Unravelling in Two-Sided Matching Markets and Similarity of Preferences

Hanna Hałaburda*

March 23, 2007

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 (which is the mechanism the literature focuses on), unravelling is more likely to occur for more similar preferences. Moreover, it 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.

1 Introduction

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

This situation, where contracting occurs long before the start of the job and before the relevant information is available, is called "unravelling". Such early matches often turn out

*Department of Economics, Northwestern University, hanna-h@northwestern.edu. I am grateful to my committee: William Rogerson, Kim-Sau Chung, Michael Whinston and Asher Wolinsky for their guidance and support throughout this research. I also thank David Goldreich, Evgeny Lyanders, Marcin Peski, Lukasz Pomorski, Balazs Szentes, and participants of the Applied Microeconomics Seminar and Students Theory Lunch Seminar at Northwestern University for discussions and comments. Financial support from the Center for the Study of Industrial Organization and from Weinberg College of Arts and Sciences is gratefully acknowledged.

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

to be inefficient when the job starts. This is because at the time of contracting, students often do not know what speciality they will want to pursue in two years, while employers do not know yet their needs, or the quality of the students. Unravelling has been recognized as a serious problem that affects many markets.[2] While measures designed to preclude it (e.g. centralized clearing houses or enforcement of uniform hiring dates) have been introduced in these markets, they have not always been successful. Unravelling still occurs in spite of such measures, for example, in the market for gastroenterologists, federal court clerks in the US, and lawyers in Canada (Roth and Xing, 1994; Avery, Jolls, Posner and Roth, 2001; Haruvy, Roth and Unver, 2006). At the same time, some other markets for entry-level professionals have never seemed to experience unravelling, for instance, markets for new professors in finance, economics or biology.

Despite extensive research in the economics literature, the causes and welfare consequences of this phenomenon are not fully understood. In particular, we have only limited understanding of why unravelling occurs in some markets but not in others. Furthermore, the basic question of how one might design an optimal matching mechanism when the potential for unravelling is taken into account remains largely unexplored. The goal of this paper is to investigate these two issues: the causes of unravelling, and the mechanism design problem in markets where unravelling is possible. It is shown that the similarity of preferences is a factor contributing to unravelling. Moreover, unravelling leads to a loss in welfare, and a Pareto-optimal mechanism cannot allow for unravelling.

To study unravelling, this paper considers a two-sided matching market populated by firms on one side and workers on the other. Agents on each side are heterogenous and have preferences over agents on the other side of the market. Their aim is to match with the best possible agent on the other side. Workers' preferences over firms are identical. The similarity of firms' preferences over workers is a comparative statics parameter. The two extreme cases are independent and identical preferences, although intermediate levels of similarity are also explored. There are two periods. Firms and workers can contract in either period, but firms only learn their preferences in the second period. Firms and workers who contract in the first period exit the market. The agents who remain in the second period participate in a mechanism that produces a matching between them. Unravelling in this model corresponds to contracting in the first period, before firms' preferences are known. Such early contracting takes place when a firm makes an offer in the first period, and this offer is accepted. This happens when contracting under uncertainty yields a higher expected payoff than the expected matching in the second period, for both the firm and the worker.

The first part of the paper investigates unravelling when the mechanism in the second period is assumed to produce the ex-post stable matching. A matching is ex-post stable if everyone prefers the match to being unmatched, and if there is no blocking pair, i.e. a worker and a firm that both strictly prefer each other than their currently matched partners. The second part of the paper allows for other mechanisms, to study the mechanism design problem in markets with potential unravelling.

The first part of the paper analyzes equilibria that arise for different levels of similarity in firms' preferences under the ex-post stable mechanism. The focus is on sequential equilibria

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

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

The second part of the paper turns to 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 when they make their decisions in the first period. The goal is to provide a Pareto-optimal outcome, from the ex-ante perspective. It turns out that any Pareto-optimal mechanism must preclude unravelling. This is because when a mechanism induces early contracting, then the expected payoff of the firms that unravel can be increased without changing any other agents' expected payoffs. Thus, a mechanism that allows for unravelling can be Pareto-improved. Furthermore, the ex-post stable mechanism – that is a mechanism that produces the ex-post stable matching among agents in the second period – is Pareto-optimal if and only if it does not induce unravelling. In some markets all Pareto-optimal mechanisms are ex-post unstable. The entire class of Pareto-optimal mechanisms is characterized for the case of identical preferences and a method for constructing a Pareto-optimal mechanism is described for cases where preferences are not identical.

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

The stability hypothesis is not the only explanation of unravelling in the literature. In Damiano, Li and Suen (2005) early contracting is the result of costly search. Li and Rosen (1998), Li and Suen (2000) and Suen (2000) point to workers' risk aversion as the main cause of the phenomenon. Although risk aversion plays an important role and may be an additional cause of early contracting, it is not a necessary condition for the phenomenon. The model

in this paper assumes risk-neutrality in order to separate incentives to unravel driven by similarity of preferences from incentives due risk-aversion.

The model predicts that unravelling occurs only in markets with substantial similarity of preferences. It may be argued that, indeed, in the markets where unravelling has been reported, employers have more similar preferences than in the markets that do not seem to unravel. For instance, on one hand medical and law students are evaluated based mainly on their grades, which are perceived by all potential employers in a similar way. In such a case, employers' preferences may be perceived as very similar. On the other hand, in disciplines such as finance, economics, or biology, students are assessed by their job market papers, which leaves more room for a subjective evaluation. This subjectivity may be one of the factors leading to differences in the way potential employers rank the candidates.

While the model specifically addresses the issue of professional job markets, it has the potential to explain unravelling in other situations, for instance the dynamics of the arrange marriage market. In the past, marriages were sometimes arranged when the parties were still in their childhood. It has been argued that such early marriages constitute unravelling (Roth and Xing, 1994). Nowadays this market no longer unravels: marriage decisions are made later on in life, with more information. This change is undoubtedly driven by many factors, one of which is likely more differentiated preferences over potential partners. In the past, the attractiveness of a potential spouse was primarily related to their wealth and social status, a signal that was easy to observe and valued in common by all interested parties. Given the high similarity of preferences, the market easily unravelled, as shown in this paper. Over time, characteristics other than wealth have become relatively more important and, consequently, differences in the way people gauge attractiveness have grown larger. It follows from the model that early marriages (unravelling) become less likely in such a situation, exactly as one observes.

