Unravelling in Two-Sided Matching Markets and Similarity of Preferences

Hanna Halaburda

Working Paper

09-068

Unravelling in Two-Sided Matching Markets and Similarity of Preferences

Hanna Hałaburda*

October 16, 2008

Abstract

This paper investigates the causes and welfare consequences of unravelling in two-sided matching markets. It shows that similarity of preferences is an important factor driving unravelling. In particular, it shows that under the ex-post stable mechanism (the mechanism that the literature focuses on), unravelling is more likely to occur when participants have more similar preferences. It also shows that any Pareto-optimal mechanism must prevent unravelling, and that the ex-post stable mechanism is Pareto-optimal if and only if it prevents unravelling.

JEL: C72; C78; D82

Key-words: two-sided matching; unravelling; similarity of preferences

1 Introduction

Hiring policy is one of the most important determinants of a firm's success. The hiring process calls for collecting information in order to choose the best individual from among the candidates. In certain markets, however, firms hire workers long before all the pertinent information is available. For instance, in the market for hospital interns before 1945, appointments have been made even as early as two years before students' graduation and the actual start of the job (Roth, 1984, 2003). A similar situation still exists in the market for federal court clerks.[1] This phenomenon of contracting long before the job begins and before relevant information is available, is called *unravelling*. Those early matches often turn out to be inefficient when the job starts.

Unravelling has been recognized as a serious problem in numerous markets.[2] Measures designed to preclude this phenomenon (such as centralized clearinghouses and enforcement of uniform hiring

---

[1] According to Haruvy, Roth and Unver (2006), "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 entry-level law and medical markets, postseason college football bowls, and fraternity and sorority rushes. For a more extensive list, see Roth and Xing (1994).

dates) have not always been successful. Unravelling prevails in certain markets because some employers see a better chance to hire their most-preferred candidates when they contract early than when they wait. Meanwhile, other markets for entry-level professionals appear never to have experienced unravelling, including markets for new professors in finance, economics and biology. Studying what factors lead to unravelling in some markets but not in others is necessary for designing better measures to prevent unravelling in markets prone to it.

Much of the existing research focuses on stability as the key to understanding unravelling. A matching is ex-post stable if every agent prefers his match to being unmatched, and if there is no blocking pair, that is, a worker and a firm that both strictly prefer each other to their assigned partners. Roth (1991) and Kagel and Roth (2000) argue that ex-post stable matching implemented upon arrival of pertinent information should preclude early contracting under uncertainty. This argument — known as the "stability hypothesis" (Roth, 1991) — is based mainly on the observation that implementing ex-post stable matching through a clearinghouse[3] 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 U.S. gastroenterology market, whose clearinghouse was abandoned in 1996 (Niederle and Roth, 2003a), and the Canadian market for new lawyers, where a large number of firms contract with students a year before graduation despite a clearinghouse (Roth and Xing, 1994). Roth and Xing (1994) also offer theoretical examples of unravelling even when ex-post stable matching is expected upon the arrival of pertinent information. There is no consensus, however, what drives those examples — what are the reasons for the stability hypothesis to fail; that is, for a clearinghouse with ex-post stable algorithm to unravel.[4]

Despite extensive discussion in the economics literature, we have only limited understanding of why unravelling occurs in some markets but not in others. And the basic question of what are the potentials and limitations of mechanisms designed for markets where unravelling is possible remains largely unexplored. This paper investigates both issues: the causes of unravelling, and mechanism design. The existing literature on unravelling had not considered the relevance of similarity of preferences. This paper shows that the similarity of preferences is an important factor contributing to unravelling. The more similar are firms' preferences, the more unravelling will occur in the market. This provides a reason for the stability hypothesis to fail: high similarity of preferences may cause an ex-post stable matching mechanism to unravel. This paper also shows that unravelling leads to a loss in welfare, and a mechanism must preclude unravelling to be Pareto-optimal. Moreover, for any market, there exists a Pareto-optimal mechanisms, which does not unravel. In markets where ex-post stable mechanism unravels, there exist unstable mechanisms that do not unravel, Pareto-improve on the ex-post stable mechanism and achieve Pareto-optimal outcome.

This paper examines a two-sided matching market populated by firms on one side and workers on the other. The agents on each side are heterogenous and they have preferences over agents on

---

[3] In a clearinghouse, firms and workers submit their preferences, and a matching among all participants is produced by an algorithm.

[4] The stability hypothesis is not the only explanation of unravelling in the literature. Other factors contributing to unravelling are congestion (Roth and Xing, 1997), exploding offers (Niederle and Roth, 2004), shocks in supply and demand (Niederle and Roth, 2003b). 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. This paper analyzes another — previously unexplored — factor contributing to unravelling: similarity of preference.

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 distinguish incentives to unravel driven by similarity of preferences from those attributable to risk aversion.

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: all workers agree on which firm is the best firm, the next-to-best or the worst firm. Firms, however, may have different preferences over the workers. The similarity of firms' preferences over workers is a comparative statics parameter; 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. The 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. In this model, contracting during the first period, before firms have learned their preferences, constitutes unravelling. Such early contracting takes place when a firm makes an offer during the first period and the offer is accepted.[5] This happens when contracting under uncertainty yields a higher expected payoff, for both the firm and the worker, than the expected matching in the second period.

The first part of the paper investigates unravelling when the mechanism in the second period is assumed to produce the ex-post stable matching. In the environment considered here there always exists a unique ex-post stable matching. It is obtained by matching the best firm with its most-preferred worker, next-best firm with its most-preferred worker from among the remaining ones etc. The focus of this part of the paper is to analyze how the nature of equilibria changes with similarity of firms' preferences, under the ex-post stable mechanism. In particular, sequential equilibria in pure strategies are explored. It is shown that the nature of these equilibria depends crucially on the level of similarity: unravelling occurs only in markets where firms' preferences are sufficiently similar. And more firms contract early in equilibrium as preferences become more similar.

With very similar preferences, many firms are likely to prefer the same workers. Once the information about rankings arrives and the ex-post stable matching is implemented by the mechanism, the best firms are matched with workers preferred by most firms, and worse firms are very likely to be matched with workers they rank low. Even before firms know the actual rankings, they are aware that once that information is available, all firms will compete for the same workers. Amid such competition, worse firms may have a better chance to hire their top candidates if they contract before rankings are known. Contracting so early presents some risk: the firm may end up with an even worse candidate. But there is also a chance that the worker hired in the first period will turn out to be one of the firm's top candidates. If the firm waits for the second period, its most desired candidates will be most likely hired by better firms, so such risk may be worth taking.

By contrast, in any market characterized by independent preferences, unravelling does not occur. Under independent preferences firms are likely to prefer different candidates. The threat to the worse firms that they will end up with their less preferred workers if they wait is not large enough to make taking the risk of early contracting worthwhile. As similarity of preferences increases, equilibria involving unravelling are more likely to occur, and it becomes less likely that "no unravelling" is an equilibrium.

The second part of the paper studies the problem of mechanism design in markets where unravelling is possible. Before the game starts, a mechanism is chosen for the second period. The mechanism is announced at the outset of the game, so that firms and workers are aware of it during the first period. The goal is to provide a Pareto-optimal outcome from the ex-ante perspective of the beginning of period 1.

It turns out that any Pareto-optimal mechanism must preclude unravelling. Whenever a mechanism induces early contracting, it is always possible to find another mechanism that precludes

---

[5] All offers made in the first period expire by the end of that period. Such offers are sometimes referred to as "exploding offers". Niederle and Roth (2004) show that exploding offers are necessary for unravelling to occur.

unravelling and Pareto-improves the outcome. This new mechanism mimics everyone's expected payoffs, except for the firms that were unravelling under the original mechanism. These firms' payoffs are strictly improved by the new mechanism. This is possible because in unravelling the firms could match with their least preferred workers. The new mechanism can assure that this does not happen. By the assumption that there are more workers than firms, it is always possible to find an allocation at which no firm is matched with its lowest-ranked worker. Thus, precluding unravelling is a necessary condition for a mechanism to be Pareto-optimal. The first part of the paper shows that the ex-post stable mechanism may unravel. When this is the case, this mechanism cannot be Pareto-optimal. However, whenever the ex-post stable mechanism does not unravel, it is Pareto-optimal. In the special case of the ex-post stable mechanism, precluding unravelling is the necessary and sufficient condition for Pareto-optimality. Furthermore, it is shown that in every market there exists a mechanism producing a Pareto-optimal outcome. In some markets, however, all Pareto-optimal mechanisms are ex-post unstable.

The model in this paper helps to explain why unravelling happens in some markets and not others. It shows that unravelling will occur only in markets characterized by substantial similarity of preferences. Indeed, it may be argued that employers have more similar preferences in markets where unravelling has been reported than in markets that do not seem to unravel. For instance, medical and law students (i.e., markets where unravelling is most prominent) are evaluated mainly on their grades, which are interpreted similarly by all potential employers. In such a case, employers' preferences are apt to be very similar. In disciplines like finance, economics, and biology, by contrast, students are assessed on the basis of their job-market papers, which leave more room for subjective evaluation. This subjectivity may contribute to differences in potential employers' rankings of candidates.

The model has the potential to explain unravelling in other situations, like the arranged-marriage market. In the past, marriages were sometimes arranged when the parties were still children.[6] Nowadays marriage decisions are made later in life, with more information in hand. This change is undoubtedly driven by many factors, but it may be argued that more differentiated preferences over potential partners is one of them. In the past, the attractiveness of a potential spouse was primarily a matter of his of her wealth and social status, both of which were easy to observe and valued similarly by all interested parties. Over time, characteristics other than wealth have become relatively more important, and differences in the ways that people gauge attractiveness have consequently grown. It follows from the model that early marriages (unravelling) become less likely when preferences are more differenciated.

Section 2 of this paper contains an illustrative example, before the model is formally presented in Section 3. Section 4 investigates unravelling under an ex-post stable mechanism. Subsection 4.1 focuses on equilibria without unravelling, while subsection 4.2 explores the existence and characteristics of equilibria with unravelling. Section 5 analyzes the problem of mechanism design in markets where unravelling is possible. Section 6 offers some concluding observations.

2   Example

Before setting up the general model, this section presents a simple example illustrating the importance of similarity of preferences.

---

[6]Such early arranged marriages are considered in the literature to be instances of unravelling (e.g. Roth and Xing, 1994).

In a market, there are 3 firms named f_1, f_2 and f_3, and 4 workers: w_1, w_2, w_3 and w_4. All workers have the same preferences over the firms. Being hired by the most preferred firm, firm 3, yields utility of 6. Being hired by firm 2 yields 5. And the worst firm, firm 1, yields 4. Remaining unmatched is worth 0.

Firms have different preferences over the workers. For instance, firm 3 may find worker w_1 to be most desirable, followed by w_2, w_3 and w_4, when firm 1 would like to hire w_2 the most, w_4 somewhat less, w_1 even less, and w_3 the least. Firms' preferences are represented by rankings — lists ordering workers from the least-preferred (lowest-ranked) to the most-preferred (highest-ranked). The preferences described above can be represented by firm 3's ranking (w_4, w_3, w_2, w_1), and firm 1's ranking (w_3, w_1, w_4, w_2). This example will consider different cases for firms' preferences. In all cases, not hiring anyone at all is the worst outcome, yielding 0. When the firm hires its lowest-ranked worker, it receives a payoff of 1. The next worker on the list yields 2. The 3rd worker on the firm's ranking list yields 3, and the highest-ranked worker yields 4.

There are two periods. Workers' preferences are commonly known from the beginning of the first period. Firms' preferences, however, are private information that the firms themselves learn only in the beginning of the second period. From the first period perspective, every possible ranking is equally likely. To illustrate the importance of similarity of firms' preferences, this example considers two cases: in the first one, all firms' rankings are identical — one ranking for all firms is drawn from among all possible rankings. In the second case, each firm's ranking is drawn independently.

