_1) + (1-\varrho)E\pi_f(\mathfrak{o}_S|G_0) + (\varrho' - \varrho)\left[E\pi_f(\mathfrak{o}_S|G_1) - E\pi_f(\mathfrak{o}_S|G_0)\right] <$$

$$< E\pi_f(\mathfrak{o}_S|\varrho)$$

This completes the proof. □

2.2.1. Equilibria without Unravelling

All firms participate in the second period mechanism in an equilibrium when either no firm makes an early offer, or all early offers are rejected. This subsection explores conditions under which there exists an equilibrium without unravelling, given that the ex-post stable mechanism, $\mathcal{M}_S$ operates in $t = 2$.

Without unravelling, $\mathcal{M}_S$ produces the ex-post stable outcome, $\mathfrak{o}_S$. This is an equilibrium, unless there exists a profitable deviation from $\mathfrak{o}_S$, i.e., there exists a firm that prefers to contract early and a worker who prefers to accept this firm's offer, given that all other firms participate in $t = 2$ mechanism. Contracting in $t = 1$ is contracting under uncertainty, as preferences of firms are not known yet. A profitable deviation exists only when both the firm and the worker are better off by contracting with deficient information than by waiting for the uncertainty to be resolved.

Consider a worker who receives an offer from firm f in period 1, when all other firms are assumed to participate in $t = 2$ mechanism. If the worker accepts the offer, he receives utility u_f. If he rejects the offer, all firms and all workers participate in the $t = 2$ matching mechanism. All the workers are a priori identical and they have an equal chance of $\frac{1}{W}$ of being matched to any of the firms in $t = 2$. Thus, a worker's expected utility from rejecting f's offer is $\frac{1}{W}\sum_{i=1}^{F} u_i$. He, therefore, accepts the offer when

$$u_f > \frac{1}{W}\sum_{i=1}^{F} u_i \tag{2.2}$$

Obviously, firm F is always accepted. Whether other firms are accepted depends on the value parameters $\mathbf{u}$ and the number of workers, W.

For any W and $\mathbf{u}$ the RHS of inequality (2.2) is constant, and u_f's are ordered to be increasing in f. Therefore, there is a cut-off point – the lowest ranked firm the offer of which will be accepted in $t = 1$. Let $\mathbb{L}_{(W,\mathbf{u})}$ denote this firm:

$$\mathbb{L}_{(W,\mathbf{u})} \equiv \min\left\{ f \,\middle|\, u_f > \frac{1}{W}\sum_{i=1}^{F} u_i \right\}$$

All firms worse than $\mathbb{L}$ will be rejected in $t = 1$. Firm $\mathbb{L}$ and all firms better than $\mathbb{L}$ will be accepted. The set of the firms that will be accepted in $t = 1$ is the *acceptance set*, $\mathcal{A}$:

$$\mathcal{A}_{(W,\mathbf{u})} \equiv \left\{\mathbb{L}_{(W,\mathbf{u})}, \ldots, F\right\}$$

Notice that at the end of period 2 there are always $W - F > 0$ workers who will be unemployed, and will receive payoff 0. Because of the threat of unemployment, for any W and any f, there exists a $\mathbf{u}$ such that $\mathcal{A}_{(W,\mathbf{u})} = \{f, \ldots, F\}$. It is possible that all firms would be accepted in $t = 1$; that is, for some W and $\mathbf{u}$, $\mathbb{L}_{(W,\mathbf{u})} = 1$. This will occur when the number of workers, W, is large enough and the high probability of unemployment makes the utility expected in $t = 2$ lower than u_1.

Incentives of firms for contracting in $t = 1$, before all the information is available, depend on the joint distribution of rankings, G_ϱ. The realization of rankings – together with the matching mechanism – determines the matching realized in $t = 2$. Firms' expected payoffs depend on this expected matching. Recall that $E\pi_f(\mathfrak{o}_S|\varrho)$ denotes firm f's expected payoff from the ex-post stable outcome under G_ϱ.

The uniform prior implies that in $t = 1$ all workers are *ex ante* the same. Thus, an offer in $t = 1$ made to any worker yields the same expected payoff. Any firm's expected payoff from early contracting – if such an offer is accepted – is

$$\pi^0 \equiv \frac{1}{W}\sum_{k=1}^{W} v_k$$

Firm f prefers early contracting to the ex-post stable matching when

$$\pi^0 > E\pi_f(\mathfrak{o}_S|\varrho) \tag{2.3}$$

Firm F never has incentives to make an offer in period 1, since in the ex-post stable outcome it always hires its most preferred worker. Other firms may have something to gain from an early offer, depending on ϱ and $\mathbf{v}$.

Example 1. *Consider firm $(F-1)$. In the ex-post stable outcome this firm gets its best worker, r_W^{F-1}, unless that worker is firm F's best worker as well. When $r_W^{F-1} \equiv r_W^F$, firm $(F-1)$ gets the next worker on its list: r_{W-1}^{F-1}. Since the probability that $r_W^{F-1} \equiv r_W^F$*

under G_ϱ *is* $\varrho + (1-\varrho)\frac{1}{W}$*, the expected payoff from the ex-post stable matching is*

$$E\pi_{F-1}(\mathfrak{o}_S \mid \varrho) = (1-\varrho)\left(1 - \frac{1}{W}\right) \cdot v_W + \left(\varrho + (1-\varrho)\frac{1}{W}\right) \cdot v_{W-1}$$

In a market with 2 *firms and* 3 *workers where* $\mathbf{v} = [1, 2, 6]$*,* $E\pi_1(\mathfrak{o}_S \mid \varrho) = \frac{14}{3}(1-\varrho) + 2\varrho$ *and* $\pi^0 = 3$*. Thus, firm* 1 *would prefer early contracting to the ex-post stable outcome when* $\varrho > \frac{1}{2}$*.* ■

The lower the firm is ranked, the lower is its expected payoff in the ex-post stable outcome (Lemma 2(1)). Thus, if firm f prefers early contracting to the ex-post stable outcome, then all firms worse than f do too. The set of all firms that prefer early contracting under G_ϱ and $\mathbf{v}$ – called the *offer set* – is an interval[10]

$$\mathcal{O}_{(\varrho,\mathbf{v})} \equiv \left\{ 1, \ldots, \mathbb{H}_{(\varrho,\mathbf{v})} \right\}$$

where $\mathbb{H}_{(\varrho,\mathbf{v})}$ is the highest ranked firm that prefers early contracting to $\mathfrak{o}_S$:

$$\mathbb{H}_{(\varrho,\mathbf{v})} \equiv \max\left\{ f \,\middle|\, \pi^0 > E\pi_f(\mathfrak{o}_S|\varrho) \right\}$$

A deviation from $\mathfrak{o}_S$ to early contracting can occur only when the offer in $t = 1$ is made and accepted. Therefore, a profitable deviation from $\mathfrak{o}_S$ is possible only when there exists a firm that prefers early contracting to the ex-post stable matching and when this firm is accepted by a worker in $t = 1$. That is, if for some f, $u_f > \frac{1}{W}\sum u_i$ and $\pi^0 > E\pi_f(\mathfrak{o}_S|\varrho)$, which is equivalent to

$$\mathcal{A}_{(W,\mathbf{u})} \cap \mathcal{O}_{(\varrho,\mathbf{v})} \neq \emptyset$$

[10]For the special cases of $\varrho = 0$ and $\varrho = 1$ considered below, the offer set is denoted by $\mathcal{O}_{(G_0,\mathbf{v})}$ and $\mathcal{O}_{(G_1,\mathbf{v})}$.

The offer set, $\mathcal{O}_{(\varrho,\mathbf{v})}$, depends on the similarity of preferences, ϱ. The following subsections show that under G_0, $\mathcal{O}_{(G_0,\mathbf{v})}$ is empty: no firm wants to contract in $t = 1$, while under G_1, $\mathcal{O}_{(G_1,\mathbf{v})}$ may be nonempty, depending on $\mathbf{v}$. For intermediate cases, $\mathcal{O}_{(\varrho,\mathbf{v})}$ increases with ϱ.

Independent Preferences, G_0

For independently distributed rankings, no firm prefers early contracting to the ex-post stable outcome. Therefore, in any market with independent preferences, there is an equilibrium without unravelling.

Lemma 3. *For any F and $W > F$, under G_0, no firm has incentive to contract in $t = 1$. That is,*

$$\forall F \quad \forall W > F \quad \forall f \qquad \pi^0 < E\pi_f(\mathfrak{o}_S|G_0)$$

Proof.

Consider the worst firm, firm 1.

$$Prob(\mathfrak{o}_S(f) = r_k^1|G_0, W) \equiv P_s(W, 1, k) = \frac{(F-1)!}{(F-W-1+k)!}\frac{(k-1)!}{W!}(W-F+1)$$

$$\text{for } k = (W-F+1), \ldots, W$$

and 0 for $k < W - F + 1$.

By induction, it can be shown that $P_s(n, 1, k) > P_s(n, 1, k')$ for $k > k'$. Therefore, distribution $P_s(n, 1, k)$ first order stochastically dominates distribution $P_0(W, 1, k) = \frac{1}{W}$ for any k, which is the distribution for early matches. Thus, $E\pi_1(\mathfrak{o}_S|G_0) > \pi^0$ in any market with G_0.

By Lemma 5(1) for any firm better than firm 1 the payoff from the ex-post stable outcome is higher. Therefore, all firms prefer to wait for $\mathfrak{o}_S$ than to unravel. □

The intuition for this result as follows. Consider the worst firm, firm 1. All other firms are matched before firm 1 in the ex-post stable outcome. If the number of workers were the same as the number of firms, $W = F$, there would be exactly one worker left for firm 1 to match with. Since the preferences are independent, this last worker may have any position in firm 1's ranking with equal probability. In such a case, the ex-post stable outcome and early contracting would yield exactly the same expected payoff for firm 1, and the firm would be indifferent. However, since $W > F$, the worst firm prefers the ex-post stable outcome to early contracting. This is because with more than one worker to choose from, firm 1 will never be matched with the worst worker, and has higher chances (than $\frac{1}{W}$) to be matched with any better worker. Moreover, by the property of the ex-post stable outcome that better firms have higher expected payoff (Lemma 2(1)), any other firm also prefers the ex-post stable outcome to early contracting.

Identical Preferences, G_1

Under identical preferences, the k-th worker of firm f is also any other firm's k-th worker. In the ex-post stable outcome, firm F gets the best worker, r^F_W, firm $(F-1)$ always gets the next best worker, r^{F-1}_{W-1}, and firm f always gets the worker ranked $(W-F+f)$, r^f_{W-F+f}. Thus, $E\pi_f(\mathfrak{o}_S|G_1) = v_{W-F+f}$.

Under G_1, condition (2.3) reduces to:

$$\frac{1}{W}\sum_{k=1}^{W} v_k > v_{W-F+f}$$

Firm f prefers to contract early rather than to wait for the ex-post stable outcome if the average value of workers is larger than v_{W-F+f}. This may be true for some values of $\mathbf{v}$. With nonempty offer set, there exists profitable deviation from $\mathfrak{o}_S$ for some acceptance sets.

Example 2 shows a market with identical preferences of firms, where there exists a profitable deviation.