The rest of the paper is organized as follows. Section 2 presents the model. Section 3 investigates unravelling under ex-post stable mechanism. Subsection 3.1 focuses on equilibria without unravelling, while subsection 3.2 explores the existence and characteristics of equilibria with unravelling. Section 4 analyzes the problem of mechanism design in markets where unravelling is possible. Section 5 concludes.

2 The Model

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

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

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

$\mathbf{u} \equiv [u_1, u_2, \ldots, u_F]$.

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

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

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

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

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

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

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

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

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

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

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

For any $\mathcal{F}$ and $\mathcal{W}$, $\boldsymbol{\mu}(\mathcal{F}, \mathcal{W})$ is the set of all possible matchings. Which of those is ex-post stable depends on firms' preferences, $\mathbf{R}^{\mathcal{F}}$. Lemma 1 below establishes that for a given preference profile there is a unique ex-post stable matching between $\mathcal{F}$ and $\mathcal{W}$.[5] It can be

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

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

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

characterized in the following way: The best firm in $\mathcal{F}$ is matched with its highest ranked worker in $\mathcal{W}$. Then, the next-best firm is matched with its highest ranked worker from among the remaining workers, etc. Every firm in $\mathcal{F}$ is matched to its highest ranked worker remaining in the pool after all the better firms in $\mathcal{F}$ have been matched.

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

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

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

Proof. The formula for the ex-post stable matching rule is obtained by the Gale-Shapley algorithm. The uniqueness follows from identical preferences of workers. Details of the proof are in the Appendix, page 29.

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

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

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

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

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

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

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

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

mechanism is incentive compatible. Therefore, the ex-post stable mechanism operating over $\mathcal{F}$ and $\mathcal{W}$ will produce the ex-post stable matching between $\mathcal{F}$ and $\mathcal{W}$.[6]

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Figure 1 illustrates how the game unfolds. Market characteristics $(F, W, \mathbf{u}, \mathbf{v}, \varrho, \mathcal{M})$ and the preferences of the workers' are commonly known all the time. At the beginning of period 1 firms simultaneously decide whether or not to make an early offer, and if so, to which worker. Every firm can make at most one offer. After the early offers are released, every worker observes the offers he has obtained, if any. He does not see offers made to other workers. Based on his beliefs about other agents' strategies, every worker presented with an offer accepts or rejects it. He may accept at most one offer. If an offer is accepted, the matched firm and worker leave the market. Firms whose offers were rejected or who did not make an offer in $t = 1$, stay in the market for $t = 2$. In period 2, firms' rankings are realized and a matching mechanism $\mathcal{M}$ operates over the agents remaining in the market at this time. The first part of the paper – Section 3 – assumes the ex-post stable mechanism in the $t = 2$. The second part – Section 4 – considers other mechanisms. There is no discounting between the periods and making offers is costless.

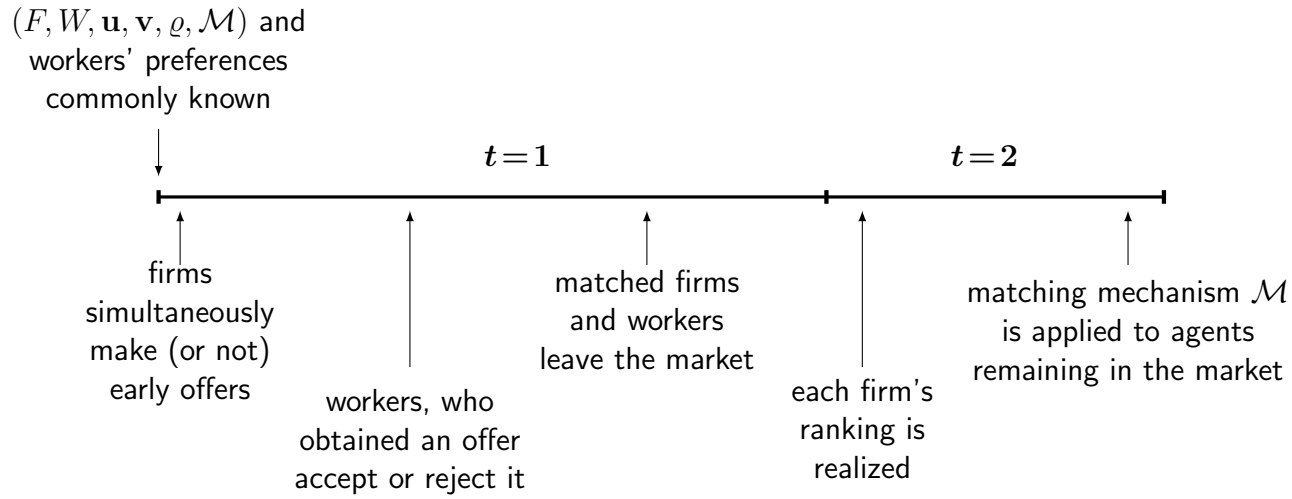


Figure 1: Timeline of the game

Under an incentive compatible mechanism, firms truthfully report their rankings in $t = 2$. Therefore, both the firms and the workers make their strategic decisions only in $t = 1$. First, every firm decides whether or not to make an offer and if so, to which worker. The analysis focuses on sequential equilibria in pure strategies, where a strategy of any firm f is $\sigma_f \in \{1, \ldots, W\} \cup \{\emptyset\}$. Since a worker can accept or reject an offer only if he has received it, a worker's strategy depends on the offers he has received. Let $\Omega_w \subset \{1, \ldots, F\}$ be the

set of firms that have made an offer to worker w in $t = 1$. Then the worker's strategy, $\sigma_w(\Omega_w) \in \Omega_w \cup \{\emptyset\}$, is the offer that he accepts. Strategy $\sigma_w(\Omega_w) = \emptyset$ means that the worker rejects all offers. Let vector $\boldsymbol{\sigma}$ be the strategy profile for all firms and workers.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$\forall\, f \in \{1, \ldots, F\} \setminus \Omega_w \quad \beta_w^*(\sigma_f|\Omega_w) = \begin{cases} 1 & for \quad \sigma_f = \sigma_f^* \\ 0 & otherwise \end{cases}$$

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

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

Offers made and accepted in period 1 constitute unravelling.

