

**Equilibria under Deferred Acceptance:
Dropping Strategies, Filled Positions,
and Welfare**

**Paula Jaramillo
Çagatay Kayi
Flip Klijn**

April 2013

Barcelona GSE Working Paper Series

Working Paper n° 686

Equilibria under Deferred Acceptance: Dropping Strategies, Filled Positions, and Welfare*

Paula Jaramillo[†] Çağatay Kayı[‡] and Flip Klijn[§]

April 3, 2013

Abstract

This paper studies many-to-one matching markets where each student is assigned to a hospital. Each hospital has possibly multiple positions and responsive preferences. We study the game induced by the student-optimal stable matching mechanism. We assume that students play their weakly dominant strategy of truth-telling.

Roth and Sotomayor (1990) showed that there can be unstable equilibrium outcomes. We prove that any stable matching can be obtained in some equilibrium. We also show that the exhaustive class of dropping strategies does not necessarily generate the full set of equilibrium outcomes. Finally, we find that the so-called ‘rural hospital theorem’ cannot be extended to the set of equilibrium outcomes and that welfare levels are in general unrelated to the set of stable matchings. Two important consequences are that, contrary to one-to-one matching markets, (a) filled positions depend on the particular equilibrium that is reached and (b) welfare levels are not bounded by the student and hospital-optimal stable matchings (with respect to the true preferences).

Keywords: many-to-one matching, deferred acceptance, Nash equilibrium, dropping strategies, filled positions, welfare.

JEL-Numbers: C78, D60.

*We thank Jordi Massó for valuable discussions.

[†]*Corresponding author.* Universidad de los Andes, Facultad de Economía, Calle 19A # 11 - 37, Bloque W, Bogotá, Colombia; e-mail: p.jaramillo26@uniandes.edu.co

[‡]Universidad del Rosario, Bogotá, Colombia. Ç. Kayı gratefully acknowledges the hospitality of Institute for Economic Analysis (CSIC) and financial support from Colciencias/CSIC (Convocatoria No: 506/2010), El Patrimonio Autónomo Fondo Nacional de Financiamiento para la Ciencia, la Tecnología y la Innovación, Francisco José de Caldas.

[§]Institute for Economic Analysis (CSIC) and Barcelona GSE, Spain. F. Klijn gratefully acknowledges the financial support from CSIC/Colciencias 2010CO0013 and Plan Nacional I+D+i (ECO2011-29847).

1 Introduction

This paper studies two-sided many-to-one matching markets. We focus on centralized markets that employ the student-optimal stable mechanism (SOSM), which is based on Gale and Shapley’s (1962) deferred acceptance algorithm. The SOSM forms the basis of many matching mechanisms used in practice. Two important instances are the New York City public school match and the National Resident Matching Program (NRMP). The latter is an entry-level matching market for hospitals and medical school graduates in the United States. In view of the NRMP, we call the agents on one side students, and the agents on the other side hospitals (with possibly multiple positions). We assume that hospitals have responsive preferences over students.¹

Dubins and Freedman (1981) and Roth (1982) showed that under the SOSM students do not have incentives to misrepresent their preferences. But hospitals typically do have incentives to do so. Kojima and Pathak (2009) studied large many-to-one markets by considering sequences of random markets with an increasing number of agents. They assumed (1) the length of students’ preference lists does not grow, (2) the number of hospital positions is bounded, (3) the number of students does not grow “much faster” than the number of hospitals, and (4) the acceptability of all students. One of their main results is that the expected proportion of hospitals that can manipulate the SOSM when others are truthful goes to zero as the number of hospitals goes to infinity. This paper aims to complement the study by Kojima and Pathak (2009). We consider many-to-one markets that do not satisfy the assumptions in Kojima and Pathak (2009) or that cannot be approximated by a sequence of infinitely growing markets. Following Kojima and Pathak (2009), we assume that students play their weakly dominant strategy of truth-telling. We study the equilibria of the matching game where the hospitals are the only strategic players.

A desirable property of a matching between students and hospitals is that it is stable with respect to the true preferences. This means that there is no individual agent who prefers to become unmatched or pair of agents who prefer to be assigned to each other (possibly disrupting their match with other agents). The set of stable matchings has a lattice structure with remarkable properties (see Roth and Sotomayor, 1990). In particular, the student-optimal stable matching is the best (worst) stable matching for the students (hospitals). Similarly, the so-called hospital-optimal stable matching is the best (worst) stable matching for the hospitals (students). Moreover, the number of filled positions in each hospital and the set of matched students do not vary across stable matchings. Nor does the set of matched students vary from one stable matching to another for any hospital with some unfilled position.

For one-to-one markets, i.e., markets in which each hospital has one position, the set of equilibrium outcomes coincides with the set of stable matchings (Gale and Sotomayor, 1985a and Roth,

¹Loosely speaking, this means that for any hospital, (a) if faced with two sets of students that differ only in one student, the hospital prefers the set of students containing the more preferred student and (b) as long as the hospital has unfilled positions, it prefers to fill a position with an “acceptable” student rather than leaving it unfilled. We give a formal definition in the next section.

1984b). An important consequence is that the remarkable properties of the set of stable matchings carry over to the set of equilibrium outcomes. Hence, the agents' welfare in equilibria has (tight) bounds that can directly be determined by the deferred acceptance algorithm. Moreover, the sets (and hence, the numbers) of matched students and filled positions are invariable, i.e., they do not depend on the particular equilibrium that is reached.