Example 2. *Consider market with 3 firms and 4 workers and with matching values vectors* $\mathbf{v} = [1,\ 2,\ 3,\ 4]$ *and* $\mathbf{u} = [4,\ 5,\ 6]$*, and with identical firms' preferences,* G_1*.*
The ex-post stable outcome is

$$\mathfrak{o}_S(f_3) = r_4^3 \quad \Longrightarrow \quad \pi_3(\mathfrak{o}_S) = 4$$

$$\mathfrak{o}_S(f_2) = r_3^2 \quad \Longrightarrow \quad \pi_2(\mathfrak{o}_S) = 3$$

$$\mathfrak{o}_S(f_1) = r_2^1 \quad \Longrightarrow \quad \pi_1(\mathfrak{o}_S) = 2$$

An early offer yields expected payoff of 2.5*. Since* $2 < 2.5 < 3$*, firm 2 has no incentive to make an early offer, but firm 1 prefers to contract in* $t = 1$ *than wait for* r_2^1 *in* $t = 1$*. That is,* $\mathcal{O}_{(G_1,\mathbf{v})} = \{1\}$*.*

A worker's expected utility from $t = 2$ *matching is* $\frac{1}{W}\sum_{f=1}^{F} u_f = \frac{15}{4} < 4 = u_1$*. It means that firm* 1*'s offer in* $t = 1$ *will be accepted by any worker. Thus,* $\mathcal{A}_{(4,\mathbf{u})} = \{1, 2, 3\}$*.*

Since $\mathcal{O}_{(G_1,\mathbf{v})} \cap \mathcal{A}_{(4,\mathbf{u})} = \{1\} \neq \emptyset$, *there exists a profitable deviation from* $\mathfrak{o}_S$ *in this market.* ■

However, a profitable deviation from $\mathfrak{o}_S$ may not exist even when firms' preferences are identical. When any firm that prefers to contract early would be rejected by a worker in $t = 2$, there is no profitable deviation. Such a market is presented in Example 3.

Example 3. *Consider a market similar to that in Example 2, with the only difference that* $\mathbf{u}' = [2, 3, 4]$. *As before,* $\mathcal{O}_{(G_1,\mathbf{v})} = \{1\}$, *but now firm* $1 \notin \mathcal{A}_{(4,\mathbf{u}')}$. *There is no profitable deviation from* $\mathfrak{o}_S$ *in this market, as* $\mathcal{O}_{(G_1,\mathbf{v})} \cap \mathcal{A}_{(4,\mathbf{u}')} = \emptyset$. ■

As the examples above illustrate, under identical preferences a profitable deviation from $\mathfrak{o}_S$ may, but does not have to exist. This can also be interpreted in terms of existence of an equilibrium without unravelling. There are markets with G_1 where there is an equilibrium without unravelling, but there also are markets where any equilibrium must exhibit unravelling.

Intermediate Similarity of Firms' Preferences

Firm F has always the same value of the ex-post stable matching: v_W. For all the other firms, the expected value of $\mathfrak{o}_S$ decreases as the similarity of preferences increases (Lemma 2(2)). As a consequence, holding other parameters of the market constant, more firms prefer early contracting as the similarity increases. That is, holding other market parameters constant, $\mathcal{O}_{(\varrho,\mathbf{v})} \subseteq \mathcal{O}_{(\varrho',\mathbf{v})}$ whenever $\varrho < \varrho'$. Therefore, if for given market parameters $(F, W, \mathbf{v}, \mathbf{u})$ there exists a profitable deviation from $\mathfrak{o}_S$ under G_ϱ, then there also exists a profitable deviation under $G_{\varrho'}$. In fact, for any market parameters

$(F, W, \mathbf{v}, \mathbf{u})$, there exists a threshold ϱ^{**} such that for any similarity higher than the threshold a profitable deviation from $\mathfrak{o}_S$ exists, but not for similarity lower than the threshold.

Lemma 4. *For any market parameters* $(F, W, \mathbf{v}, \mathbf{u})$*, there exists* $\varrho^{**} \in (0, 1]$ *s.t.*

for all $\varrho \leq \varrho^{**}$*, there exists an equilibrium without unravelling, and*

for all $\varrho > \varrho^{**}$*, there is no equilibrium without unravelling.*

Proof.

An equilibrium without unravelling exists when $\mathcal{A} \cap \mathcal{O} = \emptyset$. $\mathcal{A}$ does not depend on ϱ.

For $\varrho = 0$, $\mathcal{A} \cap \mathcal{O} = \emptyset$ by Lemma 3. If for $\varrho = 1$, $\mathcal{A} \cap \mathcal{O} = \emptyset$, then $\varrho^{**} = 1$ and for all $\varrho \in [0, 1]$ there is an equilibrium without unravelling. If for $\varrho = 1$, $\mathcal{A} \cap \mathcal{O} \neq \emptyset$, then, by monotonicity of $\mathcal{O}$ (that is the property that $\varrho < \varrho' \implies \mathcal{O}(\varrho) \subset \mathcal{O}(\varrho')$) there must exists ϱ^{**} such that $\mathcal{A} \cap \mathcal{O}(\varrho) = \emptyset$ and for any $\varrho > \varrho^{**}$, $\mathcal{A} \cap \mathcal{O}(\varrho') \neq \emptyset$. □

Workers' incentives to accept an offer in $t = 1$ do not depend on the similarity of preferences. However, firms' expected payoffs from the ex-post stable outcome decrease as the preferences become more similar. Consequently, unravelling becomes more tempting. For G_0 there are no market parameters $(F, W, \mathbf{v}, \mathbf{u})$ for which a profitable deviation from $\mathfrak{o}_S$ exists. But as similarity of preferences, ϱ, increases, there are more parameters $(F, W, \mathbf{v}, \mathbf{u})$ for which a profitable deviation exists. Thus, the result in Lemma 4 implies that as the similarity of preferences increases, profitable deviation from $\mathfrak{o}_S$ exists for a wider range of $(F, W, \mathbf{v}, \mathbf{u})$ parameters, and so "no unravelling" is not an equilibrium for a wider range of $(F, W, \mathbf{v}, \mathbf{u})$ parameters.

When for given market parameters $(F, W, \mathbf{v}, \mathbf{u})$ the threshold is $\varrho^{**} < 1$, then for high enough similarity of preferences there is no equilibrium without unravelling. In Example 2, the threshold is strictly below 1. However, when the threshold is $\varrho^{**} = 1$, there is an equilibrium without unravelling for any preferences, as in Example 3. Yet, Lemma 3 assures that for any market parameters the threshold is strictly larger than 0. That is, for a market with independent preferences, G_0, an equilibrium always exists.

2.2.2. Equilibria with Unravelling

The previous section analyzes the conditions for which in an equilibrium all firms participate in $\mathcal{M}_S$, without unravelling. But this is only one of the possible equilibrium outcomes in this game. Other equilibria may involve contracting in period 1. This section analyzes pure strategy equilibria in which some early contracting takes place.

Firms and workers that contract early exit the market before $t = 2$. In the second period, all remaining agents participate in the ex-post stable matching mechanism. In equilibrium, worker w with offers Ω_w in $t = 1$ either accepts the best offer in Ω_w or rejects all of them, depending on which of the two maximizes his expected utility. It is suboptimal for a worker to accept an offer of a firm other than the best firm in Ω_w. Therefore, for a firm that prefers to contract in $t = 1$ it is suboptimal to make an offer to the same worker as a better firm. In equilibrium all firms that want to contract early make offers to different workers.

Every equilibrium results in a set of firms that unravel, or contract early. This *equilibrium unravelling set* is denoted by $\mathcal{U}$. The remaining firms, $\{1, \ldots, F\} \smallsetminus \mathcal{U}$, participate in $t = 2$ in $\mathcal{M}_S$ with unmatched workers still present in the market at this time. The equilibrium unravelling set may be empty – such an equilibrium does not involve unravelling. There may be more than one equilibrium resulting in the same unravelling set $\mathcal{U}$. For example in one equilibrium some firm makes an early offer and is rejected, and in another equilibrium this firm does not make the early offer. Despite different strategies played, both equilibria yield the same outcome. All equilibria resulting in the same unravelling set $\mathcal{U}$ are considered to be equivalent and henceforth $\mathcal{U}$ characterizes this class of equilibria.

A property of any equilibrium is that the unravelling set, $\mathcal{U}$, is an interval, i.e. it has no "holes". For the given equilibrium unravelling set $\mathcal{U}^*$, let firm $\mathbb{H}^*$ be the highest ranked firm in $\mathcal{U}^*$, and firm $\mathbb{L}^*$ – the lowest one in $\mathcal{U}^*$. The fact that $\mathcal{U}^*$ is an interval means that all firms worse than $\mathbb{H}^*$ but better than $\mathbb{L}^*$ are in $\mathcal{U}^*$ as well.

To see why this is true, suppose, to the contrary, that in some equilibrium $\mathbb{L}^*$ and $\mathbb{H}^*$ belong to $\mathcal{U}^*$ but there is a firm f between $\mathbb{L}^*$ and $\mathbb{H}^*$ that is not in $\mathcal{U}^*$. That must be either because f prefers to wait, or because it would not be accepted in $t = 1$. But since f is lower ranked than $\mathbb{H}^*$, it prefers to contract early (as $\mathbb{H}^*$ does). And since it is better than $\mathbb{L}^*$, it would be accepted (as $\mathbb{L}^*$ is). Therefore, it can not be an equilibrium if $\mathbb{L}^*$ and $\mathbb{H}^*$ are in $\mathcal{U}^*$, and f is not. This result is formally stated in Lemma 5 below.

Thus, any nonempty $\mathcal{U}^*$ can be characterized by the best firm ($\mathbb{H}^*$) and the worst firm ($\mathbb{L}^*$) that contract early in such equilibrium: $\mathcal{U}^* \equiv \{\mathbb{L}^*, \ldots, \mathbb{H}^*\}$, for $\mathbb{L}^* \leq \mathbb{H}^*$. And an equilibrium is characterized by two conditions – one for workers and one for firms – that pin down the bounds of the equilibrium unravelling set. Given $\mathbb{H}^*$, the *equilibrium* *condition for workers* characterizes $\mathbb{L}^*$, i.e. the worst firm that would be accepted in $t = 1$. That is, given that only firms $\mathbb{H}^*$ and below would like to make early offers, workers are willing to accept only firms $\mathbb{L}^*$ and above in $t = 1$. Similarly, given $\mathbb{L}^*$, the *equilibrium condition for firms* characterizes $\mathbb{H}^*$, i.e. the best firm contracting in period 1.

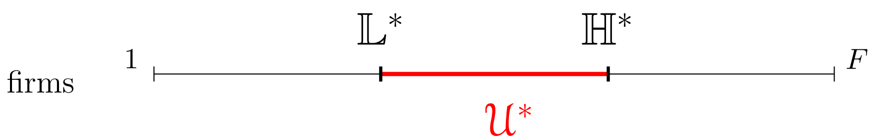


Figure 2.2. The structure of an equilibrium.

In any market there is at least one equilibrium. Consider a market $(F, W, \mathbf{u}, \mathbf{v}, \varrho, \mathcal{M}_S)$ where there exists a profitable deviation from $\mathfrak{o}_S$; that is, waiting for $\mathfrak{o}_S$ without unravelling is not an equilibrium. Then there is a set of firms $\mathcal{A}_{(W,\mathbf{u})} \cap \mathcal{O}_{(\varrho,\mathbf{v})} \neq \emptyset$ that would like to contract early and would be accepted in $t = 1$. But if those firms unravel, the expected payoff of staying to the second period decreases for better firms. This is because firms that unravel hire the workers, with a positive probability, that would be matched to the better firms in the ex-post stable outcome. When this happens, the better firms are matched with some worse workers in $t = 2$. This decrease in the expected payoff may induce some better firms to decide for early contracting, even though they initially preferred to wait for $\mathfrak{o}_S$. If more firms unravel, that also decreases expected payoff of workers in the second period. This may induce workers to accept firms in $t = 1$ that previously would not be accepted. This again may increase the number of firms that unravel. Eventually, either the process induces all the firms $1, \ldots, (F-1)$ to unravel,[11] or it reaches a "fixed" state

[11]Firm F always prefers to wait for the ex-post stable mechanism in $t = 2$, even if all the other firms unravel.

earlier. In both cases the market reaches an equilibrium with nonempty unravelling set. Thus, in every market there is at least one equilibrium. This result is formally stated in Lemma 5 below. Moreover, in a typical markets there are more than one.

A market with multiple equilibria is presented in Example 4.

Example 4. *Consider a market with 5 firms and 6 workers where* $\mathbf{u} = [2, 5, 6, 9, 10]$, $\mathbf{v} = [2, 3, 4, 5, 8, 17]$ *and firms' preferences are identical,* G_1*. In this market there are two possible unravelling sets in pure strategy equilibria:* $\mathcal{U}^* = \{3\}$ *and* $\mathcal{U}' = \{2, 3, 4\}$.

