



**HAL**  
open science

## Two-Sided Matching with (almost) One-Sided Preferences

Guillaume Haeringer, Vincent Iehlé Iehlé

► **To cite this version:**

Guillaume Haeringer, Vincent Iehlé Iehlé. Two-Sided Matching with (almost) One-Sided Preferences. *American Economic Journal: Microeconomics*, 2019, 11 (3), p.155-90. 10.1257/mic.20170115 . halshs-01513384v2

**HAL Id: halshs-01513384**

**<https://shs.hal.science/halshs-01513384v2>**

Submitted on 9 Nov 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Two-Sided Matching with (almost) One-Sided Preferences\*

Guillaume Haeringer<sup>†</sup>      Vincent Iehlé<sup>‡</sup>

September 28, 2018

**Abstract:** In a two-sided matching context we show how we can predict stable matchings by considering only one side’s preferences and the mutually acceptable pairs of agents. Our methodology consists of identifying impossible matches, i.e., pairs of agents that can never be matched together in a stable matching of any problem consistent with the partial data. We analyze data from the French academic job market for mathematicians and show that the match of about 45% of positions (and about 60% of candidates) does not depend on the preferences of the hired candidates, unobserved and submitted at the final stage of the market.

---

\*We deeply thank three anonymous referees for their insightful comments and suggestions. We also thank for their comments Atila Abdulkadiroğlu, Itai Ashlagi, Francis Bloch, Françoise Forges, Nicole Immorlica, Mar Reguant, Jacob Leshno, Vikram Manjunath, Thayer Morrill, Juan Pereyra, Baharak Rastegari, Al Roth, Daniela Saban, Jay Sethuraman, Tayfun Sönmez, Olivier Tercieux, Peter Troyan, Utku Ünver and Alexander Westkamp, and seminar audiences at Boston College, UC Davis, UCSD, Microsoft Research Cambridge, Rochester, ITAM, UCSB, Berkeley, Duke, Oslo, USC, Paris School of Economics, Vanderbilt, Bielefeld, Salerno, Barcelona, Dauphine, Caen, York and the NYC market design group for helpful comments. The first author acknowledges the support of the Barcelona GSE Research Network and of the Government of Catalonia, Spanish Ministry of Science and Innovation through the grant “Consolidated Group-C” ECO2008-04756 and the *Generalitat de Catalunya* (SGR2009-419). Part of this research was made during the first author’s visit to the Stanford Economics department, whose hospitality is gratefully acknowledged. The second author gratefully acknowledges the support of ANR-13-BSH1-0010 - DynaMITE - Appariements et Interactions Dynamiques : Théorie et Expériences (2013) and Risk Foundation (Chaire Dauphine-ENSAE-Groupama “Les particuliers face au risque”).

<sup>†</sup>Haeringer: Baruch College, Guillaume.Haeringer@baruch.cuny.edu.

<sup>‡</sup>Iehlé: Normandie Univ, UNIROUEN, CREAM, vincent.iehle@univ-rouen.fr.

*JEL classification:* C78, J41, C81

*Keywords:* *Stable matchings, Partial matching data, Hall's marriage theorem, French academic job market*

## 1. Introduction

This paper is aimed at further understanding how preferences in two-sided matching markets can shape the set of stable matchings. Stability in a matching problems is a conjunction of two requirements: no agent is matched to an unacceptable partner (*individual rationality*), and no pair of agents would prefer to be matched to one another rather than to their current mate (*absence of blocking pairs*). As the second requirement shows, stability hinges on the preferences of *both* sides of the market. In this paper we investigate how information about stable matchings can be retrieved even if we only consider the preferences of one side of the market once we know the set of mutually acceptable pairs.

To this end we consider *job market problems*, that are otherwise standard many-to-one matching problems between candidates and academic departments. In our model departments have one or several job openings (positions) to fill and are assumed to have responsive preferences over candidates.<sup>1</sup> For most of the paper we depart from this model by considering broader environments that we call *proto-matching problems*. Such problems consist of a list of mutually acceptable candidate-department pairs, and departments' preferences over candidates. Candidates' strict preferences over departments, however, are left unspecified. The main question we address is, when considering a proto-matching problem, whether we can identify pairs of agents that can never be matched together in a stable matching of any job market problem consistent with the partial data. If there does not exist *some* candidates' preference profile so that a particular candidate-department pair can be matched at *some* stable matching (for that profile), the candidate is called an *impossible match* for the department.

Our work adds to the part of the matching literature that addresses the question of extracting information from partial matching data. One of the earliest results in that line of work is the celebrated *Rural Hospital Theorem* (Roth, 1986), which allows to extract some information about *all* stable matchings from the observation of *one* stable matching.<sup>2</sup> Echenique, Lee and Shum (2013) go one step further and show how the observation of

---

<sup>1</sup>See Roth and Sotomayor (1989).

<sup>2</sup>Another celebrated result is that in marriage problems, the set of agents matched at a stable matching

stable matchings can be used to estimate agents' preferences for aggregate matchings.<sup>3</sup> Behind these results is the idea to extract information on preferences or properties about stable matchings from the observation of some stable matchings. Our contribution takes the opposite approach, as we aim at deducing information on (un)stable matchings from the partial observation of preferences. The two closest works to this paper are Rastegari *et al.* (2012), and Saban and Sethuraman (2015). Rastegari and her coauthors consider a one-to-one matching market where agents from both sides do not know their preferences and must conduct interviews to uncover them. They show that as preferences become more complete, some potential interviews may be discarded because the parties involved in such interviews can never be matched at any stable matching, for any possible completion of preferences.<sup>4</sup> Saban and Sethuraman's paper is even closer to ours, but contrary to us they focus on efficient assignments achieved by the random serial dictatorship. Like us, they investigate, upon observing agents' preferences, whether there exists an ordering of agents such that an agent will be assigned a specific object.

There are a number of environments that correspond to proto-matching problems. The first situation is when a matching has to be determined when individuals have not learned their preferences yet. For instance, this is the case for the assignment of refugees to landlords in Sweden, studied by Andersson and Ehlers (2016). In their model, at the time a matching has to be determined, the only information that is available regarding the preferences of a family of refugees is the set of acceptable landlords. Landlords, in contrast, have more precise preferences (that depend on the size of the refugee family), which can be elicited. A refugee problem is thus a proto-matching problem where landlords and refugees play the role of departments and candidates, respectively. A second class of situations that correspond to proto-matching problems (albeit similar to the case of refugees) is when individuals participating in a matching or assignment mechanism can only report dichotomous preferences. This is the case with the newly introduced college admission procedure in France, where students can only report a set of acceptable

---

is the same for all stable matchings (Roth (1984) and Gale and Sotomayor (1985)). In a similar vein Roth and Sotomayor (1989) show how the observation of different stable matchings for a college admission problem gives some information about colleges' preferences over group of students.

<sup>3</sup>See also Echenique, Lee, Shum and Yenmez (2013) and Agarwal (2015).

<sup>4</sup>See also Martínez *et al.* (2012) who, starting from a profile of incomplete preferences, identify all the possible completions thereof that give the same set of stable matchings.

programs, but each academic program has to report a strict ranking of students. Proposals by programs and students’ acceptance or rejection decisions follow a decentralized procedure.<sup>5</sup> Identifying impossible matches can thus be useful in such situations. For instance, in the case of refugees it is reasonable to assume that, once an assignment has been implemented, refugees become better able to distinguish between landlords and thus have more accurate preferences over their potential hosts. Avoiding matching impossible matches can reduce the number of blocking pairs (and rematches) that may arise once individuals learn their preferences. A third type of situation that proto-matching problems capture is having access to partial matching data. This is the case of the French academic job market for instance (which we comment in more details later in this Introduction). Unlike the Swedish case, agents on both sides of the market have strict preferences, but from the economist’s perspective it does correspond to a proto-matching because only the departments’ preferences are observable.

Our first contribution consists of characterizing impossible matches in proto-matching problems. In a proto-matching problem, the only way to know with certainty that a matching cannot be stable for any realization of candidates’ preferences is when a candidate is not matched and can form a mutually acceptable pair with a position that is either unfilled or matched to a candidate ranked lower in the preferences of the department holding that position. It is this second requirement that is crucial to characterizing impossible matches. If a candidate  $i$  is matched to a position  $d$  at some stable matching then all the candidates that are ranked higher than  $i$  for position  $d$  must be matched to some other position, otherwise the matching would not be stable. Those candidates must pass the same test, i.e., all the candidates ranked higher than them (at the positions where they are matched) must be matched to a position. Matching those additional candidates may in turn bring additional candidates to our problem, and so on and so forth. Candidate  $i$  will be an impossible match for department  $d$  when, for any possible way we match the other candidates, there is always at least one candidate in excess: a

---

<sup>5</sup>Kidney exchange is another situation where only partial “preferences” may be known. In some countries patients’ submitted preferences are dichotomous (acceptable v. unacceptable), while in reality preferences patients may have finer preferences (e.g., the age of the donor). See for instance Nicoló and Rodríguez-Álvarez (2017) who study the efficiency gain that can be realized in paired kidney exchange when patient’s “preferences” over kidneys are used instead of simple dichotomous acceptable/unacceptable preferences.

candidate that we must match, but all his acceptable departments have exhausted their positions with other candidates. In other words, the impossibility to match a candidate to a department will arise when there is a set of candidates ranked “sufficiently high” that outnumber the total number of positions of the departments acceptable for them. This implies that the existence of impossible matches is especially relevant under the following features: not all candidate-department pairs are mutually acceptable and/or the market is imbalanced, that is, there are more candidates than positions. Indeed, if there are enough positions to accommodate all candidates and if all candidate-department pairs are mutually acceptable then no candidate is an impossible match for any department.

We show that this problem is reminiscent to *Hall’s (1935) marriage theorem*, one of the very first results of matching theory. Phrased in our context, Hall’s marriage theorem identifies a sufficient and necessary condition to match a full set of candidates to their acceptable departments. Hall’s condition is not sufficient for our purposes, however, as it provides no link to stability. Part of our contribution consists of making precise what being ranked “sufficiently high” means, by carefully selecting the set of individuals for which we have to check Hall’s condition (Theorem 2). As a by-product, our result establishes a formal connection between stability and Hall’s marriage condition.<sup>6</sup>

In the second part of the paper we use the concept of impossible match to show how stable matchings can be (partially) predicted. To do so we introduce a *prediction algorithm* that identifies candidate-department pairs that are necessarily matched in every stable matching, for any realization of candidates’ preferences. Incidentally, our prediction algorithm also enables us to pinpoint candidates whose submitted preferences can have an impact on the final matching and candidates whose submitted preferences can never affect the matching that will be obtained.

To demonstrate that some candidates’ preferences may not have any impact on the final matching, we use data from the French academic job market, which is, unlike most academic job markets, centralized. The algorithm used in France is equivalent to Gale and Shapley’s (1962) Deferred Acceptance algorithm (with candidates proposing).<sup>7</sup> Thus it produces a stable matching (with respect to the submitted preferences). Our data

---

<sup>6</sup>A similar connection has been identified by Demange, Gale and Sotomayor (1986). They use Hall’s marriage conditions to show the existence of an equilibrium assignment in the context of a multi-item auction market.

<sup>7</sup>See Haeringer and Iehlé (2010).

covers the years 1999–2013 for junior positions in mathematics. For each year we observe the rankings of candidates established by the departments, from which we deduce the set of mutually acceptable pairs, but not the preferences submitted to the clearinghouse by the candidates later on.<sup>8</sup> The identification of impossible matches permits us to correctly predict, on average, nearly 45 percent of the positions. We also find that the match of about 60 percent of the candidates does not depend on the realization of candidates’ preferences.

The paper is structured as follows. In Section 2 we outline the basic model. Section 3 contains the first main contribution of this paper, i.e., the characterization of impossible matches. In Section 4 we show how the identification of impossible matches can be used to make predictions about matchings, and Section 5 illustrates our findings with an empirical analysis of the French academic job market. The main proofs are relegated to the Appendix. Additional elements and results are provided in the Web Appendix.

## 2. Preliminaries

### 1. Matchings

We consider a finite set  $I$  of candidates and a finite set  $D$  of departments, where each department  $d$  is endowed with a positive capacity  $q_d$ . The problem studied here is that of matching candidates to departments subject to their capacities. A **matching** is a mapping  $\mu : I \cup D \rightarrow 2^I \cup D$  such that, for each  $i \in I$  and each  $d \in D$ ,

- $\mu(i) \in D \cup \{i\}$ ,
- $\mu(d) \in 2^I$ ,
- $\mu(i) = d$  if, and only if,  $i \in \mu(d)$ , and
- $|\mu(d)| \leq q_d$ .

For  $v \in D \cup I$ , we call  $\mu(v)$  agent  $v$ ’s assignment ( $\mu(d)$  is possibly empty for  $d \in D$ ). For  $i \in I$ , if  $\mu(i) = d \in D$  then candidate  $i$  is **matched** to department  $d$  under  $\mu$ . If  $\mu(i) = i$  then candidate  $i$  is said to be **unmatched** under  $\mu$ .

---

<sup>8</sup>We explain in detail in Section 5 how we can deduce the set of mutually acceptable candidate-department pairs.

## 2. Job Market Problem

A **job market problem** is a 5-tuple  $(I, D, (\succ_d, q_d)_{d \in D}, (\succ_i)_{i \in I})$ , where  $I$  is a set of candidates,  $D$  a set of departments,  $\succ_d$  a preference ordering for department  $d$ ,  $q_d$  a capacity constraint for department  $d$ , and  $\succ_i$  a preference ordering for candidate  $i$ .

Each candidate  $i \in I$  has a **strict preference** relation  $\succ_i$  over the departments and the option of remaining unassigned, i.e.,  $\succ_i$  is a complete, transitive, irreflexive and asymmetric relation on  $D \cup \{i\}$ . The notation  $d \succ_i d'$  means that candidate  $i$  prefers to go to department  $d$  than department  $d'$ . A department  $d$  is **acceptable** for a candidate  $i$  under the preferences  $\succ_i$  if  $i$  prefers being matched to  $d$  to being matched to himself, i.e.,  $d \succ_i i$ , otherwise the department is said to be **unacceptable** for  $i$ .

Each department  $d \in D$  has a **strict preference** relation  $\succ_d$  over the candidates and the possibility of leaving a position unfilled, i.e.,  $\succ_d$  is a complete, transitive, irreflexive and asymmetric relation on  $I \cup \{d\}$ . We say that candidate  $i$  is **ranked higher** than candidate  $j$  at department  $d$  if  $i \succ_d j$  (and so  $j$  is ranked lower than  $i$ ). Candidate  $i$  is **acceptable** to department  $d$  if  $i \succ_d d$ .

When there is no risk of confusion, we shall simply use  $\succ$  to denote a job market problem. Given a preference relation  $\succ_j$  we denote by  $\succeq_j$  the weak relation associated to it, i.e., for each  $j \in I \cup D$ ,  $v \succeq_j v' \Leftrightarrow v \succ_j v'$  or  $v = v'$ . Given a job market problem  $\succ$  we denote by  $A_i^\succ$  the set of departments that are acceptable for  $i$ . Similarly, we denote by  $A_d^\succ$  the set of candidates that are acceptable for  $d$ .

A matching  $\mu$  is **stable** for a job market problem  $\succ$  if:

- (a) **individual rationality**, i.e., for each candidate  $i$ ,  $\mu(i) \in A_i^\succ \cup \{i\}$ , and, for each department  $d$ ,  $i \in A_d^\succ$  for each candidate  $i \in \mu(d)$ .
- (b) **no blocking pair**, i.e., there is no pair  $(i, d) \in A_d^\succ \times A_i^\succ$  such that  $d \succ_i \mu(i)$  and
  - (b.1)  $|\mu(d)| < q_d$  **or**
  - (b.2)  $i \succ_d i'$  for some  $i' \in \mu(d)$ .