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

3 Unravelling under Ex-post Stable Mechanism

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

Given that the literature focuses on ex-post stable mechanisms, this section investigates unravelling under the ex-post stable matching mechanism. Subsection 3.1 focuses on equilibria without unravelling, while Subsection 3.2 describes equilibria when unravelling occurs.

The mechanism applied to the reported rankings in $t = 2$ is assumed in this section to be the ex-post stable one, $\mathcal{M}_S$. Mechanism $\mathcal{M}_S$ is not only incentive compatible, but in

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

all equilibria it produces the ex-post stable matching among the agents remaining in $t = 2$. Unless unravelling occurs in $t = 1$, it produces the ex-post stable outcome, $\mathfrak{o}_S$.

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

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

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

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

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

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

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

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

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

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

Proof. See the Appendix, page 29.

3.1 Equilibria without Unravelling

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

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

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

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

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 W and $\mathbf{u}$ the RHS 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 the offer of which will be accepted in $t = 1$. Let $\mathbb{L}_{(W,\mathbf{u})}$ denote this firm:

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

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

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

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

Incentives of firms for contracting in $t = 1$, before all the information is available, depend on the joint distribution of rankings, G_ϱ. The realization of rankings – together with the matching mechanism – determines the matching realized in $t = 2$. Firms' expected payoffs

depend on this expected matching. Recall that $E\pi_f(\mathfrak{o}_S|\varrho)$ denotes firm f's expected payoff from the ex-post stable outcome under G_ϱ.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Independent Preferences, G_0

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

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

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

Proof. See the Appendix, page 30.

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

Identical Preferences, G_1

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

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

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

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

The ex-post stable outcome is

$$\begin{aligned} \mathfrak{o}_S(f_3) = r_4^3 \quad &\Longrightarrow \quad \pi_3(\mathfrak{o}_S) = 4 \\ \mathfrak{o}_S(f_2) = r_3^2 \quad &\Longrightarrow \quad \pi_2(\mathfrak{o}_S) = 3 \\ \mathfrak{o}_S(f_1) = r_2^1 \quad &\Longrightarrow \quad \pi_1(\mathfrak{o}_S) = 2 \end{aligned}$$

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

A worker's expected utility from $t = 2$ *matching is* $\frac{1}{W}\sum_{f=1}^{F} u_f = \frac{15}{4} < 4 = u_1$. *It means that firm* 1*'s offer in* $t = 1$ *will be accepted by any worker. Thus,* $\mathcal{A}_{(4,\mathbf{u})} = \{1, 2, 3\}$. *Since* $\mathcal{O}_{(G_1,\mathbf{v})} \cap \mathcal{A}_{(4,\mathbf{u})} = \{1\} \neq \emptyset$, *there exists a profitable deviation from* $\mathfrak{o}_S$ *in this market.* ■

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

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

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

Intermediate Similarity of Firms' Preferences

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

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

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

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

Proof. See the Appendix, page 30.

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

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

3.2 Equilibria with Unravelling

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

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

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

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

To see why this is true, suppose, to the contrary, that in some equilibrium $\mathbb{L}^*$ and $\mathbb{H}^*$

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

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

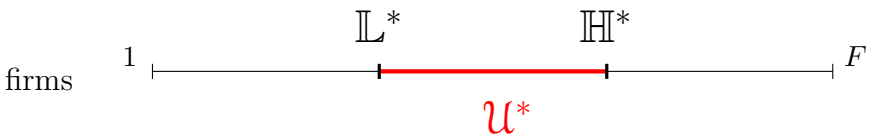


Figure 2: The structure of an equilibrium.

In any market there is at least one equilibrium. Consider a market $(F, W, \mathbf{u}, \mathbf{v}, \varrho, \mathcal{M}_S)$ where there exists a profitable deviation from $\mathfrak{o}_S$; that is, waiting for $\mathfrak{o}_S$ without unravelling is not an equilibrium. Then there is a set of firms $\mathcal{A}_{(W,\mathbf{u})} \cap \mathcal{O}_{(\varrho,\mathbf{v})} \neq \emptyset$ that would like to contract early and would be accepted in $t = 1$. But if those firms unravel, the expected payoff of staying to the second period decreases for better firms. This is because firms that unravel hire the workers, with a positive probability, that would be matched to the better firms in the ex-post stable outcome. When this happens, the better firms are matched with some worse workers in $t = 2$. This decrease in the expected payoff may induce some better firms to decide for early contracting, even though they initially preferred to wait for $\mathfrak{o}_S$. If more firms unravel, that also decreases expected payoff of workers in the second period. This may induce workers to accept firms in $t = 1$ that previously would not be accepted. This again may increase the number of firms that unravel. Eventually, either the process induces all the firms $1, \ldots, (F-1)$ to unravel,[11] or it reaches a "fixed" state earlier. In both cases the market reaches an equilibrium with nonempty unravelling set. Thus, in every market there is at least one equilibrium. This result is formally stated in Lemma 5 below. Moreover, in a typical markets there are more than one.

A market with multiple equilibria is presented in Example 4.