Most real-life applications of the SOSM are instances of many-to-one markets. For this reason we aim to determine if the extraordinary features of the equilibria in the one-to-one setting carry over to the many-to-one setting. A key difference with the one-to-one setting was established in Roth and Sotomayor (1990, Corollary 5.17) through an example of a many-to-one market with an unstable equilibrium outcome. We show that any stable matching is still the induced outcome of some equilibrium (Proposition 1). Therefore, the set of stable matchings is a (possibly strict) subset of the set of equilibrium outcomes. The fact that the two sets do not necessarily coincide does not rule out *a priori* the possibility that the set of equilibrium outcomes satisfies important structural properties as those described for the one-to-one setting. But unfortunately we show that this is not the case. Our three results on the (absence of) structure of the set of equilibrium outcomes can be summarized as follows.

Our first result is related to so-called dropping strategies (Kojima and Pathak, 2009). A dropping strategy is obtained from a hospital's true preference list by making some acceptable students unacceptable, i.e., the order of any pair of students in the hospital's submitted list is not reversed with respect to its true preferences. Kojima and Pathak (2009, Lemma 1) showed that the class of dropping strategies is strategically exhaustive: fixing the other hospitals' strategies, the match obtained from any strategy can be replicated or improved upon by a dropping strategy. Surprisingly, we show that the class of dropping strategies does not necessarily generate the full set of equilibrium outcomes (Example 1). Hence, contrary to the one-to-one setting, there can be equilibrium outcomes that are not obtained in any equilibrium that consists of dropping strategies. An important implication is that the computation of the set of equilibrium outcomes cannot be simplified by focusing only on the relatively small class of dropping strategies.

Our second result is related to the number of filled positions in the equilibria. In Example 2, we exhibit a many-to-one market in which the number of filled positions varies within the set of equilibrium outcomes. This suggests that a maldistribution of students over hospitals can be due to the particular equilibrium that is reached.

Our third result is related to the agents' welfare in the equilibria. In Example 3, we provide a many-to-one market with an equilibrium such that simultaneously for each side of the market there is an agent that is strictly worse off and another agent that is strictly better off than in any stable matching. Hence, relative to the set of stable matchings, none of the two sides of the market is necessarily favored over the other side. But maybe more importantly, welfare levels in equilibria are not bounded below nor above by the set of stable matchings. More precisely, lower and upper bounds cannot be determined through the deferred acceptance algorithm. In short, our equilibrium analysis shows that the SOSM exhibits substantial differences between the one-to-one

and the possibly more interesting many-to-one settings.

Apart from the previously mentioned papers, our study is most closely related to the work by Sotomayor (2008 and 2012) and Ma (2010). Below we comment on these papers.

Sotomayor (2008) considered many-to-one games based on stable matching mechanisms. She assumed that the hospitals are truthful² and that students are the only strategic agents. Interestingly, she showed that the hospital-optimal stable mechanism implements the set of stable matchings in Nash equilibrium. She also showed that for other stable matching mechanisms the strategic behavior of the students may lead to equilibrium outcomes that are not stable with respect to the true preferences. Finally, she proved that for any stable matching mechanism there is a strong Nash equilibrium that induces the student-optimal stable matching (under the true preferences).

Sotomayor (2012) studied many-to-one matching games based on stable matching mechanisms. She assumed that both students and hospitals are strategic agents. Her main results include a general manipulability theorem and a general impossibility theorem (which shows that if there is more than one stable matching, then some agent can profitably misrepresent its preferences). She also proved that the set of equilibrium outcomes coincides with the set of matchings in which each matched pair of agents is mutually acceptable (under the true preferences). Finally, she showed that any equilibrium in which the hospitals reveal their preferences truthfully induces a stable matching (under the true preferences).

We study the same game as Ma (2010), but our approach and, as a consequence, our findings are very different from his. A key difference is that several of his results require additional assumptions on the strategies that hospitals play in equilibrium. More precisely, Ma (2010, Theorem 6) showed that equilibria in which all hospitals play so-called truncation strategies induce matchings that are either unstable (with respect to the true preferences) or coincide with the true hospital-optimal stable matching. This result is different from our Proposition 1. However, it is not clear whether it is “reasonable” to assume that hospitals play truncation strategies when they have more than one position. One reason is that the class of truncation strategies is in general not strategically exhaustive in many-to-one markets (Kojima and Pathak, 2009). In Section 4, we discuss his results in more detail.

The remainder of the paper is organized as follows. In Section 2, we describe the many-to-one matching model with responsive preferences. Section 3, we present our results. Section 4 concludes.

²Assuming that the hospitals are truthful is not a trivial counterpart of our assumption that students are truthful under the student-optimal stable mechanism. Indeed, Roth (1985, Proposition 2) showed that for any stable mechanism, truthtelling is not a weakly dominant strategy for the hospitals. In particular, truthtelling is not even a weakly dominant strategy under the hospital-optimal stable mechanism. Nonetheless, in applications such as school choice, on which Sotomayor (2008) is based, the agents of the “many side” cannot act strategically by law. Hence, in certain settings the assumption is natural.