Given a job market problem  $\succ$ , the set of stable matchings is denoted by  $\Sigma(\succ)$ .

Notice that the way we defined stability implicitly assumes that department's preferences are **responsive** over sets of students. Following Roth (1985), let us denote by  $\succ_d^\#$  the preferences of a department  $d$  over sets of candidates. The preference  $\succ_d^\#$  is responsive



to the preference  $\succ_d$  if, for any  $I', I'' \subseteq I$ ,  $I'' \succ_d^\# I'$  whenever  $I'' = I' \cup \{v''\} \setminus \{v'\}$  for some  $v' \in I'$  and  $v'' \notin I'$  such that  $v'' \succ_d v'$ , or, if  $|I'| < q_d$ ,  $I'' \succ_d^\# I'$  whenever  $I'' = I' \cup \{v''\}$  and  $v'' \succ_d d$ .

### 3. Proto-matching problems

A **proto-matching problem** is similar to a job market problem except that preferences are summarized only by a list of mutually acceptable pairs from both sides of the market and the departments' preferences over their acceptable candidates. Formally, a proto-matching problem is given by a 5-tuple  $(I, D, (P_d, q_d)_{d \in D}, A)$ , where  $I$  is a set of candidates,  $D$  a set of departments,  $P_d$  a preference of department  $d$ ,  $q_d$  a capacity of department  $d$  and  $A \subseteq I \times D$  a set of mutually acceptable pairs  $(i, d)$  of candidates and departments. In a proto-matching problem, the preference of department  $d$ ,  $P_d$ , is a linear ordering over  $A_d$ , which is the subset of candidates that form mutually acceptable pairs with  $d$ , i.e.  $A_d := \{i \in I : (i, d) \in A\}$ . With a little abuse of terminology, we also say that  $A_d$  is the set of acceptable candidates for  $d$ . Similarly,  $A_i := \{d \in D : (i, d) \in A\}$  is the set of acceptable departments for  $i$ . For a set  $I'$  of candidates, a department belongs to  $A_{I'}$  if, and only, it is acceptable for at least one candidate in  $I'$ . For a set  $D'$  of departments, a candidate belongs to  $A_{D'}$  if, and only if, she/he is acceptable for at least one department in  $D'$ .

Note that the profile of departments' preferences over the acceptable candidates (and their capacities) is a sufficient information to describe a proto-matching problem (because  $i \in A_d$  if, and only if,  $d \in A_i$ ). Accordingly, whenever there is no risk of confusion we shall use the shorthand  $P$  to denote a proto-matching problem.

The concept of stability, as defined for job market problems, is not applicable in proto-matching problems, because candidates' preferences are not completely defined. The concept has a bite, however, when a candidate is matched to an unacceptable department; or when a candidate is unmatched and there is an acceptable department with an unfilled position or matched to another candidate ranked lower. We summarize these three situations with the following concepts.

- A matching  $\mu$  is **feasible** for  $P$  if for each  $i \in I$ ,  $\mu(i) \neq i$  implies  $(i, \mu(i)) \in A$ .
- A matching  $\mu$  is **maximal** for  $P$  if it is feasible and there is no  $(i, d) \in A$  such that

$$\mu(j) = j \text{ and } |\mu(d)| < q_d.^9$$

- A matching  $\mu$  is **comprehensive** for  $P$  if it is feasible and for each  $(i, d) \in A$ , with  $\mu(i) = d$ ,  $jP_d i$  implies  $\mu(j) \in D$ .

**Remark 1** The feasibility constraint is akin to individual rationality. The condition of maximality and comprehensiveness are respectively related to the conditions on the absence of blocking pairs, (b.1) and (b.2), but only for unmatched agents.

Given a proto-matching problem  $P$ , a job market problem  $\succ$  with the same sets of candidates and departments and identical departments' capacities is  **$P$ -compatible** if

- (a) the sets of acceptable candidates and acceptable departments in  $\succ$  generate the set of mutually acceptable pairs in  $P$ :

$$\{(i, d) \in I \times D : i \in A_d^\succ \text{ and } d \in A_i^\succ\} = A$$

- (b) for every department  $d$ , and every pair of candidates  $i, i'$  such that  $i, i' \in A_d$ :

$$i \succ_d i' \text{ if, and only if, } iP_d i'$$

Note that the condition (a) implies that  $A_j \subseteq A_j^\succ$  for each  $j \in I \cup D$ . We also observe that the same proto-matching problem can generate different job market problems (thus generate different sets of stable matchings and possibly different sets of matched candidates). We denote by  $\Theta(P)$  the set of job market problems that are  $P$ -compatible.

The next proposition connects the notions of stability, comprehensiveness and maximality in job market and proto-matching problems.

**Proposition 1** *Let  $P$  be a proto-matching problem. The following assertions are equivalent:*

- (i)  $\mu$  is a stable matching for some  $P$ -compatible job market problem  $\succ$ .
- (ii)  $\mu$  is a comprehensive and maximal matching for the proto-matching problem  $P$ .

---

<sup>9</sup>An equivalent definition is the following. A matching  $\mu$  is maximal if it is feasible and there is no feasible matching  $\mu'$  such that :  $\mu'(i) = \mu(i)$  for every  $i \in D$  such that  $\mu(i) \neq i$ , and  $\mu'(j) \neq j$  for some  $j \in D$  such that  $\mu(j) = j$ .

**Proof**  $(i) \Rightarrow (ii)$  Since  $\mu$  is individually rational for the problem  $\succ$ , every pair  $(i, d)$  such that  $\mu(i) = d$  satisfies  $i \in A_d^\succ$  and  $d \in A_i^\succ$ . From the construction of a  $P$ -compatible job problem, it follows that  $(i, d) \in A$ . Hence,  $\mu$  is feasible for  $P$ . Suppose first that  $\mu$  is not comprehensive. It must be the case that there exists  $i, j$  and  $d$  such that  $\mu(i) = i$ , and  $\mu(j) = d$  and  $iP_dj$ . Since  $\succ \in \theta(P)$ , we have  $d \succ_i i$  and  $i \succ_d j$ . So the pair  $(i, d)$  blocks  $\mu$  in the sense of condition (b.2), which contradicts  $\mu \in \Sigma(\succ)$ . If  $\mu$  is not maximal then there exists  $i$  and  $d$  such that  $\mu(i) = i$ ,  $\mu(d) < q_d$  and  $d \in A_i$ . This implies that the pair  $(i, d)$  blocks  $\mu$  in the sense of condition (b.1), which contradicts that  $\mu \in \Sigma(\succ)$ .

$(ii) \Rightarrow (i)$  Let  $\mu$  be a comprehensive and maximal matching for  $P$  (hence feasible for  $P$ ). From the definition of a  $P$ -compatible job market problem,  $\mu$  is individually rational for any  $\succ \in \theta(P)$ . Let  $\succ$  be any job market problem  $\succ \in \theta(P)$  where, for each candidate  $i$  such that  $\mu(i) \neq i$ ,  $\mu(i)$  is the most preferred department according to  $\succ_i$ . Suppose  $\mu \notin \Sigma(\succ)$ . So there exists a blocking pair  $(i, d)$ . It must be the case that  $\mu(i) = i$  from the construction of  $\succ$ . Since  $\mu$  is maximal the condition (b.1) cannot hold. Since  $\mu$  is comprehensive the condition (b.2) cannot hold. Thus  $(i, d)$  cannot form a blocking pair, a contradiction. ■

### 3. Impossible matches

The purpose of this section is to define and characterize the candidates that are **impossible matches** for some specific department in a given proto-matching problem. The question addressed by the concept of impossible match is the following: upon observing one side of the markets' preferences and the set of mutually acceptable pairs (i.e., a proto-matching problem), is there a stable matching with respect to some realization of the other side's preferences such that a particular candidate is matched to a particular department? If the answer is "no" then that candidate is said to be an impossible match for that department.

**Definition 1** Given a proto-matching problem  $P$ , a candidate  $i$  is an **impossible match for department**  $d \in A_i$  if, for each job market problem  $\succ \in \Theta(P)$ , there is no matching  $\mu \in \Sigma(\succ)$  such that  $\mu(i) = d$ .

Note that we could dispense with the requirement that  $d \in A_i$  in the definition of an impossible match. Indeed,  $d \notin A_i$  trivially prevents the existence of a stable matching

where  $i$  is matched to  $d$ , for any problem  $\succ \in \theta(P)$ . This is a trivial case that we will not address in the paper.

### 1. Introductory elements

To understand how the preferences of only one side of the market can suffice to check the stability of a matching, consider the following instances of job market problems.

**Example 1** The simplest case occurs if  $q_d$  candidates ranked higher than  $i$  at department  $d$  have no other acceptable department. Then for every job market problem  $\succ \in \Theta(P)$  and stable matching  $\mu \in \Sigma(\succ)$  either all those candidates are matched to  $d$ , or at least one of them is not matched to any department.<sup>10</sup> In either case,  $i$  cannot be matched to  $d$  at a stable matching, for otherwise one of the candidates ranked higher than  $i$  would form a blocking pair with department  $d$ .  $\square$

**Example 2** A (slightly) more complex case is the following, with 4 candidates and 4 departments. We assume that each department has only one position to offer. Let  $i_1, i_2, i_3$  and  $i_4$  be candidates whose sets of acceptable departments are  $\{d_1, d_2\}$ ,  $\{d_2, d_3\}$ ,  $\{d_3, d_4\}$ , and  $\{d_2, d_4\}$ , respectively. Table 1 describes the departments' preferences over these 4 candidates. For instance, at department  $d_2$  candidate  $i_2$  is ranked higher than candidate  $i_1$ , who is ranked higher than candidate  $i_4$ .

Table 1: Candidate  $i_1$  is an impossible match for department  $d_2$ .

$P_{d_1}$	$P_{d_2}$	$P_{d_3}$	$P_{d_4}$
$i_1$	$i_2$	$i_3$	$i_4$
	$i_1$	$i_2$	$i_3$
		$i_4$	

We claim that for each  $P$ -compatible job market problem it is impossible to obtain a stable matching where  $i_1$  is matched to  $d_2$ . To see this, note that if this is the case then  $i_2$  must be matched to  $d_3$ , for otherwise  $(i_2, d_2)$  would form a blocking pair. This

---

<sup>10</sup>There could be more than  $q_d$  candidates ranked higher than  $i$  at  $d$  and a  $P$ -compatible job market problem such that some of these candidates have department  $d$  being their most preferred department.

in turn implies that  $i_3$  is matched to  $d_4$ , which is only possible if  $i_4$  is matched to  $d_2$ , a contradiction. So, there is no preference profile and a stable matching for that preference profile where  $i_1$  is matched to  $d_2$ .  $\square$

Without entering into much detail, it is important to note that in the previous example the key property of departments  $d_2, d_3$  and  $d_4$ 's preferences to ensure that  $i_1$  is an impossible match for department  $d_2$  is not, as it may appear, that there is some sort of cycle with the relative positions of candidates  $i_2, i_3$  and  $i_4$ . As we shall see,  $i_1$  being an impossible match for  $d_2$  is rather due to the more general feature that there are (at least) 3 candidates competing for three departments and such that their relative rankings in the preferences of the departments allow them to claim the positions over  $i_1$ .

Finally, note that to establish that candidate  $i_1$  is an impossible match for department  $d_2$  only candidates  $i_2, i_3$  and  $i_4$  were needed. That is, if there were other candidates who also find one of those departments acceptable candidate  $i_1$  would remain an impossible match for  $d_2$ . For instance, we could have a candidate, say,  $i_5$  ranked higher than  $i_3$  at department  $d_3$ , or a candidate, say,  $i_6$ , ranked between  $i_4$  and  $i_3$  at department  $d_4$ , without altering our conclusion about candidate  $i_1$ .

Using Proposition 1, we have the following straightforward characterization for impossible matches (its proof is omitted).

**Proposition 2** *A candidate is an impossible match for a department in a proto-matching problem  $P$  if, and only if, there is no maximal and comprehensive matching for  $P$  such that they are matched together.*

In the next proposition we establish a Rural-Hospital type of result (Roth, 1986), when some candidates are impossible matches for some departments.

**Proposition 3** *Given a proto-matching problem  $P$ , if  $i$  is an impossible match for  $d \in A_i$  then  $d$  fills its capacity at every stable matching, for each  $P$ -compatible job market problem.*

**Proof** Suppose by way of contradiction that  $i$  is impossible for  $d \in A_i$  and that, for some  $P$ -compatible job market problem  $\succ$ , there exists a matching  $\mu \in \Sigma(\succ)$  such that  $|\mu(d)| < q_d$ . Since  $\mu$  is stable, Proposition 2 implies that  $\mu$  is maximal and comprehensive for  $P$ . From maximality it holds that  $\mu(j) \neq j$  for every candidate  $j$  such that  $d \in A_j$ . Note that  $i$  is such a candidate .

Let  $\mu'$  be the matching such that  $\mu'(j) = \mu(j)$  for each  $j \neq i$ , and  $\mu'(i) = d$ . In addition, if there exist candidates for which  $\mu(i)$  is acceptable and who are unmatched under  $\mu$ , match under  $\mu'$  the highest ranked candidate among those candidates to the department  $\mu(i)$ .

It is easy to check that  $\mu$  being comprehensive and maximal for  $P$  implies that  $\mu'$  is also maximal and comprehensive for  $P$ . From Proposition 2 it follows that  $i$  is not an impossible match for  $d$ , a contradiction. ■

**Remark 2** The impossible match property is not monotonic with respect to the preferences of the department. That is, if a candidate is an impossible match for a department then candidates ranked lower at that department are not necessarily impossible matches for that department. To see this, consider again the proto-matching problem  $P$  given in Table 1. It is easy to check that candidate  $i_1$  is also an impossible match for  $d_2$ , yet candidate  $i_4$  is not an impossible match for  $d_2$ . Indeed, let  $\succ \in \Theta(P)$  be such that  $i_1$ 's most preferred department is  $d_1$ ,  $d_2$  is the most preferred department of  $i_4$ ,  $d_3$  is the most preferred department of  $i_2$ , and  $d_4$  is the most preferred department of  $i_3$ . The matching  $\mu$  such that  $\mu(d_1) = i_1$ ,  $\mu(d_2) = i_4$ ,  $\mu(d_3) = i_2$ ,  $\mu(d_4) = i_3$  is stable for  $\succ$ .

## 2. A characterization

Part of our contribution consists of showing that we do not need to consider all candidates and all possible profiles of candidates' preferences (and compute all the corresponding stable matchings) to check whether a candidate is an impossible match for a department. Instead we are able to characterize impossible matches on the basis of the fundamentals of the proto-matching problem only. Proposition 2 provides a first baseline for the analysis of impossible matches, at the cost of extreme computations: checking whether a candidate  $i$  is an impossible match for a department  $d$  amounts to listing every comprehensive and maximal matching of the proto-matching problem  $P$  to see whether one of those assigns  $i$  to  $d$ . Another way is to consider all candidates' preference profiles and use Gusfield's algorithm (Gusfield, 1987) to identify all candidate-department pairs that are contained in at least one stable matching. While Gusfield's algorithm is polynomial-time, enumerating all possible preference profiles is clearly not solvable in polynomial time. The following result, due to Saban and Sethuraman (2015) shows that

in fact there is no hope to identify a polynomial-time algorithm for the identification of impossible matches (unless  $P = NP$ ).