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

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

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

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

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

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

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

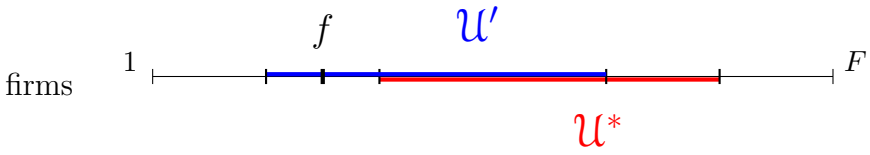
However, equilibrium unravelling sets cannot be arbitrary. For any two equilibrium unravelling sets in a given market, one needs to be fully included in the other. In particular, two equilibrium unravelling sets for the same market can not "overlap". To see why, suppose, to the contrary, that there exist two equilibrium unravelling sets $\mathcal{U}^*$ and $\mathcal{U}'$ as in Figure 3(a). There are two effects playing a role here – a "number effect" and a "position effect". The former one exists when $\mathcal{U}^*$ and $\mathcal{U}'$ are of different sizes. The latter follows from different "position" of unravelling sets among all firms. To consider only the "position effect" first, assume that $\mathcal{U}^*$ and $\mathcal{U}'$ are of the same size, but $\mathcal{U}^*$ includes better firms (on average) than $\mathcal{U}'$, as the figure shows. By the equilibrium condition for workers, firm f is not included in $\mathcal{U}^*$ because its early offer would not be accepted. But under $\mathcal{U}^*$ better firms unravel than under $\mathcal{U}'$. Thus, expected utility from staying in the market for $t = 2$ is lower for workers under $\mathcal{U}^*$ than under $\mathcal{U}'$. If under $\mathcal{U}'$ it was better for a worker to accept firm f in $t = 1$ than to wait for the expected utility in $t = 2$, it also must be so under $\mathcal{U}^*$. The "number effect" does not change this result.

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

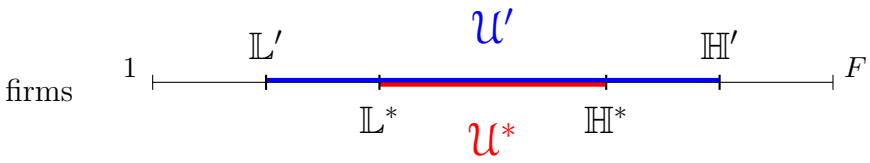
Lemma 5 *Given a market $(F, W, \mathbf{u}, \mathbf{v}, \varrho, \mathcal{M}_S)$:*

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

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



(a) An impossible configuration of multiple equilibrium unravelling sets



(b) A possible configuration of multiple equilibrium unravelling sets

Figure 3: Multiple equilibria with unravelling.

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

$$\mathbb{L}' < \mathbb{L}^* \iff \mathbb{H}^* < \mathbb{H}'$$

Proof. See the Appendix, page 30.

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

Comparative statics on ϱ

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

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

firm $(f-\eta)$ gets in $\mathfrak{o}_S$ a higher payoff than π^0. This leads to a contradiction. Thus, under independent preferences the unique equilibrium outcome is "no unravelling".

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

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

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

As ϱ increases, the maximum and minimum equilibria are more likely to be distinct. The maximum equilibrium unravelling set increases from the empty set to a non-empty one for lower ϱ than the minimum equilibrium unravelling set does. As the maximum equilibrium unravelling set increases, an equilibrium with unravelling appears in the market. Moreover, when the similarity of preferences increases, the minimum equilibrium unravelling set may also increase from the empty set to a non-empty one. When this occurs, "no unravelling" is no longer an equilibrium for high ϱ's in this market. This relation between equilibrium unravelling sets in a market, and the level of preference similarity is illustrated by Figure 4.

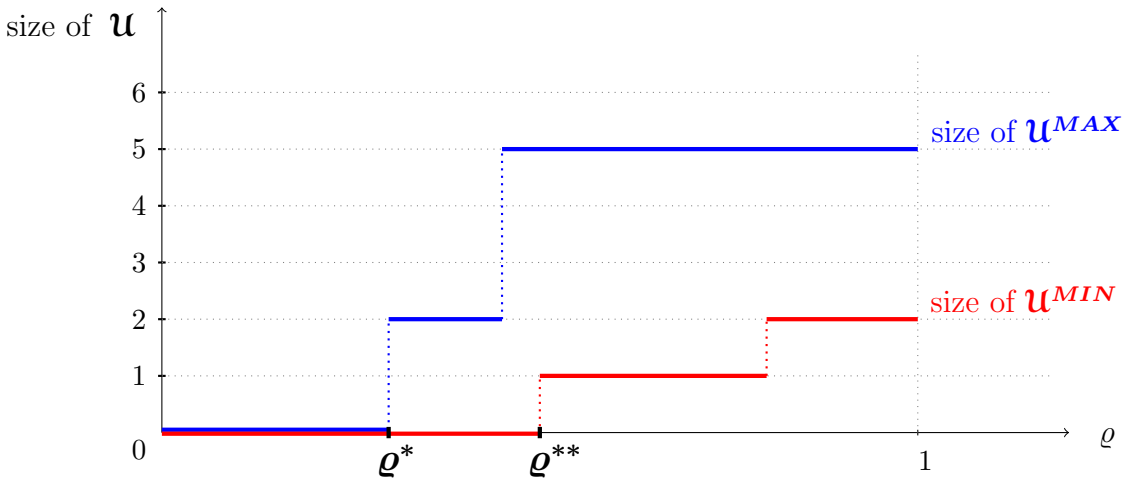


Figure 4: Relation between $\mathcal{U}^{MIN}$, $\mathcal{U}^{MAX}$ and ϱ in a typical market.

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

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

Proof. See the Appendix, page 34.

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

Discussion

In some markets that struggled with unravelling, firms called upon an institution to solve the problem.[12] It is often recommended that the institutions should apply the ex-post stable mechanism once the preferences are known. Roth (1991) and Kagel and Roth (2000), for example, argue that the ex-post stable mechanism prevents unravelling. It has, indeed, proven successful in stoping unravelling in many markets (e.g. markets for medical resident interns in the US and in the UK). However, in some markets the ex-post stable mechanism has failed to stop unravelling. For instance, in the Canadian lawyer market and in the gastroenterology market in the US, agents contracted as early as a year before information was available and the mechanism was applied. In the latter case, the clearinghouse has been subsequently abandoned in 1996, as there were too few participants waiting for its operation.