2 Model

There are two finite and disjoint sets of agents: a set of **students** S and a set of **hospitals** H . Let $I = S \cup H$ be the set of agents. We denote a generic student, hospital, and agent by s , h , and i , respectively. For each hospital h , there is an integer **quota** $q_h \geq 1$ that represents the number of positions it offers. Student s can work for at most one hospital and hospital h can hire at most q_h students. Let $q = (q_h)_{h \in H}$. For each $i \in I$, the set of potential partners of agent i is denoted by N_i . If $i \in S$, $N_i = H$ and if $i \in H$, $N_i = S$.

Each student s has a complete, transitive, and strict **preference relation** P_s over the hospitals and the prospect of “being unmatched” (or some outside option), which is denoted by s . For $h, h' \in H \cup s$,³ we write $h P_s h'$ if student s prefers h to h' ($h \neq h'$), and $h R_s h'$ if s finds h at least as good as h' , i.e., $h P_s h'$ or $h = h'$. If $h \in H$ is such that $h P_s s$, then we call h an **acceptable hospital** for student s . Let $P_S = (P_s)_{s \in S}$.

Let $h \in H$. A subset of students $S' \subseteq S$ is **feasible** for hospital h if $|S'| \leq q_h$. Let $\mathcal{F}(S, q_h) = \{S' \subseteq S : |S'| \leq q_h\}$ denote the collection of feasible subsets of students for hospital h . The element $\emptyset \in \mathcal{F}(S, q_h)$ denotes “being unmatched” (or some outside option). Hospital h has a complete, transitive, and strict **preference relation** \succ_h over $\mathcal{F}(S, q_h)$. For $S', S'' \in \mathcal{F}(S, q_h)$ we write $S' \succ_h S''$ if hospital h prefers S' to S'' ($S' \neq S''$), and $S' \succeq_h S''$ if hospital h finds S' at least as good as S'' , i.e., $S' \succ_h S''$ or $S' = S''$. Let $\succ = (\succ_h)_{h \in H}$.

Let P_h be the restriction of \succ_h to $\{\{s\} : s \in S\} \cup \{\emptyset\}$, i.e., individual students in S and being unmatched. For $s, s' \in S \cup \{\emptyset\}$, we write $s P_h s'$ if $s \succ_h s'$, and $s R_h s'$ if $s \succeq_h s'$. Let \mathcal{P}_h be the set of all such restrictions for hospital h . Agent $s \in S$ is an **acceptable student** for a hospital h with preference relation P_h if $s P_h \emptyset$. Let $P_H = (P_h)_{h \in H}$ and $\mathbf{P} = (P_S, P_H)$. Finally, for $H' \subseteq H$, let $P_{H'} = (P_{h'})_{h' \in H'}$ and $P_{-H'} = (P_h)_{h \in H \setminus H'}$.

We assume that for each hospital h , \succ_h is **responsive**, or more precisely, a responsive extension of P_h ,⁴ i.e., for each $S' \in \mathcal{F}(S, q_h)$, **(r1)** if $s \in S \setminus S'$ and $|S'| < q_h$, then $(S' \cup s) \succ_h S'$ if and only if $s P_h \emptyset$ and **(r2)** if $s \in S \setminus S'$ and $s' \in S'$, then $((S' \setminus s') \cup s) \succ_h S'$ if and only if $s P_h s'$.

A **(many-to-one) market** is given by (S, H, P_S, \succ_H, q) or, when no confusion is possible, (P_S, \succ_H) for short. Let (P_S, \succ_H) be a market. A **matching** is a function μ on the set $S \cup H$ such that each student is either matched to exactly one hospital or unmatched, i.e., for all $s \in S$, either $\mu(s) \in H$ or $\mu(s) = s$; each hospital is matched to a feasible set of students, i.e., for all $h \in H$, $\mu(h) \in \mathcal{F}(S, q_h)$; and a student is matched to a hospital if and only if the hospital is matched to the student, i.e., for all $s \in S$ and $h \in H$, $\mu(s) = h$ if and only if $s \in \mu(h)$. Let μ be a matching and $i, j \in I$. If $j \in \mu(i)$, then we say that i and j are matched to one another and that they are mates in μ . The set $\mu(i)$ is agent i 's match.

Next, we describe desirable properties of matchings. First, we are interested in a voluntary participation condition over the matchings. A matching μ is individually rational if neither a

³With some abuse of notation we often write x for a singleton $\{x\}$.

⁴See Roth (1985) and Roth and Sotomayor (1989) for a discussion about this assumption.

student nor a hospital would be better off by breaking a current match. Formally, a matching μ is **individually rational** if for each $s \in S$ and each $h \in H$, if $\mu(s) = h$, then $h P_s s$ and $s P_h \emptyset$.⁵

Second, if a student s and a hospital h are not mates in a matching μ but s would prefer to be matched to h and h would prefer to either add s or replace another student by s , then we would expect this mutually beneficial adjustment to be carried out. Formally, a student-hospital pair (s, h) is a **blocking pair** for μ if **(b1)** $h P_s \mu(s)$, and **(b2)** $[|\mu(h)| < q_h \text{ and } s P_h \emptyset]$ or $[\text{there is } s' \in \mu(h) \text{ such that } s P_h s']$.⁶ A matching is **stable** if it is individually rational and there are no blocking pairs. It is well-known that in many-to-one matching with responsive preferences, stability coincides with groupwise stability, which excludes the existence of any larger blocking coalition as well (see Roth and Sotomayor, 1990, Lemma 5.5). Let $\Sigma(\mathbf{P})$ be the set of stable matchings for market (P_S, \succ_H) . Stability does not depend on the particular responsive extensions of the agents' preferences over individual acceptable partners.⁷