**Proposition 4 (Saban and Sethuraman (2015))** *Finding whether a candidate is an impossible match is NP-complete.*

In what follows, we will narrow down the set of cases one has to consider, showing that it is sufficient to find a particular subset of candidates, called a **block**. In addition, an interesting feature of the result is that it relies on one of the earliest results of matching theory, which we rephrase now to fit our setting.

**Theorem 1 (Hall’s Marriage Theorem, 1935)** *Given a proto-matching problem  $P$ , there exists a feasible matching where every candidate is matched to a department if, and only if,*

$$\text{for each } I' \subseteq I, |I'| \leq \sum_{d \in A_{I'}} q_d . \quad (1)$$

In words, there exists a matching where every candidate can be matched to a department if, and only if, for each subset  $I'$  of candidates the number of candidates is at most equal to the total number of positions among the departments that are considered by some of the candidates in  $I'$  as acceptable. Eq. (1) is known as **Hall’s marriage condition**.<sup>11</sup> At that matching, the positions of departments are not necessarily all filled by candidates. If this is the case, then we say that the matching is **perfect**. Using Hall’s marriage theorem, one can deduce that a perfect matching exists if, and only if, Eq. (1) is satisfied together with  $|I| = \sum_{d \in A_I} q_d$ .

To incorporate the absence of blocking pair condition (b.2), or its counterpart in the proto-matching framework (comprehensiveness), we shall not apply Hall’s marriage condition to the original proto-matching problem but rather to a collection of restricted and truncated proto-matching problems.

---

<sup>11</sup>Hall’s marriage condition was originally stated for the one-to-one matching problem. It is easy to see that this condition can be extended to the case of job market problems when departments have responsive preferences by making for each department as many copies as its capacity (where each copy has only one position to offer) and making each of its copies acceptable for a candidate if the department is acceptable for that candidate.

Given a proto-matching problem  $P = (I, D, (P_d, q_d)_{d \in S}, A)$  and two sets of candidates  $\mathbf{J}$  and  $\mathbf{K}$ , with  $\mathbf{K}$  possibly empty, the proto-matching problem restricted to  $\mathbf{J}$  and truncated at  $\mathbf{K}$  is denoted by  $P(\mathbf{J}, \mathbf{K})$ . The set of acceptable pairs for  $P(\mathbf{J}, \mathbf{K})$  is

$$A^{P(\mathbf{J}, \mathbf{K})} = \{(i, d) \in A : i \in \mathbf{J} \text{ and } iP_d i' \text{ for each } i' \in \mathbf{K}\} \quad (2)$$

and the preference of department  $d$  is the ordering over  $A_d^{P(\mathbf{J}, \mathbf{K})}$  that coincides with  $P_d$  on  $A_d^{P(\mathbf{J}, \mathbf{K})}$ .<sup>12</sup>

When we consider the proto-matching problem restricted to  $\mathbf{J}$  and truncated at  $\mathbf{K}$  the set of “available” candidates and departments may not be equal to  $\mathbf{J} \setminus \mathbf{K}$  and  $A_{\mathbf{J} \setminus \mathbf{K}}$ , respectively. Indeed, it may well be that there is a candidate  $i \in \mathbf{J} \setminus \mathbf{K}$  such that there is no department  $d$  for which  $i \in A_d^{P(\mathbf{J}, \mathbf{K})}$ . Similarly, for a department  $d$  we may have  $A_d^{P(\mathbf{J}, \mathbf{K})} = \emptyset$ .

**Definition 2** For a truncated proto-matching problem  $P(\mathbf{J}, \mathbf{K})$  we say that a candidate  $i$  is **admissible** if  $i \in \cup_{d \in D} A_d^{P(\mathbf{J}, \mathbf{K})}$ . Similarly, a department  $d$  is **admissible** if  $A_d^{P(\mathbf{J}, \mathbf{K})} \neq \emptyset$ .

**Remark 3** Note that admissibility in a proto-matching problem refers to the notion of being acceptable for a department and is independent of the departments’ capacities. For instance, if for a proto-matching problem  $P(\mathbf{J}, \mathbf{K})$  there is a department  $d$  such that  $q_d = 0$ , all candidates in  $A_d^{P(\mathbf{J}, \mathbf{K})}$  consider  $d$  as acceptable (and thus are admissible), even though the capacity of department  $d$  is zero in  $P(\mathbf{J}, \mathbf{K})$ .

For sake of completeness, the meaning of the Hall’s marriage condition applied to a proto-matching problem restricted to  $\mathbf{J}$  and truncated at  $\mathbf{K}$  is precisely stated in the next definition.

**Definition 3** Given a proto-matching problem  $P$ , and two sets of candidates  $\mathbf{J}$  and  $\mathbf{K}$ , with  $\mathbf{K}$  possibly empty, Hall’s marriage condition holds in  $P(\mathbf{J}, \mathbf{K})$  if

$$\text{for each } I' \subseteq \cup_{d \in D} A_d^{P(\mathbf{J}, \mathbf{K})}, |I'| \leq \sum_{d \in A_{I'}^{P(\mathbf{J}, \mathbf{K})}} q_d. \quad (3)$$

---

<sup>12</sup>Note that the proto-matching problem  $P(\mathbf{J}, \emptyset)$  is simply the problem  $P$  restricted to the set of candidates  $\mathbf{J}$  and  $P(I, \mathbf{K})$  is the problem  $P$  truncated at  $\mathbf{K}$ .



Given a proto-matching problem  $P$  and a department  $d \in D$ , we denote by  $\overline{P}^d$  the same problem as  $P$  except that one position of  $d$  is removed, i.e. department  $d$ 's capacity is  $q_d - 1$ . Note that  $\overline{P}^d$  is a proto-matching problem. Given a pair  $(i, d) \in A$  and a subset  $\mathbf{J} \subseteq I \setminus \{i\}$ , we denote by  $\mathbf{J}_{i,d} \subseteq \mathbf{J}$  the set of candidates in  $\mathbf{J}$  that are ranked higher than  $i$  by  $d$ , i.e.,  $\mathbf{J}_{i,d} = \{j \in \mathbf{J} : j P_d i\}$ .

**Definition 4** Given a proto-matching problem  $P = (I, D, (P_d, q_d)_{d \in D}, A)$ , a **block at**  $(i, d) \in A$  is a nonempty set  $\mathbf{J} \subseteq I \setminus \{i\}$ , with  $d \in A_{\mathbf{J}}$ , such that the following conditions hold:

1. There exists a set  $\mathbf{J}' \subseteq \mathbf{J}$  such that  $|\mathbf{J}'| = \sum_{\hat{d} \in A_{\mathbf{J}}} q_{\hat{d}}$  and Hall's marriage condition holds in  $P(\mathbf{J}', \emptyset)$ ;
2. For each  $\mathbf{K} \subseteq \mathbf{J} \setminus \mathbf{J}_{i,d}$  such that  $|\mathbf{K}| = |\mathbf{J}| - \sum_{\hat{d} \in A_{\mathbf{J}}} q_{\hat{d}} + 1$ , Hall's marriage condition does not hold in  $\overline{P}^d(\mathbf{J}, \mathbf{K})$ .

We state now the main result of the paper.

**Theorem 2** *A candidate  $i$  is an impossible match for a department  $d$  in a proto-matching problem  $P$  if, and only if,  $P$  admits a block at  $(i, d)$ .*

The result provides additional insights on impossible matches. As illustrated by the examples in the sub-section 1, an impossible match  $i$  for  $d$  is by no way determined by the sole ranking of  $d$  or the way  $i$  is ranked elsewhere by his acceptable departments.<sup>13</sup> Nevertheless we do not need to consider the whole market. Our main result shows that the impossible match property can be captured *locally* through a block of candidates, at least one of them being ranked by  $d$ . More precisely, we prove in the Web Appendix

<sup>13</sup>In Example 2, there is a block at  $(i_1, d_2)$  formed by the set of candidates  $\mathbf{J} = \{i_2, i_3, i_4\}$ . Condition 1 is indeed satisfied for  $\mathbf{J}' = \mathbf{J}$  since the three candidates in  $\mathbf{J}'$  can be clearly matched to the three available positions in the set of admissible departments  $\{d_2, d_3, d_4\}$ . To check condition 2, we first observe that the sets  $\mathbf{K}$  satisfying the cardinality condition are necessarily singletons since  $|\mathbf{J}| - \sum_{\hat{d} \in A_{\mathbf{J}}} q_{\hat{d}} + 1 = 1$ . In addition,  $i_2$  cannot belong to  $\mathbf{K}$  since  $i_2 \in \mathbf{J}_{i_1, d_2}$ . Hence, it remains to check Hall's marriage condition (3) in  $\overline{P}^{d_2}(\mathbf{J}, \mathbf{K})$  with  $\mathbf{K}$  equal to  $\{i_3\}$  or  $\{i_4\}$ . If  $\mathbf{K} = \{i_3\}$ , then the two remaining admissible candidates in  $\overline{P}^{d_2}(\{i_2, i_3, i_4\}, \{i_3\})$  are  $i_2$  and  $i_4$ . They cannot both be matched to an admissible department since  $d_3$  is not admissible in  $\overline{P}^{d_2}(\{i_2, i_3, i_4\}, \{i_3\})$  and the capacity of the admissible department  $d_2$  in  $\overline{P}^{d_2}(\{i_2, i_3, i_4\}, \{i_3\})$  is equal to 0. A similar reasoning applies to the case  $\mathbf{K} = \{i_4\}$ . Hence Hall's marriage condition (3) cannot be satisfied for every  $\mathbf{K}$ .

(Proposition A.1) that the conditions 1. and 2. in the definition of a block imply there are at least  $q_d$  candidates in  $\mathbf{J}$  ranked higher than  $i$  at department  $d$ , i.e.,  $|\mathbf{J}_{i,d}| \geq q_d$ . Next, Hall’s condition (3) holds in  $P(\mathbf{J}', \emptyset)$  for some  $\mathbf{J}'$  such that  $|\mathbf{J}'| = \sum_{\hat{d} \in A_{\mathbf{J}}} q_{\hat{d}}$  is tantamount to saying that there exists a perfect matching between the departments  $A_{\mathbf{J}}$  and the candidates  $\mathbf{J}'$ , i.e., there exists a matching where all candidates  $\mathbf{J}'$  are matched to a department and all departments exhaust their capacities with candidates in  $\mathbf{J}'$ . Second, condition 2 implies in particular that the ability to obtain a perfect match is tight in the following sense: If we consider the problem  $\bar{P}^d$  (the problem identical to  $P$  except that one position is removed from  $d$ ) then there is no perfect match between the departments  $A_{\mathbf{J}}$  and any subset  $\mathbf{J}' \subseteq \mathbf{J}$  that includes all the candidates in  $\mathbf{J}_{i,d}$ , and such that the matching is comprehensive for  $P(\mathbf{J}, \emptyset)$ .<sup>14</sup>

The Appendix provides several examples to illustrate the notion of block and its role in the determination of impossible matches. We extend our formal analysis in Section A of the Web Appendix. We introduce there several variants of the notion of block and an algorithm for finding blocks, that is, finding impossible matches.

#### 4. Matching prediction

We now investigate what the identification of impossible matches entails. Our claim is that such an identification can be used to predict matchings, thereby measuring the extent to which one side’s preferences matter in the determination of a final matching. Identifying impossible matches will permit us to single out two sets of candidates. On the one hand, we shall have the set of candidates for whom the strict preferences over departments is irrelevant for their final match. For them, no strict preferences, even partial, can be deduced from the data. On the other hand, we shall have candidates for whom the strict ordering in their (submitted) preference lists matters for the final match, but only their (relative) top choice can be sometimes recovered.

Our strategy is the following. Given an observed proto-matching problem, deleting candidates identified as impossible matches (from the departments’ rankings for which they are impossible matches) creates a *reduced* proto-matching problem. In this new problem, a candidate who is ranked top by some department, and ranked only once, must

---

<sup>14</sup>The notion of block echoes a strong version of market imbalance in the spirit of Ashlagi, Kanoria and Leshno (2017). See Section 6.

either have begun with only one acceptable department or have all but one department deemed an impossible match. It is straightforward to show that the candidate is *necessarily* matched to that department, irrespectively of the way candidates rank departments in their preferences.

If the observed data also contains the final (realized) matching we can go one step further by considering the assignments of market stars, that is, candidates that are always top ranked by all acceptable departments in the reduced problem. For those candidates, there is no way to infer their final match on the basis of the sole proto-matching problem. Note, however, that their match trivially indicates their top choice (within the remaining choices). Once market stars' choices are taken into account new impossible matches may be identified and so on and so forth. This is the baseline of our general procedure.

Let  $P$  be a proto-matching problem. The **reduced proto-matching problem of  $P$** , denoted by  $R(P)$ , is the proto-matching problem where impossible matches have been deleted from the rankings of the positions for which they are impossible matches, and where the lists of acceptable candidates or departments are made mutually consistent. Formally, the set of mutually acceptable pairs is

$$A^{R(P)} = A \setminus \{(i, d) : i \text{ is an impossible match for } d \text{ in } P\}$$

and for each pair  $i, j \in A_d^{R(P)}$ ,  $iR(P)_d j$  if and only if  $iP_d j$ . We shall denote by  $R(P)_d$  the ranking list of  $d$  in the proto-matching problem  $R(P)$ .

The next result motivates the construction of the predicting algorithm that we present hereafter.

**Proposition 5** *Let  $P$  be a proto-matching problem. For every  $P$ -compatible job market problem  $\succ$  we have:*

1. *Candidate  $i$  is matched to department  $d$  under every matching that is stable for  $\succ$  if, and only if,  $A_i^{R(P)} = \{d\}$  and  $|\{j : jP_d i\}| < q_d$*
2. *Candidate  $i$  is unmatched under every matching that is stable for  $\succ \in \theta(P)$  if, and only if,  $A_i^{R(P)} = \emptyset$ .*

In other words, Proposition 5 says that a candidate's match is always the same for every stable matching of every  $P$ -compatible job market problem if, and only if, the candidate falls into one of the two cases mentioned in the Proposition.

Table 2: A predicted matching

$P_{d_1}$	$P_{d_2}$
$i_1$	$i_2$
$i_3$	$i_3$
$i_2$	$i_1$

Even though Proposition 5 seems straightforward, deleting matches is not innocuous for the property of impossible matches. Deleting iteratively impossible pairs one after the other can alter the matching predictions we can make, by declaring erroneously some pairs *not* impossible. To see this, consider the proto-matching problem depicted in Table 2. It is easy to see that  $i_3$  is an impossible match for  $d_1$  and  $d_2$ , and  $i_2$  is an impossible match for  $d_1$  (and  $i_1$  for  $d_2$ ). So  $R(P)_{d_1} = i_1$  and  $R(P)_{d_2} = i_2$ . The unique comprehensive matching for  $P$  that fills the available positions is  $\mu(i_1) = d_1$  and  $\mu(i_2) = d_2$  (indicated by the square boxes in Table 2), thus it is the stable matching of any compatible job market problem. Note, however, that if we delete  $i_3$  at  $d_1$  and  $d_2$  then  $i_2$  is no longer an impossible match for  $d_1$  (and  $i_1$  for  $d_2$ ).

The reduction procedure can be generalized if the economist has access to the final match of the candidates. The initial data is a proto-matching problem  $P$  and a matching  $\mu$  for  $P$ . If the matching is not observed, then the matching  $\mu$  used in the algorithm is the empty matching, i.e.,  $\mu(i) = i$  for each candidate  $i$ .<sup>15</sup>

### Predicting Algorithm

**Input:** A pair  $(P, \mu)$ , where  $\mu$  is a matching for  $P$ .

**Step 1** Identify and remove all impossible match pairs in  $P$  and construct  $P^1 := R(P)$ .

**Step h** For  $h \geq 2$

**h.1** Identify the set  $S^h \subseteq \{i \in I \mid \mu(i) \neq i\}$  of all candidates  $i$  such that (a)  $i$  has at least 2 acceptable positions in  $P^h$ ; and (b)  $i$  is ranked among the top  $q_d$  candidates in every department  $d \in A_i^{P^h}$  in the problem  $P^h$ .