The present model provides an explanation for why the ex-post stable matching mechanism failed to stop unravelling in some markets. In some markets with high similarity of preferences all equilibria under the ex-post stable mechanism involve early contracting. In those markets only an ex-post *unstable* matching mechanism can prevent unravelling.

4 Mechanism Design

The previous section investigated issue of unravelling when the mechanism operating in $t = 2$ is restricted to be the ex-post stable one.

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

This section explores the problem of mechanism design in markets where unravelling may occur. A social planner chooses a mechanism for the second period. It is assumed that agents cannot renege on the matching produced by the mechanism in $t = 2$, but they can contract in $t = 1$. The mechanism is announced at the onset of the game. Firms decide about their early offers, and workers decide whether to accept such offers, knowing what mechanism will be in operation the next period. All agents that did not contract in $t = 1$ participate in the mechanism in $t = 2$. The goal of the social planner is to provide a Pareto-optimal outcome, from the ex-ante perspective.

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 $\mathfrak{o}$ is

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

The previous section considered a special case of an outcome function – the ex-post stable outcome, $\mathfrak{o}_S$.

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

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

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

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

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

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

and

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

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

An outcome is be said to be *complete* if it assigns every firms to a worker with probability 1. An outcome is be said to be *anonymous* if the assignments of firms to workers are based only on firms' rankings but not an workers' identities.[13] As will be explained below, this section restricts itself to considering mechanisms that induce complete anonymous

[13] An example of an anonymous match is when firm F is matched with its most preferred worker, r_W^F, no matter what is the worker's identity. An example of a not-anonymous match is when firm 1 is matched with worker 1, no matter what are the firms' preferences.

outcomes. It is straightforward to see that the expected utility of every worker is the same under any complete anonymous outcome and is given by

$$EU_w(\mathfrak{o}) = \frac{1}{W}\sum_{f=1}^{F} u_f$$

Notice that under an anonymous complete outcome, $\mathfrak{o}$, the sum of expected utilities of all workers is

$$\sum_{w=1}^{W} EU_w(\mathfrak{o}) = \sum_{f=1}^{F} u_f$$

Thus, it is not possible for any outcome $\mathfrak{o}'$ (anonymous or not) to strictly increases some workers' expected utility without making some other workers' worse off. That is, an not-anonymous outcome can not Pareto-dominate an anonymous complete outcome.

Therefore, the definition of strict Pareto-dominance takes on a simpler form when only complete anonymous outcomes are considered. A matching outcome $\mathfrak{o}'$ *strictly Pareto-dominates* outcome $\mathfrak{o}''$ when

$$\forall\ f \quad E\pi_f(\mathfrak{o}') \geq E\pi_f(\mathfrak{o}'')$$

and

$$\exists\ f \quad E\pi_f(\mathfrak{o}') > E\pi_f(\mathfrak{o}'')$$

And a (complete anonymous) outcome $\mathfrak{o}$ is Pareto-optimal if there is no other (complete, anonymous) outcome that Pareto-dominates $\mathfrak{o}$ in the sense of the simpler definition.

Recall that $\boldsymbol{\sigma}$ denotes vector of strategies in $t = 1$ for all agents. For any mechanism $\mathcal{M}$, let $\boldsymbol{\Sigma}^{\mathcal{M}}$ denote the set of all equilibrium strategies for $\mathcal{M}$. A pair $(\mathcal{M}, \boldsymbol{\sigma})$, where $\boldsymbol{\sigma} \in \boldsymbol{\Sigma}^{\mathcal{M}}$ is called a *mechanism-equilibrium* pair. This section only considers mechanisms that are incentive compatible and will assume that all firms truthfully report their rankings in the second period.[14] Therefore, a mechanism-equilibrium pair $(\mathcal{M}, \boldsymbol{\sigma})$ uniquely determines an outcome $\mathfrak{o}_{(\mathcal{M},\boldsymbol{\sigma})}$.

As for outcomes, define a mechanism to be *complete* if it assigns a worker to every firm that participates in the mechanism. Similarly, define a mechanism to be *anonymous* if it produces a matching based only on firms' reported rankings and ignoring workers' identities. Clearly, the outcome induced by a mechanism-equilibrium pair is complete and anonymous if the mechanism is complete and anonymous, and the vector of strategies is anonymous. This paper restricts itself to considering mechanisms-equilibrium pairs $(\mathcal{M}, \boldsymbol{\sigma})$ where $\mathcal{M}$ is complete and anonymous and $\boldsymbol{\sigma}$ is anonymous.

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

[14] By the revelation principle, this is without loss of generality.

$\mathfrak{O}_{(\mathcal{M},\boldsymbol{\sigma})}$. Clearly, any unconstrained Pareto-optimal pair $(\mathcal{M},\boldsymbol{\sigma})$ is also Pareto-optimal in the constrained sense.

The following proposition presents the main result of this section. It shows that if a mechanism-equilibrium pair $(\mathcal{M},\boldsymbol{\sigma})$ exhibits unravelling, it cannot be constrained Pareto-optimal.

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

Proof. Suppose, to the contrary, that $\mathcal{M}$ produces in equilibrium $\boldsymbol{\sigma}$ a non-empty unravelling set $\mathcal{U}^{\mathcal{M}} \neq \emptyset$, and that $(\mathcal{M},\boldsymbol{\sigma})$ is constrained Pareto-optimal. Then consider the following mechanism $\mathcal{M}'$:

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

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

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

There is an equilibrium without unravelling under $\mathcal{M}'$. This is because all firms in $\mathcal{U}^{\mathcal{M}}$ prefer to wait for $\mathcal{M}'$ than to unravel given that other firms wait for $t=2$. Since firms outside $\mathcal{U}^{\mathcal{M}}$ did not unravel when some firms were contracting early – either because they preferred not to, or because they would not be accepted in $t=1$ – they do not unravel when all other firms wait for $t=2$. For firms that preferred not to unravel under $(\mathcal{M},\boldsymbol{\sigma})$ the value of waiting increased, as there are more workers in $t=2$ when no firm unravels. Also for the workers, the value of waiting is higher when no firm unravels. So, the firms that were not accepted under $(\mathcal{M},\boldsymbol{\sigma})$ are also not accepted under $\mathcal{M}'$. Therefore, no unravelling occurs. Denote the equilibrium without unravelling by $\boldsymbol{\sigma}'$.