Firms' condition for $\mathcal{U}^*$ *is as follows: Knowing that firm 3 or better would be accepted in* $t = 1$*, firm* 3 *prefers the early contracting, but firm* 4 *prefers to wait for the ex-post stable matching – without firm* 3 *– in* $t = 2$*. Workers' condition for* $\mathcal{U}^*$ *says that knowing that firms 5 and 4 prefer to participate in* $t = 2$ *matching, a worker accepts firm* 3 *but not firm* 2*, in* $t = 1$*. Matching with firm* 2 *yields lower utility for a worker than the expectations over* $t = 2$*, even without firm* 3*.*

Conditions for $\mathcal{U}'$ *are calculated in a similar fashion.* ■

In Example 4 both equilibrium unravelling sets were nonempty. But it does not need to be so. The following example shows a market with multiple equilibria – some with unravelling, while others without unravelling.

Example 5. *Consider a market similar to the one in Example 4, with the only difference that* $\mathbf{u'} = [1, 6, 7, 13, 14]$*. In such a market there are also exactly 2 equilibrium unravelling sets. One is the same a before,* $\mathcal{U}' = \{2, 3, 4\}$*, but the other is* $\mathcal{U}^* = \emptyset$.

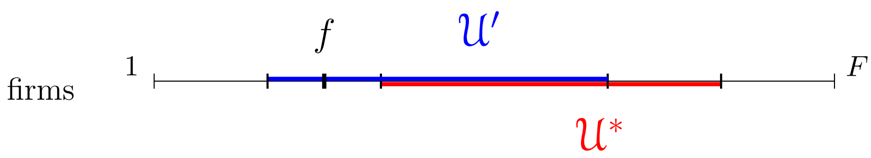
That $\mathcal{U}^* = \emptyset$ *is an equilibrium unravelling set is verified by showing that* $\mathcal{O}_{(G_1,\mathbf{v})} \cap \mathcal{A}_{(6,\mathbf{u'})} = \emptyset$*. Since utility from* $\mathfrak{o}_S$ *expected by a worker is* $\frac{1}{6}\sum u_f = 7\frac{1}{6}$*, the acceptance set* *in this market is* $\mathcal{A}_{(6,\mathbf{u}')} = \{4,\, 5\}$. *As the expected payoff of an early offer is* $\frac{1}{6}\sum v_k = 6.5$, *the offer set is* $\mathcal{O}_{(G_1,\mathbf{v})} = \{1,\, 2,\, 3\}$. *Therefore,* $\mathcal{O}_{(G_1,\mathbf{v})} \cap \mathcal{A}_{(6,\mathbf{u}')} = \emptyset$, *i.e.* $\mathcal{U}^* = \emptyset$ *is an equilibrium unravelling set.* ■

However, equilibrium unravelling sets cannot be arbitrary. For any two equilibrium unravelling sets in a given market, one needs to be fully included in the other. In particular, two equilibrium unravelling sets for the same market can not "overlap". To see why, suppose, to the contrary, that there exist two equilibrium unravelling sets $\mathcal{U}^*$ and $\mathcal{U}'$ as in Figure 2.3(a). There are two effects playing a role here – a "number effect" and a "position effect". The former one exists when $\mathcal{U}^*$ and $\mathcal{U}'$ are of different sizes. The latter follows from different "position" of unravelling sets among all firms. To consider only the "position effect" first, assume that $\mathcal{U}^*$ and $\mathcal{U}'$ are of the same size, but $\mathcal{U}^*$ includes better firms (on average) than $\mathcal{U}'$, as the figure shows. By the equilibrium condition for workers, firm f is not included in $\mathcal{U}^*$ because its early offer would not be accepted. But under $\mathcal{U}^*$ better firms unravel than under $\mathcal{U}'$. Thus, expected utility from staying in the market for $t = 2$ is lower for workers under $\mathcal{U}^*$ than under $\mathcal{U}'$. If under $\mathcal{U}'$ it was better for a worker to accept firm f in $t = 1$ than to wait for the expected utility in $t = 2$, it also must be so under $\mathcal{U}^*$. The "number effect" does not change this result.

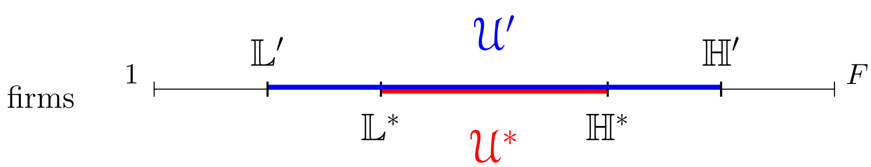
The following lemma summarizes properties of equilibria in an arbitrary market with the ex-post stable matching mechanism, $(F, W, \mathbf{u}, \mathbf{v}, \varrho, \mathcal{M}_S)$.

Lemma 5. *Given a market* $(F, W, \mathbf{u}, \mathbf{v}, \varrho, \mathcal{M}_S)$:

(1) **(convexity of unravelling set)** *In any equilibrium, the equilibrium unravelling set,* $\mathcal{U}^*$, *is an interval.*



(a) An impossible configuration of multiple equilibrium unravelling sets



(b) A possible configuration of multiple equilibrium unravelling sets

Figure 2.3. Multiple equilibria with unravelling.

(2) **(existence of pure strategy equilibrium)** *There exists an equilibrium in pure strategies.*

(3) **(multiple equilibria)** *If there are two equilibrium unravelling sets, $\mathfrak{U}^*$ and $\mathfrak{U}'$ where $\mathfrak{U}^* \neq \mathfrak{U}'$, then $\mathfrak{U}^* \subsetneq \mathfrak{U}'$. Moreover, if both unravelling sets are nonempty, $\mathfrak{U}^* = \{\mathbb{L}^*, \ldots, \mathbb{H}^*\}$ and $\mathfrak{U}' = \{\mathbb{L}', \ldots, \mathbb{H}'\}$ then*

$$\mathbb{L}' < \mathbb{L}^* \quad \Longleftrightarrow \mathbb{H}^* < \mathbb{H}'$$

Proof. The proof proceeds separately for each point:

(1) **(interval)** ***Proof.***

Assume, to the contrary, that there exists a firm f s.t. $\mathbb{H}^* > f > \mathbb{L}^*$ and $f \notin \mathfrak{U}^*$. If the firm is not in $\mathfrak{U}^*$, it must be either because it prefers to wait, or because it would not be accepted in $t = 1$.

Since $\mathbb{L}^*$ is accepted in $t = 1$, given all the other firms in $\mathfrak{U}^*$ contracting early, then $EU(t = 2 | \mathfrak{U}^* \setminus \{\mathbb{L}^*\}) < u_{\mathbb{L}^*}$. But if an acceptable firm contracts in $t = 1$, the expected utility of $t = 2$ matching decreases for workers:

$$EU(t = 2|\,\mathcal{U}^* \setminus \{\mathbb{L}^*\}) < u_{\mathbb{L}^*} \quad \Longrightarrow \quad EU(t = 2|\,\mathcal{U}^*) < EU(t = 2|\,\mathcal{U}^* \setminus \{\mathbb{L}^*\})$$

Moreover, $u_{\mathbb{L}^*} < u_f$. Thus, $EU(t = 2|\,\mathcal{U}^*) < u_f$. Therefore, f would be accepted by a worker in $t = 1$.

Since firm $\mathbb{H}^*$ prefers contracting early, given all the other firms in $\mathcal{U}^*$ contracting in $t = 1$,

$$E\pi_{\mathbb{H}^*}(\mu_S^{\mathcal{F},\mathcal{W}}|\,\mathcal{U}^* \setminus \{\mathbb{H}^*\}) < \pi^0$$

Since lower ranked firms get lower expected payoff in the ex-post stable matching,

$$E\pi_f(\mu_S^{\mathcal{F},\mathcal{W}}|\mathcal{U}^* \setminus \{\mathbb{H}^*\}) < E\pi_{\mathbb{H}^*}(\mu_S^{\mathcal{F},\mathcal{W}}|\,\mathcal{U}^* \setminus \{\mathbb{H}^*\})$$

Since when better firms unravel, the expected payoff from $t = 2$ matching decreases:

$$E\pi_f(\mu_S^{\mathcal{F},\mathcal{W}}|\,\mathcal{U}^*) < E\pi_f(\mu_S^{\mathcal{F},\mathcal{W}}|\mathcal{U}^* \setminus \{\mathbb{H}^*\})$$

Together, it yields $E\pi_f(\mu_S^{\mathcal{F},\mathcal{W}}|\,\mathcal{U}^*) < \pi^0$. Therefore, f prefers to contract in $t = 1$ than to wait.

Hence, the contradiction. This proves part (1) of Lemma 5. □

(2) **(existence)** ***Proof.***

Given $\mathbb{H}^*$, the *equilibrium condition for workers* (**CW**) characterizes $\mathbb{L}^*$, i.e. the worst firm that would be accepted in $t = 1$.

With firms $\mathbb{L}, \ldots, \mathbb{H}$ contracting in $t = 1$, a worker's expected payoff from matching in $t = 2$ is

$$EU\big(t = 2 \,|\, \{\mathbb{L}, \ldots, \mathbb{H}\}\big) = \frac{1}{W - \mathbb{H} + \mathbb{L} - 1}\left(\sum_{f=1}^{\mathbb{L}-1} u_f + \sum_{f=\mathbb{H}+1}^{F} u_f\right)$$

The worst firm that would be accepted in $t = 1$, given $\mathbb{H}^*$, is characterized by two inequalities:[12]

$$u_{\mathbb{L}^*} > \frac{1}{W - \mathbb{H}^* + \mathbb{L}^* - 1}\left(\sum_{f=1}^{\mathbb{L}^*-1} u_f + \sum_{f=\mathbb{H}^*+1}^{F} u_f\right)$$

(**CW**) and

$$u_{\mathbb{L}^*-1} \leq \frac{1}{W - \mathbb{H}^* + \mathbb{L}^* - 1}\left(\sum_{f=1}^{\mathbb{L}^*-1} u_f + \sum_{f=\mathbb{H}^*+1}^{F} u_f\right)$$

The first part of equilibrium condition (**CW**) indicates that if firms $\mathbb{L}^* + 1, \ldots, \mathbb{H}^*$ contract early, then firm $\mathbb{L}^*$ would also be accepted in $t = 1$. The second part of the condition assures that $\mathbb{L}^*$ is the lowest ranked firm that would be accepted in $t = 1$, given $\mathbb{H}^*$. That is, lower firm $\mathbb{L}^* - 1$ is not accepted in $t = 1$, when firms $\mathbb{L}^*, \ldots, \mathbb{H}^*$ contract early.

[12] More precisely, the first inequality is

$$u_{\mathbb{L}^*} > \frac{1}{W - \mathbb{H}^* + \mathbb{L}^*}\left(\sum_{f=1}^{\mathbb{L}^*} u_f + \sum_{f=\mathbb{H}^*+1}^{F} u_f\right)$$

but if all $u_{\mathbb{L}^*}$ terms are moved to the LHS, the equivalent inequality is as above.

Given $\mathbb{L}^*$, the *equilibrium condition for firms* (**CF**) characterizes $\mathbb{H}^*$, i.e. the best firm contracting in period 1. Two inequalities constitute this condition:

$$E\pi_{\mathbb{H}^*}\big(\mathcal{M}_S|\mathcal{U}(\mathbb{L}^*, \mathbb{H}^* - 1)\big) < \pi^0$$

(**CF**) and

$$E\pi_{\mathbb{H}^*+1}\big(\mathcal{M}_S|\mathcal{U}(\mathbb{L}^*, \mathbb{H}^*)\big) \geq \pi^0$$

where $E\pi_f(\mathcal{M}_S|\mathcal{U})$ represents expected payoff of firm f from the ex-post stable matching mechanism in $t = 2$, given that firms in $\mathcal{U}$ contract in $t = 1$, and $\mathcal{U}(\mathbb{L}, \mathbb{H})$ is a shorthand for $\mathcal{U} = \{\mathbb{L}, \ldots, \mathbb{H}\}$. In case the of $\mathbb{H}^* = \mathbb{L}^*$, $\mathcal{U}(\mathbb{L}^*, \mathbb{H}^* - 1) = \emptyset$.