---

<sup>15</sup>The algorithm and our results also work if the matching  $\mu$  is only partially observed. If the matching of candidate  $i$  (resp. department  $d$ ) is not observed then it suffices to set  $\mu(i) = i$  (resp.  $\mu(d) = d$ ).

**h.2** If  $S^h$  is empty STOP. Otherwise, define the proto-matching problem  $\tilde{P}^h$  such that:

- \* for each  $i \in S^h$  and  $d \in A_i^{P^h}$  such that  $\mu(i) \neq d$ ,  $(i, d)$  is removed from  $A^{P^h}$ , i.e.,

$$A^{\tilde{P}^h} := A^{P^h} \setminus \bigcup_{i' \in S^h} \{(i', d') : d' \in A_{i'}^{P^h} \text{ and } \mu(i') \neq d'\}$$

- \* for each  $d \in D$ , the ranking list  $\tilde{P}_d^h$  is the ordering over  $A_d^{\tilde{P}^h}$  that coincides with  $P^h$  on  $A_d^{\tilde{P}^h}$ .

**h.3** Construct  $P^{h+1} := R(\tilde{P}^h)$  and go to Step  $h + 1$ .

Note that if we do not observe the final matching then the algorithm stops at step 2.2. The algorithm implicitly partitions candidates in three sets:

- Candidates for which we will predict their matches for the proto-matching problem: Any candidate who does *not* belong to the sets  $S^2, S^3, \dots$ , but there is a problem  $P^h$  in which the candidate is ranked by only one department, say,  $d$ , and is among the top  $q_d$  candidates in that department's ranking.
- Candidates for which we use their observed match, i.e., they belong to the sets  $S^2, S^3, \dots$
- All the other candidates.

For each step  $h \geq 2$ , the candidates in  $S^h$  play a particular role. A candidate  $i \in S^h$  is called a **market star** at step  $h$ , that is,  $i$  is a matched candidate under  $\mu$  and is ranked among the top  $q_d$  candidates in all the acceptable departments  $d \in A_i^{P^h}$ , with  $|A_i^{P^h}| \geq 2$ . In that case, we say that a market star is *extracted* in the algorithm. By construction, the algorithm makes sure that the match of such candidate is only used once, i.e., for every  $h, h' \geq 2$ ,  $S^h \cap S^{h'} = \emptyset$ . Adding the fact that there is a finite number of candidates and departments, we deduce that the algorithm eventually stops after a finite number of steps.

**Definition 5** In the Predicting Algorithm, the match of a candidate  $i$  is **predicted** at step  $h$  if there is no  $h' \neq h$  such that  $i \in S^{h'}$  and  $h$  is the smallest index such that either:

- (a)  $A_i^{P^h} = \{d\}$  for some  $d$  and  $|\{j : jP_d^h i\}| < q_d$ ; or
- (b)  $A_i^{P^h} = \emptyset$ .

In case (a) the **prediction** is  $d$  and it is **consistent** with  $\mu$  if  $\mu(i) = d$ . Similarly, in case (b) the prediction is  $i$  and it is consistent with  $\mu$  if  $\mu(i) = i$ .

The Predicting Algorithm can use any matching  $\mu$  as an input. This implies that the predictions used made through the algorithm are not necessarily consistent. However, if the matching  $\mu$  is stable for a job market problem  $\succ$  that is  $P$ -compatible then predictions turn out to be consistent.<sup>16</sup>

**Proposition 6** *Let  $(P, \mu)$  be a pair such that  $\mu$  is a stable matching of an unobserved job market problem  $\succ \in \Theta(P)$ . All the predictions of the algorithm with input  $(P, \mu)$  are consistent with  $\mu$ .*

Since at Step 1 the Predicting Algorithm does not use market stars we obtain the following corollary: the final matching of candidates whose match is predicted in Step 1 does not depend on the profile of candidates' preferences.

**Corollary 1** *Let  $(P, \mu)$  be a pair such that  $\mu$  is a stable matching of an unobserved job market problem  $\succ \in \Theta(P)$ . If the match of a candidate is predicted in Step 1 of the Predicting Algorithm, then the match of that candidate is the same, for every stable matching of every job market problem in  $\Theta(P)$ .*

The predictions obtained from the Predicting Algorithm obviously depend on the match of the market stars, i.e., the candidates from the sets  $S^1, S^2, \dots$ . We can show that the predictions of our algorithm remain the same for any preference profile of the candidates that does not alter the top choices of market stars in a certain way. To see this, suppose that  $\mu$  (the input of the algorithm) is a stable matching for some problem  $\succ$ . We will observe that if a candidate  $i$  is a market star at some step  $h$  ( $i \in S^h$ ) then  $\mu(i)$  is necessarily candidate  $i$ 's top choice among the departments in  $A_i^{P^h}$ . Consider now the job

---

<sup>16</sup>The market stars are extracted consistently in the algorithm. That is, if  $i$  is a market star at step  $h$  then  $\mu(i)$  is still an acceptable department for the candidate  $i$  in the problem  $P^h$ , and remains thus the unique acceptable department in the next steps. This can be deduced from Lemma 4 in the Appendix, which shows that  $\mu(i)$  is never deleted from the list of acceptable departments of  $i$ .

market problem  $\succ' = (\succ'_i, \succ_{-i})$ , with  $\succ'_i$  such that  $\mu(i)$  is also the top choice among the departments in  $A_i^{P_h}$ , and let  $\mu'$  be a stable matching for  $\succ'$ . As the next proposition shows, running the Predicting Algorithm with input  $(P, \mu)$  or  $(P, \mu')$  yield the same predictions, and the market stars identified throughout steps 2, 3, ... are identical.

To prove this, we need to define a subset of the compatible job market problems. Given  $(P, \mu)$ , let  $\Theta(P, \mu)$  be the set of job market problems  $\succ' \in \Theta(P)$  such that, for every market star  $i$  at some step  $h$ ,

$$\mu(i) \succeq'_i d \text{ for all } d \in A_i^{P_h}.$$

**Proposition 7** *Let  $(P, \mu)$  be a pair such that  $\mu$  is a stable matching of an unobserved job market problem  $\succ \in \Theta(P)$  and consider the Predicting Algorithm with input  $(P, \mu)$ . Then for every problem  $\succ' \in \Theta(P, \mu)$  and matching  $\mu'$  stable for  $\succ'$ ,*

1. *The set of market stars is the same with input  $(P, \mu)$  and  $(P, \mu')$ .*
2.  *$\mu'(i) = \mu(i)$  for every candidate  $i$  whose match is predicted and for every market star  $i$  ( $i \in S^h$  for some  $h \geq 2$ ).*

*In addition, the Predicting Algorithm allows for a partial identification of the market stars' preferences, that is, the unobserved job market problem  $\succ$  must belong to  $\Theta(P, \mu)$ .*

The last statement of Proposition 7 simply says that, for every market star  $i$  at step  $h$  of the algorithm,  $\mu(i)$  is  $i$ 's most preferred department among those remaining at step  $h$  (and acceptable for  $i$ ).

The main feature to be deduced from Proposition 7 is stated explicitly in the next corollary.

**Corollary 2** *If the match of a candidate is predicted in the Predicting Algorithm then the order in which that candidate ranks the acceptable departments is irrelevant for the final match of the candidate.*

Finally, note that the Predicting Algorithm does not fully clear the market. Some departments and candidates are not considered at any step of the algorithm. Typically, this means that the remaining candidates are not top ranked in the remaining acceptable departments of the final proto-matching problem. Those assignments are the most difficult to sort out in the market. From Proposition 7, we only know that the preferences of the remaining candidates affect neither the matches of the stars and nor the predictions.

## 5. Empirical analysis

### 1. *The French academic job market*

We consider in this paper junior positions, called *Maître de Conférences*. They correspond to a lectureship in the U.K., or to an assistant professorship in the US. Junior positions in France are civil servant, tenured positions.<sup>17</sup> We first outline the general timeline of the French job market, and then comment on several aspects. There is much to say about this market, but to avoid unnecessary digressions we only present the aspects that are relevant for this paper (more details can be found in the Web Appendix).

Unlike, for instance, the US academic job market, most of the hiring procedure in French universities is centralized.<sup>18</sup> During the hiring season all departments first submit to the Ministry of Higher Education an ordered list of candidates (after interviewing them, usually in May), and then candidates are invited to submit a preference list over the positions that ranked them (usually in June). The final step is made by the ministry, where an algorithm is run to compute a match. Haeringer and Iehlé (2010) showed that the algorithm used in France is equivalent to the Deferred Acceptance algorithm with candidates proposing. That is, the matching computed is stable, and thus comprehensive in the context of proto-matching problems.

### 2. *Market for mathematicians*

In 1998 a small group of young mathematicians set up a web site, *Opération Postes*, inviting recruiting committees to announce the lists of candidates to be interviewed as well as the rankings of candidates that will be submitted to the clearinghouse (the ministry), as soon as these would be decided.<sup>19</sup> The community of mathematicians was very responsive and the web site quickly became a central tool in the job market.<sup>20</sup> The data for each position (interviewees list and rankings) is usually uploaded by the the chairs of the

---

<sup>17</sup>Tenure is made official after one year of service, but it is extremely rare that tenure is not confirmed.

<sup>18</sup>Since 2008 there is a “decentralized” market running parallel to the regular, centralized market, but with much less job openings. See the Web Appendix for more details.

<sup>19</sup><http://postes.smai.emath.fr/>.

<sup>20</sup>The web site is now supported by the main professional mathematical societies in France: the French Mathematical Society (SMF), the Society for Applied and Industrial Mathematics (SMAI, which hosts the web site), and the French Society of Statistics (SFdS).



recruiting committees themselves (and if not, by a member of the committee). On average, about 90–95 percent of the job openings’ interview lists and rankings are available.<sup>21</sup> The data of *Opération Postes* is public, although not in a format that makes it immediately usable for any analysis. There are many misspellings, and we sometimes found confusions between the married and maiden names of some female candidates. By cross-referencing the data with other sources we were able to compose a clean dataset.<sup>22</sup>

We also collected for each year the assignment of candidates to departments. This assignment is computed by the Ministry of Higher Education by using candidate’s submitted preference lists over the departments and the rankings of candidates established by the recruiting committees.<sup>23</sup>

### 3. Data analysis

Our data does not contain the preferences submitted by the candidates in June. Such lists are considered confidential by the Ministry of Higher Education; access to this data has been denied. It is nevertheless reasonable to assume that whenever a candidate appears on the ranking for a position that position is also *acceptable* for the candidate, i.e., he would prefer to be matched to that position than not being matched to any position. There are several observations that uphold this **acceptability assumption** for that market.

First, a candidate can only be ranked for a position if he has been interviewed, *on campus*, for that position. Since transportation and lodging costs accrue to the candidates it is unlikely that a candidate would bear the cost of applying for a position he would

---

<sup>21</sup>The missing data usually concerns overseas departments (e.g, La Réunion, Martinique or New Caledonia). The web site also invites the recruiting committees to submit the date when the interview list will be decided, the date of the interviews, and the date when the ranking will be decided. We discarded the positions that are reserved for administrative transfer (*postes à la mutation*), i.e., positions that are only open to candidates that are already *Maître de Conférences*. Such positions represent only a small fraction of the total number of openings, most positions being reserved for rookies.

<sup>22</sup>To be able to apply to a job opening candidates must first pass an evaluation by a national committee. The lists of candidates that have passed that evaluation (without spelling errors) is published by the Ministry of Higher Education and is freely available at <https://www.legifrance.gouv.fr/initRechJO.do>.

<sup>23</sup>Unfortunately, the assignment computed by the Ministry is not published. However, all of the assignments could have been easily collected manually using web searches, looking at the departments’ lists of faculty, submitted rankings (hired candidates have to be ranked by the departments), and candidates résumés (some departments also have an archive of the candidates they hired the previous years).

eventually refuse.<sup>24</sup> The timeline of the job market is another element to sustain our assumption. Interviews take place in May for a position that starts in September. That is, candidates still looking for a job have usually few outside options (if any). Finally, the vast majority of candidates have French degrees (and are French), and mathematics departments in France rank well according to international standards. Adding the fact that the departments offer tenured positions, such jobs are very attractive for mathematicians.

Using the observed matching we can also check, to some extent, whether our acceptability assumption is satisfied. Since we know that the final matching must be comprehensive (stable), any candidate  $i$  ranked for a position above the candidate that has been hired must be hired by another department. If this is not the case, then we must conclude that candidate  $i$  declared that position unacceptable in June. So from the observation of the matching we can refine the candidates' sets of acceptable departments.<sup>25</sup>

After carrying out this necessary exercise, our analysis in this paper will consider the rankings submitted by the recruiting committees, where a candidate  $i$  ranked by a department  $d$  means that  $(i, d)$  is a mutually acceptable pair. So our data is a real-life example of proto-matching problems and we can use the Predicting Algorithm from Section 4. Before describing our results, we first observe that the predicted matches we obtain are always correct.

Table 3 together with Table 3 of the Web Appendix give a good picture of the complexity of the problems. In 2013, for instance, there are 73 positions and 193 candidates ranked at least once in one of them; the length of the ranking is between 5 and 7 in 75 percent of the positions. For each year in our dataset, we first identify all the pairs that form impossible matches. Once all such pairs  $(c, d)$  are identified we simply delete candidate  $c$  from the ranking list of department  $d$ . From this reduced market we count the number of positions' matchings that are predicted. Table 3 reports for the years 1999 to 2013, in this order: the number of job openings, the number of candidates that ap-

---

<sup>24</sup>Interviews are not “flyouts” where the candidate can meet his (potentially) future colleagues and have a better knowledge of the department. The interview lasts only 20 or 30 minutes, and the candidate leaves the department quickly after the interview.

<sup>25</sup>The number of such violations and violators for every year is given in the Web Appendix, Table 5. Most of those candidates turn out to be also participating to the job market for another discipline (e.g., physics or biology), obtaining a job in a research institution (e.g., CNRS, INRIA, etc.) or getting a job at a foreign institution.

pear in the rankings, the number of predictions of candidates’ matches and the number of predictions of positions’ matches (the percentage in parenthesis relates the number of predictions to the numbers of candidates and job openings, respectively). The last two columns (with the “total” term) are obtained when running our Predicting Algorithm. The input matching we used in the algorithm is the observed matching, i.e., the matching computed by the Ministry of Higher Education.<sup>26</sup>

We are not aiming to identify what drives departments’ strategies when establishing their rankings of candidates, a future research project will address that question. Nevertheless, Table 3 is illustrating the *impact* that department’s strategies have on final outcomes (together with the set of mutually acceptable partners). Indeed, recall that in the French market *all* departments establish their rankings of candidates *before* candidates are called to submit their preferences. Table 3 shows that, on average, nearly 45 percent of the positions are awarded independently of the way candidates rank the positions in their preferences. If we consider the (observed) choices of market stars to refine our analysis, we end up with almost 60 percent of the positions’ matches being predicted. In other words, the way candidates rank their choices have little impact on the final matching. Table 3 is also instructive, from candidates’ viewpoint. On average, we are able to predict *ex ante* the matches of 60 percent of the candidates, be they matched to a department or not. According to Corollary 2, for 78 percent of the candidates, the way each of them ranks his acceptable departments is irrelevant for his final match (all other things remaining equal), and up to 91 percent in 2012.