Every firm in $\mathcal{U}^{\mathcal{M}}$ has a strictly higher expected payoff in $\mathfrak{O}_{(\mathcal{M}',\boldsymbol{\sigma}')}$ then in $\mathfrak{O}_{(\mathcal{M},\boldsymbol{\sigma})}$. All the other firms have exactly the same expected payoff in both outcomes. Therefore, $\mathfrak{O}_{(\mathcal{M}',\boldsymbol{\sigma}')}$ Pareto-dominates $\mathfrak{O}_{(\mathcal{M},\boldsymbol{\sigma})}$, and thus $(\mathcal{M},\boldsymbol{\sigma})$ could not have been Pareto-optimal.□

Proposition 2 establishes a necessary condition for constrained Pareto-optimality of $(\mathcal{M},\boldsymbol{\sigma})$. Given this condition, the following corollary shows that the set of constrained Pareto-optimal and the set of unconstrained Pareto-optimal mechanism-equilibrium pairs is the same.

Corollary 1 *A mechanism-equilibrium pair is constrained Pareto-optimal if and only if it is unconstrained Pareto-optimal.*

Proof. It follows form the definitions that every unconstrained Pareto-optimal $(\mathcal{M}, \boldsymbol{\sigma})$ is also constrained Pareto-optimal.

For the proof in the opposite direction, suppose to the contrary that $(\mathcal{M}, \boldsymbol{\sigma})$ is constrained Pareto-optimal, but it is not unconstrained Pareto-optimal. That is, $\mathfrak{O}_{(\mathcal{M},\boldsymbol{\sigma})}$ is not Pareto-optimal. Then there must exists an outcome $\mathfrak{O}'$ that Pareto-dominates $\mathfrak{O}_{(\mathcal{M},\boldsymbol{\sigma})}$, i.e.

$$\forall\, f \quad E\pi_f(\mathfrak{O}') \geq E\pi_f(\mathfrak{O}_{(\mathcal{M},\boldsymbol{\sigma})})$$

and

$$\exists\, f \quad E\pi_f(\mathfrak{O}') > E\pi_f(\mathfrak{O}_{(\mathcal{M},\boldsymbol{\sigma})})$$

Now, consider a mechanism $\mathcal{M}'$ that produces $\mathfrak{O}'$ when all agents participate. Notice that since $(\mathcal{M}, \boldsymbol{\sigma})$ is constrained Pareto-optimal, then by Proposition 2, the pair does not exhibit unravelling. Then, there exists an equilibrium without unravelling under $\mathcal{M}'$, denoted by $\boldsymbol{\sigma}^0 \in \boldsymbol{\Sigma}^{\mathcal{M}'}$. This is because the mechanism is not changing the acceptance set. So, the firms that would not be accepted under $\mathcal{M}$ would not be accepted under $\mathcal{M}'$, either. And the firms that did not want to contract early under $\mathcal{M}$, do not want to contract under $\mathcal{M}'$ as they are as well or better off under $\mathcal{M}'$. Thus, the outcome $\mathfrak{O}_{(\mathcal{M}',\boldsymbol{\sigma}^0)} = \mathfrak{O}'$ Pareto-dominates the outcome $\mathfrak{O}_{(\mathcal{M},\boldsymbol{\sigma})}$. Therefore, $(\mathcal{M}, \boldsymbol{\sigma})$ cannot be constrained Pareto-optimal. □

Since in this model the two notions of Pareto-optimality are equivalent, the remainder of the paper does not distinguish between them and uses a common term "Pareto-optimal".

Notice that a mechanism in a Pareto-optimal pair $(\mathcal{M}, \boldsymbol{\sigma})$ does not need to be ex-post stable. In particular, when the ex-post stable mechanism unravels, it can not be Pareto-optimal. Moreover, any $(\mathcal{M}_S, \boldsymbol{\sigma})$ that does not exhibit unravelling is Pareto-optimal. It has been already established in the literature that the ex-post stable outcome is always Pareto-optimal.[15] When the ex-post stable mechanism does not unravel, it produces the ex-post stable outcome and, thus, it is Pareto-optimal.

Corollary 2 *A mechanism-equilibrium pair with the ex-post stable mechanism,* $(\mathcal{M}_S, \boldsymbol{\sigma})$ *is Pareto-optimal if and only if there is no unravelling in* $\boldsymbol{\sigma}$*.*

Proposition 2 provides a necessary condition for Pareto-optimality. It is that $(\mathcal{M}, \boldsymbol{\sigma})$ can not exhibit unravelling. However, it is not a sufficient condition. In order to characterize the set of all Pareto-optimal mechanism-equilibrium pairs, one needs to identify the necessary and sufficient conditions.

Suppose that all firms wait for $t = 2$, and then outcome $\mathfrak{O}$ is produced. It is said that, in a given market, there exist *a profitable deviation from* $\mathfrak{O}$ when there is a firm and a worker that would prefer to contract in $t = 1$, given that all the other agents wait for the second period.[16] Let $\mathcal{M}^{\mathfrak{O}}$ be any mechanism that produces outcome $\mathfrak{O}$ whenever all agents participate in the mechanism. If there does not exist a profitable deviation from $\mathfrak{O}$ in a market, then there is an equilibrium without unravelling under mechanism $\mathcal{M}^{\mathfrak{O}}$. Denote this equilibrium by $\boldsymbol{\sigma}^0 \in \boldsymbol{\Sigma}^{\mathcal{M}^{\mathfrak{O}}}$.

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

[16] Section 3.1 investigated a profitable deviation from $\mathfrak{O}_S$. A similar analysis can be conducted for any outcome.

For a given market, let $\mathfrak{O}^*$ be the set of outcomes such that any $\mathfrak{o} \in \mathfrak{O}^*$ is Pareto-optimal and there does not exists a profitable deviation from $\mathfrak{o}$. Then, the following corollary characterizes the set of all Pareto-optimal mechanism-equilibrium pairs in a market.