Firms and workers can contract in the first period, even though the firms' preferences are not yet known at that time. If a firm and a worker contract in the first period, they leave the market. In the second period the ex-post stable mechanism operates over all firms and workers that are still in the market. This mechanism produces the ex-post stable matching over the remaining agents. With workers' preferences being identical, there is always a unique ex-post stable matching obtained in a following way: the best firm in the market is matched with its highest-ranked worker from among the available ones, then the next-best firm is matched with its highest-ranked worker from the remaining ones and so on.

Firms' preferences are identical. Consider what happens when all firms wait for the ex-post stable matching. When firms' preferences are identical, the best firm, firm 3, will get its (and every other firm's) most-preferred worker, and receive payoff of 4. Firm 2 will be matched with the 3rd worker in the ranking (i.e., almost the most-preferred worker) and receive payoff of 3. By then the 4th and the 3rd workers in firm 1's ranking (the highest-ranked workers) will not be available, so it will be matched with the 2nd worker in its ranking. Firm 1 receives payoff of 2. These payoffs can be established already in the first period, before the common ranking is realized and it is known which worker is the most-preferred one etc.

From the first period perspective, every worker is equally likely to be in any place in a firm's ranking. Thus, a firm's expected payoff from contracting in the first period is the average value of all the workers, i.e. 2.5. For firm 1 this payoff is greater than what the firm will get by waiting for the ex-post stable mechanism ($2.5 > 2$). Thus, firm 1 has incentive to contract early (but notice that firm 2 does not). Firm 1 would also be accepted in the first period. Because every ranking is equally likely, all workers have the same chance to be matched to any firm in the second period (or remain unmatched): any worker will be matched with firm 3, firm 2, firm 1 or remain unmatched, each with probability $\frac{1}{4}$. Thus a worker's expected value of waiting until the second period is $\frac{15}{4}$. Accepting firm 1 in the first period yields 4, which is better. Thus, there will be unravelling, as firm 1 wants to contract in the first period and is accepted.

Firms' preferences are independent. Firms' preferences are independent when each ranking is drawn from the uniform distribution independently of the others. Now the probability that firm 3's

most-preferred worker is the same as firm 2's most-preferred worker is $\frac{1}{4}$. If all firms wait for the mechanism in the second period, the ex-post stable matching always assigns firm 3 to its most-preferred worker, yielding payoff of 4. Firm 2 is matched with its highest-ranked worker available after firm 3 is matched. With probability $\frac{1}{4}$ firm 2's most-preferred worker was assigned to firm 3, and firm 2 is matched with the 3rd worker in its ranking, receiving payoff of 3. Otherwise, it is matched with its most-preferred worker and receives 4. Thus, firm 2's expected payoff is $\frac{15}{4}$. The worst firm, firm 1, is then matched with its highest-ranked worker among the remaining ones. With probability $\frac{1}{4} \cdot \frac{1}{3} \cdot 2 = \frac{1}{6}$ two highest-ranked workers of firm 1 (workers 4th and 3rd in the ranking) have been assigned to better firms, and firm 1 is matched with the 2nd worker in its ranking, receiving payoff of 2. With probability $\frac{1}{4} \cdot (1 - \frac{1}{3}) \cdot 2 = \frac{1}{3}$ the most-preferred worker of firm 1 was assigned to either of the better firms, but not the 3rd worker. So firm 1 is matched with the 3rd worker, and receives 3. With the remaining probability of $\frac{1}{2}$ neither of the better firms was assigned to firm 1's most-preferred worker, and firm 1 receives payoff of 4. Altogether, waiting for the ex-post stable mechanism yields firm 1 expected payoff of $\frac{1}{6} \cdot 2 + \frac{1}{3} \cdot 3 + \frac{1}{2} \cdot 4 = 3.333$. Under independent preferences, firm 1 does not have incentive to contract early, as $3.333 > 2.5$. Thus, unravelling will not occur when firms' preferences are independent.

3   The Model

To investigate unravelling, I construct a two-stage game between two types of agents: firms and workers. Firms and workers can contract during the first stage. If they do, they leave the market. In the second stage, the remaining agents are matched by a mechanism. The game — described in this section — is represented in Figure 1.

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, $W > F$.[7] Each firm has exactly one position to fill, and each worker can take at most one job.

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

Workers have identical preferences over firms: all workers consider firm F the most desirable, 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. Workers prefer being hired by the worst firm to not being 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. Firm f's preferences are described by its ranking, denoted by

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

which is an ordered list of length W, where r_1^f represents the lowest-ranked (least-desired) worker, and r_W^f represents the highest-ranked (most-desired) 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 $\mathcal{F}$, let $\mathbf{R}^{\mathcal{F}}$ be the corresponding vector of the rankings of the firms in $\mathcal{F}$.

The value to firm f of being matched to worker r_k^f is v_k.[8] It is better to hire the worst worker than to keep a vacancy, i.e., $0 < v_1 < v_2 < \ldots < v_W$. Let $\mathbf{v} \equiv [v_1, v_2, \ldots, v_W]$. The matching value

---

[7]Because the model assumes that it is always better to be matched with a worker than to keep a vacancy, with $W < F$ independent preferences would not be possible.

[8]The assumption that every firm has the same value of being matched with k-th worker on its list is needed for clarity of exposition. The general results remain true for differing matching values (see footnote 19).

vectors, $\mathbf{u}$ and $\mathbf{v}$, are publicly known.[9] 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.

Definition 1. *A* **matching** *between $\mathcal{F}$ and $\mathcal{W}$ is a function $\mu^{\mathcal{F},\mathcal{W}} : \mathcal{F} \to \mathcal{W} \cup \{\emptyset\}$ that assigns every worker to at most one firm. That is, for any two firms f and f' in $\mathcal{F}$ such that $f \neq f'$*

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

Expression $\mu^{\mathcal{F},\mathcal{W}}(f) = \emptyset$ means that in matching $\mu^{\mathcal{F},\mathcal{W}}$, firm f is not matched with any worker. When $\mu^{\mathcal{F},\mathcal{W}}(f) = w \in \mathcal{W}$, then firm f is matched with worker w. 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 $\mu^{\mathcal{F},\mathcal{W}}$.

Much of the literature emphasizes the importance of ex-post stability in matching. Roth (1991) and Kagel and Roth (2000), for example, argue that ex-post stable matching implemented after the arrival of relevant information should preclude early contracting. The notion of ex-post stability[10] was introduced by Gale and Shapley (1962). A matching is called *ex-post unstable* if it results in a firm and a worker who would prefer to be matched to each other than to remain in their current matches. A matching is called *ex-post stable* if it is not ex-post unstable.

For any $\mathcal{F}$ and $\mathcal{W}$, let $\boldsymbol{\mu}(\mathcal{F}, \mathcal{W})$ denote the set of all possible matchings between $\mathcal{F}$ and $\mathcal{W}$. Which of them is ex-post stable depends on firms' preferences, $\mathbf{R}^{\mathcal{F}}$. A well established result in the literature (e.g., Gusfield and Irving (1989) or Roth and Sotomayor (1990)) states that in the environment where workers' preferences are identical, for any given firms' preference profile there exists a unique ex-post stable matching between $\mathcal{F}$ and $\mathcal{W}$.[11] It can be characterized in the following way: The best firm — the firm most desired by workers — 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, and so on. 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. Let $\mu_S^{\mathcal{F},\mathcal{W}}(\mathbf{R}^{\mathcal{F}})$ denote the ex-post stable matching between $\mathcal{F}$ and $\mathcal{W}$ under firms' rankings $\mathbf{R}^{\mathcal{F}}$. 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 such matching.

A matching is defined between a subset of firms and a subset of workers. The special case of a matching between *all* firms and *all* workers describes an outcome in the market. 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 o_S — is the ex-post stable matching between all workers, $\{1, \ldots, W\}$, and all firms, $\{1, \ldots, F\}$, in the market: $\mathsf{o}_S \equiv \mu_S^{\{1,\ldots,W\},\{1,\ldots,F\}}$. I drop $\mathbf{R}$ from the notation, keeping in mind that ex-post stable matching depends on rankings. In the ex-post stable outcome, o_S, firm F is matched with its most-preferred worker, r_W^F. Firm $(F-1)$ is then matched with its most-preferred worker excluding $w \equiv r_W^F$, who has been already 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. Since the ex-post stable outcome is unique for every market, any other matching outcome is ex-post unstable. In

---

[9]For the purposes of this analysis, v_k and u_f do not need to be the precise values of a match; it is sufficient if they are the expected values. The actual values may be realized only after the match is made.

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

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

particular in the case where there is ex-post stable matching between subsets $\mathcal{F}$ and $\mathcal{W}$, but the rest of firms and workers are matched in some other way (e.g. at random), the matching outcome in the market is not ex-post stable.

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. *A **matching mechanism**, $\mathcal{M}$, is a function that maps $\mathcal{F}$, $\mathcal{W}$, and firms' reported rankings, $\widehat{\mathbf{R}}^{\mathcal{F}}$, to a lottery over all matchings between $\mathcal{F}$ and $\mathcal{W}$. That is*

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

where $Lottery\Big(\boldsymbol{\mu}(\mathcal{F}, \mathcal{W})\Big)$ is an element of the set of all possible lotteries on $\boldsymbol{\mu}(\mathcal{F}, \mathcal{W})$.

A matching mechanism is *incentive compatible* if no firm benefits from misreporting its preferences. A mechanism is called *ex-post stable* — and denoted $\mathcal{M}_S$ — if it applies ex-post stable matching to the reported rankings with probability 1. It is easy to confirm 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 ex-post stable matching between $\mathcal{F}$ and $\mathcal{W}$.[12]

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

With W workers there are $W!$ possible rankings. Denote as $\mathfrak{R}$ the set of all 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.[13] Two special cases — of identical preferences and independent preferences — are defined below.

Let G_1 be the joint distribution where all firms' rankings are identical and the marginal distribution of any individual ranking is uniform on $\mathfrak{R}$. That is, every ranking in $\mathfrak{R}$ is drawn with equal probability of $\frac{1}{W!}$ and all firms will have the same ranking.

Let G_0 be the joint distribution such that any firm's ranking is drawn from the uniform distribution independently of any other firms' rankings. That is, any combination of firms' rankings is drawn with equal probability of $\left(\frac{1}{W!}\right)^F$.

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

Definition 3. *For $\rho \in [0, 1]$,*
$$G_\rho = \rho G_1 + (1 - \rho) G_0$$

---

[12]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 $r_W^f, \ldots, r_{W-F+f}^f$ are relevant in producing ex-post stable matching. Under the ex-post stable mechanism, misreporting this portion of its ranking would make the firm strictly worse. Misreporting the rest of the ranking is irrelevant for the equilibrium outcome. Therefore, under the ex-post stable mechanism, the unique equilibrium outcome is ex-post stable matching between the agents that participate in the mechanism.

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

The parameter ρ is a measure of preference similarity[14] and will be a comparative statics parameter in the analysis below. Preferences are said to be *more similar* under $G_{\rho'}$ than under G_ρ when $\rho' > \rho$. Since ρ completely characterizes G_ρ, the two are used interchangeably.

The marginal distributions are uniform under both G_1 and G_0, and also under G_ρ. Therefore, prior beliefs in period 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 ($k = 1, \dots, W$) in a given firm's ranking with equal probability.