The first part of equilibrium condition (**CF**) says that firm $\mathbb{H}^*$ prefers to contract early, given that firms $\mathbb{L}^*, \ldots, (\mathbb{H}^* - 1)$ do. The second part of the condition assures that $\mathbb{H}^*$ is the highest ranked firm that wants to contract early. That is, better firm $(\mathbb{H}^* + 1)$, prefers to wait for $t = 2$, given that firms $\mathbb{L}^*, \ldots, \mathbb{H}^*$ unravel. As Lemma 6 indicates, this also means that – given $\mathbb{L}^*$ – all firms worse than $\mathbb{H}^*$ prefer to contract early, and all firms better than $\mathbb{H}^*$ prefer to wait.

To continue with the proof of part (2) of Lemma 5, one needs to establish results presented in following lemmas.

Lemma 6. *For any* $\mathbb{L}^*$ *and any* $f \geq \mathbb{L}^*$

$$E\pi_{f+1}\big(\mathcal{M}_S|\mathcal{U}(\mathbb{L}^*, f)\big) \geq E\pi_f\big(\mathcal{M}_S|\mathcal{U}(\mathbb{L}^*, f - 1)\big)$$

where equality holds for $\varrho = 0$*, and strict inequality holds otherwise.*

Lemma 7. (equilibrium with unravelling) *There exists an equilibrium with nonempty unravelling set* $\mathcal{U}^* = \{\mathbb{L}^*, \ldots, \mathbb{H}^*\}$ *if and only if*

(cf) given $\mathbb{L}^*$*,* $\mathbb{H}^*$ *satisfies condition* (**CF**)*, and*

(cw) given $\mathbb{H}^*$*,* $\mathbb{L}^*$ *satisfies condition* (**CW**)*.*

In a market there exists an equilibrium with nonempty unravelling set $\mathcal{U}^* = \{\mathbb{L}^*, \ldots, \mathbb{H}^*\}$ if and only if given $\mathbb{L}^*$, $\mathbb{H}^*$ satisfies (**CF**) and given $\mathbb{H}^*$, $\mathbb{L}^*$ satisfies (**CW**).

Lemma 8. *In a given market* $(F, W, \mathbf{u}, \mathbf{v}, \varrho, \mathcal{M})$*, for any* $\mathbb{H}^*$*, if there exists* $\mathbb{L}^* \leq \mathbb{H}^*$ *that satisfies condition* (**CW**)*, it is unique.*

Proof.

Assume, to the contrary, that two distinct $\mathbb{L}^*$ and $\mathbb{L}'$, lower than $\mathbb{H}^*$, satisfy (**CW**) given $\mathbb{H}^*$ Without loss of generality, $\mathbb{L}' < \mathbb{L}^* < \mathbb{H}^*$, i.e. $u_{\mathbb{L}'} < u_{\mathbb{L}^*}$. Then,

$$u_{\mathbb{L}^*} > \frac{1}{W - \mathbb{H}^* + \mathbb{L}^* - 1}\left(\sum_{f=1}^{\mathbb{L}^*-1} u_f + \sum_{f=\mathbb{H}^*+1}^{F} u_f\right) \equiv EU(t = 2 \,|\, \mathcal{U}(\mathbb{L}^*, \mathbb{H}^*))$$

and

$$u_{\mathbb{L}^*-1} < \frac{1}{W - \mathbb{H}^* + \mathbb{L}^* - 1}\left(\sum_{f=1}^{\mathbb{L}^*-1} u_f + \sum_{f=\mathbb{H}^*+1}^{F} u_f\right)$$

as well as

$$u_{\mathbb{L}'} > \frac{1}{W - \mathbb{H}^* + \mathbb{L}' - 1}\left(\sum_{f=1}^{\mathbb{L}'-1} u_f + \sum_{f=\mathbb{H}^*+1}^{F} u_f\right) \equiv EU(t = 2 \,|\, \mathcal{U}(\mathbb{L}', \mathbb{H}^*))$$

and

$$u_{\mathbb{L}'-1} < \frac{1}{W - \mathbb{H}^* + \mathbb{L}' - 1}\left(\sum_{f=1}^{\mathbb{L}'-1} u_f + \sum_{f=\mathbb{H}^*+1}^{F} u_f\right)$$

That is

$$u_{\mathbb{L}'-1} < EU(t=2\,|\,\mathcal{U}(\mathbb{L}',\mathbb{H}^*)) < u_{\mathbb{L}'} \leq u_{\mathbb{L}^*-1} < EU(t=2\,|\,\mathcal{U}(\mathbb{L}^*,\mathbb{H}^*)) < u_{\mathbb{L}^*} \tag{2.4}$$

Let $n' = W - \mathbb{H}^* + \mathbb{L}' - 1$ and $\Delta\mathbb{L} = \mathbb{L}^* - \mathbb{L}' > 0$. Then $W - \mathbb{H}^* + \mathbb{L}^* - 1 = n' + \Delta\mathbb{L}$ and

$$\begin{aligned} EU(t=2\,|\,\mathcal{U}(\mathbb{L}^*,\mathbb{H}^*)) &= \frac{1}{n'+\Delta\mathbb{L}}\left(\sum_{f=1}^{\mathbb{L}'-1} u_f + \sum_{f=\mathbb{L}'}^{\mathbb{L}^*-1} u_f + \sum_{f=\mathbb{H}^*+1}^{F} u_f\right) \\ EU(t=2\,|\,\mathcal{U}(\mathbb{L}',\mathbb{H}^*)) &= \frac{1}{n'}\left(\sum_{f=1}^{\mathbb{L}'-1} u_f + \sum_{f=\mathbb{H}^*+1}^{F} u_f\right) \end{aligned}$$

So

$$(n'+\Delta\mathbb{L})\cdot EU(t=2\,|\,\mathcal{U}(\mathbb{L}^*,\mathbb{H}^*)) - \sum_{f=\mathbb{L}'}^{\mathbb{L}^*-1} u_f \;=\; n'\cdot EU(t=2\,|\,\mathcal{U}(\mathbb{L}',\mathbb{H}^*)) \qquad \Longleftrightarrow$$

$$\Longleftrightarrow \qquad n'(EU(t=2\,|\,\mathcal{U}(\mathbb{L}',\mathbb{H}^*)) - EU(t=2\,|\,\mathcal{U}(\mathbb{L}^*,\mathbb{H}^*))) \;=\; \sum_{f=\mathbb{L}'}^{\mathbb{L}^*-1}(EU(t=2\,|\,\mathcal{U}(\mathbb{L}^*,\mathbb{H}^*)) - u_f)$$

From the equilibrium condition for $\mathcal{U}^*$, $EU(t=2\,|\,\mathcal{U}(\mathbb{L}^*,\mathbb{H}^*)) > u_{\mathbb{L}^*-1}$. Thus all terms in the sum on the RHS are positive. Therefore, $EU(t=2\,|\,\mathcal{U}(\mathbb{L}',\mathbb{H}^*)) > EU(t=2\,|\,\mathcal{U}(\mathbb{L}^*,\mathbb{H}^*))$. But from (2.4), $EU(t=2\,|\,\mathcal{U}(\mathbb{L}',\mathbb{H}^*)) < EU(t=2\,|\,\mathcal{U}(\mathbb{L}^*,\mathbb{H}^*))$. Thus, a contradiction. This proves Lemma 8 □

Lemma 9. *In a given market,* $(F, W, \mathbf{u}, \mathbf{v}, \varrho, \mathcal{M})$*, for any* $\mathbb{L}^*$*, if there exists* $\mathbb{H}^* \geq \mathbb{L}^*$ *satisfying the equilibrium condition for firms* (**CF**)*, it is unique.*

Proof.

Assume, to the contrary, that two distinct $\mathbb{H}^*$ and $\mathbb{H}'$, higher than $\mathbb{L}^*$, satisfy condition (**CF**) given $\mathbb{L}^*$. Without loss of generality, $\mathbb{H}' > \mathbb{H}^* \geq \mathbb{L}^*$. By Lemma 6

$$E\pi_{\mathbb{H}'+1}(\mu_S^{\mathcal{F},\mathcal{W}} \,|\, \mathcal{U}(\mathbb{L}^*, \mathbb{H}')) > E\pi_{\mathbb{H}'}(\mu_S^{\mathcal{F},\mathcal{W}} \,|\, \mathcal{U}(\mathbb{L}^*, \mathbb{H}'-1)) \geq$$

$$\geq E\pi_{\mathbb{H}^*+1}(\mu_S^{\mathcal{F},\mathcal{W}} \,|\, \mathcal{U}(\mathbb{L}^*, \mathbb{H}^*)) > E\pi_{\mathbb{H}^*}(\mu_S^{\mathcal{F},\mathcal{W}} \,|\, \mathcal{U}(\mathbb{L}^*, \mathbb{H}^*-1))$$

(If there exists at least one $\mathbb{H}$ firm, it must be that $\varrho > 0$, because for G_0 there is no firm satisfying (**CF**) for any $\mathbb{L}^*$. Therefore the strict inequalities.)

Since π^0 is constant, either $\mathbb{H}'$ does not satisfy the first part of the condition, or $\mathbb{H}^*$ does not satisfy the second part. They can not both satisfy the condition.

This proves Lemma 9. □

With those lemmas at hand, the proof of Lemma 5 part (2) can be completed. Notice that there exists an equilibrium without unravelling, $\mathcal{U} = \emptyset$, if and only if $\mathcal{A}_{(W,\mathbf{u})} \cap \mathcal{O}_{(\varrho,\mathbf{v})} = \emptyset$.

If in a market there exists an equilibrium without unravelling, the existence of a pure strategy equilibrium is satisfied. To proof that in the case that $\mathcal{A} \cap \mathcal{O} \neq \emptyset$, there always exists an equilibrium in pure strategies with unravelling, notice following.

There is a unique $\mathbb{L}$ satisfying condition (**CW**) for given $\mathbb{H}$. Moreover, this $\mathbb{L}$ is decreasing as $\mathbb{H}$ increases. Similarly, there is a unique $\mathbb{H}$ satisfying condition (**CF**) for given $\mathbb{L}$. This $\mathbb{H}$ increases as $\mathbb{L}$ decreases.

Now, since $\mathcal{A} \cap \mathcal{O} \neq \emptyset$, $\mathbb{L}_{(F,\mathbf{u})} \leq \mathbb{H}_{(\varrho,\mathbf{v})}$. Now, let $\mathbb{H}^1 = \mathbb{H}(\mathbb{L}_{(F,\mathbf{u})}$ and $\mathbb{L}^1 = \mathbb{L}(\mathbb{H}_{(\varrho,\mathbf{v})})$, and further $\mathbb{H}^{i+1} = \mathbb{H}(\mathbb{L}^i)$ and $\mathbb{L}^{i+1} = \mathbb{L}(\mathbb{H}^i)$. Because of the monotonicity result above, it must be that there exists $\mathbb{H}^*$ and $\mathbb{L}^*$ such that $\mathbb{H}^*(\mathbb{L}^*)$ and $\mathbb{L}^*(\mathbb{H}^*)$. Thus, an equilibrium exists. This completes the proof of part (2) of Lemma 5. □

(3) **multiple equilibria**: ***Proof.***

With Lemmas 9 and 8, it remains to show that "overlapping equilibria", i.e. $\mathbb{L}^* > \mathbb{L}'$ and $\mathbb{H}^* > \mathbb{H}'$, are not possible.

Assume, to the contrary that there exist two "overlapping" equilibria. Let $\Delta\mathbb{H} = \mathbb{H}^* - \mathbb{H}' > 0$, and $\Delta\mathbb{L} = \mathbb{L}^* - \mathbb{L}' > 0$.