A mechanism assigns a matching to each market. We assume that quotas are commonly known by the agents (because, for instance, the quotas are determined by law).⁸ Therefore, the only information that the mechanism asks from the agents are their preferences over the other side of the market. Many real-life centralized matching markets employ mechanisms that only ask for the preferences $P = (P_i)_{i \in I}$ over individual partners, i.e., they do not depend on the particular responsive extensions. Throughout the paper, we focus on this class of mechanisms. Hence, a mechanism φ assigns a matching $\varphi(P)$ to each profile P . We often denote agent i 's match $\varphi(P)(i)$ by $\varphi_i(P)$. A mechanism φ is stable if for each P , $\varphi(P) \in \Sigma(P)$. In this paper we focus on a well-known stable mechanism that is employed in many real-life markets: the **student-optimal stable mechanism (SOSM)** φ^S , which is based on Gale and Shapley's (1962) deferred acceptance algorithm.⁹ Let $P = (P_S, P_H)$ be a profile (of preferences over individual agents). Then, $\varphi^S(P)$ is computed as follows:

STEP 1: Each student s proposes to the hospital that is ranked first in P_s (if there is no such hospital then s remains unmatched). Each hospital h considers its proposers and tentatively assigns its q_h positions to these students one at a time following the preferences P_h . All other proposers are rejected.

STEP k , $k \geq 2$: Each student s that is rejected in Step $k - 1$ proposes to the next hospital in his list P_s (if there is no such hospital then s remains single). Each hospital h considers the students that were tentatively assigned a position at h in Step $k - 1$ together with its new proposers. Hospital h tentatively assigns its q_h positions to these students one at a time following the preferences P_h .

⁵Alternatively, by responsiveness condition (r1), a matching μ is individually rational if for each $s \in S$ and each $h \in H$, if $\mu(s) = h$, then $h P_s s$ and $\mu(h) \succ_h (\mu(h) \setminus s)$.

⁶By responsiveness conditions (r1) and (r2), (b2) is equivalent to $[|\mu(h)| < q_h \text{ and } \mu(h) \cup s \succ_h \mu(h)]$ or $[\text{there is } s' \in \mu(h) \text{ such that } (\mu(h) \setminus s') \cup s \succ_h \mu(h)]$.

⁷Note that the set of stable matchings does not depend on the agents' orderings of the (individual) unacceptable partners either.

⁸Contrary to Sönmez (1997), we assume that quotas cannot be manipulated.

⁹We refer to Roth (2008) for an account on the history and applications of the deferred acceptance algorithm.

All other proposers are rejected.

The algorithm terminates when no student is rejected. Then, all tentative matches become final and $\varphi^S(P)$ is the resulting matching.¹⁰ Gale and Shapley (1962, p.14) proved that $\varphi^S(P)$ is stable with respect to P . Moreover, $\varphi^S(P)$ is called the student-optimal stable matching since it is the best (worst) stable matching for the students (hospitals) with respect to P (Gale and Shapley, 1962, Theorem 2 and Roth and Sotomayor, 1990, Corollary 5.32).¹¹

Under φ^S , it is a weakly dominant strategy for the students to reveal their true preferences (Roth, 1985, Theorem 5*). Since we focus on φ^S , we assume that students are truthful and that hospitals are the only strategic agents. Henceforth we fix and suppress P_S , which in particular leads to the notation $\varphi^S(P_H)$ and $\Sigma(P_H)$.

A **strategy** is an (ordered) preference list of a subset of students.¹² More precisely, for each hospital h , \mathcal{P}_h is the set of strategies and $\mathcal{P} \equiv \prod_{h \in H} \mathcal{P}_h$ is the set of strategy profiles. A **game** is a quadruple $(H, \mathcal{P}, \varphi^S, \succ_H)$ where H is the set of players, \mathcal{P} is the set of strategy profiles, φ^S is the outcome function, and the outcome is evaluated through the (true) preference relations \succ_H . A result due to Roth (1982, Theorem 3) implies that submitting its true preferences is in general not a weakly dominant strategy for a hospital.

A dropping strategy of a hospital is an ordered list obtained from its true ordered list of acceptable students by removing some acceptable students, i.e., the order of any pair of students in the hospital's submitted list is not reversed with respect to its true preferences (Kojima and Pathak, 2009). Formally, for a hospital h with preferences P_h over individual students, P'_h is a **dropping strategy** if for any students $s, s' \in S$, [if $s R'_h s' R'_h \emptyset$, then $s R_h s' R_h \emptyset$]. Clearly, any truncation strategy is a dropping strategy.

A profile of strategies Q is a **(Nash) equilibrium** of the game $(H, \mathcal{P}, \varphi^S, \succ_H)$ if for each hospital h and each strategy Q'_h , $\varphi_h^S(Q) \succeq_h \varphi_h^S(Q'_h, Q_{-h})$. Let $\mathcal{E}(H, \mathcal{P}, \varphi^S, \succ_H)$ be the set of equilibria. Finally, let $\mathcal{O}(H, \mathcal{P}, \varphi^S, \succ_H) = \{\mu \in \mathcal{M} : \mu = \varphi^S(Q) \text{ and } Q \in \mathcal{E}(H, \mathcal{P}, \varphi^S, \succ_H)\}$ be the set of equilibrium outcomes. When no confusion is possible, a game $(H, \mathcal{P}, \varphi^S, \succ_H)$ and its corresponding sets of equilibria and equilibrium outcomes are denoted by \succ_H , $\mathcal{E}(\succ_H)$, and $\mathcal{O}(\succ_H)$, respectively. A special case of many-to-one markets are marriage markets where for each $h \in H$, $q_h = 1$. In that case, the previous notation can be replaced by P_H , $\mathcal{E}(P_H)$, and $\mathcal{O}(P_H)$, respectively.