**Remark 4** The French academic job market is clearly unbalanced: there are more candidates than positions available. In a recent contribution, Ashlagi, Kanoria and Leshno (2017) explore the impact of such imbalances, showing that agents on the short side of the market (the departments in our case) tend to be matched with their top choices. This is confirmed by Table 4 in the Web Appendix, showing that nearly half of the hired candidates are ranked either top or second by the department that hire them.

---

<sup>26</sup>We consider in the Predicting Algorithm a less severe definition of stars, that we call pseudo stars (see Section D.2 of the Web Appendix). Proposition 7 remains valid. Also, across all years, the Predicting algorithm stops after 5 to 9 steps, the average number of steps is 6.8 (i.e., 5.8 iterations of steps h.1, h.2 and h.3).

Table 3: Predictions

Year	# Positions	# Candidates	# Cand pred.	# Pos. pred	# Cand. total pred.	# Pos. total pred.
1999	121	250	101 (40.4%)	36 (29.8%)	154 (61.6%)	52 (43.0%)
2000	89	199	128 (64.3%)	44 (49.4%)	159 (79.9%)	58 (65.2%)
2001	86	194	121 (62.4%)	39 (45.3%)	168 (86.6%)	60 (69.8%)
2002	70	164	105 (64.0%)	35 (50.0%)	141 (86.0%)	50 (71.4%)
2003	97	213	132 (62.0%)	46 (47.4%)	171 (80.3%)	64 (66.0%)
2004	73	186	131 (70.4%)	37 (50.7%)	161 (86.6%)	48 (65.8%)
2005	89	207	122 (58.9%)	37 (41.6%)	157 (75.8%)	53 (59.6%)
2006	128	268	140 (52.2%)	52 (40.6%)	209 (78.0%)	79 (61.7%)
2007	115	261	169 (64.8%)	56 (48.7%)	221 (84.7%)	79 (68.7%)
2008	111	213	115 (54.0%)	45 (40.5%)	141 (66.2%)	55 (49.5%)
2009	127	250	121 (48.4%)	43 (33.9%)	157 (62.8%)	59 (46.5%)
2010	114	249	125 (50.2%)	39 (34.2%)	150 (60.2%)	47 (41.2%)
2011	121	273	144 (52.7%)	49 (40.5%)	223 (81.7%)	78 (64.5%)
2012	83	213	152 (71.4%)	49 (59.0%)	194 (91.1%)	64 (77.1%)
2013	73	193	147 (76.2%)	44 (60.3%)	169 (87.6%)	54 (74.0%)
Average	99.8	222.2	130.2 (59.5%)	43.4 (44.8%)	171.7 (77.9%)	60 (61.6%)

#### 4. Simulations

One may wonder what drives the number of impossible matches we can identify and, more generally, the number of matchings that can be predicted. It turns out that across the 15 years of data a large number of candidates are ranked only once. Table 4 (first row) shows the average rank distribution for the years 1999-2015. Undoubtedly, the high proportion of candidates being ranked only once or twice (almost 70 percent) has a great impact on the number of impossible matches present in the ranking profiles. Indeed, a candidate ranked only once is a trivial singleton block, so all candidates ranked below him or her are impossible matches. Since the number of impossible matches is a key parameter to determine the number of matches that can be predicted, we may legitimately conjecture that our prediction rate is due to the skewness of the distribution. If we consider only hired candidates (second line of Table 4), the distribution is still less skewed, though.

Table 4: Rank distribution (average 1999–2013)

# times ranked	1	2	3	4	5	6	7 or more
All candidates	48.3%	20.2%	12.5%	7.2%	4.4%	2.6%	5.7%
Hired candidates	31.7%	20.6%	15.8%	11.2%	6.9%	4.7%	9.2%

There are two variables that influence the existence of blocks (and thus of impossible matches), beyond the number of times each candidate is ranked. The first variable is the relative ranking of each candidate. A profile where most of the candidates being ranked first are also ranked only once is clearly more predictable than another profile where most of the candidates being ranked only once are at the bottom of the rank lists. The second variable relates to the overlap between the rank lists in terms of the sets of candidates being ranked. To see this, consider two candidates being ranked only twice. If these two candidates are ranked by the same departments they constitute a block for any candidate ranked below them and for the candidates ranked between them. However, they are less likely to be part of a block if they are not ranked by the same departments.

To understand the impact of each of these parameters we ran two types of simulations, for each year in our dataset. The first type of simulation (called *reshuffled rankings*) simply consists of reshuffling randomly departments' preferences (independently), thereby maintaining the overlaps between the rank lists. For the second type of simulations

(called *simulated rankings*) we restrict to the profiles such that the rank distribution of the candidates (first line of Table 4) and the distribution of the length of departments' rankings are the same as for the year under consideration. Within this subset of profiles we pick one at random.

For each year, we ran 10,000 simulations for each specification (simulated rankings and reshuffled rankings). Figure 1 shows the averages (over 10,000 runs) of the percentage of positions being predicted for both randomizations protocols and the prediction using the original rankings.<sup>27</sup>

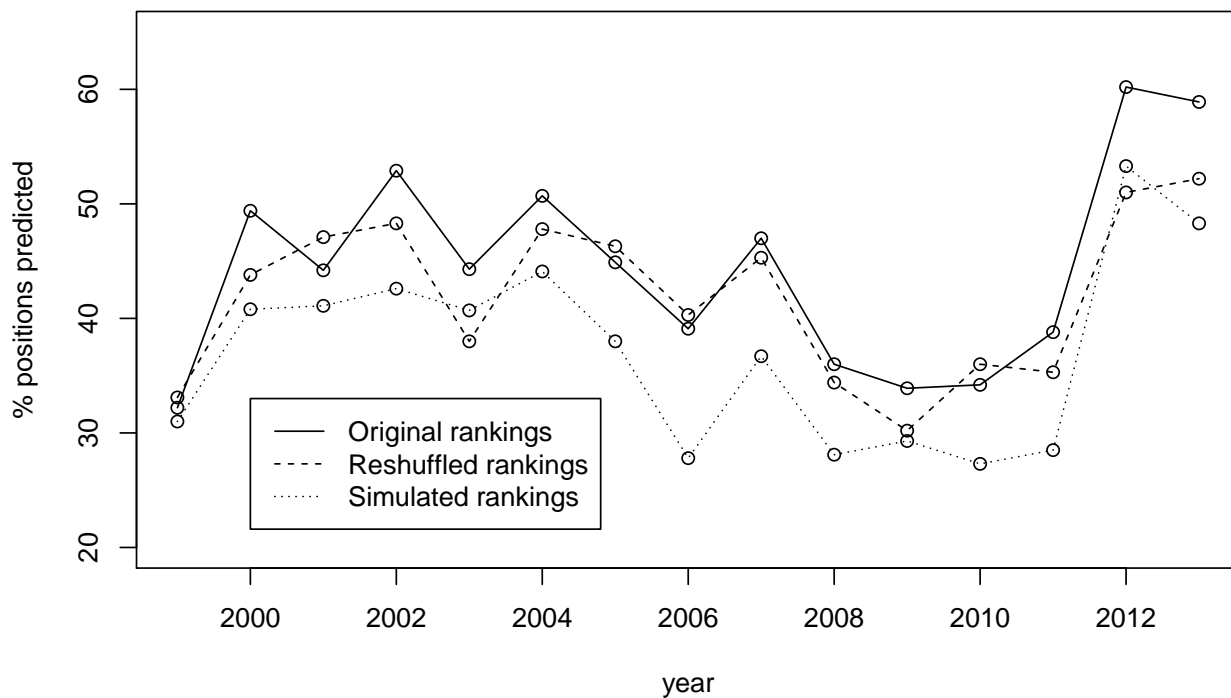


Figure 1: Prediction rate

<sup>27</sup>To make the comparison meaningful between observed and simulated rankings we did not withdraw the candidates that violate the acceptability assumption (we did withdraw these candidates for the results in Table 3). Table 8 in the Web Appendix presents, for each year, the average percentage of positions predicted for the random and reshuffled rankings, and the percentage of predicted positions for the original rankings when we do not withdraw the candidates violating the acceptability assumption.

Reshuffled rankings (resp. the original rankings) allow on average about 15 percent (resp. 22 percent) more predictions than simulated rankings. This difference indicates that the high level of prediction we observe in the original data is mostly due to the overlaps in departments ranking. Perhaps more surprising is the difference in the prediction rates of the reshuffled and original rankings, which is on average only of 5.8 percent. This suggests that the way departments rank the candidates are not as important as who they rank.

## 6. Concluding remarks

Our main result, Theorem 2, shows that identifying impossible matches relies on a combination of two factors: relative rankings and mutual acceptability.

The role of relative rankings shows up in the second condition of a block when we truncate, below some candidates, the proto-matching problem restricted to a subset of candidates (the block). This is precisely how we are able to connect Hall’s marriage conditions to stability. However, extracting a precise pattern of rankings that would guarantee or not the existence of impossible matches seems unattainable. Our examples and simulations indeed illustrate how ambiguous can be the impact of rankings on the existence of impossible matches. For instance, increasing the correlation between some departments’ rankings can create new impossible matches while making some candidates no longer impossible matches for some departments.

The set of mutually acceptable pairs obviously also plays an important role in the definition of a block (for both the first and second conditions). This set not only determines the imbalance between each side of the market but also specifies the overlaps of the mutually acceptable pairs, which is a key implicit element in the definition of a block. The presence of many unacceptable pairs implies that there can be “local” imbalances that create bottlenecks when matching certain candidate-department pairs, even if the total number of positions exceeds the total number of candidates. That is, local competition matters.<sup>28</sup>

Our results are thus of particular relevance when at least one side of the market submits short preference lists, which is a salient feature of most (if not all) real-life matching markets. There are various explanations for this stylized fact. For instance, some matching

---

<sup>28</sup>We thank a referee for suggesting this viewpoint.

markets involve a large number of participants or take place on a non-trivial geographic extent, making it virtually impossible to expect complete preference lists by all participants. Also, some matching institutions may constrain agents with respect to the length of their preference lists (e.g., position choice in New York City, university admission in Spain or the French academic job market).<sup>29</sup>

Analyzing the effects of competition on stable outcomes is a direction taken by Ashlagi, Kanoria and Leshno (2017). Unlike us, they assume that every candidate-department pair is acceptable (i.e., complete preference lists) and specify preferences of each side (random, iid).<sup>30</sup> They show that the slightest imbalance between the number of agents on the two sides of the market is sufficient to get a set of stable matchings that is essentially unique. Our contribution suggests that the global imbalance condition is not any more sufficient when the preference lists are incomplete. Instead, Hall’s marriage conditions, i.e., local imbalance conditions, might provide the adequate tool for capturing local competition effects.

## A. Appendix

### 1. Examples

The next examples show that the requirements of Definition 4 are all independent from each other.

**Example 3** Let  $I = \{i_0, i_1, i_2, i_3\}$  and  $d = \{d_0, d_1, d_2, d_3, d_4\}$ , and assume that for each department  $d \in D$ ,  $q_d = 1$ . The admissible departments for each candidate are given by  $A_{i_0} = \{d_0\}$ ,  $A_{i_1} = \{d_0, d_1, d_2\}$ ,  $A_{i_2} = A_{i_3} = \{d_1, d_2, d_3, d_4\}$ . The departments’ preferences are described in Table 5.

The matching  $\mu$  such that  $\mu(i_0) = d_0$ ,  $\mu(i_h) = d_h$  for  $h = 1, 2, 3$  is comprehensive and maximal, so from Proposition 2 candidate  $i_0$  is not an impossible match for department  $d_0$ . In this example condition 1 of the definition of a block is not satisfied, because for

---

<sup>29</sup>Abdulkadiroğlu, Pathak and Roth (2005), Haeringer and Klijn (2009), Calsamiglia, Haeringer and Klijn (2010), Pathak and Sönmez (2013) or Bodoh-Creed (2013) constitute additional accounts on the length of submitted preference lists in centralized matching markets.

<sup>30</sup>In contrast, our approach to candidates’ potential preferences over positions is distribution-free, i.e., we do not make any assumption on the distribution over candidates’ preference profiles.



Table 5: Condition 1 not satisfied — too few candidates

$P_{d_0}$	$P_{d_1}$	$P_{d_2}$	$P_{d_3}$	$P_{d_4}$
$i_1$	$i_2$	$i_3$	$i_2$	$i_3$
$i_0$	$i_3$	$i_2$	$i_3$	$i_2$
	$i_1$	$i_1$		

every set  $\mathbf{J} \subseteq I \setminus \{i_0\}$  it holds that  $|\mathbf{J}| < \sum_{d \in A_{\mathbf{J}}} q_d$ . In this case, condition 2 is satisfied but has clearly no bite since, for each  $\mathbf{J} \subseteq I \setminus \{i_0\}$  with  $d_0 \in A_{\mathbf{J}}$  (e.g.,  $\mathbf{J} = \{i_1\}$ ), there cannot be a set  $\mathbf{K} \subseteq \mathbf{J} \setminus \mathbf{J}_{i_0, d_0}$  that satisfies the cardinality requirement of condition 2.  $\square$

**Example 4** Let  $I = \{i_0, i_1, \dots, i_7\}$ ,  $D = \{d_0, d_1, d_2, d_3\}$ , and assume that  $q_d = 1$  for each  $d \neq d_4$ , and  $q_{d_3} = 4$ . The departments' preferences are described in Table 6.

Table 6: Condition 1 not satisfied — no perfect match

$P_{d_0}$	$P_{d_1}$	$P_{d_2}$	$P_{d_3}$
$i_1$	$i_2$	$i_3$	$i_3$
$i_0$	$i_1$	$i_2$	$i_4$
	$i_5$	$i_6$	$i_7$

Candidate  $i_0$  is not an impossible match for department  $d_0$ : the matching  $\mu$  such that  $\mu(i_0) = d_0$ ,  $\mu(i_h) = d_h$  for  $h = 1, 2, 3$ , is comprehensive. Consider the set  $\mathbf{J} = \{i_1, \dots, i_7\}$ . So,  $|\mathbf{J}| = 7 = \sum_{d \in A_{\mathbf{J}}} q_d$ . For this set  $\mathbf{J}$ , however, condition 1 is not satisfied because there cannot be a perfect match between any set  $\mathbf{J}' \subseteq \mathbf{J}$  and the departments in  $A_{\mathbf{J}}$  (because, for each matching,  $d_3$  does not fill its capacity).

Unlike Example 3, condition 2 of the definition of a block does bite. To see this, note that we need to consider all sets  $\mathbf{K} \subseteq \mathbf{J} \setminus \{i_1\}$  such that  $|\mathbf{K}| = |\mathbf{J}| - \sum_s q_d + 1 = 1$ . If  $\mathbf{K} = \{i_2\}$  Hall's marriage condition (3) does not hold in  $\overline{P}^{d_0}(\mathbf{J}, \mathbf{K})$  because we cannot match  $i_1$ .<sup>31</sup> If  $\mathbf{K} = \{i_3\}$ , then again Hall's marriage condition does not hold in  $\overline{P}^{d_0}(\mathbf{J}, \mathbf{K})$ . Indeed, either  $i_1$  or  $i_2$  must remain unmatched.

---

<sup>31</sup>Note that the reason Hall's condition does not hold in  $\overline{P}^{d_0}(\mathbf{J}, \mathbf{K})$  is because  $i_1$  cannot be matched even though  $i_1$  is admissible at  $d_0$ , i.e.,  $d_0$  have zero capacity does not preclude  $i_1$ 's admissibility at  $d_0$  (see Remark 3).