Corollary 3 *A mechanism-equilibrium pair $(\mathcal{M}, \boldsymbol{\sigma})$ is Pareto-optimal in a market if and only if*

(1) $\mathcal{M} \equiv \mathcal{M}^{\mathfrak{o}}$ for some $\mathfrak{o} \in \mathfrak{O}^*$, i.e., whenever all agents participate in $\mathcal{M}$, it produces outcome $\mathfrak{o} \in \mathfrak{O}^*$;

(2) $\boldsymbol{\sigma} \equiv \boldsymbol{\sigma}^0 \in \boldsymbol{\Sigma}^{\mathcal{M}^{\mathfrak{o}}}$, i.e., $\boldsymbol{\sigma}$ is an equilibrium without unravelling.

For the special case of identical preferences, the set of Pareto-optimal outcomes can be given more precisely. In this special case, all firms' rankings are the same. An outcome is Pareto-optimal under identical preferences if and only if it assigns every firm to one of the top F workers with probability 1. To see why any such outcome is Pareto-optimal, notice that the sum of all firms expected payoffs in any such outcome, $\mathfrak{o}$, is

$$\sum_{f=1}^{F} E\pi_f(\mathfrak{o}) = \sum_{f=1}^{F} v_{W-F+1}$$

It is not possible to achieve a higher sum of payoffs under identical preferences, as at most one firm can be matched to each worker. Therefore, it is not possible to increase any firms expected payoff without decreasing the expected payoff of some other firm.

To see why only an outcome that assigns the top F workers to the firms with probability 1 is Pareto-optimal, consider an outcome that matches a firm with a worker outside the top F with a positive probability. Then there is a worker from the top F workers who remains unmatched. Another outcome that in such a case matches that firm with this worker (with no other changes) Pareto-dominates the outcome where the worker remains unmatched.

Corollary 4 *Under identical preferences, $(\mathcal{M}, \boldsymbol{\sigma})$ is Pareto-optimal if and only if*

(1) there is no unravelling in the equilibrium $\boldsymbol{\sigma} \in \boldsymbol{\Sigma}^{\mathcal{M}}$, and

(2) the mechanism assigns top F workers to the F firms in the market with probability 1, given that there is no unravelling.

There does not seem to be an equally simple way to describe the set of all Pareto-optimal outcomes, and therefore the set of all mechanism-equilibrium pairs, for arbitrary preferences. However, the following proposition guarantees that in any market there exists a Pareto-optimal mechanism-equilibrium pair.

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

Proof. See the Appendix, page 35.

This result is proven by construction. If in a market there exists an equilibrium without unravelling under the ex-post stable mechanism, then the statement of the proposition is clearly satisfied. A Pareto-optimal mechanism is constructed for the markets where all equilibria under $\mathcal{M}_S$ unravel, i.e. $\mathcal{U}^{MIN} \neq \emptyset$. Below, the construction for the case of identical preferences is presented. The construction for arbitrary similarity of preferences is delegated to the Appendix.

Under identical preferences, firm f's k-th ranked worker is any firm's k-th ranked worker. Consider the following mechanism $\mathcal{M}^{\mathcal{A}}$:

(1) Every firm $f \in \mathcal{U}^{MIN}$ is (tentatively) matched with a random worker. Some firms get workers worse than r_{W-F+1} with a positive probability.

(2) All other firms $f \in \{1, \ldots, F\} \setminus \mathcal{U}^{MIN}$ are matched according to the ex-post stable matching rule over the remaining workers.

(3) For every firm assigned to a worker worse than r_{W-F+1}, there is an unmatched worker better than r_{W-F}. These firms are matched with such workers at random.

Firstly, notice that there is an equilibrium without unravelling under $\mathcal{M}^{\mathcal{A}}$. All firms in $\mathcal{U}^{MIN}$ get a strictly higher expected payoff under $\mathcal{M}^{\mathcal{A}}$ than in early contracting. This precludes such firms from unravelling. All other firms get exactly the same expected payoff, and their incentives did not change. Those firms that did not want to contract early under $\mathcal{M}_S$, do not want to unravel under $\mathcal{M}^{\mathcal{A}}$ either. Firms that wanted to unravel under $\mathcal{M}_S$ but would not be accepted, would not be accepted under $\mathcal{M}^{\mathcal{A}}$ either. This is because waiting yields higher expected utility for workers under $\mathcal{M}^{\mathcal{A}}$ as more firms participate in the mechanism.

5 Conclusions

This paper has investigated the causes and welfare consequences of unravelling in two-sided matching markets by considering a two period model where firms learn information about their preferences over workers at the beginning of the second period. It is assumed that firms and workers have the ability to make and accept offers in the first period if they wish to, and that a clearinghouse mechanism is used in the second period to assign remaining firms to workers. Unravelling is said to occur when offers are both made and accepted in the first period.

The first part of the paper explores the issue of unravelling given that the ex-post stable mechanism operates in the second period, which is the clearinghouse mechanism that most of the existing literature focuses on. It shows that unravelling becomes more likely as firms' preferences over workers grow more similar. This is because when preferences of firms are very similar, worse firms can be matched with their most preferred worker only if they contract with them early.

The second part of the paper considers the mechanism design problem of choosing a Pareto-optimal mechanism, given that firms and workers can choose to contract in the first period if they wish to. The main result shows that a necessary condition for a mechanism to

be Pareto-optimal is that it does not induce unravelling. Thus, without loss of generality, a social planner can restrict himself to considering mechanisms that do not induce unravelling. It is also shown that the ex-post stable mechanism is Pareto-optimal if and only if it does not unravel. Finally, the paper characterizes the set of all Pareto optimal mechanisms for the case of identical preferences and shows how to construct a Pareto optimal mechanism for cases where preferences are not identical.