¹⁰Note that the deferred acceptance algorithm does not depend on the particular responsive extensions.

¹¹By switching the roles of students and hospitals in the deferred acceptance algorithm the hospital-optimal stable mechanism φ^H is obtained. For each profile P , $\varphi^H(P)$ is called the hospital-optimal stable matching since it is the best (worst) stable matching for the hospitals (students) with respect to P .

¹²The listed students are interpreted as the acceptable students. The other students are unacceptable and, since we focus on stable mechanisms, their relative ordering is irrelevant.

3 Results

We first discuss the relation of the set of stable matchings with the set of equilibrium outcomes. It was shown by Gale and Sotomayor (1985a, Theorem 2) that for each marriage market P_H , each stable matching can be obtained in some equilibrium, i.e., $\Sigma(P_H) \subseteq \mathcal{O}(P_H)$. Gale and Sotomayor (1985a, Theorem 3) and Roth (1984b, Theorem, p. 386) also showed the inverse inclusion, i.e., $\mathcal{O}(P_H) \subseteq \Sigma(P_H)$. Therefore, for any marriage market P_H , $\mathcal{O}(P_H) = \Sigma(P_H)$.

The result of Gale and Sotomayor (1985a, Theorem 2) can be extended to many-to-one markets. To the best of our knowledge the extended result has not been mentioned in the literature.

Proposition 1. *Let \succ_H be a many-to-one market. Then, any stable matching is obtained in some equilibrium that consists of dropping strategies. In particular, $\Sigma(P_H) \subseteq \mathcal{O}(\succ_H)$ and $\mathcal{O}(\succ_H) \neq \emptyset$.*

Proof. Let $\mu \in \Sigma(P_H)$. For each hospital $h \in H$, let Q_h be the dropping strategy that only lists the students in $\mu(h)$. Let $Q = (Q_h)_{h \in H}$. It is easy to check that $\varphi^S(Q) = \mu$. It remains to verify that $Q \in \mathcal{E}(\succ_H)$. Suppose $Q \notin \mathcal{E}(\succ_H)$. Let $h \in H$ and Q'_h be such that $\varphi_h^S(Q'_h, Q_{-h}) \succ_h \varphi_h^S(Q)$. Let $Q' = (Q'_h, Q_{-h})$.

Assume $\varphi_h^S(Q') \subseteq \varphi_h^S(Q)$. Then, since $\varphi_h^S(Q)$ only contains acceptable students (with respect to the true preferences), responsiveness condition r1 implies $\varphi_h^S(Q) \succeq_h \varphi_h^S(Q')$, which contradicts the choice of Q'_h . Hence, $\varphi_h^S(Q') \setminus \varphi_h^S(Q) \neq \emptyset$.

Let $\Delta := \varphi_h^S(Q') \setminus \varphi_h^S(Q)$. Let $s \in \Delta$. We prove that s ranks hospital h higher than $\varphi_s^S(Q)$. If $\varphi_s^S(Q) = s$, then from $s \in \varphi_h^S(Q')$ and the deferred acceptance algorithm applied to Q' it follows that s finds h acceptable (and hence s ranks h higher than $\varphi_s^S(Q)$). Suppose $\varphi_s^S(Q) \in H$. At profile Q' the only two hospitals that find s acceptable are h and $\varphi_s^S(Q)$. Since hospital $\varphi_s^S(Q)$ plays the same strategy at Q and Q' , the number of students it finds acceptable at Q' does not exceed its quota. Therefore, the fact that $s \in \varphi_h^S(Q')$ implies that s ranks hospital h higher than $\varphi_s^S(Q)$.

We conclude that each agent $i \in \Delta \cup \{h\}$ strictly prefers $\varphi_i^S(Q')$ to $\varphi_i^S(Q)$ (according to its true preferences). Moreover, the coalition $\Delta \cup \{h\}$ can “enforce” matching $\varphi^S(Q')$ from matching $\varphi^S(Q)$: hospital h can accept the students in Δ (by possibly breaking up with some other students), and each student in Δ can accept working for hospital h (by possibly breaking up with some other hospital). Hence, $\Delta \cup \{h\}$ is a blocking coalition (with respect to the true preferences) for matching $\varphi^S(Q)$. Then, from Roth and Sotomayor (1990, Lemma 5.5) it follows that $\varphi^S(Q) = \mu$ is not stable (with respect to the true preferences), which contradicts $\mu \in \Sigma(P_H)$. Hence, $\Sigma(P_H) \subseteq \mathcal{O}(\succ_H)$. Since $\Sigma(P_H) \neq \emptyset$, $\mathcal{O}(\succ_H) \neq \emptyset$. \square

Roth and Sotomayor (1990, Corollary 5.17) exhibited an example of a many-to-one market \succ_H with an unstable equilibrium outcome. Proposition 1 implies that for this market \succ_H the set of stable matchings is a strict subset of the set of equilibrium outcomes, i.e., $\Sigma(P_H) \subsetneq \mathcal{O}(\succ_H)$.¹³