For $\boldsymbol{\Delta}\mathbb{L} > \boldsymbol{\Delta}\mathbb{H}$. That is, size of $\mathcal{U}^*$ is larger than size of $\mathcal{U}'$. It must be then that firm $\mathbb{H}^*$ prefers $t = 1$, given that $\mathbb{H}^* - \mathbb{L}^*$ firms unravel. But the same firm prefers to wait for $t = 2$ if $\mathbb{H}' - \mathbb{L}' + 1$ firms unravel. But following lemma shows that as more firms (worse than $\mathbb{H}^*$) unravel, the expected reward from $t = 2$ for firm $\mathbb{H}^*$ is decreasing. Hence, $\mathbb{H}^*$ also prefers $t = 1$ when $\mathbb{H}' - \mathbb{L}' + 1$ firms unravel. So, $\mathcal{U}'$ is not an equilibrium.

For the rest of the proof following lemma will be useful.

Lemma 10. *Holding the market constant, as more firms worse than f contract in $t = 1$, the expected payoff from $t = 2$ match for firm f is decreasing. I.e.,*

$$E\pi_f(\mu_S^{\mathcal{F},\mathcal{W}}|G, \eta) < E\pi_f(\mu_S^{\mathcal{F},\mathcal{W}}|G, \eta + 1)$$

For $\mathbf{\Delta}\mathbb{L} < \mathbf{\Delta}\mathbb{H}$. Since $\mathbb{L}^* > \mathbb{L}'$ and with Lemma 10,

$$u_{\mathbb{L}'-1} < EU(t=2|\mathcal{U}') < u_{\mathbb{L}'} \leq u_{\mathbb{L}^*+1} < EU(t=2|\mathcal{U}^*) < u_{\mathbb{L}^*}$$

So

$$\begin{aligned} EU(t=2|\mathcal{U}') &< EU(t=2|\mathcal{U}^*) \\ \frac{\sum_{f=1}^{\mathbb{L}'-1} u_f + \sum_{\mathbb{H}'+1}^{F} u_f}{W - \mathbb{H}' + \mathbb{L}' - 1} &< \frac{\sum_{f=1}^{\mathbb{L}^*-1} u_f + \sum_{\mathbb{H}^*+1}^{F} u_f}{W - \mathbb{H}^* + \mathbb{L}^* - 1} \end{aligned}$$

which cannot be true, given $(\sum_{f=\mathbb{H}'}^{\mathbb{H}^*-1} u_f)/\Delta\mathbb{H} > (\sum_{\mathbb{L}'+1}^{\mathbb{L}^*} u_f)/\Delta\mathbb{L}$ and $\Delta\mathbb{H} > \Delta\mathbb{L}$.

Hence, a contradiction. This completes the proof of part (3) of Lemma 5. □

The last property of multiple equilibria allows to draw conclusions about how increasing similarity of preferences drives the changes in equilibrium outcomes.

2.3. Comparative statics on ϱ

This subsection investigates how equilibrium unravelling sets in a market $(F, W, \mathbf{u}, \mathbf{v}, \varrho, \mathcal{M}_S)$ change with the similarity of preferences, ϱ, when other market parameters are held constant. It is shown that, in general, equilibrium unravelling – as measured by the size of $\mathcal{U}^*$ – weakly increases with the similarity of preferences.

In any market with independent preferences all equilibria result in no unravelling. By Lemma 3, there always exists an equilibrium without unravelling for G_0. But there also is no other equilibrium outcome possible. Suppose, to the contrary, that there is an equilibrium with a nonempty unravelling set $\mathcal{U}^* \neq \emptyset$. Then for any firm f in $\mathcal{U}^*$ early contracting must yield a higher payoff than waiting for $t = 2$. Let η denote the number of firms better than f that unravel. Under independent preferences there is no difference for firm f in $t = 2$ if another firm contracts early with a random worker or if it picks its best worker before f in the ex-post stable matching. When η firms worse than f contract early, it has the same effect on f's payoff as if η more firms were choosing their best worker before f in the ex-post stable matching. Therefore, firm f's payoff of waiting when η worse firms unravel is the same as the payoff of firm $(f - \eta)$ in $\mathfrak{o}_S$ (without unravelling). But Lemma 3 says that even firm $(f - \eta)$ gets in $\mathfrak{o}_S$ a higher payoff than π^0. This leads to a contradiction. Thus, under independent preferences the unique equilibrium outcome is "no unravelling".

In a given market, $\mathcal{U}^* = \emptyset$ is an equilibrium unravelling set if and only if there is no profitable deviation from $\mathfrak{o}_S$ in this market. Therefore, Lemma 4 implies that as ϱ increases, equilibria with $\mathcal{U}^* = \emptyset$ exist for a smaller range of market parameters $(F, W, \mathbf{u}, \mathbf{v})$.

By the property of multiple equilibrium unravelling sets (Lemma 5(3)), every equilibrium unravelling set in a given market (if there is more than one) has a different number of firms contracting early. Thus, for any market, all equilibria can be ordered by the size of $\mathcal{U}^*$. The *maximum* equilibrium ($\mathcal{U}^{MAX}$) and the *minimum* equilibrium ($\mathcal{U}^{MIN}$) can be distinguished. The former is the class of equilibria with maximum unravelling, i.e. the largest $\mathcal{U}^*$, and the latter is the class of equilibria with minimum unravelling, i.e. the smallest $\mathcal{U}^*$. It may happen for a market that $\mathcal{U}^{MAX} \equiv \mathcal{U}^{MIN}$, that is, all equilibria in this market result in the same unravelling set. For instance, in any market with G_0, $\mathcal{U}^{MAX} \equiv \mathcal{U}^{MIN} = \emptyset$.

As similarity of preferences increases, both minimum and maximum equilibrium unravelling sets increase. Let $\mathcal{U}(\varrho)$ be an equilibrium unravelling set in a market with similarity of preferences ϱ. Then, holding other market parameters constant, $\mathcal{U}^{MIN}(\varrho) \subseteq \mathcal{U}^{MIN}(\varrho')$ and $\mathcal{U}^{MAX}(\varrho) \subseteq \mathcal{U}^{MAX}(\varrho')$ whenever $\varrho < \varrho'$.

As ϱ increases, the maximum and minimum equilibria are more likely to be distinct. The maximum equilibrium unravelling set increases from the empty set to a non-empty one for lower ϱ than the minimum equilibrium unravelling set does. As the maximum equilibrium unravelling set increases, an equilibrium with unravelling appears in the market. Moreover, when the similarity of preferences increases, the minimum equilibrium unravelling set may also increase from the empty set to a non-empty one. When this occurs, "no unravelling" is no longer an equilibrium for high ϱ's in this market. This relation between equilibrium unravelling sets in a market, and the level of preference similarity is illustrated by Figure 2.4.

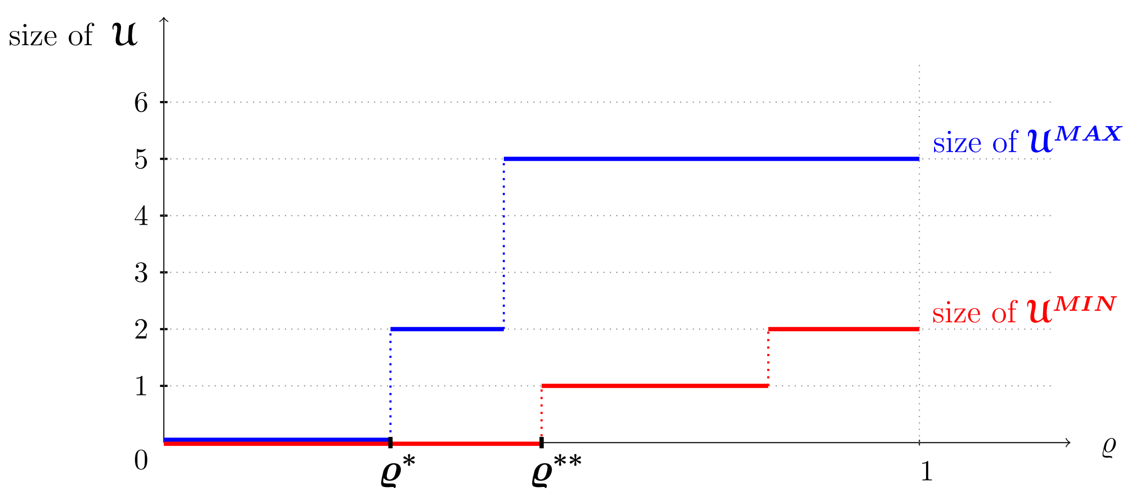


Figure 2.4. Relation between $\mathcal{U}^{MIN}$, $\mathcal{U}^{MAX}$ and ϱ in a typical market.

Proposition 1. *Under $\mathcal{M}_S$, for any market parameters $F, W, \mathbf{u}, \mathbf{v}$, there exist ϱ^* and ϱ^{**} such that $0 < \varrho^* \leq \varrho^{**} \leq 1$ and*

$$\begin{aligned} \varrho \in [0, \varrho^*] &\implies \mathcal{U}^{MAX}(\varrho) = \emptyset \\ \varrho \in (\varrho^*, \varrho^{**}] &\implies \mathcal{U}^{MIN}(\varrho) = \emptyset \quad \& \quad \mathcal{U}^{MAX}(\varrho) \neq \emptyset \\ \varrho \in (\varrho^{**}, 1] &\implies \mathcal{U}^{MIN}(\varrho) \neq \emptyset \end{aligned}$$

Proof.

First, following lemma needs to be established.

Lemma 11. *In any market with G_0, $(F, W, \mathbf{u}, \mathbf{v}, G_0)$, the only equilibrium outcome is $\mathcal{U}^* = \emptyset$.*

Proof.

Assume, to the contrary, that there is an equilibrium with $\mathcal{U}^* \neq \emptyset$ under G_0. Then for any $f \in \mathcal{U}^*$, it must be that $E\pi_f(\mathfrak{o}_S|G_0, \eta) < E\pi(t = 1)$. But $E\pi_f(\mathfrak{o}_S|G_0, \eta) \equiv E\pi_{f-\eta}(\mathfrak{o}_S|G_0)$ (for $f - \eta \geq 1$, which is always satisfied). By Theorem 3, $\forall\, i \geq 1 \quad E\pi_i(\mathfrak{o}_S|G_0) > E\pi(t = 1)$. So, it must also be true for $i \equiv f - \eta$. Therefore, contradiction. □

The rest of the proof of Proposition 1 follows from the fact that $\mathcal{U}^{MIN} \subseteq \mathcal{U}^{MAX}$ and from the monotonicity of $\mathcal{O}(\varrho, \mathbf{v})$ in ϱ. □

For any market parameters, there are thresholds ϱ^* and ϱ^{**} such that for preference similarity lower than ϱ^* all equilibrium outcomes involve no unravelling; for preference similarity between ϱ^* and ϱ^{**} there are equilibrium outcomes with unravelling and without unravelling; and for similarity higher than ϱ^{**} all equilibrium outcomes involve nonempty unravelling set. In extreme cases, the thresholds may be equal to 1. When $\varrho^{**} = 1$, then the interval $(\varrho^{**}, 1]$ is empty, and for any similarity of preferences there exists an equilibrium without unravelling. Similarly, when $\varrho^* = 1$, than in all equilibria for any preference similarity there is no unravelling. Moreover, ϱ^* must be strictly greater than 0. That means that for any market parameters, if the preference similarity is sufficiently low, all equilibria preclude unravelling.

2.4. Discussion

This chapter has shown that similarity of preferences is an important factor driving unravelling. In markets where firms' preferences are very similar, there is a strong competition for the same workers once the information arrives. Worse firms see their only chance to be matched with their most preferred workers in contracting early, with less information. One of the more surprising results of the analysis presented above is that stable matching imposed in the second period (after information about preferences is revealed) may not stop unravelling if preferences are similar enough. Thus, the stability hypothesis that much of the previous research concentrates on does not hold in general.

Although this chapter contributes to identifying causes of unravelling, it does not offer a way to stop it. This important issue is analyzed in the next chapter.