Appendix

Proof of Lemma 1 (page 6)

Proof. The proof follows in two steps: (i) $\mu_S^{\mathcal{F},\mathcal{W}}$ is a ex-post stable matching over $\mathcal{F}$ and $\mathcal{W}$, (ii) there is no other ex-post stable matching.

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

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

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

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

$$\exists\, f \in \mathcal{F} \text{ s.t. } \mu'(f) \not\equiv \mu_S^{\mathcal{F},\mathcal{W}}(f) \tag{3}$$

But it can not be true for the best firm in $\mathcal{F}$. If μ' does not match the best firm with its most desirable worker in $\mathcal{W}$, then this firm and this worker form a blocking pair. Also, (3) can not be true for the next best firm. If μ' does not match the next best firm with its most desirable worker available after the best firm got matched, this firm and this worker form a blocking pair. Similarly for any $f \in \mathcal{F}$. Thus, if μ' is a ex-post stable matching, it must be the same as $\mu_S^{\mathcal{F},\mathcal{W}}$, i.e. $\forall\, f \in \mathcal{F}, \mu'(f) = \mu_S^{\mathcal{F},\mathcal{W}}(f)$. □

Proof of Lemma 2 (page 11)

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

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

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

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

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

Let $\varrho' > \varrho$, then

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

This completes the proof.□

Proof of Lemma 3 (page 14)

Proof. Consider the worst firm, firm 1.

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

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

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

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

Proof of Lemma 4 (page 15)

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

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

Proof of Lemma 5 (page 18)

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

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

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

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

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

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

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

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

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

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

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

Hence, the contradiction. □

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

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

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

The worst firm that would be accepted in $t=1$, given $\mathbb{H}^*$, is characterized by two inequalities:[17]

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

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

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

$$E\pi_{\mathbb{H}^*}\big(\mathcal{M}_S|\mathcal{U}(\mathbb{L}^*,\mathbb{H}^*-1)\big) < \pi^0$$
$$\text{and} \qquad (\mathbf{CF})$$
$$E\pi_{\mathbb{H}^*+1}\big(\mathcal{M}_S|\mathcal{U}(\mathbb{L}^*,\mathbb{H}^*)\big) \ge \pi^0$$

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

The first part of equilibrium condition (**CF**) says that firm $\mathbb{H}^*$ prefers to contract early, given that firms $\mathbb{L}^*,\ldots,(\mathbb{H}^*-1)$ do. The second part of the condition assures that $\mathbb{H}^*$ is the highest

[17] More precisely, the first inequality is

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

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

ranked firm that wants to contract early. That is, better firm $(\mathbb{H}^*+1)$, prefers to wait for $t=2$, given that firms $\mathbb{L}^*,\ldots,\mathbb{H}^*$ unravel. As Lemma 6 indicates, this also means that – given $\mathbb{L}^*$ – all firms worse than $\mathbb{H}^*$ prefer to contract early, and all firms better than $\mathbb{H}^*$ prefer to wait.

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

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

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

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

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

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

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

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

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

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

and

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

as well as

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

and

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

That is

$$u_{\mathbb{L}'-1}<EU(t=2\,|\,\mathcal{U}(\mathbb{L}',\mathbb{H}^*))<u_{\mathbb{L}'}\leq u_{\mathbb{L}^*-1}<EU(t=2\,|\,\mathcal{U}(\mathbb{L}^*,\mathbb{H}^*))<u_{\mathbb{L}^*} \qquad (4)$$

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

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

So

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

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

From the equilibrium condition for $\mathcal{U}^*$, $EU(t=2\,|\,\mathcal{U}(\mathbb{L}^*,\mathbb{H}^*)) > u_{\mathbb{L}^*-1}$. Thus all terms in the sum on the RHS are positive. Therefore, $EU(t=2\,|\,\mathcal{U}(\mathbb{L}',\mathbb{H}^*)) > EU(t=2\,|\,\mathcal{U}(\mathbb{L}^*,\mathbb{H}^*))$. But from (4), $EU(t=2\,|\,\mathcal{U}(\mathbb{L}',\mathbb{H}^*)) < EU(t=2\,|\,\mathcal{U}(\mathbb{L}^*,\mathbb{H}^*))$. Thus, a contradiction. □

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

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

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

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

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

Notice that there exists an equilibrium without unravelling, $\mathcal{U} = \emptyset$, if and only if $\mathcal{A}_{(W,\mathbf{u})} \cap \mathcal{O}_{(\varrho,\mathbf{v})} = \emptyset$.

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

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

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

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

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

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

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

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

For $\Delta\mathbb{L} < \Delta\mathbb{H}$. Since $\mathbb{L}^* > \mathbb{L}'$,

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

So

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

which cannot be true, given $(\sum_{f=\mathbb{H}'}^{\mathbb{H}^*-1} u_f)/\Delta\mathbb{H} > (\sum_{\mathbb{L}'+1}^{\mathbb{L}^*} u_f)/\Delta\mathbb{L}$ and $\Delta\mathbb{H} > \Delta\mathbb{L}$. Hence, a contradiction. This completes the proof of part (3) of Lemma 5. □

Proof of Proposition 1 (page 20)

Proof.

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

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

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

Proof of Proposition 3 (page 26)

Proof. The ex-post stable mechanism is a Pareto-optimal mechanism when it does not unravel. Thus, consider a market where every equilibrium under $\mathcal{M}_S$ unravels, i.e. $\mathcal{U}^{MIN} \neq \emptyset$.

For such cases, consider following mechanism $\mathcal{M}^{\mathcal{A}}$:

(1) All firms $f \in \mathcal{U}^{MIN}$ draw a random number out of $\{1, \ldots, W\}$.

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

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

This mechanism is incentive compatible. Moreover, there exists an equilibrium without unravelling under $\mathcal{M}^{\mathcal{A}}$. Denote this equilibrium by $\boldsymbol{\sigma}^{\mathcal{A}}$. The mechanism-equilibrium pari $(\mathcal{M}^{\mathcal{A}}, \boldsymbol{\sigma}^{\mathcal{A}})$ is Pareto-optimal. □

References

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

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

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

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

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

[6] 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.

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

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

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

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

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

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

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

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

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