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

Figure 1 illustrates how the game unfolds. Market characteristics $(F, W, \mathbf{u}, \mathbf{v}, \rho, \mathcal{M})$ and workers' preferences are commonly known at any time. At the beginning of period 1, firms simultaneously decide whether or not to make an early offer, and if so, to which worker. Each firm can make at most one offer. After the early offers are released, each worker observes the offers he has received, if any. He does not see offers made to other workers. Every worker presented with an offer accepts or rejects it, based on his beliefs about other agents' strategies. 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 period 1, remain in the market for period 2. In period 2, firms' rankings are realized and a matching mechanism $\mathcal{M}$ operates on the agents remaining in the market. Section 4 of the paper assumes the ex-post stable mechanism in period 2. Section 5 considers other mechanisms. There is no discounting between the periods and making offers is costless.
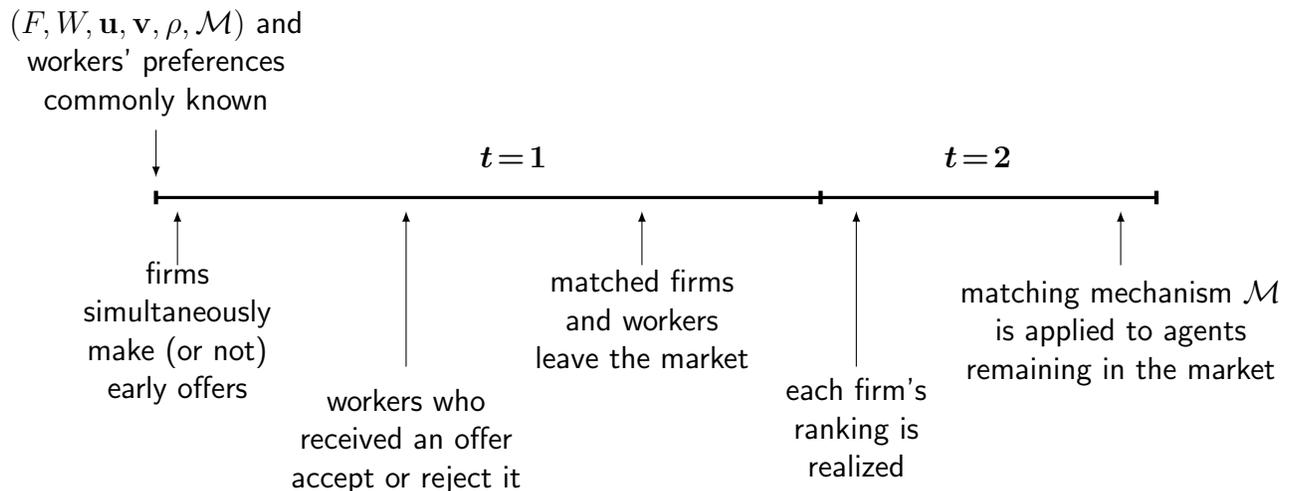
Figure 1: Timeline of the game

This paper considers only incentive compatible mechanisms.[15] Under an incentive compatible mechanism, firms truthfully report their rankings in period 2. Therefore, both firms and workers make their strategic decisions only in period 1. First, every firm decides whether or not to make

---

[14]Similarity of preferences, as measured by ρ is similar to the concept of correlation. However, correlation for rankings is not well defined. Since preferences are expressed as rankings, *rankings* and *preferences* are used interchangeably.

[15]Since the revelation principle holds in this environment, this is without loss of generality.

an offer and if so, to which worker. The analysis focuses on sequential equilibria in pure strategies, where the 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 period 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.

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 $(F, W, \mathbf{u}, \mathbf{v}, \rho, \mathcal{M})$, firms' realized rankings $\mathbf{R}$, and the strategies played by all agents in the market. The payoff expected by firm f at the beginning of the game depends on market characteristics, f's strategy and its beliefs about other agents' strategies. 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 to the current analysis are retained.

For a given market, let π_f be firm f's payoff and $E\pi_f(\sigma_f)$ — its expected payoff from playing strategy σ_f. Similarly, let U_w be worker w's utility, and $EU_w(\sigma_w \,|\, \beta_w(\boldsymbol{\sigma}_{-w}))$ – his expected utility from playing strategy σ_w.

A definition of sequential equilibrium applied to this model yields following characterization of equilibrium.

In the game with market $(F, W, \mathbf{u}, \mathbf{v}, \rho, \mathcal{M})$, a profile of strategies and system of beliefs constitute a sequential equilibrium when

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

    (f) in its only information set, firm $f \in \{1, \ldots, F\}$ chooses σ_f^* that maximizes its expected payoff, i.e.

$$E\pi_f(\sigma_f^*) \geq E\pi_f(\sigma_f) \qquad \forall \, \sigma_f \in \{1, \ldots, W\} \cup \{\emptyset\}$$

    (w) in each information set Ω_w, 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.

$$EU_w(\sigma_w^* \,|\, \Omega_w) \geq EU_w(\sigma_w \,|\, \Omega_w) \qquad \forall \, \sigma_w \in \{\Omega_w\} \cup \{\emptyset\}$$

(2) beliefs are consistent with the strategies played.

The beliefs are consistent with the strategies played on the equilibrium path. Firms make their decisions simultaneously at the beginning of the game. They cannot observe anything off the equilibrium path. Workers observe only the set of their own offers when making a decision to accept or reject. There are only two possible events that a worker may observe off-the-equilibrium path: when he receives an offer he did not expect, and when he does not receive an offer he expected. A property of sequential equilibrium determines in a unique way, sequentially rational beliefs and strategies even at the decision nodes not reached on the equilibrium path. In particular, in a sequential equilibrium, when a worker receives an offer he did not expect, he updates his beliefs only about the firm that made him the off-equilibrium offer. Now he believes that the firm made him an offer, instead of making it to some other worker or making no offer at all. However, this off-equilibrium offer does not change the worker's beliefs about any other firm.

In the case that a worker did not receive an offer he expected, he now beliefs that the firm has taken the best strategy available after excluding this worker. Again, this off-the-equilibrium event does not change the worker's beliefs about other firms.

Offers made and accepted in period 1 constitute unravelling.

Definition 4 (unravelling). Unravelling *is a situation where in equilibrium some firms and workers contract in the first period, before firms know their own preferences.*

4 Unravelling under Ex-Post Stable Mechanism

Ex-post stable matching is considered desirable in the extant literature. It has been proposed that an ex-post stable mechanism prevents unravelling (Roth 1991, Kagel and Roth 2000). It has also been argued that the ex-post stable outcome maximizes social welfare (Bulow and Levin 2006). Moreover, the ex-post stable mechanism is often adopted by clearinghouses introduced to prevent unravelling. The mechanism is chosen for clearinghouses either independently, as in the market for medical residents in 1952, or on the recommendation of economists, as in the case of the Boston public schools in 2005.[16] It can also be argued that ex-post stable matching is one of the equilibria in a decentralized market (a market without a clearinghouse) after information about preferences becomes available.

Given that the literature focuses on ex-post stable mechanisms, this section investigates unravelling under the ex-post stable matching mechanism. Subsection 4.1 focuses on equilibria without unravelling, while Subsection 4.2 describes equilibria when unravelling occurs.

This section assumes that the mechanism applied to the reported rankings in period 2 is the ex-post stable mechanism, $\mathcal{M}_S$. Mechanism $\mathcal{M}_S$ is not only incentive compatible, but in all equilibria it also produces ex-post stable matching among the agents remaining in period 2. Unless unravelling occurs in period 1, it produces the ex-post stable outcome, o_S.

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 period 1 chooses between u_f — a sure payoff for accepting the offer — and a lottery in period 2, in which he might be matched to a better firm or a worse firm or even remain unmatched. For a firm, early contracting yields expected payoff of the average value of workers, due to the uniform prior. The alternative is the ex-post stable matching in period 2, where better firms may be matched with firm f's most preferred workers.

When a firm expects the ex-post stable outcome in period 2, its expected payoff depends on its own position and the level of similarity of preferences in the market. The ex-post stable outcome has two properties that are of particular interest here. One is that lower-ranked firms receive lower expected payoffs in 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 period 2, firm f gets its most-preferred worker that remains in the pool after all better firms $i > f$ have been matched. Because fewer workers are left for worse firms, it is more likely that such firms' most-preferred workers are already matched. For this reason, worse firms are more likely to prefer early contracting under $\mathcal{M}_S$ than better firms.

To unravel, therefore, firms need to be good enough to be accepted in period 1 and bad enough to want to contract early. The result is "unravelling in the middle" — the phenomenon that in a typical market it is not the best or worst firms but firms "in the middle" that unravel. In special cases, firms at the extremes of the spectrum also contract early. It is possible to find equilibria in which any firm (except the best one) unravels.

---

[16]See Kimberly Atkins, "Committee OKs new school assignment plan", *Boston Herald*, July 21, 2005.

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

Let $E\pi_f(\mathsf{o}_S|\rho)$ denote firm f's expected payoff in the ex-post stable outcome in a given market. Then the following lemma summarizes the properties of o_S.

Lemma 1 (properties of o_S).

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

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

$$\rho < \rho' \quad \implies \quad E\pi_f(\mathsf{o}_S \,|\, \rho) > E\pi_f(\mathsf{o}_S \,|\, \rho')$$

Proof. See the Appendix, page 28.

4.1   Equilibria without Unravelling

An equilibrium has no unravelling when either no firm makes an early offer, or all early offers are rejected. This subsection explores conditions under which such an equilibrium exists.

Without unravelling, the ex-post stable mechanism $\mathcal{M}_S$ produces the ex-post stable outcome, o_S. Contracting in period 1 is contracting under uncertainty, in that firms' preferences are not known yet. A profitable deviation exists only when both the firm and the worker are better off contracting with deficient information than waiting for the uncertainty to be resolved.

Consider a worker who receives an offer from firm f in period 1, when in equilibrium all firms are expected to participate in the period 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 period 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 period 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

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

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 given market, the right-hand side of inequality (1) is constant, and u_f's are ordered to be increasing in f. Therefore, there is a cut-off point: the lowest-ranked firm whose offer will be accepted in period 1, as a deviation from the ex-post stable outcome. Let $L^0_{(W,\mathbf{u})}$ denote this firm:

$$L^0_{(W,\mathbf{u})} \equiv \min \left\{ f \,\middle|\, u_f > \frac{1}{W} \sum_{i=1}^{F} u_i \right\}$$

All firms worse than L^0 will be rejected in period 1. Firm L^0 and all firms better than L^0 will be accepted. The set of firms that will be accepted in period 1 is the *acceptance set*: $\left\{ L^0_{(W,\mathbf{u})}, \ldots, F \right\}$

Notice that at the end of period 2 there are always $W - F > 0$ workers who are unemployed, and who receive payoff 0. Because of the threat of unemployment, for any W and any f there exists a $\mathbf{u}$ such that the acceptance set includes f, i.e., $L^0_{(W,\mathbf{u})} \leq f$. In particular, it is possible that all firms would be accepted in period 1; that is, there exist W and $\mathbf{u}$ such that $L^0_{(W,\mathbf{u})} = 1$. This will occur when the number of workers, W, is large enough and the high probability of unemployment makes the utility expected in period 2 lower than u_1.

The incentives for firms to contract in period 1, before all relevant information is available, depend on the joint distribution of rankings, G_ρ. The realization of rankings — together with the matching mechanism — determines the outcome realized in period 2. Firms' expected payoffs depend on this expected outcome. Recall that $E\pi_f(\mathbf{o}_S|\rho)$ denotes firm f's expected payoff from the ex-post stable outcome under G_ρ.