In some markets that struggled with unravelling, firms called upon an institution to solve the problem.[13] It is often recommended that the institutions should apply the ex-post stable mechanism once the preferences are known. Roth (1991) and Kagel and Roth (2000), for example, argue that the ex-post stable mechanism prevents unravelling. It has, indeed, proven successful in stoping unravelling in many markets (e.g. markets for medical resident interns in the US and in the UK). However, in some markets the ex-post stable mechanism has failed to stop unravelling. For instance, in the Canadian lawyer market and in the gastroenterology market in the US, agents contracted as early as a year before information was available and the mechanism was applied. In the latter case, the clearinghouse was subsequently abandoned in 1996, as there were too few participants waiting for its operation.

The present model provides an explanation for why the ex-post stable matching mechanism failed to stop unravelling in some markets. In some markets with high similarity of preferences all equilibria under the ex-post stable mechanism involve early contracting. In those markets only an ex-post *unstable* matching mechanism can prevent unravelling. The next chapter focuses on finding mechanisms that improve situation on markets where unravelling is possible.

[13] E.g., National Resident Matching Program established for the US medical residents market, Judicial conferences for federal court clerkship market, or Articling Student Matching Program for entry-level lawyer positions in Canada. See Roth and Xing (1994) for an extensive list.

CHAPTER 3

Mechanism Design

This chapter turns to the problem of mechanism design in markets where unravelling is possible. Despite extensive research on unravelling in the economics literature, the welfare consequences of this phenomenon are not fully understood. In particular, the question of how one might design an optimal matching mechanism for markets where unravelling may occur remains largely unexplored. Policy recommendations in the present economics literature are overshadowed by the stability hypothesis, which implies that whenever there is a problem with unravelling, the market should introduce the ex-post stable clearinghouse. However, the previous chapter establishes that the ex-post stable mechanism in the second period does not always solve the problem of early contracting. This finding is supported by the experience of markets for new gastroenterologists and for new lawyers in Canada, where stable clearinghouses did not stop unravelling.

In the literature one may find accounts of markets participants who intuitively feel that unravelling is a socially harmful situation. Kozinski (1991, p.1710)[1] describes the federal court clerks market:

> To be sure, we would all prefer to know *precisely* how a particular student will do during the full six semesters he spends in law school. If a decision could magically be delayed until after graduation, we would have all of an applicant's grades as well as the potential input of a large battery

[1]Quoted after Roth and Xing (1994), p.1002

> of law school professors. Also, we could be better informed about the student's performance in various extracurricular activities. Did she do an excellent job as a law review editor? Did she publish and, if so, what does the product look like? Did he compete in moot court and, if so, how high did he place?... All of these would be mighty helpful hints when picking clerks.

Doig and Munday (1969)[2] have similar remarks about UK medical markets:

> The filling of preregistration house-officer posts by private arrangements between consultants and students is considered unsatisfactory by most students and by an increasing number of consultants. This practice allows consultants to offer appointments to promising students early in their clinical years and often causes subsequent regret on the part of those who accept when they are later offered more desirable posts; [p.1250]

However, it is not certain that in the markets that unravel such situation may be improved. This chapter shows a precise way in which unravelling is socially harmful. It shows that unravelling is Pareto-inefficient, for a broad class of mechanisms with a property that one would expect from a mechanism in a real market.

This property requires the mechanism to match firms and workers based on preference rankings, not based on workers' identities. A mechanism that satisfies this property is called an *anonymous* mechanism. For example, the ex-post stable mechanism is anonymous, as are all mechanisms ever to be used in real markets.

[2]Quoted after Roth (1991), p.416

It is shown that when an anonymous mechanism induces early contracting, then the outcome may be Pareto-improved by employing some other mechanism that would not unravel. Thus, any anonymous Pareto-optimal mechanism must preclude unravelling.

Moreover, the ex-post stable matching mechanism is Pareto-optimal if and only if it does not induce unravelling. This chapter also demonstrates that in every market there always exists a mechanism that produces a Pareto-optimal outcome. Thus, in the markets where ex-post stable clearinghouse unravels, there exists an ex-post *unstable* mechanism that stops unravelling and improves the welfare of the market participants.

3.1. Notions of Pareto-Optimality

An outcome is a function from the profile of rankings to randomization over matchings between all firms and all workers. Recall from Chapter 2 that $\boldsymbol{\mu}\big(\{1,\ldots,F\},\{1,\ldots,W\}\big)$ is the set of all possible matchings between all firms and all workers. Then an outcome $\mathfrak{o}$ is

$$\mathfrak{o} : \mathbf{R} \mapsto Rand\Big(\boldsymbol{\mu}\big(\{1,\ldots,F\},\{1,\ldots,W\}\big)\Big)$$

The previous chapter considered a special case of an outcome function – the ex-post stable outcome, $\mathfrak{o}_S$. This is the outcome when all agents participate in the ex-post stable mechanism, $\mathcal{M}_S$.[3] This chapter examines also other outcomes and mechanisms.

Firm f's payoff from an outcome depends on the realized rankings, $\mathbf{R}$, and is denoted by $\pi_f(\mathfrak{o}|\mathbf{R})$. The ex-ante expected payoff of an outcome is the expectation over all possible ranking realizations. The payoff and the expected payoff depend also on market

[3]When unravelling occurs under the ex-post stable mechanism, the resulting outcome is other than $\mathfrak{o}_S$.

characteristics, especially on the probability distribution of ranking profiles.[4] Let $E\pi_f(\mathfrak{o})$ be firm f's expected payoff from outcome $\mathfrak{o}$, then

$$E\pi_f(\mathfrak{o}) = \sum_{\mathbf{R} \in \mathfrak{R}^F} \pi_f(\mathfrak{o}|\mathbf{R}) \cdot Prob(\mathbf{R}|\varrho)$$

Similarly, let $U_w(\mathfrak{o}|\mathbf{R})$ be worker w's utility from outcome $\mathfrak{o}$, given the realized rankings, and $EU_w(\mathfrak{o})$ – worker w's expected utility of outcome $\mathfrak{o}$, then

$$EU_w(\mathfrak{o}) = \sum_{\mathbf{R} \in \mathfrak{R}^F} U_w(\mathfrak{o}|\mathbf{R}) \cdot Prob(\mathbf{R}|\varrho)$$

An outcome $\mathfrak{o}'$ *strictly Pareto-dominates* outcome $\mathfrak{o}''$ when

$$\Big(\forall\, f \quad E\pi_f(\mathfrak{o}') \geq E\pi_f(\mathfrak{o}'') \qquad \textit{and} \qquad \forall\, w \quad EU_w(\mathfrak{o}') \geq EU_w(\mathfrak{o}'')\Big)$$

and

$$\Big(\exists\, f \quad E\pi_f(\mathfrak{o}') > E\pi_f(\mathfrak{o}'') \qquad \textit{or} \qquad \exists\, w \quad EU_w(\mathfrak{o}') > EU_w(\mathfrak{o}'')\Big)$$

A matching outcome $\mathfrak{o}$ is *Pareto-optimal* in a given market when there does not exist an outcome in that market that strictly Pareto-dominates $\mathfrak{o}$.

The goal of the social planner is to obtain the best outcome in the Pareto sense. He designs a mechanism, which is in operation in the second period. The mechanism is known already in period 1, when firms and workers make their decisions about early contracting. Given mechanism $\mathcal{M}$ in period 2, all firms need to decide whether to contract early, and

[4] The probability distribution of ranking profiles depends on ϱ. Other market characteristics also play a role in calculating the payoffs. However, since this chapter always considers outcomes and mechanisms for a given market (no comparative statistics), there is no need to use notation where payoffs and utilities are conditional on market characteristics.

firms that participate in $\mathcal{M}$ need to choose whether to report their true preferences. By the revelation principle, it may be assumed that $\mathcal{M}$ is incentive compatible and that preferences are reported truthfully. For an incentive compatible mechanism in $t = 2$, an equilibrium under $\mathcal{M}$ is described by the first-period strategies of agents. Recall that $\boldsymbol{\sigma}$ denotes a vector of strategies in $t = 1$ for all agents. A mechanism may possibly implement many equilibria. For example, it was demonstrated that the game with the ex-post stable mechanism usually has multiple equilibria. Let $\boldsymbol{\Sigma}^{\mathcal{M}}$ be the set of all possible equilibria under mechanism $\mathcal{M}$. A pair $(\mathcal{M}, \boldsymbol{\sigma})$, where $\boldsymbol{\sigma} \in \boldsymbol{\Sigma}^{\mathcal{M}}$, is called a *mechanism-equilibrium* pair. A mechanism-equilibrium pair $(\mathcal{M}, \boldsymbol{\sigma})$ determines a unique outcome $\mathfrak{o}_{(\mathcal{M},\boldsymbol{\sigma})}$.

A mechanism-equilibrium pair $(\mathcal{M}, \boldsymbol{\sigma})$ is *unconstrained Pareto-optimal* when it produces a Pareto-optimal outcome; that is, when $\mathfrak{o}_{(\mathcal{M},\boldsymbol{\sigma})}$ is Pareto-optimal. However, a social planner is constrained to inducing outcomes by a mechanism. A mechanism-equilibrium pair $(\mathcal{M}, \boldsymbol{\sigma})$ is *constrained Pareto-optimal* when there is no other mechanism-equilibrium pair $(\mathcal{M}', \boldsymbol{\sigma}')$ such that its outcome $\mathfrak{o}_{(\mathcal{M}',\boldsymbol{\sigma}')}$ strictly Pareto-dominates $\mathfrak{o}_{(\mathcal{M},\boldsymbol{\sigma})}$. Clearly, any unconstrained Pareto-optimal pair $(\mathcal{M}, \boldsymbol{\sigma})$ is also Pareto-optimal in the constrained sense.

Define a mechanism to be *anonymous* if it assigns workers to firms based only on firms' rankings, but not on workers' identities. For example, the ex-post stable mechanism, $\mathcal{M}_S$, is anonymous. Define a vector of strategies, $\boldsymbol{\sigma}$, to be anonymous if any firm that contracts with a worker in period 1, selects a worker at random, paying no attention to his identity. A mechanism-equilibrium pair $(\mathcal{M}, \boldsymbol{\sigma})$ is anonymous when $\mathcal{M}$ and $\boldsymbol{\sigma}$ are anonymous.

It seems reasonable to consider anonymous mechanism-equilibrium pairs. A priori all workers are the same, except for their identities. But in $t = 2$ they differ in their positions in firms' rankings. It seems that all realistic mechanisms would match firms and workers based on the rankings, not identities. Moreover, if a firm matches with a worker in $t = 1$, the early contracting reflects the firm's expectations about the worker's place in the firm's ranking. But that does not depend on worker's identity, as expectations are the same about all the workers. In this context, the identity does not matter either for early contracting in $t = 1$, or for a mechanism in $t = 2$.

Notice that ex-ante expected utility of every worker is the same under an anonymous mechanism-equilibrium pair is the same. This utility depends on the set of firms that get matched under the mechanism-equilibrium pair. Any worker has equal probability to be matched to any of those firms, or to remain unmatched.

It is said that a mechanism-equilibrium pair $(\mathcal{M}, \boldsymbol{\sigma})$ *exhibits unravelling* when there is a positive probability that an early offer is both made and accepted under the vector of strategies $\boldsymbol{\sigma}$.

3.2. Pareto-Optimality and Unravelling

The following proposition presents the main result of this chapter. It shows that if an anonymous mechanism-equilibrium pair $(\mathcal{M}, \boldsymbol{\sigma})$ exhibits unravelling, then there is another anonymous mechanism-equilibrium pair that produces a better outcome, in the Pareto sense. That is, when an anonymous $(\mathcal{M}, \boldsymbol{\sigma})$ exhibits unravelling, it cannot be constrained Pareto-optimal.