It remains then to consider candidates in  $\{i_4, i_5, i_6, i_7\}$ . Routine examination shows that for any such candidate we cannot match all admissible candidates in  $\overline{P}^{d_0}(\mathbf{J}, \mathbf{K})$ , which proves the claim.  $\square$

**Example 5** Let  $I = \{i_0, i_1, i_2, i_3, i_4\}$ ,  $D = \{d_0, d_1, d_2, d_3\}$  and assume that  $q_d = 1$  for each  $d \in D$ . The departments' preferences are depicted in Table 7. In this example candidate  $i_0$  is not an impossible match for  $d_0$ . Here the only possible set  $\mathbf{J}$  satisfying condition 1 is  $\{i_1, i_2, i_3, i_4\}$ . Since  $|\mathbf{J}| = 4 = \sum_{d \in A_{\mathbf{J}}} q_d$ , we have to consider sets  $\mathbf{K}$  such that  $|\mathbf{K}| = 1$ . To see that condition does not hold it suffices to consider  $\mathbf{K} = \{i_4\}$ . The matching  $\mu$ , where  $\mu(i_1) = d_1$ ,  $\mu(i_2) = d_2$  and  $\mu(i_3) = d_3$  is such that all admissible candidates in  $\overline{P}^{d_0}(\mathbf{J}, \{i_4\})$ , i.e.,  $i_1, i_2$  and  $i_3$ , are matched to a department. That is, Hall's marriage condition (3) is satisfied in  $\overline{P}^{d_0}(\mathbf{J}, \{i_4\})$ .  $\square$

Table 7: Condition 2 not satisfied

$P_{d_0}$	$P_{d_1}$	$P_{d_2}$	$P_{d_3}$
$i_1$	$i_2$	$i_3$	$i_3$
$i_0$	$i_1$	$i_2$	$i_4$

## 2. Preliminaries before the proof of Theorem 2

Let  $P = (I, D, (P_d, q_d)_{d \in D}, A)$  be a proto-matching problem. For our purposes it is convenient to describe it by a non-directed graph  $G^P(V, E)$  where the set of vertices is  $V = I \cup D$  and the set of edges is  $E = \{(v, v') \in V \times V : v \in A_{v'}\}$ . A **path**  $\pi$  in the graph  $G^P(V, E)$  is a (finite) sequence of pairwise distinct vertices  $(v_1, v_2, \dots, v_k)$  such that for each  $h = 1, \dots, k - 1$  we have  $(v_h, v_{h+1}) \in E$ . Given a matching  $\mu$ , an edge  $(v, v')$  is **free** if  $v$  is not matched to  $v'$  under the matching  $\mu$ . Otherwise the edge is **matched**. A path  $\pi \in G^P(V, E)$  is an **alternating path for a matching**  $\mu$  if it alternates matched and free edges. Given a matching  $\mu$ , a vertex  $v$  is **neutral** if either  $v \in I$  and  $\mu(v) = v$ , or  $v \in D$  and  $|\mu(v)| < q_v$ . A path  $\pi = (v_1, \dots, v_k)$  in  $G^P(V, E)$  is an **augmenting path** for a matching  $\mu$  if it is an alternating path for  $\mu$  where both  $(v_1, v_2)$  and  $(v_{k-1}, v_k)$  are free edges and where both  $v_1$  and  $v_k$  are neutral vertices. A feasible matching  $\mu$  is **maximum** in  $P$  if there does not exist a feasible matching  $\mu'$  in  $P$  that matches more candidates than  $\mu$ , i.e.,  $|\{i \in I : \mu(i) \neq i\}| \geq |\{i \in I : \mu'(i) \neq i\}|$  for any feasible matching  $\mu'$  in

$P$ . A well-known result by Berge (1957) establishes that a matching is maximum if, and only if, there is no augmenting paths. We rephrase Berge’s theorem in our context.<sup>32</sup>

**Theorem 3 ((adapted from) Berge, 1957)** *Let  $P$  and  $\mu$  be respectively a proto-matching problem and a feasible matching in  $P$ . The matching  $\mu$  is maximum for  $P$  if, and only if, there is no augmenting path in  $G^P(V, E)$  for  $\mu$ .*

Let  $\mu$  be a matching and let  $\pi = (i_1, d_1, i_2, d_2, \dots, i_k, d_k)$  be an augmenting path for  $\mu$ . Without loss of generality we consider in the remainder only augmenting paths starting with a (neutral) candidate and ending with a (neutral) department.<sup>33</sup> The matching  $\mu' = \mu \oplus \pi$  is the matching obtained when updating  $\mu$  with the path  $\pi$ , i.e.,  $\mu'(i) = \mu(i)$  for each  $i \notin \pi$  and  $\mu'(i_h) = d_h$  for each  $i_h \in \pi$ ,  $h = 1, \dots, k$ .

**Lemma 1** *Let  $\mu$  be a comprehensive matching in  $P$ . If  $\mu$  is not a maximum matching then there exists an augmenting path  $\pi$  such that  $\mu' = \mu \oplus \pi$  is comprehensive in  $P$ .<sup>34</sup>*

**Proof** Let  $\mu$  be a comprehensive matching for  $P$  and suppose that  $\mu$  is not a maximum matching. So, by Berge’s Theorem there exists an augmenting path for  $\mu$ . Let  $\pi = (i_1, d_1, i_2, \dots, d_k)$  be the minimum length augmenting path, and if there are several augmenting paths of the same length as  $\pi$  assume that  $\pi$  is one of those paths such that  $iP_{d_1}i_1 \Rightarrow \mu(i) \neq i$ . Let  $\mu' = \mu \oplus \pi$ . Note that  $\mu'$  matches a strict superset of agents than  $\mu$ . We claim that  $\mu'$  is comprehensive for  $P$ . To see this, suppose that  $\mu'$  is not comprehensive. So, there exists  $j$  such that  $\mu'(j) = j$  and  $j'$  such that  $jP_{\mu'(j')}j'$ . From the choice of  $\pi$ ,  $j' \neq i_1$ . Since  $\mu$  is comprehensive there exists  $h > 1$  such that  $j' = i_h$  (and thus  $\mu'(j') = d_h$ ). Consider now the path  $\pi' = (j, d_h, i_{h+1}, \dots, d_k)$ . Clearly,  $\pi'$  is an augmenting path. Since  $h > 1$ , the contradiction comes from the assumption of minimality of  $\pi$ . ■

**Remark 5** The proof of Lemma 1 suggests an algorithm to construct a maximum **and** comprehensive matching, which works as follows. Start with the empty matching, which

---

<sup>32</sup>Berge’s original theorem is stated for one-to-one matching problems. It is straightforward to extend Berge’s result to a many-to-one matching problem by adequately adapting the definition of a neutral vertex as we did.

<sup>33</sup>Let  $\pi = (v_1, v_2, \dots, v_k)$  be an augmenting path for some matching  $\mu$ , and suppose that  $v_1 \in D$ . Then it must be that  $v_k \in I$ . We can then consider the path  $\pi' = (v_k, v_{k-1}, \dots, v_2, v_1)$ , which is also, by definition, an augmenting path.

<sup>34</sup>Andersson and Ehlers (2016) use a similar result in a different context.

is trivially comprehensive. Now it suffices to identify augmenting paths of the smallest length and select the smallest path with the same criterion we chose  $\pi$  in the proof. This operation can be performed in polynomial time with a Breadth First Search algorithm.<sup>35</sup> It follows then that finding a maximum and comprehensive matching can be done in polynomial time.

### 3. Proof of Theorem 2

**Remark 6** To prove Theorem 2 it is more convenient to rely on the notion of **maximum** matching, which has been defined in the previous sub-section. More precisely, we will use the existence of a maximum and comprehensive matching to show the necessary condition of Theorem 2 (i.e., construction of a block). Note that a maximum matching is necessarily maximal.

#### *Necessity*

Let  $i_0$  be an impossible match for  $d_0$  in the proto-matching problem  $P$ . We need to show that there exists a block  $\mathbf{J}$  at  $(i_0, d_0)$ , with  $i_0 \notin \mathbf{J}$ . Let  $\widehat{P}$  be the problem that is identical to  $P$  except that in  $\widehat{P}$  candidate  $i_0$  finds only acceptable  $d_0$ .<sup>36</sup> Observe that if  $\mu$  is such that  $\mu(i_0) = d_0$  and  $\mu$  is maximal and comprehensive for  $P$  (resp.  $\widehat{P}$ ) then  $\mu$  is also maximal and comprehensive for  $\widehat{P}$  (resp.  $P$ ). So from Proposition 2  $i_0$  is an impossible match for  $d_0$  in  $P$  if, and only if,  $i_0$  is an impossible match for  $d_0$  in  $\widehat{P}$ . So it is without loss of generality that we can assume  $P$  being such that  $\mathbf{A}_{i_0} = \{\mathbf{d}_0\}$ .

Let  $\mu$  be a maximum and comprehensive matching in  $P$  such that  $A_{i_0} = \{d_0\}$ . By Remark 5, such a matching exists. So by Proposition 2 and the assumption that  $A_{i_0} = \{d_0\}$ ,  $\mu(i_0) = i_0$ . Hence,  $|\mu(d_0)| = q_{d_0}$ , otherwise  $\mu$  would not be maximum. The next procedure constructs explicitly a block at  $(i_0, d_0)$ . To start with, define the following sets  $D^0$  and  $\mathbf{J}^1$ ,

$$D^0 = \{d \in D : |\mu(d)| < q_d\}$$

$$\mathbf{J}^1 = \{i \in I : \mu(i) \neq i\} \setminus A_{D^0}.$$

---

<sup>35</sup>See Lovász and Plummer (1986).

<sup>36</sup>The problem  $\widehat{P}$  is such that  $\widehat{P}_{d_0} = P_{d_0}$  (and thus  $A_{d_0}^{\widehat{P}} = A_{d_0}$ ), and for each  $d \neq d_0$ ,  $A_d^{\widehat{P}} = A_d \setminus \{i_0\}$  and for each  $i, i' \neq i_0$ ,  $iP_d i'$  if, and only if,  $i\widehat{P}_d i'$ .

In other words,  $D^0$  is the set of departments that do not fill their capacity under the maximum and comprehensive matching  $\mu$ , and the set  $\mathbf{J}^1$  is the set of all matched candidates minus the candidates for department in  $D^0$ . For  $h \geq 1$ , we define recursively the following sets,

$$D^h = \left\{ d \in D \setminus \left( \cup_{h' < h} D^{h'} \right) : |\mu(d) \cap \mathbf{J}^h| < q_d \right\}$$

$$\mathbf{J}^{h+1} = \mathbf{J}^h \setminus A_{D^h}.$$

So, the set  $D^1$  is the set of departments that are acceptable by candidates in  $\mathbf{J}^1$  and that do not fill their capacity with candidates in  $\mathbf{J}^1$ . The set  $\mathbf{J}^2$  is then the set of candidates in  $\mathbf{J}^1$  minus the candidates acceptable for departments in  $D^1$ .

Observe that at each step we withdraw candidates, i.e.,  $\mathbf{J}^{h+1} \subseteq \mathbf{J}^h$ , for  $h \geq 1$ . The inclusion is strict until we eventually reach a step  $\ell$  such that  $\mathbf{J}^{\ell+1} = \mathbf{J}^\ell$ , which occurs when there is a step  $\ell$  such that  $D^\ell = \emptyset$ . We show later in the proof that  $\mathbf{J}^\ell \neq \emptyset$ . Note also that  $A_{\mathbf{J}^{h+1}} \cap D^h = \emptyset$ , for  $h \geq 0$ , and  $D^h \cap D^{h'} = \emptyset$  whenever  $h \neq h'$ ,  $h, h' < \ell$ .

We claim that

$$A_{\mathbf{J}^\ell} = D \setminus \left( \cup_{h < \ell} D^h \right). \quad (4)$$

To see this, observe first that  $A_{\mathbf{J}^{h+1}} \cap D^h = \emptyset$ , for each  $h \geq 0$ . Hence we only need to show that if  $d \in D \setminus \left( \cup_{h < \ell} D^h \right)$  then  $d \in A_{\mathbf{J}^h}$ , for each  $1 \leq h \leq \ell$ . Let  $h$  be such that  $1 \leq h \leq \ell$ . Since  $d \notin \cup_{h' < h} D^{h'}$ , it holds that  $d \in D \setminus \left( \cup_{h' < h} D^{h'} \right)$ . Since  $d \notin D^h$ , it must be necessarily true that  $|\mu(d) \cap \mathbf{J}^h| = q_d$ . Hence,  $d \in A_{\mathbf{J}^h}$ .

Let  $\mathbf{T}$  be the set of candidates, except  $i_0$ , who are unmatched under  $\mu$  but consider acceptable at least one department in  $A_{\mathbf{J}^\ell}$ ,

$$\mathbf{T} = \{i \in I \setminus \{i_0\} : \mu(i) = i \text{ and } A_i \cap A_{\mathbf{J}^\ell} \neq \emptyset\}$$

*Claim:* If  $\mathbf{T} \neq \emptyset$  then  $A_{\mathbf{T}} \subseteq A_{\mathbf{J}^\ell}$ .<sup>37</sup>

Suppose by way of contradiction that  $A_{\mathbf{T}} \not\subseteq A_{\mathbf{J}^\ell}$ . So there exists  $d \in A_{\mathbf{T}}$  such that  $d \notin A_{\mathbf{J}^\ell}$ . Hence, there is a candidate, say,  $i$ , such that  $i \in \mathbf{T}$  and  $d \in A_i$ , and, by Eq. (4),  $d \in D^h$  for some  $h < \ell$ . Since  $\mu$  is a maximum matching and  $\mu(i) = i$  it must be that  $|\mu(d)| = q_d$ , and thus  $d \notin D^0$ . So  $h \geq 1$ .

---

<sup>37</sup>It could be that  $\mathbf{T} = \emptyset$ . This is the case for instance if  $P_{d_0} = i_1, i_0$ ,  $P_d = i_2, i_1$  and  $q_{d_0} = q_{d_1} = 1$  with  $I = \{i_0, i_1, i_2\}$  and  $D = \{d_0, d_1\}$ . What is necessary for the proof is that, if  $\mathbf{T} \neq \emptyset$ , there is no department in  $A_{\mathbf{T}}$  that is not in  $A_{\mathbf{J}^\ell}$ .

We then have  $|\mu(d) \cap \mathbf{J}^{h'}| = q_d$  for each  $h' < h$ , and, from the definition of  $D^h$ ,  $|\mu(d) \cap \mathbf{J}^h| < q_d$ . Therefore, there exists a candidate, say,  $i_h$ , such that  $\mu(i_h) = d$ ,  $i_h \notin \mathbf{J}^h$  and  $i \in \mathbf{J}^{h'}$ . Since  $\mathbf{J}^h = \mathbf{J}^{h-1} \setminus A_{D^{h-1}}$  we have  $i_h \in A_{D^{h-1}}$ . It follows that there is a department  $d_h \in D^{h-1}$  such that  $i_h \in A_{d_h}$ . Note that  $D^h \cap D^{h-1} = \emptyset$  implies  $d_h \neq d$ . We have shown that  $d \in A_i$ ,  $\{d, d_h\} \subseteq A_{i_h}$ , so  $(i, d, i_h, d_h)$  is an alternating path.