The uniform prior implies that in period 1 all workers are ex ante the same. Thus an offer made to any worker in period 1 yields the same expected payoff. Any firm's expected payoff from early contracting — that is, if its offer is accepted — is

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

Firm f prefers early contracting to ex-post stable outcome when

$$\pi^0 > E\pi_f(\mathbf{o}_S|\rho) \tag{2}$$

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 most-preferred worker, r^{F-1}_W, unless that worker is firm F's most-preferred worker as well. When $r^{F-1}_W \equiv r^F_W$, firm $F-1$ gets the next worker on its list: r^{F-1}_{W-1}. Since the probability that $r^{F-1}_W \equiv r^F_W$ under G_ρ is $\rho + (1-\rho)\frac{1}{W}$, the expected payoff from the ex-post stable matching is*

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

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

The lower-ranked the firm, the lower its expected payoff in the ex-post stable outcome (Lemma 1(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 $\left\{ 1, \ldots, H^0_{(\rho,\mathbf{v})} \right\}$, where $H^0_{(\rho,\mathbf{v})}$ is the highest-ranked firm[17] that prefers early contracting to $\mathbf{o}_S$:

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

---

[17]For the special cases of $\rho = 0$ and $\rho = 1$ considered below, the corresponding firm H^0 is denoted by $H^0_{(G_0,\mathbf{v})}$ and $H^0_{(G_1,\mathbf{v})}$. Similarly, the expected payoff of the ex-post stable outcome is denoted by $E\pi_f(\mathbf{o}_S|G_0)$ and $E\pi_f(\mathbf{o}_S|G_1)$ for those special cases.

A deviation from $\mathbf{o}_S$ to early contracting can occur only when the offer in period 1 is made and accepted. Therefore, a profitable deviation from $\mathbf{o}_S$ is possible only when there exists a firm that prefers early contracting to the ex-post stable matching and when this firm's offer is accepted in period 1. That is, when there exists a firm that belongs to both the acceptance set and the offer set. This happens when the two sets have nonempty intersection, i.e. when for some f, $u_f > \frac{1}{W}\sum u_i$ and $\pi^0 > E\pi_f(\mathbf{o}_S|\rho)$, which means that

$$L^0_{(W,\mathbf{u})} \leq H^0_{(\rho,\mathbf{v})}$$

The H^0, and thus the offer set, depend on the similarity of preferences, ρ. The following subsections show that under independent preferences, G_0, the offer set is empty: no firm wants to contract in period 1, i.e. $H^0_{(G_0,\mathbf{v})} < 1$. Under identical preferences, G_1, by contrast, there may be firms willing to contract early, depending on $\mathbf{v}$. For intermediate cases, $H^0_{(\rho,\mathbf{v})}$ increases with ρ.

Independent Preferences, G_0

For independently distributed rankings, no firm prefers early contracting to the ex-post stable outcome. That is, $H^0_{(G_0,\mathbf{v})} < 1$ for any vector $\mathbf{v}$; thus the offer set is empty. Therefore, in any market with independent preferences, there is an equilibrium without unravelling.

Lemma 2. *For any F, $\mathbf{v}$ and $W > F$, if the preferences are independent, G_0, then $H^0_{(G_0,\mathbf{v})} < 1$. That is,*

$$\forall F, \mathbf{v}, W \quad s.t. \ W > F \qquad \pi^0 < E\pi_f(\mathbf{o}_S|G_0) \quad \forall f$$

Proof. See the Appendix, page 28.

The intuition for this result is 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 remaining worker may occupy 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 the case because with more than one worker to choose from, firm 1 will never be matched with its least-preferred 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 1(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 in firm f's ranking 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 $W-F+f$, i.e. r^f_{W-F+f}. Thus, $E\pi_f(\mathbf{o}_S|G_1) = v_{W-F+f}$.

Under G_1, condition (2) reduces to:

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

Firm f prefers to contract early rather than 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 firms and some values of $\mathbf{v}$. With a nonempty offer set, it is always possible to find an acceptance set for which there exists profitable deviation from o_S.

Example 2 — which is a formalized part of the example in Section 2 — shows a market characterized by identical preferences of firms, where there exists a profitable deviation.

Example 2. *Consider a 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*

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

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

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

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 period 1 than wait for* r_2^1 *in period 2. That is,* $H^0_{(G_1,\mathbf{v})} = 1$.

A worker's expected utility from period 2 matching is $\frac{1}{W}\sum_{f=1}^{F} u_f = \frac{15}{4} < 4 = u_1$. *This means that firm 1's offer in period 1 will be accepted by any worker. Thus,* $L^0_{(4,\mathbf{u})} = 1$ *and the acceptance set is* $\{1, 2, 3\}$. *Since the acceptance and the offer sets overlap at* $H^0_{(G_1,\mathbf{v})} = L^0_{(4,\mathbf{u})} = 1$, *there exists a profitable deviation from* o_S *in this market.* ∎

However, a profitable deviation from 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 period 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,* $H^0_{(G_1,\mathbf{v})} = 1$, *but now firm 1 does not belong to the acceptance set, as* $L^0_{(4,\mathbf{u}')} = 2$. *There is no profitable deviation from* o_S *in this market, as* $H^0_{(G_1,\mathbf{v})} < L^0_{(4,\mathbf{u})}$, *i.e., the offer and the acceptance set do not intersect.* ∎

As these above illustrate, under identical preferences a profitable deviation from o_S may but need not exist. This can also be interpreted in terms of existence of an equilibrium without unravelling: There are markets characterized by G_1, in which there exists an equilibrium without unravelling, but there also are markets with G_1 in which any equilibrium must exhibit unravelling.

Intermediate Similarity of Firms' Preferences

Firm F always has the same value of the ex-post stable matching: v_W. For all the other firms, the expected value of o_S decreases as similarity of preferences increases (Lemma 1(2)). As a consequence, holding other parameters of the market constant, more firms prefer early contracting as similarity of preferences increases. That is, holding other market parameters constant, $H^0_{(\rho,\mathbf{v})} \leq H^0_{(\rho',\mathbf{v})}$ whenever $\rho < \rho'$. Therefore, if for given market parameters $(F, W, \mathbf{v}, \mathbf{u})$ there exists a profitable deviation from o_S under G_ρ, then there also exists a profitable deviation under $G_{\rho'}$. For any market parameters $(F, W, \mathbf{v}, \mathbf{u})$, in fact, there exists a threshold ρ^{**} such that a profitable deviation from o_S exists for any similarity higher than the threshold but not for similarity lower than the threshold.

Lemma 3. *For any market parameters $(F, W, \mathbf{v}, \mathbf{u})$, there exists $\rho^{**} \in (0, 1]$ s.t.*

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

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

Proof. See the Appendix, page 29.

Workers' incentives to accept an offer in period 1 do not depend on similarity of preferences. However, firms' expected payoffs from 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 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 3 implies that as similarity of preferences increases, profitable deviation from o_S exists for a wider range of $(F, W, \mathbf{v}, \mathbf{u})$ parameters, and thus 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 $\rho^{**} < 1$, then for sufficiently similar preferences there is no equilibrium without unravelling. In Example 2, the threshold is strictly below 1. However, when the threshold is $\rho^{**} = 1$, there is an equilibrium without unravelling for any preferences, as in Example 3. Yet Lemma 2 assures that for any market parameters the threshold is strictly larger than 0. That is, for a market characterized by independent preferences, G_0, an equilibrium without unravelling always exists. This follows from the fact that under G_0, the offer set is always empty.

4.2   Equilibria with Unravelling

The previous section analyzed the conditions under which there exists an equilibrium in which 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 period 2. The remaining agents participate in the ex-post stable matching mechanism. In equilibrium, worker w who receives offers Ω_w in period 1 either accepts the best offer in Ω_w or rejects all of them, depending on which of the two options maximizes his expected utility. It is suboptimal for a worker to accept an offer from a firm other than the best firm in Ω_w. For a firm that prefers to contract in period 1, therefore, it is suboptimal to compete with a better firm for the same worker. In equilibrium firms that want to contract early make offers to different workers.

For any given equilibrium, define *equilibrium unravelling set* as the set of firms that contract early in this equilibrium, and denote this set by $\mathcal{U}$. The remaining firms, $\{1, \ldots, F\} \setminus \mathcal{U}$, participate in $\mathcal{M}_S$ in period 2 with workers still remaining in the market. The equilibrium unravelling set may be empty; such an equilibrium does not involve unravelling. More than one equilibrium may result in the same unravelling set $\mathcal{U}$. For example in one equilibrium a given firm makes an early offer and is rejected, and in another equilibrium the same firm does not make an early offer at all. Despite different strategies, 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.

It is a property of any equilibrium that the unravelling set is an interval, that is, $\mathcal{U}$ has no "holes." For the given equilibrium unravelling set $\mathcal{U}^*$, let firm H^* be the highest-ranked firm in $\mathcal{U}^*$,

and firm L^* — the lowest-ranked in $\mathcal{U}^*$. The fact that $\mathcal{U}^*$ is an interval means that all firms worse than H^* but better than L^* belong to $\mathcal{U}^*$ as well.

To understand why this is true, suppose to the contrary that in some equilibrium L^* and H^* belong to $\mathcal{U}^*$ but that a firm f between L^* and H^* is not in $\mathcal{U}^*$. This must be the case either because f prefers to wait or because it would not be accepted in period 1. But because f is lower-ranked than H^*, it prefers to contract early (as H^* does). And because it is better than L^*, it would be accepted (as L^* is). Therefore, it cannot be an equilibrium if L^* and H^* are in $\mathcal{U}^*$ and f is not. This result is formally stated in Lemma 4(1) below.

Thus, any nonempty $\mathcal{U}^*$ can be characterized by the best firm (H^*) and the worst firm (L^*) that contract early in such equilibrium: $\mathcal{U}^* \equiv \{L^*, \ldots, H^*\}$, for $L^* \le H^*$. 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 H^*, the *equilibrium condition for workers* characterizes L^*, i.e., the worst firm that would be accepted in period 1. That is, given that only firms H^* and below would want to make early offers, workers are willing to accept only firms L^* and above in period 1. Similarly, given L^*, the *equilibrium condition for firms* characterizes H^*, i.e., the best firm contracting in period 1.
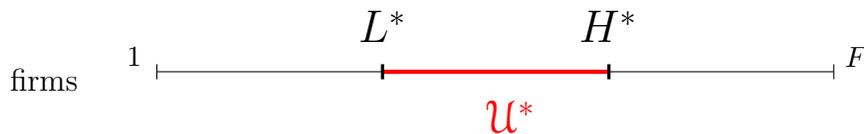

Figure 2: The structure of an equilibrium

In any market there is at least one equilibrium. Consider a market $(F, W, \mathbf{u}, \mathbf{v}, \rho, \mathcal{M}_S)$ where there exists a profitable deviation from o_S; that is, waiting for o_S without unravelling is not an equilibrium. That means that there is a number of firms, $\{L^0_{(W,\mathbf{u})}, \ldots, H^0_{(\rho,\mathbf{v})}\} \neq \emptyset$, that would like to contract early and would be accepted in period 1, even if no other firm would. As those firms unravel, the expected payoff of waiting until the second period decreases for better firms. This is the case because firms that unravel may end up hiring workers who would be matched to better firms in the ex-post stable outcome. When this happens, better firms are matched with some worse workers in period 2. This decrease in the expected payoff may induce some better firms to opt for early contracting, even though they initially preferred to wait for o_S. As more firms unravel, the expected payoff for workers staying for the second period also decreases, inducing workers to accept firms in period 1 that they initially would not accept. This circumstance again may increase the number of firms that unravel. Eventually, either the process induces all the firms $1, \ldots, F-1$ to unravel,[18] or it reaches a "fixed" state earlier. In both cases the market reaches an equilibrium with a nonempty unravelling set. Thus, in every market there is at least one equilibrium. This result is formally stated in Lemma 4(2) below. Moreover, in a typical market there is more than one equilibrium.

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* $\mathcal{U}' = \{2, 3, 4\}$.

---

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

Firms' condition for $\mathcal{U}^*$ is as follows: Knowing that firm 3 or better would be accepted in period 1, firm 3 prefers early contracting, but firm 4 prefers to wait for the ex-post stable matching — without firm 3 — in period 2. Workers' condition for $\mathcal{U}^*$ specifies that, knowing that firms 5 and 4 prefer to participate in period 2 matching, a worker accepts firm 3 but not firm 2 in period 1. Matching with firm 2 yields lower utility for a worker than the expectations for period 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 this does not need to be the case. The following example shows a market with multiple equilibria, some with unravelling and others without.