Proposition 2. *For any anonymous mechanism-equilibrium pair $(\mathcal{M}, \boldsymbol{\sigma})$ which exhibits unravelling, there exists an anonymous a pair $(\mathcal{M}', \boldsymbol{\sigma}')$ such that it does not exhibit unravelling and outcome $\mathsf{o}_{(\mathcal{M}', \boldsymbol{\sigma}')}$ strictly Pareto-dominates outcome $\mathsf{o}_{(\mathcal{M}, \boldsymbol{\sigma})}$.*

Proof.

Consider an anonymous $(\mathcal{M}, \boldsymbol{\sigma})$ such that $\mathcal{M}$ produces in equilibrium $\boldsymbol{\sigma}$ a non-empty unravelling set $\mathcal{U}^{\mathcal{M}} \neq \emptyset$. Now, consider the following mechanism $\mathcal{M}'$:

(1) To all firms in $\mathcal{U}^{\mathcal{M}}$, $\mathcal{M}'$ tentatively assigns a random worker from the set of all workers. This mimics the unravelling outcome for those firms. Notice that with probability $\frac{1}{W}$ a firm is assigned to its worst worker.

(2) All other firms are matched according to $\mathcal{M}$. These firms get the same expected payoff as under $(\mathcal{M}, \boldsymbol{\sigma})$. For these firms it is the final match.

(3) (the "worst workers correction") For all firms in $\mathcal{U}^{\mathcal{M}}$ that got matched to their worst workers, $\mathcal{M}'$ substitutes these workers with workers still remaining in the pool. That is feasible, because after all firms are matched there is at least one worker still in the pool. For firm f tentatively matched with its worker r_1^f, any of the remaining workers is better than the tentative match. This way all firms tentatively matched with their worst workers can improve their payoff. When there are no more firms in $\mathcal{U}^{\mathcal{M}}$ that are matched to their worst worker the algorithm stops and the matching is finalized.

Notice that $\mathcal{M}'$ is an incentive compatible mechanism. To see that, recall that $\mathcal{M}$ is incentive compatible. For all the firms that did participate in the original $\mathcal{M}$, the incentives to truthfully report their preferences did not change. Firms in $\mathcal{U}^{\mathcal{M}}$ that participate in $\mathcal{M}'$ cannot gain by misreporting their preferences, and they can even loose if they misreport their lowest ranked worker.

Notice also that $\mathcal{M}'$ is anonymous, since $\mathcal{M}$ is anonymous.

There is an equilibrium without unravelling under $\mathcal{M}'$. This is because all firms in $\mathcal{U}^{\mathcal{M}}$ prefer to wait for $\mathcal{M}'$ than to unravel given that other firms wait for $t = 2$. Since firms outside $\mathcal{U}^{\mathcal{M}}$ did not unravel when some firms were contracting early – either because they preferred not to, or because they would not be accepted in $t = 1$ – they do not unravel when all other firms wait for $t = 2$ under $\mathcal{M}'$. For firms that preferred not to unravel under $(\mathcal{M}, \boldsymbol{\sigma})$ the value of waiting does not change under $\mathcal{M}'$. For the workers, the value of waiting for the mechanism is strictly higher when no firm unravels (since $\mathcal{M}'$ is anonymous, additional firms participating in the mechanism yield higher expected utility to every worker). So, the firms that were not accepted under $(\mathcal{M}, \boldsymbol{\sigma})$ are not accepted under $\mathcal{M}'$ either. Therefore, no unravelling occurs. Denote the equilibrium without unravelling by $\boldsymbol{\sigma}'$.

Notice that since $(\mathcal{M}, \boldsymbol{\sigma})$ is anonymous, $(\mathcal{M}', \boldsymbol{\sigma}')$ is anonymous as well. And since every firm that is matched to a worker under $(\mathcal{M}, \boldsymbol{\sigma})$ is also matched under $(\mathcal{M}', \boldsymbol{\sigma}')$, the expected payoff of every worker does not change. Every firm in $\mathcal{U}^{\mathcal{M}}$ has a strictly higher expected payoff in $\mathfrak{o}_{(\mathcal{M}', \boldsymbol{\sigma}')}$ then in $\mathfrak{o}_{(\mathcal{M}, \boldsymbol{\sigma})}$. All the other firms have exactly the same expected payoff in both outcomes. Therefore, $\mathfrak{o}_{(\mathcal{M}', \boldsymbol{\sigma}')}$ Pareto-dominates $\mathfrak{o}_{(\mathcal{M}, \boldsymbol{\sigma})}$. □

3.3. The Role of Anonymity

To see what role the assumption of anonymity plays in Proposition 2, consider following example.

Example 6. *Consider a non-anonymous mechanism-equilibrium pair* $(\mathcal{M}, \boldsymbol{\sigma})$ *where* $\mathcal{M}$ *assigns firm 1 to worker 1, firm 2 to worker 2, etc., and in* $\boldsymbol{\sigma}$ *firm* F *contracts early with worker* F *and all other firms wait for* $\mathcal{M}$*. Under the outcome produced by this* $(\mathcal{M}, \boldsymbol{\sigma})$*, worker with index number* i *ex-ante expects utility*

$$EU_i(\mathfrak{o}_{(\mathcal{M},\boldsymbol{\sigma})}) = \begin{cases} u_i & \textit{for} \quad i = 1, \ldots, F \\ 0 & \textit{for} \quad i = F+1, \ldots, W \end{cases}$$

Each firm's ex-ante expected payoff is

$$E\pi_f(\mathfrak{o}_{(\mathcal{M},\boldsymbol{\sigma})}) = \frac{1}{W}\sum_{k=1}^{W} v_k \qquad \forall\, f$$

Notice that, despite unravelling in equilibrium, this outcome cannot be Pareto-improved. Especially, it cannot be improved by the equilibrium without unravelling. ■

Anonymity is an important assumption for the result in Proposition t2. However, it seems reasonable to consider anonymous mechanisms as the most plausible for implementation in real markets.

3.4. Other Results on Pareto-Optimality

Proposition 2 establishes that no unravelling is a necessary condition for constrained Pareto-optimality of an anonymous $(\mathcal{M}, \boldsymbol{\sigma})$. In particular, when the ex-post stable mechanism – which is anonymous – unravels, it cannot be constrained Pareto-optimal. Moreover, for the ex-post stable mechanism, any $(\mathcal{M}_S, \boldsymbol{\sigma})$ that does not exhibit unravelling is unconstrained Pareto-optimal. It has been already established in the literature that the ex-post stable outcome is always Pareto-optimal.[5] When the ex-post stable mechanism does not unravel, it produces the ex-post stable outcome and, thus, it is unconstrained (and constrained) Pareto-optimal.

Corollary 1. *A mechanism-equilibrium pair with the ex-post stable mechanism,* $(\mathcal{M}_S, \boldsymbol{\sigma})$ *is constrained and unconstrained Pareto-optimal if and only if there is no unravelling in* $\boldsymbol{\sigma}$*.*

For markets where $\mathcal{M}_S$ does not unravel, the market achieves an unconstrained Pareto-optimal outcome. However, there are markets where $\mathcal{M}_S$ always unravels. Proposition 2 implies that in those markets there must exist an ex-post unstable mechanism that does not unravel and Pareto-improves the unravelling outcome.

The following Proposition 3 is even stronger. It guarantees that in any market there exists an unconstrained Pareto-optimal mechanism-equilibrium pair, i.e. one that produces a Pareto-optimal outcome.

Proposition 3. *For any market, there exists a mechanism* $\mathcal{M}$ *and an equilibrium* $\boldsymbol{\sigma} \in \boldsymbol{\Sigma}^{\mathcal{M}}$ *such that* $(\mathcal{M}, \boldsymbol{\sigma})$ *is unconstrained Pareto-optimal.*

Proof.

Consider a mechanism $\mathcal{M}$ that first assigns all participating firms a number from between 1 and F at random. Then, the mechanism works in the same way as the ex-post stable mechanism but the order in which firms get their workers is based on the randomly assigned numbers, not on their position in the market.

[5] E.g., see Roth and Sotomayor (1991).

Notice that this mechanism is anonymous. It is also incentive compatible, as the ex-post stable mechanism is. Moreover, there exists an equilibrium without unravelling. If all agents participate in the mechanism than all firms have higher expected payoff from the mechanism than from unravelling. Thus, no firm wants to unravel when no other firm unravels. Denote the no-unravel equilibrium by $\boldsymbol{\sigma}$.

Now, notice that $(\mathcal{M}, \boldsymbol{\sigma})$ produces an Pareto-optimal outcome. The sum of workers' expected utilities and the sum of firms' expected payoffs are the same under $(\mathcal{M}, \boldsymbol{\sigma})$ as under $\mathfrak{o}_S$. Since $\mathfrak{o}_S$ is Pareto-optimal, so $\mathfrak{o}_{(\mathcal{M},\boldsymbol{\sigma})}$ must be: In both outcomes it is not possible to increase expected payoff of one agent without decreasing it for some other agent on the same side of the market. □

However, the Pareto-optimal outcome as constructed in the proof of the proposition above, does not Pareto-improve on $(\mathcal{M}_S, \boldsymbol{\sigma})$ which exhibits unravelling.

Nonetheless, using a similar constructing approach it may be shown that whenever the ex-post stable mechanism unravels, it can be modified in such a way that it Pareto-improves on the unravelling $\mathcal{M}_S$, and it achieves a Pareto-optimal outcome.

Observation 1. *If $(\mathcal{M}_S, \boldsymbol{\sigma})$ exhibits unravelling, there exists an unconstrained Pareto-optimal $(\mathcal{M}', \boldsymbol{\sigma}')$ that strictly Pareto-improves on $(\mathcal{M}_S, \boldsymbol{\sigma})$.*

This result is shown by construction.

When $(\mathcal{M}_S, \boldsymbol{\sigma})$ unravels, it produces a non-empty unravelling set, $\mathcal{U} \neq \emptyset$. Now, consider the following mechanism, denoted $\mathcal{M}^{\mathcal{A}}$:

(1) All firms $f \in \mathcal{U}$ draw a random number out of $\{1, \dots, W\}$.[6]

(2) All other firms $f \in \{1, \dots, F\} \smallsetminus \mathcal{U}$ in order from the highest ranked to the lowest ranked get the highest number that is still available.

(3) Firms' get matched with their best available worker in the order of their numbers – starting from the one with the highest number. That is, the firm with the highest number is treated as firm F in the ex-post stable matching, the firm with the second highest number is treated as firm $(F-1)$ in the ex-post stable matching, and so on, until the firm with the lowest number, which is treated as firm 1 in the ex-post stable matching.

This mechanism is incentive compatible, as the ex-post stable mechanism is. Moreover, there exists an equilibrium without unravelling under $\mathcal{M}^{\mathcal{A}}$. Denote this equilibrium by $\boldsymbol{\sigma}^{\mathcal{A}}$. Let $\mathfrak{o}^{\mathcal{A}}$ be the outcome of $(\mathcal{M}^{\mathcal{A}}, \boldsymbol{\sigma}^{\mathcal{A}})$. Outcome $\mathfrak{o}^{\mathcal{A}}$ is Pareto-optimal. To see that, notice that the sum of expected payoffs to firms and the sum of expected utilities of workers are the same under $\mathfrak{o}^{\mathcal{A}}$ as under $\mathfrak{o}_S$. Since $\mathfrak{o}_S$ is Pareto-optimal, than $\mathfrak{o}^{\mathcal{A}}$ is as well: In both outcomes it is not possible to increase expected payoff of one agent without decreasing it for some other agent.

Moreover, outcome $\mathfrak{o}^{\mathcal{A}}$ Pareto-improves on the outcome $\mathfrak{o}_{(\mathcal{M}_S, \boldsymbol{\sigma})}$. Since $(\mathcal{M}^{\mathcal{A}}, \boldsymbol{\sigma}^{\mathcal{A}})$ is anonymous, expected utility of workers is the same under both outcomes. Firms in $\mathcal{U}$ clearly have higher expected utility under $\mathfrak{o}^{\mathcal{A}}$. Other firms' expected payoff does not change. This is because the numbers that $\mathcal{U}$-firms draw in step (1) are mathematically equivalent – from the payoff point of view – to the workers gone from the market in $t = 1$ when they early match with $\mathcal{U}$-firms.