From  $d_h \in D^{h-1}$  we deduce as before the existence of a candidate, say  $i_{h-1}$  such that  $i_{h-1} \notin \mathbf{J}^{h-1}$  and  $\mu(i_{h-1}) = d_h$ . So  $i_{h-1} \neq i$  because  $\mu(i) = i$ . Since  $i_h \in \mathbf{J}^{h-1}$ ,  $i_{h-1} \neq i_h$ . Hence,  $(i, d, i_h, d_h, i_{h-1})$  is an alternating path. Continuing this way we eventually end up with a department  $d_1 \in D^0$  and an alternating path  $\pi = (i, d, i_h, d_h, i_{h-1}, \dots, i_1, d_1)$ . However,  $d_1 \in D^0$  implies that  $|\mu(d_1)| < q_{d_1}$ , and thus  $\pi$  is an augmenting path. Berge's Theorem then implies that  $\mu$  is not maximum, a contradiction. So  $A_{\mathbf{T}} \subseteq A_{\mathbf{J}^\ell}$ .  $\square$

*Claim:*  $\mu(d_0) \subseteq \mathbf{J}^\ell$ .

Suppose by way of contradiction that there exists a candidate, say,  $i$  such that  $i \in \mu(d_0)$  and  $i \notin \mathbf{J}^\ell$ . So  $i \in A_{D^h}$  for some  $h < \ell$ . So there is a department  $d \in D^h$  such that  $i \in A_{d_h}$ . Observe that  $(d_0, i, d)$  is an alternating path. We can use the same arguments used for the previous claim to construct an alternative path  $(i, d, i_h, d_{h-1}, \dots, i_2, d_1)$  with  $d_1 \in D^0$ . Therefore, since  $\mu(i) = d_0$ ,  $i_0 \in A_{d_0}$  and  $\mu(i_0) = i_0$ ,  $(i_0, d_0, i, d, i_h, d_{h-1} \dots i_2, d_1)$  is an augmenting path. Again applying Berge's Theorem we deduce that  $\mu$  is not a maximum matching, a contradiction. So,  $\mu(d_0) \subseteq \mathbf{J}^\ell$ .  $\square$

Let  $\mathbf{J} = \mathbf{J}^\ell \cup \mathbf{T}$ . We claim that  $\mathbf{J}$  is a block at  $(i_0, d_0)$ .

Since  $|\mu(d_0)| = q_{d_0}$ ,  $\mu(d_0) \subseteq \mathbf{J}^\ell \subseteq \mathbf{J}$  implies  $\mathbf{J} \neq \emptyset$ . By construction  $|\mu(d)| = q_d$  for each  $d \in A_{\mathbf{J}^\ell}$  and  $\mu(i) \neq i$  for each  $i \in \mathbf{J}^\ell$ . Moreover,  $A_{\mathbf{J}^\ell} = A_{\mathbf{J}}$ , so  $\mu$  is a perfect matching between  $\mathbf{J}^\ell$  and  $A_{\mathbf{J}}$ , i.e., the first condition for a block is met with the set  $\mathbf{J}$  —where  $\mathbf{J}^\ell$  plays the role of  $\mathbf{J}'$  in Definition 4.

It remains to show that condition 2 of a block also holds. Suppose by way of contradiction that there exists a non-empty set  $\overline{\mathbf{K}}$ , with  $\overline{\mathbf{K}} \subseteq \mathbf{J} \setminus \mathbf{J}_{i_0, d_0}$  and  $|\overline{\mathbf{K}}| = |\mathbf{J}| - \sum_{d \in A_{\mathbf{J}}} q_d + 1$ , such that Hall's condition (3) is satisfied in  $\overline{P}^{d_0}(\mathbf{J}, \overline{\mathbf{K}})$ . This is tantamount to saying that there exists a matching  $\tilde{\mu}$  for  $\overline{P}^{d_0}(\mathbf{J}, \overline{\mathbf{K}})$  such that all admissible candidates are matched to a department. We construct the following matching  $\mu'$  for  $P$ . For each  $i \notin \mathbf{J}$ , let  $\mu'(i) = \mu(i)$ . For each  $i \in \mathbf{J}$  that is admissible in  $\overline{P}^{d_0}(\mathbf{J}, \overline{\mathbf{K}})$ , let  $\mu'(i) = \tilde{\mu}(i)$ . Finally, for each  $i \in \mathbf{J}$  that is not admissible in  $\overline{P}^{d_0}(\mathbf{J}, \overline{\mathbf{K}})$  let  $\mu'(i) = i$ .<sup>38</sup> Note that, since  $i_0 \notin \mathbf{J}$

---

<sup>38</sup>Note that the students in  $\mathbf{J}$  that are not admissible in  $\overline{P}^{d_0}(\mathbf{J}, \overline{\mathbf{K}})$  include the students in  $\overline{\mathbf{K}}$  but may

and  $\mu(i_0) = i_0$ ,  $\mu'(i_0) = i_0$ . By construction, all the candidates in  $\mathbf{J}$  that are matched under  $\mu'$  are matched to a department in  $A_{\mathbf{J}}$ . Also, from the construction of  $\mathbf{J}$ , there is no candidate in  $I \setminus \mathbf{J}$  that is matched to a department in  $A_{\mathbf{J}}$ . The following arguments will also make that clear.

*Claim:*  $\mu'$  is a comprehensive matching for  $P$ .

By construction,  $\mu'$  is a matching for  $P$  and it is feasible since  $\mu$  and  $\tilde{\mu}$  are both feasible. To demonstrate the second requirement of comprehensiveness, suppose by way of contradiction that there exists a department, say,  $d$  and two candidates, say,  $i$  and  $j$ , such that  $\mu'(j) = d$ ,  $\mu'(i) = i$  and  $iP_dj$ . We consider the following two cases.

*Case 1:*  $j \notin \mathbf{J}$ . Hence,  $\mu'(j) = \mu(j)$  and thus  $\mu(j) = d$ . Suppose that  $i \in \mathbf{J}$ . Since  $i$  is admissible for  $d$  in  $P$ ,  $d \in A_i$  and thus  $d \in A_{\mathbf{J}}$ . From the construction of  $\mathbf{J}$ ,  $\mu(d) \in \mathbf{J}$ . Hence,  $j \in \mathbf{J}$ , a contradiction. So  $i \notin \mathbf{J}$ . Hence  $\mu'(i) = \mu(i)$  and thus  $\mu(i) = i$ . This implies that  $\mu$  is not comprehensive for  $P$ , a contradiction.

*Case 2:*  $j \in \mathbf{J}$ . Since  $\mu'(j) \neq j$  it must be that  $\tilde{\mu}(j) = \mu'(j)$ . If  $i \in \mathbf{J}$  we then have that  $i$  is admissible in  $\overline{P}^{d_0}(\mathbf{J}, \overline{\mathbf{K}})$ . Therefore we should have  $\mu'(i) = \tilde{\mu}(i) \neq i$ . This is a contradiction with the initial assumption that  $\mu'(i) = i$ , so we have  $i \notin \mathbf{J}$ . Hence,  $\mu'(i) = \mu(i)$ . Since  $j \in \mathbf{J}$  we have, from  $\mu'(j) = d$ ,  $d \in A_{\mathbf{J}^e}$ . Thus,  $i \in \mathbf{T}$ . Since  $\mathbf{T} \subset \mathbf{J}$  we obtain  $i \in \mathbf{J}$ , a contradiction.  $\square$

Before concluding, we need two additional properties satisfied by  $\mu'$ .

Consider the candidates in  $Z = \{i : iP_{d_0}i_0\}$ . Note that we do not necessarily have  $Z = \mathbf{J}_{i_0, d_0}$ .<sup>39</sup> We show that all those candidates are matched to a department under  $\mu'$ . Let  $i$  be any candidate in  $Z$ . If  $i \in \mathbf{J}$ ,  $i$  is admissible in  $\overline{P}^{d_0}(\mathbf{J}, \mathbf{K})$ . So  $\tilde{\mu}(i) \neq i$  and thus  $\mu'(i) \neq i$ . Suppose now that  $i \notin \mathbf{J}$ . If  $\mu(i) \neq i$  then  $\mu'(i) = \mu(i)$  and thus  $\mu'(i) \neq i$ . If  $\mu(i) = i$  then  $i \in \mathbf{T}$  and thus  $i \in \mathbf{J}$ , a contradiction.

Second, we claim that  $|\mu'(d_0)| \leq q_{d_0} - 1$ . To see this, take any  $i \in \mu'(d_0)$ . This means that either  $i \notin \mathbf{J}$  (and thus  $\mu'(i) = \mu(i)$ ), or  $i \in \mathbf{J}$  (and thus  $\mu'(i) = \tilde{\mu}(i)$ ). The former case contradicts the earlier claim that  $\mu(d_0) \subseteq \mathbf{J}^\ell$  (because  $\mathbf{J}^\ell \subseteq \mathbf{J}$ ). So  $\mu'(d_0) = \tilde{\mu}(d_0)$ . From  $\tilde{\mu}(d_0) \subseteq \mathbf{J}$ ,  $|\tilde{\mu}(d_0)| \leq q_{d_0} - 1$ . We have thus  $|\mu'(d_0)| \leq q_{d_0} - 1$ .

Define  $\mu''$  as follows. For each candidate  $i \neq i_0$ ,  $\mu''(i) = \mu'(i)$ , and  $\mu''(i_0) = d_0$ . This

---

also include some students in  $\mathbf{J} \setminus \overline{\mathbf{K}}$ .

<sup>39</sup>Take for instance  $P_{d_0} = i_1, i_2, i_0$ ,  $P_{d_1} = i_3, i_1$  and  $P_{d_2} = i_2$ ,  $q_{d_0} = q_{d_1} = 1$  and  $q_{d_2} = 2$ . With our construction we have  $\mathbf{J} = \{i_1, i_3\}$ , and thus  $\mathbf{J}_{i_0, d_0} = \{i_1\}$  and  $Z = \{i_1, i_2\}$ .

matching is well defined since  $|\mu'(d_0)| \leq q_{d_0} - 1$ . It is also comprehensive in  $P$  because  $i \in Z$  implies  $\mu''(i) \neq i$ . If  $\mu''$  is not maximal, then consider the departments that do not exhaust their capacities under  $\mu''$  and fill these departments with the best ranked and unmatched remaining candidates (if any). That is, we can assume without loss of generality that  $\mu''$  is maximal. So, by Proposition 2,  $i_0$  is not an impossible match for  $d_0$ , a contradiction. It follows that condition 2 of a block holds and thus  $\mathbf{J}$  is a block at  $(i_0, d_0)$ .

### Sufficiency

The next two lemmata will be useful to prove the sufficiency part of Theorem 2.

**Lemma 2** *Let  $\mathbf{J}$  be a block at  $(i, d)$  and let  $\mu$  be a comprehensive matching in  $\overline{P}^d(\mathbf{J}, \emptyset)$  where  $\mu(j) \neq j$  for each  $j \in \mathbf{J}_{i,d}$ . Then*

$$|\{j \in \mathbf{J} : \mu(j) = j\}| > |\mathbf{J}| - \sum_{\hat{d} \in A_{\mathbf{J}}} q_{\hat{d}} + 1 \quad (5)$$

**Proof** Let  $r = |\mathbf{J}| - \sum_{\hat{d} \in A_{\mathbf{J}}} q_{\hat{d}} + 1$ . Since  $\mu$  is comprehensive it is feasible for  $\overline{P}^d$ , and there are at most  $\sum_{\hat{d} \in A_{\mathbf{J}}} q_{\hat{d}} - 1$  candidates in  $\mathbf{J}$  that are matched to a department under  $\mu$ . It follows that there exists a set  $\mathbf{K}$  made by  $r$  unmatched candidates. Hence,  $\mu$  is also a well defined feasible matching for  $\overline{P}^d(\mathbf{J}, \mathbf{K})$ . By assumption, it also holds that  $\mathbf{K} \subseteq \mathbf{J} \setminus \mathbf{J}_{i,d}$ . Since  $\mathbf{J}$  is a block, Hall's marriage condition (3) does not hold for  $\overline{P}^d(\mathbf{J}, \mathbf{K})$ , i.e., for every feasible matching  $\mu'$  for the problem  $\overline{P}^d(\mathbf{J}, \mathbf{K})$  there exists an admissible candidate  $j \in \mathbf{J} \setminus \mathbf{K}$  such that  $\mu'(j) = j$ . Therefore, there exists  $j \in \mathbf{J} \setminus \mathbf{K}$  such that  $\mu(j) = j$ . So, there are at least  $r + 1$  candidates in  $\mathbf{J}$  who are unmatched under  $\mu$ , as was to be proved. ■

**Lemma 3** *Let  $\mathbf{J}$  be a block at  $(i, d)$  and let  $\mu$  be a comprehensive matching for  $\overline{P}^d(\mathbf{J}, \emptyset)$  such that  $\mu(j) \neq j$  for each  $j \in \mathbf{J}_{i,d}$ . Then there exists a matching  $\mu'$  that is comprehensive for  $\overline{P}^d(\mathbf{J}, \emptyset)$  such that*

$$|\{j \in \mathbf{J} : \mu'(j) = j\}| = |\{j \in \mathbf{J} : \mu(j) = j\}| - 1. \quad (6)$$

and  $\mu'(j) \neq j$  for each  $j \in \mathbf{J}_{i,d}$ .



**Proof** From condition 1 of a block we can match  $\sum_{\hat{d} \in A_{\mathbf{J}}} q_{\hat{d}}$  candidates of  $\mathbf{J}$  in  $P(\mathbf{J}, \emptyset)$ . So we can match  $\sum_{\hat{d} \in A_{\mathbf{J}}} q_{\hat{d}} - 1$  candidates of  $\mathbf{J}$  in  $\overline{P}^d(\mathbf{J}, \emptyset)$ .

Let  $\mu$  be a matching satisfying the conditions of the Lemma. Using Lemma 2 we have

$$\begin{aligned} |\{j \in \mathbf{J} : \mu(j) = j\}| &= |\mathbf{J}| - |\{j \in \mathbf{J} : \mu(j) \neq j\}| > |\mathbf{J}| - \sum_{\hat{d} \in A_{\mathbf{J}}} q_{\hat{d}} + 1 \\ \iff |\{j \in \mathbf{J} : \mu(j) \neq j\}| &< \sum_{\hat{d} \in A_{\mathbf{J}}} q_{\hat{d}} - 1. \end{aligned}$$

Hence,  $\mu$  is not a maximum matching in  $\overline{P}^d(\mathbf{J}, \emptyset)$ . Since  $\mu$  is comprehensive in  $\overline{P}^d(\mathbf{J}, \emptyset)$ , there exists from Lemma 1 an augmenting path  $\pi$  such that  $\mu' = \mu \oplus \pi$  is comprehensive in  $\overline{P}^d(\mathbf{J}, \emptyset)$ . Clearly,  $\mu'$  satisfies Eq. (6). From the definition of an augmenting path,  $\mu(j) \neq j$  implies  $\mu'(j) \neq j$ . That is, we have also  $\mu'(j) \neq j$  for each  $j \in \mathbf{J}_{i,d}$ .  $\blacksquare$

Let  $\mathbf{J}$  be a block at  $(i_0, d_0)$ . Suppose that there exists a matching  $\tilde{\mu}$  comprehensive for  $P$  such that  $\tilde{\mu}(i_0) = d_0$ . Let  $\mu$  be the matching such that  $\mu(i) = \tilde{\mu}(i)$  for each  $i \in \mathbf{J}$ . Clearly,  $\mu$  is a feasible matching for  $\overline{P}^d(\mathbf{J}, \emptyset)$ . Moreover,  $\mu$  is comprehensive for  $\overline{P}^d(\mathbf{J}, \emptyset)$ . Indeed, if  $\mu$  is not comprehensive then there exist  $i, j \in \mathbf{J}$  such that  $\mu(i) = i$ ,  $\mu(j) \neq j$  and  $iP_{\mu(j)}j$ . But then  $\tilde{\mu}$  would not be comprehensive for  $P$ . Note also that  $\tilde{\mu}$  comprehensive for  $P$  and  $\tilde{\mu}(i_0) = d_0$  imply that, for each  $i \in \mathbf{J}$  such that  $iP_{d_0}i_0$ ,  $\mu(i) \neq i$ .