Example 5. *Consider a market similar to that in Example 4, with the sole difference that $\mathbf{u}' = [1, 6, 7, 14, 15]$. In such a market there are also exactly 2 equilibrium unravelling sets. One is the same as 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 $H^0_{(G_1, \mathbf{v})} < L^0_{(6, \mathbf{u}')}$. Since the utility from $\mathbf{o}_S$ expected by a worker is $\frac{1}{6}\sum u_f = 7\frac{1}{6}$, so $L^0_{(6, \mathbf{u}')} = 4$ and the acceptance set in this market is $\{4, 5\}$. As the expected payoff of an early offer is $\frac{1}{6}\sum v_k = 6.5$, so $H^0_{(G_1, \mathbf{v})} = 3$ and the offer set is $\{1, 2, 3\}$. Therefore, the offer and the acceptance sets do not intersect, i.e., there exists an equilibrium where $\mathcal{U}^* = \emptyset$ is the 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 cannot "overlap." To understand why, suppose to the contrary that there exist two equilibrium unravelling sets $\mathcal{U}^*$ and $\mathcal{U}'$ as in Figure 3(a). Notice that $\mathcal{U}^*$ includes better firms (on average) than $\mathcal{U}'$, and so worse firms (on average) stay for the period 2 under $\mathcal{U}^*$ than under $\mathcal{U}'$. By the equilibrium condition for workers, firm f is not included in $\mathcal{U}^*$ because its early offer would not be accepted. But since better firms unravel under $\mathcal{U}^*$ than under $\mathcal{U}'$, the expected utility for workers of staying in the market for $t = 2$ is lower under $\mathcal{U}^*$ than under $\mathcal{U}'$. If under $\mathcal{U}'$ it was better for a worker to accept firm f in period 1 than to wait for the expected utility in period 2, that must also be the case under $\mathcal{U}^*$. Lemma 4(3) states this result formally.
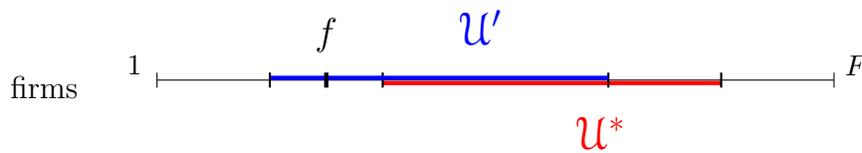
(a) An impossible configuration of multiple equilibrium unravelling sets
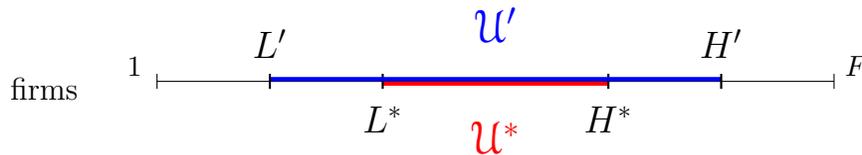
(b) A possible configuration of multiple equilibrium unravelling sets

Figure 3: Multiple equilibria with unravelling

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

Lemma 4. *Given a market* $(F, W, \mathbf{u}, \mathbf{v}, \rho, \mathcal{M}_S)$*:*

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

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

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

$$L' < L^* \iff H^* < H'$$

Proof. See the Appendix, page 29.

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

Comparative statics on ρ

This subsection investigates how equilibrium unravelling sets in a market $(F, W, \mathbf{u}, \mathbf{v}, \rho, \mathcal{M}_S)$ change with the level of 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.

If a market is characterized by independent preferences, all equilibria result in no unravelling. By Lemma 2, there always exists an equilibrium without unravelling for G_0. The following observation also says that no other equilibrium outcome is possible.

Observation. *In any market with* G_0*, the unique equilibrium outcome is "no-unravelling."*

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 period 2. Let* η *denote the number of firms better than* f *that unravel. Under independent preferences it makes no difference for firm* f *in period 2 if another firm contracts early with a random worker or picks its most-preferred worker before* f *in the ex-post stable matching. When* η *firms worse than* f *contract early, the effect on* f's *expected payoff is the same as if* η *more firms were choosing their most-preferred worker before* f *in the ex-post stable matching. Therefore, firm* f's *payoff from waiting when* η *worse firms unravel is the same as the payoff for firm* $f - \eta$ *in* $\mathbf{o}_S$ *(without unravelling). But Lemma 2 says that in* $\mathbf{o}_S$ *even firm* $f - \eta$ *gets 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 characterized by arbitrary preferences, $\mathcal{U} = \emptyset$ is an equilibrium unravelling set if and only if there is no profitable deviation from $\mathbf{o}_S$ in this market. Therefore, Lemma 3 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 4(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 in a market that $\mathcal{U}^{MAX} \equiv \mathcal{U}^{MIN}$,

that is, that 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 $\mathcal{U}(\rho)$ be an equilibrium unravelling set in a market characterized by similarity of preferences ρ. Then, holding other market parameters constant, $\mathcal{U}^{MIN}(\rho) \subseteq \mathcal{U}^{MIN}(\rho')$ and $\mathcal{U}^{MAX}(\rho) \subseteq \mathcal{U}^{MAX}(\rho')$ whenever $\rho < \rho'$.

As ρ increases, the maximum and minimum equilibria are more likely to be distinct. The maximum equilibrium unravelling set increases from empty to non-empty at 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 similarity of preferences increases, the minimum equilibrium unravelling set may also increase empty to non-empty. When this occurs, "no unravelling" is no longer an equilibrium in markets with high ρ. This relationship between equilibrium unravelling sets in a market and the level of preference similarity is illustrated by Figure 4.
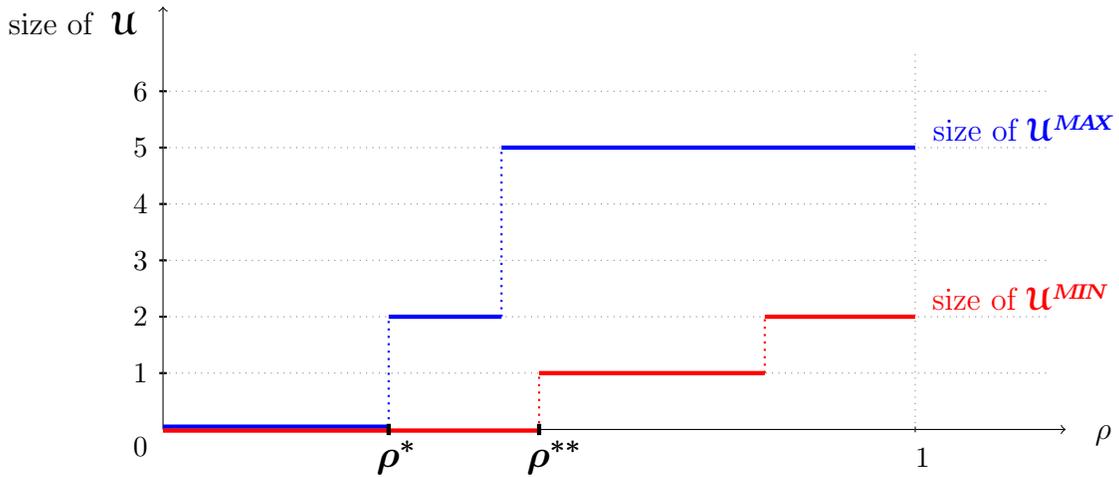
Figure 4: The relationship of $\mathcal{U}^{MIN}$, $\mathcal{U}^{MAX}$ and ρ in a typical market

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

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

Proof. See the Appendix, page 35.

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 similarity between ρ^* and ρ^{**}, there are equilibrium outcomes with unravelling and without unravelling; and for similarity higher than

ρ^{**}, all equilibrium outcomes involve a nonempty unravelling set. In extreme cases, the thresholds may be equal to 1. When $\rho^{**} = 1$, the interval $(\rho^{**}, 1]$ is empty, and for any similarity of preferences there exists an equilibrium without unravelling. Similarly, when $\rho^* = 1$, then in all equilibria for any preference similarity there is no unravelling. Moreover, ρ^* must be strictly greater than 0. This means that for any market parameters, all equilibria preclude unravelling if the preference similarity is sufficiently low.[19]

Discussion

This section has shown that similarity of preferences is an important factor driving unravelling. In markets where firms' preferences are very similar, strong competition for the same workers ensues when pertinent information becomes available. Lower-ranked firms view contracting early, with less information as their best chance of being matched with their most-preferred workers.

In some markets that have struggled with unravelling, firms have called on an outside institution, or created one, to solve the problem.[20] It is often recommended that such institutions apply the ex-post stable mechanism once 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 halting unravelling in many markets (e.g. markets for medical interns in the US and the UK). In other markets, however, the ex-post stable mechanism has failed to stop unravelling. For instance, in the Canadian lawyer market and the gastroenterology market in the US, agents contracted as early as a year before pertinent information was available and the mechanism was applied. In the latter case, the clearinghouse was subsequently abandoned in 1996, due to too few participants.

Proposition 1 explains why the ex-post stable matching mechanism has sometimes failed to stop unravelling. In some markets characterized by high similarity of preferences, there exist equilibria with and without unravelling. In those markets the ex-post stable clearinghouse could successfully stop unravelling, as it could help to coordinate on the no-unravelling equilibrium. In other markets, however, all equilibria under the ex-post stable mechanism involve early contracting. In such markets, introducing the ex-post stable clearinghouse will not stop unravelling. But it may be possible to find *unstable* matching mechanisms that can prevent unravelling in those markets. Section 5 shows that unravelling is always Pareto-suboptimal, and that there exist Pareto-optimal mechanisms for markets were unravelling is possible.

5 Mechanism Design

The previous section investigated unravelling when the mechanism operating in period 2 is fixed to be the ex-post stable one. This section turns to the problem of mechanism design in markets where unravelling may occur. Clearly, some of the mechanisms constructed in this section would be hard to apply in practice. However, they show that it is possible to find a mechanism that would fix

---

[19]This analysis assumes that matching value vector, $\mathbf{v}$, is identical for all firms. When firms have different values, it is still true that a firm is more willing to contract early if it is ranked lower in the market than if it is ranked higher. And more firms want to contract early (and are accepted) as preferences become more similar. However, with different matching values it is no longer true that the unravelling set is an interval, and that it can be easily characterized by L and H.

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

unravelling in markets where ex-post stable mechanism fails. These existence results are intended as a starting point for searching for other mechanism, easier to implement in practice.

A social planner chooses a mechanism for the second period; the mechanism is announced at the onset of the game. It is assumed that agents cannot renege on the matching produced by the mechanism in period 2, but that they can contract in period 1. Firms decide about their early offers, and workers decide whether to accept such offers, in full awareness of the mechanism that will operate in the next period. All agents that did not contract in period 1 participate in the mechanism in period 2. The goal of the social planner is to provide a Pareto-optimal outcome, from the ex-ante perspective.

This section shows that unravelling is Pareto-inefficient, for a broad class of mechanisms characterized by 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 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 used in real markets.

It is shown here that when an anonymous mechanism induces early contracting, 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 section also demonstrates that in every market there always exists a mechanism that produces a Pareto-optimal outcome. In the markets where the ex-post stable clearinghouse unravels, there exists an ex-post *unstable* mechanism that will stop unravelling and improve the welfare of the market participants.

5.1 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 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 o is

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

The previous section considered a special case of an outcome function — the ex-post stable outcome, o_S. This is the outcome in the case when all agents participate in the ex-post stable mechanism, $\mathcal{M}_S$.[21] This section also examines other outcomes and mechanisms.

Firm f's payoff from an outcome depends on the realized rankings, $\mathbf{R}$, and is denoted by $\pi_f(\mathsf{o}|\mathbf{R})$. The ex-ante expected payoff of an outcome is the expectation over all possible ranking realizations. The payoff and the expected payoff also depend on market characteristics, especially the probability distribution of ranking profiles.[22] Let $E\pi_f(\mathsf{o})$ be firm f's expected payoff from outcome o, then

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

---

[21]When unravelling occurs under the ex-post stable mechanism, the resulting outcome is other than o_S.