[6]Notice that the number is drawn from the index numbers of workers, not only firms (there is more workers).

3.5. Conclusions

This study investigates the causes and welfare consequences of unravelling in two-sided matching markets. It considers a two period model where firms learn about their preferences over workers at the beginning of the second period. It is assumed that firms and workers have the ability to make and accept offers in the first period if they wish to, and that a clearinghouse mechanism is used in the second period to assign the remaining firms to workers. Unravelling is said to occur when offers are both made and accepted in the first period. Notice that firms that choose to contract early do so with no information about which workers are more preferred than others.

Chapter 2 explores the issue of unravelling given that the ex-post stable mechanism operates in the second period. The ex-post stable matching is the clearinghouse mechanism that most of the existing literature focuses on. Chapter 2 shows that unravelling becomes more likely as firms' preferences over workers grow more similar. This is because when preferences of firms are very similar, worse firms can be matched with their most preferred worker only if they contract with them early. Despite insufficient information in the first period, it may be worth for such firms to bear the risk and contract early. Firms most likely to unravel are those "in the middle" – bad enough to prefer the uncertainty of early contracting, and good enough to be accepted.

Chapter 3 investigates the impact of different mechanisms on the equilibrium outcome and its welfare consequences. The goal of this mechanism design analysis is to characterize Pareto-optimal mechanisms given that firms and workers can choose to contract in the first period if they wish to. The main result says that a necessary condition for an anonymous mechanism to be Pareto-optimal is that it does not induce unravelling. Any anonymous mechanism that induces unravelling is Pareto-inefficient. In particular, the ex-post stable matching mechanism is Pareto-optimal if and only if it does not unravel.

Another result of Chapter 3 demonstrates that in every market there exists a mechanism that produces a Pareto-optimal outcome. In markets where the ex-post stable clearinghouse unravels, it is an ex-post unstable mechanism that achieves Pareto-optimality.

These findings are particularly noteworthy given the importance that the literature assigns to stability. In some circumstances, an ex-post unstable mechanism that precludes unravelling is actually preferable from the policy standpoint.

References

[1] C. Avery, C. Jolls, R.A. Posner and A.E. Roth (2001), “The Market for Federal Judicial Law Clerks,” *University of Chicago Law Review*, Vol. 68, No. 3, 793-902.

[2] M. Balinsky and T. Sonmez (1999), “A Tale of Two Mechanisms: Student Placement,” *Journal of Economic Theory*, Vol. 84, 73-94.

[3] G.S. Becker (1973), “A Theory of Marriage: Part I,” *The Journal of Political Economy*, Vol. 81, No. 4, 813-846.

[4] T.C. Bergstom and M. Bagnoli (1993), “Courtship as a Waiting Game,” *The Journal of Political Economy*, Vol. 101, No. 1, 185-202.

[5] J. Bulow and J. Levin (2003), “Matching and Price Competition,” working paper, Stanford University

[6] V.P. Crawford (1991), “Comparative Statics in Matching Markets,” *Journal of Economic Theory*, Vol. 54, 389-400.

[7] V.P. Crawford and E.M. Knoer (1981), “Job Matching with Heterogeneous Firms and Workers,” *Econometrica*, Vol. 49, No. 2, 437-450.

[8] E. Damiano and R. Lam (2005), “Stability in dynamic matching markets,” *Games and Economic Behavior*, Vol. 52, 34-53.

[9] E. Damiano, H. Li and W. Suen (2005), “Unravelling of Dynamic Sorting,” *Review of Economic Studies*, Vol. 72, 10571076.

[10] G. Demange and D. Gale (1985), “The Strategy Structure of Two-Sided Matching Markets,” *Econometrica*, Vol. 53, No. 4, 873-888.

[11] L.E. Dubins and D.A. Freedman (1981), “Machiavelli and the Gale-Shapley Algorithm,” *The American Mathematical Monthly*, Vol. 88, No. 7, 485-494.

[12] G. Frechette, A.E. Roth and M.U. Unver (2006), "Unraveling Yields Inefficient Matchings: Evidence from Post-Season College Football Bowls," working paper.

[13] D. Gale (2001), "The Two-Sided Matching Problem. Origin, Developement and Current Issues," *International Game Theory Review*, Vol. 3, Nos. 2 & 3, 237-252.

[14] D. Gale and L.S. Shapley (1962), "College Admissions and the Stability of Marriage," *The American Mathematical Monthly*, Vol. 69, No. 1, 9-15.

[15] D. Gale and M. Sotomayor (1985a), "Ms. Machiavelli and the Stable Matching Problem," *The American Mathematical Monthly*, Vol. 92, No. 4, 261-268.

[16] D. Gale and M. Sotomayor (1985b), "Some Remarks on the Stable Matching Problem," *Discrete Applied Mathematics*, Vol. 11, 223-232.

[17] D. Gusfield and R.W. Irving (1989), *The Stable Marriage Problem. Structure and Algorithms*, The MIT Press.

[18] E. Haruvy, A.E. Roth and M.U. Unver (2006), "The Dynamics of Law Clerk Matching: An Experimental and computational Investigation of Proposals for Reform of the Market," *Journal of Economic Dynamics and Control*, Vol. 30, No. 3, 457-486.

[19] J.W. Hatfield and P.R. Milgrom (2005), "Matching with Contracts," *The American Economic Review*, Vol. 95, No. 4, 913-935.

[20] J.H. Kagel and A.E. Roth (2000), "The Dynamics of Reorganization in Matching Markets: A Laboratory Experiment Motivated by a Natural Experiment," *Quaterly Journal of Economics*, Vol. 115, No. 1, 201-235.

[21] T. Kara and T. Sonmez (1996), "Nash Implementation of Matching Rules," *Journal of Economic Theory*, Vol. 68, 425-439.

[22] A.S. Kelso, Jr. and V.P. Crawford (1982), "Job Matching, Coalition Formation, and Gross Substitutes," *Econometrica*, Vol. 50, No. 6, 1483-1504.

[23] C.N. McKinney, M. Niederle and A.E. Roth (2005), "The Collapse of a Medical Labor Clearinghouse (and Why Such Failures are Rare)," *American Economic Review*, 2005, Vol. 95, No. 3, 878-89.

[24] A. Kozinsky (1991), "Confessions of a Bad Apple," *The Yale Law Journal*, Vol. 100, No. 6, 1707-1730.

[25] H.B. Leonard (1983), "Elicitation of Honest Preferences for the Assignment of Individuals to Positions," *Journal of Political Economy*, Vol. 91, No. 3, 461-479.

[26] P. Legros and A.F. Newman (2002), "Monotone Matching in Perfect and Imperfect Worlds," *Review of Economic Studies*, Vol. 69, 925-945.

[27] H. Li and S. Rosen (1998), "Unravelling in Matching Markets," *American Economic Review*, Vol. 88, No. 3, 371-387.

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

[29] S. Mongell and A.E. Roth (1986), "A Note on Job Matching with Budget Constraints," *Economics Letters*, Vol. 21, No. 2, 135-138.

[30] S. Mongell and A.E. Roth (1991), "Sorority Rush as a Two-Sided Matching Mechanism," *The American Economic Review*, Vol. 81, No. 3, 441-464.

[31] D.T. Mortensen (1982), "Property Rights and Efficiency in Matin, Racing, and Related Games," *The American Economic Review*, Vol. 72, No. 5, 968-979.

[32] D.T. Mortensen and C.A. Pissarides (1994), "Job Creation and Job Destruction in the Theory of Unemployment," *Review of Economic Studies*, Vol. 61, 397-415.

[33] M. Niederle and A.E. Roth (2003), "Unravelling Reduces Mobility in a Labor Market: Gastroenterology with and without a Centralized Match," *Journal of Political Economy*, Vol. 111, No. 6, 1342.

[34] M. Niederle and A.E. Roth (2004), "Market Culture: How Norms Governing Exploding Offers Affect Market Performance“, NBER working paper 10256, January 2004.

[35] M. Niederle and A.E. Roth (2005), "The Gastroenterology Fellowship Market: Should There Be a Match?," *AEA Papers and Proceedings*, Vol. 95, No. 2, 372-375.

[36] L.F. Oberdorfer and M.N. Levy (1992), "On Clerkship Selection: A Reply to the Bad Apple," *The Yale Law Journal*, Vol. 101, No. 5, 1097-1108.

[37] A.E. Roth (1982), "The Economics of Matching: Stability and Incentives," *Mathematics of Operations Research*, Vol. 7, No. 4, 617-628.

[38] A.E. Roth (1985a), "Common and Conflicting Interests in Two-Sided Matching Markets," *European Economic Review*, Vol. 27, 75-96.

[39] A.E. Roth (1985b), "The College Admissions Problem is not Equivalent to the Marriage Problem," *Journal of Economic Theory*, Vol. 36, No. 2, 277-288.

[40] A.E. Roth (1984), "The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory," *The Journal of Political Economy*, Vol. 92, No. 6, 991-1016.

[41] A.E. Roth (1989), "Two-Sided Matching with Incomplete Information About Others' Preferences," *Games and Economic Behavior*, Vol. 1, No. 2, 191-209.

[42] A.E. Roth (1991), "A Natural Experiment in the Organization of Entry-Level Labor markets: Regional Markets for New Physicians and Surgeons in the United Kingdom," *The American Economic Review*, Vol. 81, No. 3, 415-440.

[43] A.E. Roth and E. Peranson (1999), "The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Econonmic Design," *The American Economic Review*, Vol. 89, No. 4, 748-780.

[44] A.E. Roth and U.G. Rothblum (1999), "Truncation Strategies in Matching Markets – in Search of Advice for Participants," *Econometrica*, Vol. 67, No. 1, 21-43.

[45] A.E. Roth and M. Sotomayor (1989), "The College Admissions Problem Revisited," *Econometrica*, Vol. 57, No. 3, 559-570.

[46] A.E. Roth and M. Sotomayor (1990), *Two-sided Matching: A Study in Game-Theoretic Modeling and Analysis*, Cambridge University Press 1990.

[47] A.E. Roth and J.H. Vande Vate (1990), "Random Paths to Stability in Two-Sided Matching," *Econometrica*, Vol. 58, No. 6, 1475-1480.

[48] A.E. Roth and X. Xing (1994), "Jumping the Gun: Imperfections and Institutions Related to the Timing in Market Transactions," *The American Economic Review*, Vol. 84, No. 4, 992-1044.

[49] A.E. Roth and X. Xing (1997), "Turnaround Time and Bottlenecks in Market Clearing: Decentralized Matching in the Market for Clinical Psychologists," *Journal of Political Economy*, Vol. 105, No. 2, 284-329.

[50] L.S. Shapley and M. Shubik (1971), "The Assignment Game I: The Core," *International Journal of Game Theory*, Vol. 1, No. 1, 111-130.

[51] R. Shimer and L. Smith (2000), "Assortative Matching and Search," *Econometrica*, Vol. 68, No. 2, 343-369.

[52] T. Sonmez (1996), "Implementation in Generalized Matching Problems," *Journal of Mathematical Economics*, Vol. 26, 429-439.

[53] T. Sonmez (1999), "Can Pre-arranged Matches Be Avoided in Two-Sided Matching Markets?," *Journal of Economic Theory*, Vol. 86, 148-156.

[54] W. Suen (2000), "A Competitive Theory of Equilibrium and Disquilibrium Unravelling in Two-Sided Matching," *RAND Journal of Economics*, Vol. 31, No. 1, 101-120.

[55] M.U. Unver (2001), "Backward Unravelling Over Time: The Evolution of Strategic Behavior in the Entry Level British Medical Labor Markets," *Journal of Economic Dynamic and Control*, Vol. 25, 1039-1080.

[56] M.U. Unver (2005), "On the Survival of Some Unstable Two-Sided Matching Mechanisms," *International Journal of Game Theory*, Vol. 33, 239-254.

[57] P.M. Wald (1990), "Selecting Law Clerks," *Michigan Law Review*, Vol. 89, No. 1, 152-163.