¹³The market in Roth and Sotomayor (1990, Corollary 5.17) has a unique stable matching. The set of equilibrium

It is well-known that the set of stable matchings can be computed in polynomial time (see for instance, Gusfield and Irving, 1989). An important computational question is whether there is a fast algorithm to compute the (possibly strictly larger) set of equilibrium outcomes as well. In fact, it would be useful to be able to restrict equilibrium computations to a natural class of strategies. A natural candidate is the class of dropping strategies: Kojima and Pathak (2009, Lemma 1) showed that the class of dropping strategies is strategically exhaustive in the following sense. For any hospital, fixing the other hospitals' strategies, the match obtained from any strategy can be replicated or improved upon by a dropping strategy.¹⁴ Surprisingly, in Example 1 we show that the class of dropping strategies does not necessarily generate the full set of equilibrium outcomes. Hence, contrary to the one-to-one setting, there can be equilibrium outcomes that are not obtained in any equilibrium that consists of dropping strategies.¹⁵ Even though this result does not discard the existence of a fast algorithm for finding all equilibrium outcomes, it does tell us that one cannot simply restrict attention to the relatively small class of dropping strategies.

Example 1. (Dropping strategies.)

Consider a many-to-one market (P_S, \succ_H) with 4 students, 2 hospitals, and preferences over individual partners P given by the columns in Table 1. For instance, $P_{h_1} = s_1, s_2, s_3, s_4$. The hospitals have quota 2.

Students				Hospitals	
s_1	s_2	s_3	s_4	h_1	h_2
h_2	h_1	h_1	h_1	s_1	s_4
h_1	h_2	h_2	h_2	s_2	s_1
				s_3	s_2
					s_3

Table 1: Preferences P in Example 1

One easily verifies that the unique stable matching for P is given by

$$\varphi^S(P) : \begin{array}{cc} h_1 & h_2 \\ | & | \\ \{s_2, s_3\} & \{s_1, s_4\} \end{array}$$

which is the boxed matching in Table 2.

outcomes consists of the stable matching and the unstable equilibrium outcome discussed in Roth and Sotomayor (1990, Corollary 5.17). Calculations are available upon request.

¹⁴Jaramillo et al. (2012) proved that the class of dropping strategies is even exhaustive in many-to-many markets.

¹⁵Proposition 1 shows that any stable matching can be obtained in some equilibrium that consists of dropping strategies. Hence, any equilibrium outcome that cannot be obtained through dropping strategies is unstable. But not each equilibrium that consists of dropping strategies induces a stable outcome; see, for instance, Roth and Sotomayor (1990, Corollary 5.17) or Examples 2 and 3.

Consider the strategies $Q_{h_1} = s_1, s_2, s_4, s_3$ and $Q_{h_2} = s_4, s_2, s_3$ for hospitals h_1 and h_2 . Routine computations¹⁶ show that $Q = (Q_{h_1}, Q_{h_2})$ is an equilibrium and that it induces the matching

$$\varphi^S(Q) : \begin{array}{cc} & h_1 & h_2 \\ & | & | \\ \{s_1, s_2\} & & \{s_3, s_4\} \end{array}$$

which is the boldfaced matching in Table 1.

Clearly, Q_{h_1} is not a dropping strategy. It is easy but cumbersome to check that matching $\varphi^S(Q)$ cannot be obtained in any equilibrium that consists of dropping strategies.¹⁷ Hence, dropping strategies do not generate the full set of equilibrium outcomes. \diamond

An important issue in many practical situations (such as the National Resident Matching Program) is the number of matched students or, equivalently, the number of filled positions. In order to study this question we relate it to the so-called “rural hospital theorem,” which for many-to-one markets can be formally stated as follows.¹⁸ Let \succ_H be a market and let \mathcal{M} be a set of matchings (discussed below).

Rural Hospital Theorem (for \mathcal{M}). Let $\mu, \mu' \in \mathcal{M}$. Then,

R1S. for each $s \in S$, $\mu(s) = s \iff \mu'(s) = s$;

R1H. for each $h \in H$, $|\mu(h)| = |\mu'(h)|$;

R2. for each $h \in H$, $|\mu(h)| < q_h \implies \mu(h) = \mu'(h)$.

For many-to-one markets with responsive preferences and $\mathcal{M} = \Sigma(P_H)$, R1S and R1H were first proved by Roth (1984a, Theorem 9) and Gale and Sotomayor (1985b, p. 225); R2 is due to Roth (1986).¹⁹ Loosely speaking, these results imply that *if all agents are truth-telling* then neither the number of matched agents (R1S and R1H) nor the set of students that take up positions in “underdemanded” hospitals (R2) can be changed by switching from one stable mechanism to another.

Unfortunately, in the SOSM truth-telling is not a weakly dominant strategy for all agents. Therefore, a more interesting question would be whether it is possible to establish the rural hospital theorem for $\mathcal{M} = \mathcal{O}(\succ_H)$. Such a result would imply that the number or set of matched students and filled positions does not depend on the particular equilibrium that is reached.

For marriage markets, the question is immediately settled in the affirmative. This follows from the above results on the rural hospital theorem with $\mathcal{M} = \Sigma(P_H)$ and the fact that for any marriage market P_H , $\mathcal{O}(P_H) = \Sigma(P_H)$. Unfortunately, as the following example shows, the rural hospital

¹⁶From Kojima and Pathak (2009, Lemma 1) it follows that it suffices to only consider dropping strategies for possible profitable deviations.