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

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

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

An outcome o' *strictly Pareto-dominates*[23] outcome o'' when

$$\left(\forall f \quad E\pi_f(\mathsf{o}') \geq E\pi_f(\mathsf{o}'') \qquad and \qquad \forall w \quad EU_w(\mathsf{o}') \geq EU_w(\mathsf{o}'') \right)$$

and

$$\left(\exists f \quad E\pi_f(\mathsf{o}') > E\pi_f(\mathsf{o}'') \qquad or \qquad \exists w \quad EU_w(\mathsf{o}') > EU_w(\mathsf{o}'') \right)$$

A matching outcome o is *Pareto-optimal* in a given market when there does not exist an outcome in that market that strictly Pareto-dominates 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 already known 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 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 period 2, an equilibrium under $\mathcal{M}$ is described by the first-period strategies of agents. Recall that $\boldsymbol{\sigma}$ denotes a vector of period 1 strategies 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 $\mathsf{o}_{(\mathcal{M},\boldsymbol{\sigma})}$.

A mechanism-equilibrium pair $(\mathcal{M}, \boldsymbol{\sigma})$ is *unconstrained Pareto-optimal* when it produces a Pareto-optimal outcome, that is, when $\mathsf{o}_{(\mathcal{M},\boldsymbol{\sigma})}$ is Pareto-optimal. However, a social planner is constrained to inducing outcomes by means of 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 $\mathsf{o}_{(\mathcal{M}',\boldsymbol{\sigma}')}$ strictly Pareto-dominates $\mathsf{o}_{(\mathcal{M},\boldsymbol{\sigma})}$. 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, not on workers' identities. For example, the ex-post stable mechanism, $\mathcal{M}_S$, is anonymous. Notice that in this context anonymity refers only to the workers' side. Identities of the firms — and thus preferences of workers — are commonly known. Define a vector of strategies, $\boldsymbol{\sigma}$, as anonymous if any firm that contracts with a worker in period 1 selects a worker at random, ignoring his identity.[24] 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 period 2 they differ in their positions in firms' rankings. It appears that all realistic mechanisms would match firms and workers based on the rankings, not on identities. Moreover, if a firm matches with a worker in period 1, that early contracting reflects the firm's expectations about the worker's eventual place in its ranking. But it does not depend

---

[23]All notions of Pareto-optimality here are from ex-ante perspective. Ex-post, every realized matching is Pareto-optimal as long as all firms are matched to some worker.

[24]It is assumed, however, that no two firms that want to contract in period 1 make offer to the same worker.

on the worker's identity, as expectations are identical for all workers. In this context, the worker's identity does not matter either for early contracting in period 1, or for a mechanism in period 2.

Notice that under an anonymous mechanism-equilibrium pair, every worker has the same ex-ante expected utility at the beginning of the first period. This utility depends on the set of firms that get matched under the mechanism-equilibrium pair. All workers have the same probability of being matched to any of those firms or of remaining 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 $\boldsymbol{\sigma}$.

5.2   Pareto-Optimality and Unravelling

The following proposition presents the main result of this section: 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 $(\mathcal{M}, \boldsymbol{\sigma})$ that exhibits unravelling, there exists an anonymous mechanism-equilibrium pair $(\mathcal{M}', \boldsymbol{\sigma}')$ such that it does not exhibit unravelling and that 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 least-preferred 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 "least-preferred workers correction") For all firms in $\mathcal{U}^{\mathcal{M}}$ that were matched to their least-preferred workers, $\mathcal{M}'$ replaces these workers with workers still remaining in the pool. This 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 preferable to the tentative match. This way, all firms tentatively matched with their least-preferred workers can improve their payoff. When there are no more firms in $\mathcal{U}^{\mathcal{M}}$ that are matched to their least-preferred worker, the algorithm stops and the matching is finalized.

Notice that $\mathcal{M}'$ is an incentive compatible mechanism. To understand how, recall that $\mathcal{M}$ is incentive compatible. For all 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 lose if they misreport their least-preferred worker.

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

There is an equilibrium without unravelling under $\mathcal{M}'$. This is the case because all firms in $\mathcal{U}^{\mathcal{M}}$ prefer to wait for $\mathcal{M}'$ rather than to unravel given that other firms wait for period 2. Since firms outside $\mathcal{U}^{\mathcal{M}}$ did not unravel when some other firms were contracting early — either because they preferred not to or because they would not be accepted in period 1 — they do not unravel when all other firms wait for period 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 to every worker does not change. Every firm in $\mathcal{U}^{\mathcal{M}}$ has a strictly higher expected payoff in $\mathsf{O}_{(\mathcal{M}', \boldsymbol{\sigma}')}$ than in $\mathsf{O}_{(\mathcal{M}, \boldsymbol{\sigma})}$. All the other firms have exactly the same expected payoff in both outcomes. Therefore, $\mathsf{O}_{(\mathcal{M}', \boldsymbol{\sigma}')}$ Pareto-dominates $\mathsf{O}_{(\mathcal{M}, \boldsymbol{\sigma})}$. $\qquad\square$

5.3 The Role of Anonymity

To understand the role that 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*

$$EU_i(\mathsf{O}_{(\mathcal{M}, \boldsymbol{\sigma})}) = \begin{cases} u_i & for \quad i = 1, \dots, F \\ 0 & for \quad i = F+1, \dots, W \end{cases}$$

Each firm's ex-ante expected payoff is

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

Notice that, despite unravelling in equilibrium, this outcome cannot be Pareto-improved. In particular, it cannot be improved by the equilibrium without unravelling. This is because any change in allocation would make some worker worse off. ∎

Anonymity is an important assumption for the result in Proposition 2. However, it also seems reasonable to consider anonymous mechanisms the most plausible for implementation in real markets.

5.4 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.[25] 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 $\boldsymbol{\sigma}$.*

---

[25]E.g., see Roth and Sotomayor (1990).

When $\mathcal{M}_S$ 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 that Pareto-improves the unravelling outcome.

Proposition 3 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 $\boldsymbol{\sigma} \in \Sigma^{\mathcal{M}}$ such that $(\mathcal{M}, \boldsymbol{\sigma})$ is unconstrained Pareto-optimal.*

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

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

Now, notice that $(\mathcal{M}, \boldsymbol{\sigma})$ produces a 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 they are under o_S. Since o_S is Pareto-optimal, so must $\mathsf{o}_{(\mathcal{M}, \boldsymbol{\sigma})}$ be: in both outcomes it is impossible to increase the expected payoff for one agent without decreasing it for some other agent on the same side of the market. $\qquad\square$

Thus, if $(\mathcal{M}_S, \boldsymbol{\sigma})$ exhibits unravelling, the social planner can implement another mechanism that achieves a Pareto-optimal outcome. Notice, however, that the Pareto-optimal outcome as constructed in the proof of the proposition above does not Pareto-improve on a $(\mathcal{M}_S, \boldsymbol{\sigma})$ that exhibits unravelling.

Nonetheless, following observation shows that whenever $(\mathcal{M}_S, \boldsymbol{\sigma})$ exhibits unravelling, the mechanism may be modified in such a way that not only is there no unravelling and the outcome is Pareto-optimal but the outcome also strictly Pareto-improves upon $(\mathcal{M}_S, \boldsymbol{\sigma})$. The modification of the mechanism uses a construction similar to the proof of Proposition 3.

Observation. *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 $\{1, \ldots, W\}$.[26]

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

(3) Firms get matched with their most-preferred available worker in order of their numbers, starting with the highest number. That is, the firm with the highest number is treated as

---

[26]Notice that the number is drawn from the index numbers of workers, not just firms. (There are more workers than firms.)

firm F in the ex-post stable matching, the firm with the second-highest number is treated as firm $(F - 1)$, and so on, ending with the firm with the lowest number, which is treated as firm 1 in the ex-post stable matching.

This mechanism is incentive compatible, as is the ex-post stable mechanism. Moreover, there exists an equilibrium without unravelling under $\mathcal{M}^{\mathcal{A}}$. Denote this equilibrium by $\boldsymbol{\sigma}^{\mathcal{A}}$. Let $\mathsf{o}^{\mathcal{A}}$ be the outcome of $(\mathcal{M}^{\mathcal{A}}, \boldsymbol{\sigma}^{\mathcal{A}})$. Outcome $\mathsf{o}^{\mathcal{A}}$ is Pareto-optimal. To understand why, notice that the sum of expected payoffs to firms and the sum of expected utilities to workers are the same under $\mathsf{o}^{\mathcal{A}}$ as under o_S. Since o_S is Pareto-optimal, $\mathsf{o}^{\mathcal{A}}$ is as well: in both outcomes it is impossible to increase the expected payoff for one agent without decreasing it for some other agent.

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

6   Conclusions

This study investigates the causes and welfare consequences of unravelling in two-sided matching markets. It considers a two-period model in which firms receive pertinent information about workers and specify preferences over them at the beginning of the second period. It is assumed that firms and workers can make and accept offers during the first period if they wish to, and that a clearinghouse mechanism is used in the second period to assign workers to the remaining firms. 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 in the absence of information about which workers are most-preferred.

Section 4 explores the issue of unravelling when the ex-post stable mechanism operates in the second period. Ex-post stable matching is the clearinghouse mechanism that most of the existing literature focuses on. Section 4 shows that unravelling becomes more likely as firms' preferences over workers grow more similar. This is the case because when firms' preferences are very similar, lower-ranked firms can be matched with their most-preferred worker only by contracting with them early. Despite insufficient information in the first period, it may be worthwhile for such firms to bear the risk and contract early. The firms most likely to unravel are those "in the middle" — bad enough to prefer the uncertainty of early contracting but good enough to be accepted.

Section 5 investigates the impact of different mechanisms on the equilibrium outcome and their welfare consequences. The goal of this 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 demonstrates 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-suboptimal. In particular, the ex-post stable matching mechanism is Pareto-optimal if and only if it does not unravel.

Another result of Section 5 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 a policy standpoint.

Appendix

Proof of Lemma 1 (page 12)

First, note that the probability that firm f is matched with its kth worker in the ex-post stable outcome under independent preferences is:

$$P(W, f, k) \equiv \frac{(F - f)!}{(F - W - f + k)!} \frac{(k - 1)!}{W!}(W - F + f)$$

This formula is derived by applying combinatorics formulas.[27]

(1) *Proof.* The probability that firm $f - 1$ gets its worker $k > W - F + f$ is

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

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

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

(2) *Proof.* $E\pi_f(\mathbf{o}_S|\rho) = \sum_{k=1}^{W} v_k \cdot Prob(\mathbf{o}_S(f) = r_k^f|\rho) = \rho \cdot E\pi_f(\mathbf{o}_S|G_1) + (1 - \rho)E\pi_f(\mathbf{o}_S|G_0)$
because $Prob(\mathbf{o}_S(f) = r_k^f|\rho) = \rho \cdot Prob(\mathbf{o}_S(f) = r_k^f|G_1) + (1 - \rho)Prob(\mathbf{o}_S(f) = r_k^f|G_0)$

$$E\pi_f(\mathbf{o}_S|G_0) = \sum_{k=W-F+f}^{W} v_k \cdot P(W, f, k) > v_{W-F+f} \sum_{k=W-F+f}^{W} \cdot P(W, f, k) = v_{W-F+f} = E\pi_f(\mathbf{o}_S|G_1)$$

Let $\rho' > \rho$, then

$$E\pi_f(\mathbf{o}_S|\rho') =$$

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

$$< E\pi_f(\mathbf{o}_S|\rho)$$

This completes the proof. □

Proof of Lemma 2 (page 14)

Proof. Consider the worst firm, firm 1.