Let  $r = |\mathbf{J}| - \sum_{\hat{d} \in A_{\mathbf{J}}} q_{\hat{d}} + 1$ . For any matching  $\bar{\mu}$ , let  $K(\bar{\mu}) = \{i \in \mathbf{J} : \bar{\mu}(i) = i\}$ . By Lemma 2,  $|K(\mu)| > r$ . So by Lemma 3 there exists a matching, say,  $\mu_1$ , such that  $\mu_1$  is comprehensive for  $\overline{P}^{d_0}(\mathbf{J}, \emptyset)$  and  $\mu_1(i) \neq i$  for each  $i \in \mathbf{J}$  such that  $iP_{d_0}i_0$  and  $|K(\mu_1)| < |K(\mu)|$ . By Lemma 2  $|K(\mu_1)| > r$ , so by Lemma 3 there exists a matching, say,  $\mu_2$ , that satisfies the properties required by Lemma 2 and 3.

Again by Lemma 2,  $|K(\mu_2)| > r$ , so by Lemma 3 there exists  $\mu_3$  that satisfies the properties required by Lemma 2 and 3, and such that  $|K(\mu_3)| < |K(\mu_2)|$ . Repeating the process yields  $|K(\mu)| = \infty$ , a contradiction with  $I$  being a finite set. Hence there is no comprehensive matching  $\tilde{\mu}$  in  $P$  such that  $\tilde{\mu}(i_0) = d_0$ . From Proposition 2 we deduce then that  $i_0$  is an impossible match for  $d_0$   $\blacksquare$

#### 4. Proofs of Section 4

**Proof of Proposition 5** 1. ( $\Leftarrow$ ) Suppose by way of contradiction that there is  $A_i^{R(P)} = \{d\}$  for some department  $d$ , yet for some  $\tilde{\varphi} \in \Theta(P)$  and matching  $\tilde{\mu}$  stable for

$\tilde{\succ}$ ,  $\tilde{\mu}(i) \neq d$ . Suppose first that  $\tilde{\mu}(i) = i$ . Note that  $d \in A_i^{R(P)}$  implies  $d \in A_i$ . Since  $\tilde{\mu}$  is stable for  $\tilde{\succ}$ , it is obviously comprehensive for  $P$  because  $\tilde{\succ} \in \Theta(P)$ . Furthermore,  $i$ 's rank in  $P_d$  being lower or equal to  $q_d$  implies that  $|\tilde{\mu}(d)| < q_d$ , which contradicts that  $\tilde{\mu}$  satisfies the condition (b.1) of stability. So  $\tilde{\mu}(i) = d'$  for some  $d' \neq d$  (and thus  $d' \in A_i$ ). By construction,  $d' \notin A_i^{R(P)}$  implies that  $i$  is an impossible match for  $d'$ , which contradicts  $\tilde{\mu}(i) = d'$ .

( $\Rightarrow$ ) Suppose that for every  $\tilde{\succ} \in \Theta(P)$  and every matching  $\tilde{\mu}$  stable for  $\tilde{\succ}$ ,  $\tilde{\mu}(i) = d$ . To begin with, suppose that  $i$  is not among the top  $q_d$  candidates in  $P_d$ . Let  $\succ \in \Theta(P)$  be such that  $d$  is the top choice of each candidate  $j$  whose rank in  $P_d$  is lower or equal to  $q_d$ . Clearly, if  $\mu$  is stable for  $\succ$  then  $\mu(j) = d$  for any candidate  $j$  ranked lower or equal to  $q_d$  in  $P_d$ . So  $i$  is not always matched to  $d$  at any stable matching, for any  $P$ -compatible job market problem. Suppose now that there exists  $d' \neq d$  such that  $d' \in A_i^{R(P)}$ . So  $i$  is not an impossible match for  $d'$ , and thus there exists by definition  $\succ \in \Theta(P)$  and a matching  $\mu$  stable for  $\succ$  such that  $\mu(i) = d'$ , which contradicts our initial assumption.

2. If  $A_i^{R(P)} = \emptyset$  then  $i$  is an impossible match for each department  $d \in A_i$ . Hence, candidate  $i$  cannot be matched to any department in  $A_i$ , under any stable matching for  $\succ \in \Theta(P)$ . Conversely, if  $i$  is always unmatched under any stable matching then  $i$  is an impossible match for any  $d \in A_i$ , which implies  $A_i^{R(P)} = \emptyset$ .

3. Since the construction of the reduced problem  $R(P)$  is the same for any  $P$ -compatible job market problem, the result follows directly from the definition of impossible matches. ■

The next lemma will be key to prove Proposition 6.

**Lemma 4** *Let  $P$  be a proto-matching problem and  $\succ \in \Theta(P)$  a job market problem. Let  $\mu$  be a matching stable for  $\succ$ . The following two statements hold, for any step  $h \geq 1$  of the Predicting Algorithm with input  $(P, \mu)$ :*

1. *If a candidate  $i$  is deleted from  $P_d^h$  or  $\tilde{P}_d^h$ , then  $d \neq \mu(i)$ .*
2. *The matching  $\mu$  is maximal and comprehensive for  $P^h$*

**Proof**

1. The first statement directly follows from the definition of the Predicting Algorithm. More precise arguments can be found in the proof of statement 2.

2. Comprehensiveness for step  $h = 1$  directly follows from the stability of  $\mu$ . Weakly non-wastefulness for step  $h = 1$  follows immediately from the condition (b.1) of stability. Observe that if  $\mu$  is a comprehensive and weakly non-wasteful matching for  $P$ , then  $\mu$  is obviously a maximal and comprehensive matching for any  $P'$  that is obtained from  $P$  by deleting from some  $P_d$  a candidate  $i \in A_d$  such that  $\mu(i) \neq d$ . Repeating for all impossible matches in  $P$  gives statement (2) for  $h = 1$ .

Consider the beginning of Step  $h$  and suppose as induction hypothesis that  $\mu$  is maximal and comprehensive for the proto-matching  $P^h$ . By construction if a candidate  $i \in S^h$  is deleted from  $\tilde{P}_d^h$  then  $\mu(i) \neq d$ . It follows that  $\mu$  is maximal and comprehensive for  $\tilde{P}^h$  after Step  $h.2$ . It remains to check Step  $h.3$ . A candidate  $i$  is deleted from the department  $d$ 's ranking if, and only if,  $i$  is an impossible match for  $d$  in  $\tilde{P}^h$ . So  $\mu$  being comprehensive for  $\tilde{P}^h$ , deletion from  $d$  implies that  $d \neq \mu(i)$ . It follows again that  $\mu$  is still a well defined matching for  $P^{h+1}$ , and also maximal and comprehensive, as was to be proved. Statement (2) is thus proved. ■

**Proof of Proposition 6** Suppose that  $\mu(i) = d$ , yet the prediction for  $i$  is  $d' \neq d$  or  $i$ . Since  $\mu(i) = d$  it must be the case that  $d \in A_i^{P^h}$  for every  $h$  by Lemma 4 (no deletion), which is applicable since  $\mu$  is stable. Since the prediction is either  $d'$  or  $i$ , there exists some iteration  $h'$  where  $A_i^{P^{h'}} = \{d'\}$  or  $A_i^{P^{h'}} = \{i\}$ . Hence,  $d \notin A_i^{P^{h'}}$ , which is a contradiction. Suppose now that  $\mu(i) = i$ , yet the prediction is  $d$ , predicted at some step  $h$ . Since  $d$  is a prediction, the department is acceptable at step  $h$  for candidate  $i$ ,  $d \in A_i^{P^h}$ , and  $|\{j : j P_d^h i\}| < q_d$ . By Lemma 4,  $\mu$  is comprehensive for the problem  $P^h$ , so  $\mu(i) = i$  implies that no candidate ranked below  $i$  in  $P_d^h$  is matched to  $d$  under  $\mu$ . Again by Lemma 4, candidates  $j$  such  $\mu(j) = d$  cannot be deleted from  $P_d^{h'}$ , at any step  $h' \leq h$ . So we must have  $|\mu(d)| < q_d$ , which contradicts that  $\mu$  is maximal for  $P$ . ■

**Proof of Proposition 7** The statements of the proposition are deduced from the following lines of arguments.

Denote by  $P^h$  and  $P'^h$  the proto-matching problems obtained at step  $h$  in the algorithm with input  $(P, \mu)$  and  $(P, \mu')$ , respectively. Similarly, for  $h \geq 2$ , let  $S^h$ , and  $T^h$ , denote the set of market stars for  $(P, \mu)$  and  $(P, \mu')$ , respectively. We first show by induction

that  $P^h = P'^h$  for every  $h \geq 1$ . By construction it is clear that  $P^1 = P'^1$  since neither the realized matching, nor the realization of the preferences are used at this step. Suppose that  $P^h = P'^h$ , for some  $h \geq 2$ . We show that  $P^{h+1} = P'^{h+1}$ .

At step  $h.1$ , notice that  $S^h = T^h$ , for otherwise there would be  $i \in S^h \setminus T^h$  or  $i \in T^h \setminus S^h$ . Assume without loss of generality that  $i \in S^h \setminus T^h$ . the construction of the sets  $S^h$  and  $T^h$ , together with  $P^h = P'^h$ , imply that  $\mu'(i) = i$ . But since  $i \in S^h$ ,  $|\{j : jP_d^h i\}| = |\{j : jP'_d^h i\}| < q_d$  for any  $d \in A_i^{P^h} = A_i^{P'^h}$ . Therefore, there are strictly less than  $q_d$  candidates ranked higher than  $i$  in  $P_d$  and matched to  $d$  under  $\mu'$ . So  $\mu'$  does not satisfy the condition (b.1), which contradicts the stability of  $\mu'$ .

At step  $h.2$ , we note that if  $j \in S^h$  then  $\mu(j) \in A_j^{P^h}$ .<sup>40</sup> Stability implies that  $j$  can claim any of the seats in the departments  $A_j^{P^h}$  and that  $\mu(j)$  is the top choice in the set  $A_j^{P^h}$  according to  $\succ_j$ . The exact same argument should apply to the problem  $\succ'$  (recall that  $P^h = P'^h$ ). Since the top choice is the same department by construction,  $j$  obtains the same department under  $\mu'$ . So  $\mu(j) = \mu'(j)$  for every  $j \in S^h$ . It also follows that the deletions are the same in both cases. So  $\tilde{P}^h = \tilde{P}'^h$ . As we proceed with step  $h.3$ , the impossible matches are clearly the same in both problems. So  $P^{h+1} = R(\tilde{P}^h) = R(\tilde{P}'^h) = P'^{h+1}$ , as was to be proved.

Since the sequence of sets of market stars and reduced proto-matching problems are the same with inputs  $(P, \mu)$  and  $(P, \mu')$ . So any prediction obtained with  $(P, \mu)$  is obtained with  $(P, \mu')$ , and vice-versa, that is,  $\mu(i) = \mu'(i)$  for any candidate  $i$  whose match is predicted. ■

---

<sup>40</sup>It can be deduced from Lemma 4, which shows that  $\mu(j)$  cannot be deleted from the list of acceptable departments of  $j$ .

## References

- Abdulkadiroğlu, Atila, Parag A. Pathak and Alvin E. Roth.** 2005. “The New York City High School Match.” *American Economic Review, Papers and Proceedings*, 95: 364–367.
- Agarwal, Nikhil** 2015. “An Empirical Model of the Medical Match,” *American Economic Review* **105**, 1939–78.
- Andersson, Tommy and Lars Ehlers.** 2016. “Assigning Refugees to Landlords in Sweden: Stable Maximum Matching,” mimeo.
- Ashlagi, Itai, Yash Kanoria and Jacob Leshno.** 2017. “Unbalanced Random Matching Markets: the Stark Effect of Competition,” *Journal of Political Economy* **125**, 69–98.
- Berge, Claude.** 1957. “Two Theorems in Graph Theory,” *Proceedings of the National Academy of the United States of America*, **43**, 842–844.
- Bodoh-Creed, Aaron L.** 2013. “Risk and Conflation in Matching Markets,” mimeo.
- Calsamiglia, Caterina, Guillaume Haeringer and Flip Klijn.** 2010. “Constrained School Choice: An Experimental Study,” *American Economic Review*, **100**, 1860–1874
- Demange, Gabrielle, David Gale and Marilda Sotomayor.** 1986 “Multi-item Auctions,” *Journal of Political Economy*, **94**, 863–872.
- Echenique, Federico, SangMok Lee, and Matthew Shum.** 2013 “Partial Identification in Two-Sided Matching Models,” *Advances in Econometrics*, **31**, 117–139.
- Echenique, Federico, SangMok Lee, Matthew Shum and M. Bumin Yenmez.** 2013. “The Revealed Preference Theory of Stable and Extremal Stable Matchings,” *Econometrica*, **81**, 153–171.
- Gale, David and Lloyd S. Shapley.** 1962. “College Admissions and the Stability of Marriage,” *American Mathematical Monthly*, **69**, 9–15.
- Gale, D. and M. Sotomayor.** 1985. “Some Remarks on the Stable Matching Problem,” *Discrete Applied Mathematics*, **11**, 223–232.
- Gusfield, Dan.** 1987. “Three fast algorithms for four problems in stable marriage.” *SIAM Journal on Computing*, **16**, 111–128.
- Haeringer, Guillaume and Vincent Iehlé.** 2010. “Enjeux Stratégiques du Concours de Recrutement des Enseignants-Chercheurs,” *Revue Économique*, **61**, 697–721. In French.

- Haeringer, Guillaume and Flip Klijn.** 2009. “Constrained School Choice,” *Journal of Economic Theory*, **144**, 1921–1947.
- Hall, Philip.** 1935. “On Representatives of Subsets,” *Journal of the Mathematical London Society*, **10**, 26–30.
- Lovász, L. and Michael D. Plummer.** 1986. “Matching Theory,” *Annals of Discrete Mathematics*, North-Holland Mathematics Studies, volume 29.
- Martínez, Ruth, Jordi Massó, Alejandro Neme and Jorge Oviedo.** 2012. “On the Invariance of the Set of Core Matchings with Respect to Preference Profiles,” *Games and Economic Behavior* **74**, 588–600.
- Nicoló, Antonio and Carmelo Rodríguez-Álvarez.** 2017. “Age-based preferences in paired kidney exchange,” *Games and Economic Behavior* **102**, 508–524.
- Pathak, Parag A. and Tayfun Sönmez.** 2013. “School Admissions Reform in Chicago and England: Comparing Mechanisms by their Vulnerability to Manipulation,” *American Economic Review*, **103**, 80–106.
- Rastegari, Baharak, Anne Condon, Nicole Immorlica and Kevin Leyton-Brown.** 2012. “Two-sided Matching with Partial Information,” Proceedings of the fourteenth ACM conference on Electronic Commerce.
- Roth, A. E.** 1984. “The Evolution of the Labor Market for Medical Interns and Resident: A Case Study in Game Theory,” *Journal of Political Economy*, **92**, 991–1016.
- Roth, Alvin E.** 1985. “The College Admissions Problem is Not Equivalent to the Marriage Problem,” *Journal of Economic Theory*, **36**, 277–288.
- Roth, Alvin E.** 1986. “On the Allocation of Residents to Rural Hospitals: A General Property of Two-Sided Matching Markets,” *Econometrica*, **54**, 425–427.
- Roth, Alvin E. and Marilda Sotomayor.** 1989. “The College Admission Revisited,” *Econometrica*, **57**, 559–570.
- Saban, Daniela and Jay Sethuraman.** 2015. “The Complexity of Computing the Random Priority Allocation Matrix,” *Mathematics of Operations Research*, **40**, 1005–1014.