¹⁷Moreover, $\mathcal{O}(\succ_H) = \{\varphi^S(P), \varphi^S(Q)\}$. Calculations are available upon request.

¹⁸The theorem owes its name to the perceived maldistribution of students over rural hospitals in the National Resident Matching Program. See Roth (1986) for further details on this clearinghouse.

¹⁹One easily verifies that R2 implies R1H. We refer to Klijn and Yazıcı (2012) for an overview of results on the rural hospital theorem in more general settings.

theorem does not necessarily hold for $\mathcal{M} = \mathcal{O}(\succ_H)$ if hospitals have a quota larger than 1. An important consequence is that the number of filled positions depends on the particular equilibrium that is reached.

Example 2. (Filled positions.)

Consider a many-to-one market (P_S, \succ_H) with 4 students, 2 hospitals, and preferences over individual partners P given by the columns in Table 2. The hospitals have quota 2. For instance, $P_{h_1} = s_1, s_2, s_3$ and student s_4 is not acceptable for hospital h_1 . Moreover, assume that $\{s_1\} \succ_{h_1} \{s_2, s_3\}$. Note that this assumption does not contradict the responsiveness of h_1 's preferences.²⁰

Students				Hospitals	
s_1	s_2	s_3	s_4	h_1	h_2
h_2	h_1	h_1	h_2	s_1	s_2
h_1	h_2	h_2		s_2	s_3
				s_3	s_1
					s_4

Table 2: Preferences P in Example 2

One easily verifies that the unique stable matching for P is given by

$$\varphi^S(P) : \begin{array}{cc} h_1 & h_2 \\ | & | \\ \{s_2, s_3\} & \{s_1, s_4\} \end{array}$$

which is the boxed matching in Table 2.

Consider the dropping strategies $Q_{h_1} = s_1$ and $Q_{h_2} = s_2, s_3, s_1, s_4$ for hospitals h_1 and h_2 . Routine computations show that $Q = (Q_{h_1}, Q_{h_2})$ is an equilibrium and that it induces the matching

$$\varphi^S(Q) : \begin{array}{cc} h_1 & h_2 \\ | & | \\ \{s_1\} & \{s_2, s_3\} \end{array}$$

which is the boldfaced matching in Table 2.

Clearly, the number of filled positions at $\varphi^S(P)$ and $\varphi^S(Q)$ is different: $|\varphi_{h_1}^S(Q)| = 1 \neq 2 = |\varphi_{h_1}^S(P)|$. In other words, R1S and R1H (and hence R2) do not hold for $\mathcal{M} = \mathcal{O}(\succ_H)$.²¹ \diamond

²⁰But it does play a crucial role in the number of filled positions in the equilibria.

²¹It is easy but cumbersome to verify that $\mathcal{O}(\succ_H) = \{\varphi^S(P), \varphi^S(Q), \bar{\mu}\}$, where

$$\bar{\mu} : \begin{array}{cc} h_1 & h_2 \\ | & | \\ \{s_1, s_2\} & \{s_3, s_4\} \end{array}$$

is the third (unstable) equilibrium outcome. Calculations are available upon request.

In Examples 1 and 2 and the example in Roth and Sotomayor (1990, Corollary 5.17) students are weakly worse off in any unstable equilibrium outcome relative to any stable equilibrium outcome. The next example shows that this is not true in general. More precisely, Example 3 shows that there can even exist an equilibrium outcome such that simultaneously for each side of the market there is an agent that is strictly worse off and another agent that is strictly better off than in any stable matching. Obviously, this is only possible for many-to-one markets, since for any marriage market P_H , $\mathcal{O}(P_H) = \Sigma(P_H)$.

Example 3. (Welfare.)

Consider a many-to-one market (P_S, \succ_H) with 6 students, 3 hospitals, and preferences over individual partners P given by the columns in Table 3. The hospitals have quota 2. Moreover, assume that $\{s_1, s_4\} \succ_{h_2} \{s_5, s_6\}$. Note that this assumption does not contradict the responsiveness of h_2 's preferences.

Students						Hospitals		
s_1	s_2	s_3	s_4	s_5	s_6	h_1	h_2	h_3
h_2	h_1^*	h_1	h_3	h_2^*	h_2	s_1^*	s_4^*	s_5
h_1^*		h_3^*	h_2^*	h_3	h_3^*	s_2^*	s_5^*	s_6^*
						s_3	s_6	s_3^*
							s_1	s_4

Table 3: Preferences P in Example 3

One easily verifies that the student-optimal stable matching $\mu := \varphi^S(P)$ is given by

$$\mu : \begin{array}{ccc} & h_1 & h_2 & h_3 \\ & | & | & | \\ & \{s_1, s_2\} & \{s_5, s_6\} & \{s_3, s_4\} \end{array}$$

which is the boxed matching in Table 3. The only other stable matching in $\Sigma(P)$ is given by

$$\mu^* : \begin{array}{ccc} & h_1 & h_2 & h_3 \\ & | & | & | \\ & \{s_1, s_2\} & \{s_4, s_5\} & \{s_3, s_6\} \end{array}$$

which is the matching marked with * in Table 3. The difference between μ and μ^* is in the matches of students s_4 and s_6 .