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

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

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

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

By Lemma 1(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 $\mathbf{o}_S$ rather than to unravel. □

---

[27]Derivation of this formula is available at author's website.

Proof of Lemma 3 (page 16)

Proof. Fix an arbitrary market parameters $(W, F, \mathbf{u}, \mathbf{v})$. An equilibrium without unravelling exists when $H^0 < L^0$. L^0 does not depend on ρ. H^0 is monotonic in ρ, i.e.

$$\rho < \rho' \implies H^0_{(\rho, \mathbf{v})} \leq H^0_{(\rho', \mathbf{v})}$$

This follows from the fact that $E\pi_f(\mathsf{O}_S \,|\, \rho)$'s decrease as ρ increases, by Lemma 1(2).

By Lemma 2, for $\rho = 0$, $H^0_{(G_0, \mathbf{v})} < 1$, so $H^0_{(G_0, \mathbf{v})} < L^0$ always holds.

For $\rho = 1$ it must be either $H^0_{(G_1, \mathbf{v})} < L^0$ or $L^0 \leq H^0_{(G_1, \mathbf{v})}$. If $H^0_{(G_1, \mathbf{v})} < L^0$, then $\rho^{**} = 1$ and for all $\rho \in [0, 1]$ there exists an equilibrium without unravelling.

If $L^0 \leq H^0_{(G_1, \mathbf{v})}$, then by the monotonicity of H^0 there must exists ρ^{**} such that

$$H^0(\rho) < L^0 \quad \text{for all} \quad \rho \leq \rho^{**}$$
$$\text{and} \quad L^0 \leq H^0(\rho) \quad \text{for all} \quad \rho > \rho^{**}$$

$\square$

Proof of Lemma 4 (page 19)

(1) **interval property**:

Proof. Assume, to the contrary, that there exists a firm f such that $H^* > f > L^*$ and $f \notin \mathcal{U}^*$. If the firm is not in $\mathcal{U}^*$, it either prefers to wait or would not be accepted in period 1.

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

$$EU(t = 2 | \mathcal{U}^* \setminus \{L^*\}) < u_{L^*} \implies EU(t = 2 | \mathcal{U}^*) < EU(t = 2 | \mathcal{U}^* \setminus \{L^*\})$$

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

Since firm H^* prefers to contract early, given all the other firms in $\mathcal{U}^*$ contracting in period 1,

$$E\pi_{H^*}(\mathcal{M}_S | \mathcal{U}^* \setminus \{H^*\}) < \pi^0$$

where $E\pi_f(\mathcal{M}_S | \mathcal{U})$ represents the expected payoff to firm f from the ex-post stable mechanism operating in period 2 given that firms in $\mathcal{U}$ unravel.

Notice that[28]

$$E\pi_{H^*}(\mathcal{M}_S | \mathcal{U}^* \setminus \{H^*\}) \geq E\pi_f(\mathcal{M}_S | \mathcal{U}^* \setminus \{L^*\})$$

Since the expected payoff from period 2 matching decreases when an additional worse firm unravels:[29]

$$E\pi_f(\mathcal{M}_S | \mathcal{U}^*) < E\pi_f(\mathcal{M}_S | \mathcal{U}^* \setminus \{L^*\})$$

Together, these inequalities yield $E\pi_f(\mathcal{M}_S | \mathcal{U}^*) < \pi^0$. Therefore, f prefers to contract in period 1 rather than wait.

Hence, the contradiction. $\square$

---

[28]This is for the following reason. Let n^H be the number of firms better than H^*. Let $n^{\mathcal{U}}$ be the number of firms in $\mathcal{U}^*$. And let n' be the number of firms better than f which *do not belong* to $\mathcal{U}^*$; notice that $n' \geq n^H$. When firms in $\mathcal{U}^* \setminus \{H^*\}$ unravel, H^* is in the second period mechanism the $(n^H + 1)$-th firm from the top with $n^{\mathcal{U}} - 1$ firms unravelling. When firms in $\mathcal{U}^* \setminus \{L^*\}$ unravel, f is in the second period the $(n' + 1)$-th firm from the top with $n^{\mathcal{U}} - 1$ firms unravelling. That is no better position than H^*'s in the former situation.

[29]This is because this worse firm can match with a "too-good" worker, which would cause f to match with a worse worker than if this worse firm had not unraveled. There is no potential benefits for f from a worse firm unravelling.

(2) **existence of pure strategy equilibrium**: Given H^*, the *equilibrium condition for workers* (**CW**) characterizes L^*, i.e., the worst firm that would be accepted in period 1.

With firms $L, \ldots, H$ contracting in period 1, a worker's expected payoff from matching in period 2 is

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

The worst firm that would be accepted in period 1, given H^*, is characterized by two inequalities:[30]

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

and (**CW**)

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

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

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

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

and (**CF**)

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

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

The first part of equilibrium condition (**CF**) says that firm H^* prefers to contract early, given that firms $L^*, \ldots, (H^* - 1)$ do so. The second part of the condition assures that H^* is the highest-ranked firm that wants to contract early. That is, better firm (H^*+1) prefers to wait until period 2, given that firms $L^*, \ldots, H^*$ unravel. Because of the monotonicity of the expected payoffs in the ex-post stable matching, this also means that, given L^*, all firms worse than H^* prefer to contract early, and all firms better than H^* prefer to wait.

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

(cf) given L^*, H^* satisfies condition (**CF**), and

(cw) given H^*, L^* satisfies condition (**CW**).

---

[30]More precisely, the first inequality is

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

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

If in a market there exists an equilibrium without unravelling, i.e. $L^0_{(F,\mathbf{u})} > H^0_{(\rho,\mathbf{v})}$, then the existence of a pure strategy equilibrium is satisfied.

To prove the existence for the case when $L^0_{(F,\mathbf{u})} \leq H^0_{(\rho,\mathbf{v})}$, notice the following:

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

Those properties are formally stated in following lemmas.

Lemma 5. *In a given market* $(F, W, \mathbf{u}, \mathbf{v}, \rho, \mathcal{M})$*, for any firm* H^**, there exists unique* $L(H^*)$ *that satisfies condition* (**CW**) *under* H^**.*

Proof. Consider a market $(F, W, \mathbf{u}, \mathbf{v}, \rho, \mathcal{M})$ and an arbitrary firm H^*.

It must be that either

(A) there exists firm $i \leq H^*$ s.t. $u_i > \frac{1}{W - H^* + i - 1} \left(\sum_{f=1}^{i-1} u_f + \sum_{f=H^*+1}^{F} u_f \right)$, or

(B) for all firms $i \leq H^*$, $u_i \leq \frac{1}{W - H^* + i - 1} \left(\sum_{f=1}^{i-1} u_f + \sum_{f=H^*+1}^{F} u_f \right)$.

In case (B) no firm worse than H^* would be accepted in $t = 1$, if firms better than H^* do not unravel. So, there is no firm contracting early. Condition (**CW**) for this special case takes on the form

$$ u_L > \frac{1}{W} \sum_{f=1}^{F} u_f $$

$$ u_{L-1} \leq \frac{1}{W} \sum_{f=1}^{F} u_f $$

Firm L that satisfies this condition is $L^0_{(F,\mathbf{u})}$ (and it must be that $L^0_{(F,\mathbf{u})} > H^*$). Thus for this case $L(H^*)$ exists and is unique.

In case (A), again, it must be that either

(i) there exists a firm $j < i$ s.t. $u_j \leq \frac{1}{W - H^* + j - 1} \left(\sum_{f=1}^{j-1} u_f + \sum_{f=H^*+1}^{F} u_f \right)$, or

(ii) for all firms $j < i$, $u_j > \frac{1}{W - H^* + j - 1} \left(\sum_{f=1}^{j-1} u_f + \sum_{f=H^*+1}^{F} u_f \right)$

In case (ii) all firms worse than H^* (including H^*) would be accepted in $t = 1$. For this special case condition (**CW**) takes on a reduced form

$$ u_1 > \frac{1}{W - H^*} \sum_{f=H^*+1}^{F} u_f $$

since firm 1 is the lowest-ranked firm that is accepted in $t = 1$ given that H^* is the highest-ranked firm making offers in $t = 1$. Thus, $L(H^*)$ in this case exists and is unique.

For the remaining case (Ai), $L(H^*)$ exists because

$$ u_i > \frac{1}{W - H^* + i - 1} \left(\sum_{f=1}^{i-1} u_f + \sum_{f=H^*+1}^{F} u_f \right) \iff u_i > \frac{1}{W - H^* + i} \left(\sum_{f=1}^{i} u_f + \sum_{f=H^*+1}^{F} u_f \right) $$

To prove that $L(H^*)$ is unique, assume — to the contrary — that two distinct L^* and L', lower-ranked than H^* (but higher-ranked than firm 1), satisfy (**CW**) given H^*. Without loss of generality, $L' < L^* \leq H^*$, i.e., $u_{L'} < u_{L^*}$. Then,

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

and

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

as well as

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

and

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

That is,

$$u_{L'-1} \leq EU(t = 2 \,|\, \mathcal{U}(L', H^*)) < u_{L'} \leq u_{L^*-1} \leq EU(t = 2 \,|\, \mathcal{U}(L^*, H^*)) < u_{L^*} \tag{3}$$

Notice, however, that $\{L^*, H^*\}$ has on average better firms than $\{L', H^*\}$; i.e. $\frac{\sum_{f=L^*}^{H^*} u_f}{H^* - L^* + 1} > \frac{\sum_{f=L'}^{H^*} u_f}{H^* - L' + 1}$. So the firms left for $t = 2$ are on average worse under $\{L^*, H^*\}$ than under $\{L', H^*\}$; i.e. $EU(t = 2 \,|\, \mathcal{U}(L', H^*)) > EU(t = 2 \,|\, \mathcal{U}(L^*, H^*))$.[31] Hence, a contradiction. $\square$

Lemma 6. *If* $H' > H^*$ *then* $L(H') < L(H^*)$.

Proof. Let $L' \equiv L(H')$ and $L^* \equiv L(H^*)$.
Assume, to the contrary, $L' > L^*$; so $u_{L'} > u_{L^*}$ and $\Delta L = L' - L^* > 0$. The corresponding

---

[31] To show this formally, let $n' = W - H^* + L' - 1$ and $\Delta L = L^* - L' > 0$. Then $W - H^* + L^* - 1 = n' + \Delta L$ and

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

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

So

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

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

From the equilibrium condition for $\mathcal{U}^*$, $EU(t = 2 \,|\, \mathcal{U}(L^*, H^*)) > u_{L^*-1}$. Thus all terms in the sum on the RHS are positive. Therefore, $EU(t = 2 \,|\, \mathcal{U}(L', H^*)) > EU(t = 2 \,|\, \mathcal{U}(L^*, H^*))$.

(**CW**) conditions are

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

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

and

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

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

Thus,

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

Let $\Delta H = H' - H^* > 0$ and $\gamma = H^* - L' + 1$. Then

$$EU(t = 2 \,|\, \mathcal{U}(L^*, H^*)) = \frac{1}{W - \gamma - \Delta L} \left(\sum_{f=1}^{L^*-1} u_f + \sum_{f=H^*+1}^{H'} u_f + \sum_{f=H'+1}^{F} u_f \right) \tag{4}$$

$$EU(t = 2 \,|\, \mathcal{U}(L', H')) = \frac{1}{W - \gamma - \Delta H} \left(\sum_{f=1}^{L^*-1} u_f + \sum_{f=L^*}^{L'-1} u_f + \sum_{f=H'+1}^{F} u_f \right) \tag{5}$$

From (4)

$$\sum_{f=1}^{L^*-1} u_f + \sum_{f=H'+1}^{F} u_f = EU(t = 2 \,|\, \mathcal{U}(L^*, H^*))(W - \gamma - \Delta L) - \sum_{f=H^*+1}^{H'} u_f$$

And from (5)

$$\sum_{f=1}^{L^*-1} u_f + \sum_{f=H'+1}^{F} u_f = EU(t = 2 \,|\, \mathcal{U}(L', H'))(W - \gamma - \Delta H) - \sum_{f=L^*}^{L'-1} u_f$$

So,

$$EU(t = 2 \,|\, (L^*, H^*))(W - \gamma - \Delta L) - \sum_{f=H^*+1}^{H'} u_f =$$

$$= EU(t = 2 \,|\, \mathcal{U}(L', H'))(W - \gamma - \Delta H) - \sum_{f=L^*}^{L'-1} u_f$$

$$\left(EU(t = 2 \,|\, (L^*, H^*)) - EU(t = 2 \,|\, \mathcal{U}(L', H')) \right)(W - \gamma) =$$

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

33

$$\Big(EU(t=2\,|\,(L^*,H^*)) - EU(t=2\,|\,\mathcal{U}(L',H'))\Big)(W - \gamma - \Delta L) =$$

$$= \sum_{f=H^*+1}^{H'} u_f - \sum_{f=L^*}^{L'-1} u_f - \underbrace{EU(t=2\,|\,\mathcal{U}(L',H'))(\Delta H - \Delta L)}_{(\blacktriangle)}$$

Formula ($\blacktriangle$) is always positive. This is because for all $f \in \{L^*, \dots, L'-1\}$, $u_f < EU(t=2\,|\,(L',H'))$ and for all $f \in \{H^*+1, \dots, H'\}$, $u_f > EU(t=2\,|\,(L',H'))$. Thus

$$\sum_{H^*+1}^{H'} u_f > \Delta H \cdot EU(t=2\,|\,(L',H')) \quad \text{and} \quad \sum_{L^*}^{L'-1} u_f < \Delta L \cdot EU(t=2\,|\,(L',H'))$$

And combined:

$$\sum_{H^*+1}^{H'} u_f - \sum_{L^*}^{L'-1} u_f > EU(t=2\,|\,(L',H'))(\Delta H - \Delta L)$$

$\square$

Similar properties are also true for the firms' condition (**CF**).

Lemma 7. *In a given market* $(F, W, \mathbf{u}, \mathbf{v}, \rho, \mathcal{M})$, *for any firm* L^*, *there exists unique* $H(L^*)$ *that satisfies condition* (**CF**) *under* L^*.

Proof. First, notice that for any L^* and any $f \geq L^*$

$$E\pi_{f+1}\big(\mathcal{M}_S|\mathcal{U}(L^*, f)\big) \geq E\pi_f\big(\mathcal{M}_S|\mathcal{U}(L^*, f-1)\big) \tag{6}$$

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

This must be true because of the following reasoning: Firm $f+1$ has $F-(f+1)$ firms better than itself, and $f-L^*+1$ firms that unravel. Firm f has $F-f$ firms better than itself, and $f-1-L^*+1 = f-L^*$ firms that unravel. Under $\rho = 0$, it is as if firm $f+1$ was $f+1-(f-L^*+1) = L^*$ firm in $\mathbf{O}_S$. Similarly, firm f is as if it was $f-(f-L^*) = L^*$ firm in $\mathbf{O}_S$. Thus under $\rho = 0$ their payoffs are the same. Under independent preferences, if firm f' — higher-ranked than firm f — unravels, it results in exactly the same expected payoff to firm f as if both would wait until the ex-post stable matching. Under $\rho = 1$, it is better to have one more firm unravel rather than to have one firm that is higher ranked and doesn't unravel and for sure will be matched with a more preferred worker. Thus, under $\rho = 1$ the inequality holds strictly. Because other ρ's are the convex combination of $\rho = 0$ and $\rho = 1$, the strict inequality in (6) holds for $\rho > 0$.

Now, given L^*, it must be that either

 (A) there exists firm $f \geq L^*$ such that $E\pi_f(\mathcal{M}_S\,|\,\mathcal{U} = \emptyset) < \pi^0$, or

 (B) for all $f \geq L^*$, $E\pi_f(\mathcal{M}_S\,|\,\mathcal{U} = \emptyset) \geq \pi^0$.

In the latter case, no firm higher ranked than L^* would prefer to contract early. In such a case there would be no unravelling. Condition (**CF**) in this special case takes on the form

$$E\pi_H(\mathcal{M}_S\,|\,\mathcal{U} = \emptyset) < \pi^0$$
$$E\pi_{H+1}(\mathcal{M}_S\,|\,\mathcal{U} = \emptyset) \geq \pi^0$$

Firm H that satisfies this condition is $H^0_{(\rho, \mathbf{v})}$ (and it must be that $H^0_{(\rho, \mathbf{v})} < L^*$). Thus, for this case $H(L^*)$ exists and is unique.

Now, consider the case (A). We know that the best firm never wants to contract early, even when all the other firms are unravelling (this follows from the assumption that $W > F$). Thus,

$$E\pi_F(\mathcal{M}_S \,|\, \mathcal{U}(L^*, F - 1)) \geq \pi^0$$

Together with the monotonicity condition (6), there must exist a unique H^* such that satisfies condition (**CF**) given L^*. This is $H(L^*)$. $\qquad \square$

Lemma 8. *If $L' > L^*$ then $H(L') < H(L^*)$.*

Proof. Assume, to the contrary, that $H(L') > H(L^*)$. Then the proof follows the proof of Lemma 6. $\qquad \square$

Given Lemmas 5–8, we can proceed to prove the existence of equilibrium in pure strategies for the case where $L^0_{(F,\mathbf{u})} \leq H^0_{(\rho,\mathbf{v})}$. Let $H^1 = H(L^0_{(F,\mathbf{u})})$ and $L^1 = L(H^0_{(\rho,\mathbf{v})})$, and further $H^{i+1} = H(L^i)$ and $L^{i+1} = L(H^i)$.

Because of the monotonicity and uniqueness results in the lemmas above, it must be that there exist H^* and L^* such that $H^*(L^*)$ and $L^*(H^*)$. To see that, notice that by the monotonicity results in Lemmas 6 and 8: $L^1 = L(H^0) \leq L^0$ and $H^1 = H(L^0) \geq H^0$. Further, $L^2 = L(H^1) \leq L(H^0) = L^1$ since $H^1 \geq H^0$. And so for any i, $L^{i+1} \leq L^i$ and $H^{i+1} \geq H^i$. Either this iteration stops for some i where $H^{i+1} = H^i$ and $L^{i+1} = L^i$ and so $H^i = H^* = H(L^*)$ and $L^i = L^* = L(H^*)$, i.e H^* and L^* are equilibrium values; or it continues until it reaches the extreme case $L^* = 1$ and $H^* = H(L^*)$.

Thus, an equilibrium exists.

(3) **multiple equilibria**:

Proof. This property follows directly from Lemmas 5–8.

Lemmas 5 and 7 show that it cannot be that two different equilibrium unravelling sets would have the same L or the same H.

Lemmas 6 and 8 show that "overlapping" equilibrium unravelling sets are not possible.

Thus the only possibility for two different unravelling sets is for one to be completely included in the other.

This completes the proof of Lemma 4. $\qquad \square$

Proof of Proposition 1 (page 20)

Proof. First, notice that in any market with G_0, $(F, W, \mathbf{u}, \mathbf{v}, G_0)$, the only equilibrium outcome is $\mathcal{U}^* = \emptyset$.

To see why it must be so, 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 $E\pi_f(\mu_S^{\mathcal{F},\mathcal{W}}|G_0, \eta) < E\pi(t = 1)$, where η is the number of firms that unravel and are higher-ranked than f. But $E\pi_f(\mu_S^{\mathcal{F},\mathcal{W}}|G_0, \eta) \equiv E\pi_{f-\eta}(\mathbf{o}_S|G_0)$ (for $f - \eta \geq 1$, which is always satisfied). By Lemma 2, for any $i \geq 1$, $E\pi_i(\mathbf{o}_S|G_0) > E\pi(t = 1)$. So it must also be true for $i \equiv f - \eta$. Therefore, a contradiction and thus the only equilibrium outcome under G_0 is $\mathcal{U}^* = \emptyset$.

The rest of the proof follows from the fact that $\mathcal{U}^{MIN} \subseteq \mathcal{U}^{MAX}$ and monotonicity of $H^0(\rho, \mathbf{v})$ and $E\pi_f(\mu_S^{\mathcal{F},\mathcal{W}}|\rho)$ in ρ. The part for ρ^{**} is proven in Lemma 3. For ρ^*, notice that for any market parameters $(F, W, \mathbf{u}, \mathbf{v})$ under identical preferences, G_1, it must be that either $\mathcal{U}^{MAX} = \emptyset$, or $\mathcal{U}^{MAX} \neq \emptyset$. In the former case, $\rho^* = 1$ satisfies the Proposition.

In the latter case, let $\mathcal{U}^{MAX} = \{L^{MAX}, H^{MAX}\}$. From condition (**CF**) must be that

$$E\pi_{H^{MAX}}(\mathcal{M}_S \,|\, \mathcal{U}^{MAX} \setminus \{H^{MAX}\}, G_1) < \pi^0$$

By monotonicity of $E\pi_f$ in ρ, for $\rho < 1$

$$E\pi_{H^{MAX}}(\mathcal{M}_S \mid \mathcal{U}^{MAX} \smallsetminus \{H^{MAX}\}, \rho) > E\pi_{H^{MAX}}(\mathcal{M}_S \mid \mathcal{U}^{MAX} \smallsetminus \{H^{MAX}\}, G_1)$$

And we also know that

$$E\pi_{H^{MAX}}(\mathcal{M}_S \mid \mathcal{U}^{MAX} \smallsetminus \{H^{MAX}\}, G_0) > \pi^0$$

Thus, there must exist a threshold value ρ' such that

$$E\pi_{H^{MAX}}(\mathcal{M}_S \mid \mathcal{U}^{MAX} \smallsetminus \{H^{MAX}\}, \rho) \begin{cases} < \pi^0 & \text{if } \rho > \rho' \\ \geq \pi^0 & \text{if } \rho \leq \rho' \end{cases}$$

i.e. for similarity of preferences lower than ρ', H^{MAX} does not belong to $\mathcal{U}^{MAX}$.

Similarly, there exists a threshold ρ'' such that firm $H^{MAX} - 1$ does not belong to $\mathcal{U}^{MAX}$ for $\rho \leq \rho''$. And so on. Thus, there must be a threshold value ρ^* such that there is no firm that belongs to $\mathcal{U}^{MAX}$ under $\rho \leq \rho^*$, but $\mathcal{U}^{MAX}$ is nonempty for similarity of preferences higher than ρ^*.

Values of ρ^* and ρ^{**} may be the same or different. But by definitions of $\mathcal{U}^{MIN}$ and $\mathcal{U}^{MAX}$ it is not possible that $\rho^* > \rho^{**}$.

Thus, Proposition 1 holds. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad\square$

References

[1] K. Atkins, Committee OKs new school assignment plan, Boston Herald, July 21, 2005.

[2] 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.

[3] J. Bulow and J. Levin (2006), Matching and Price Competition, American Economic Review, Vol. 96, No. 3, 652-668.

[4] E. Damiano, H. Li and W. Suen (2005), Unravelling of Dynamic Sorting, Review of Economic Studies, Vol. 72, 1057-1076.

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

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

[7] 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.

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

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

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

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

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

[13] M. Niederle and A.E. Roth (2003b), The Collapse of a Medical Clearinghouse (And Why Such Failures are Rare), NBER Working Paper No. W9467.

[14] M. Niederle and A.E. Roth (2004), Market Culture: How Norms Governing Exploding Offers Affect Market Performance, NBER working paper, February 2004.

[15] 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.

[16] 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, American Economic Review, Vol. 81, No. 3, 415-440.

[17] A.E. Roth (2003), *Matching and Allocation in Medicine and Health Care*, transcript of a talk at the National Academy of Engineering, June 16, 2003.

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

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

[20] 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.

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