Consider the dropping strategies $Q_{h_1} = s_2, s_3$, $Q_{h_2} = s_4, s_1$, and $Q_{h_3} = s_5, s_6, s_4$ for hospitals h_1 , h_2 , and h_3 . It is easy to verify that $Q = (Q_{h_1}, Q_{h_2}, Q_{h_3})$ is an equilibrium and that it induces the matching

$$\mu' := \varphi^S(Q) : \begin{array}{ccc} & h_1 & h_2 & h_3 \\ & | & | & | \\ & \{s_2, s_3\} & \{s_1, s_4\} & \{s_5, s_6\}. \end{array}$$

which is the boldfaced matching in Table 3.

Consider the unstable equilibrium outcome μ' . It is easy to see that for each side of the market there is an agent that is strictly worse off than in any stable matching and there is another agent that is strictly better off than in any stable matching. Indeed, for the hospitals' side we observe that $\mu(h_1) = \mu^*(h_1) \succ_{h_1} \mu'(h_1)$, but $\mu'(h_3) \succ_{h_3} \mu^*(h_3) \succ_{h_3} \mu(h_3)$. And similarly, for the students' side we find that $\mu(s_5) = \mu^*(s_5) P_{s_5} \mu'(s_5)$, but $\mu'(s_1) P_{s_1} \mu(s_1) = \mu^*(s_1)$. \diamond

Relative to the set of stable matchings, the unstable equilibrium outcome in Example 3 does not favor any of the two sides of the market. It also shows that welfare levels are no longer necessarily bounded below nor above by the welfare levels obtained in the set of stable matchings. In other words, unlike the one-to-one setting, welfare levels in equilibrium cannot be bounded through the direct computation of the student and hospital-optimal stable matchings.

In summary, our equilibrium analysis shows that the widely used student-optimal stable mechanism exhibits substantial differences between the one-to-one and the possibly more interesting many-to-one settings.

Remark 1. Hospital h 's responsive preference relation \succ_h is **additive** if there is a utility function $u : \{s \in S : s P_h \emptyset\} \rightarrow \mathbb{R}$ such that for $S', S'' \in \mathcal{F}(S, q_h)$ with $S', S'' \succeq_h \emptyset$, [$S' \succ_h S''$ if and only if $\sum_{s \in S'} u(s) > \sum_{s \in S''} u(s)$]. One easily verifies that in each of our three examples the hospitals' preference relations are additive. This strengthens our negative results since they even hold on the strictly smaller domain of additive preference relations.²² \diamond

4 Concluding Remarks

We have studied the same game as Ma (2010). He suggested that in an environment with low information about the preferences of the other hospitals it is plausible that hospitals employ truncation strategies. A truncation strategy of a hospital is an ordered list obtained from its true ordered list of acceptable students by making a tail of acceptable students unacceptable (Roth and Vande Vate, 1991). Formally, for a hospital h with preferences P_h over individual students, P'_h is a **truncation strategy** if for any students $s, s' \in S$, [if $s R'_h s' R'_h \emptyset$, then $s R_h s' R_h \emptyset$] and [if $s' P'_h \emptyset$ and $s P_h s'$, then $s P'_h \emptyset$].²³ In the context of one-to-one markets, Roth and Vande Vate (1991, Theorem 2) showed that the class of truncation strategies is strategically exhaustive. Ma (2010, Theorem 4) proved that in *one-to-one* markets the set of outcomes induced by the equilibria that consist of truncation strategies collapses to the true hospital-optimal stable matching.

Under some minor additional assumptions, Ma (2010, Theorem 6) extended his result to *many-to-one* markets: the equilibria in which all hospitals play truncation strategies induce matchings

²²Note that responsiveness does not imply additivity. For instance, consider $S = \{s_1, s_2, s_3, s_4, s_5\}$ and a hospital h with quota 3. Then, there are responsive preference relations \succ_h such that each of the students in S is acceptable, $\{s_1, s_4\} \succ_h \{s_2, s_3\}$, and $\{s_2, s_3, s_5\} \succ_h \{s_1, s_4, s_5\}$. Any such \succ_h is not additive.

²³Clearly, any truncation strategy is a dropping strategy.

that are either unstable (with respect to the true preferences) or coincide with the true hospital-optimal stable matching. This result is very different from our Proposition 1. However, it is not clear whether it is “reasonable” to focus on truncation strategies when hospitals’ quotas are larger than 1. One argument is that the class of truncation strategies is in general not strategically exhaustive in many-to-one markets (Kojima and Pathak, 2009, Example 1 and Footnote 18).

Kojima and Pathak (2009, Lemma 1) showed that the class of dropping strategies is strategically exhaustive for many-to-one markets. Note that the equilibrium strategies in Proposition 1 and Example 3 are dropping strategies. The equilibrium strategies in Example 2 even pertain to the smaller class of truncation strategies. This observation strengthens the negative results established in our Examples 2 and 3.²⁴

Finally, Ma (2010, Theorem 9) additionally assumed that hospitals in the many-to-one setting can be restricted to submit only truncation strategies and shows that (i) equilibria exist and (ii) each hospital is weakly better off than at the true hospital-optimal stable matching. Our Example 3 shows that one cannot dispense with the truncation assumption: at the equilibrium outcome μ' , hospital h_1 is strictly worse off than at the true hospital-optimal stable matching μ^* .

An interesting but open question is a characterization of the set of equilibria or equilibrium outcomes. One easily verifies that in general not any individually rational matching can be obtained as an equilibrium outcome. Example 3 also shows that in general the set of equilibrium outcomes is not a subset of the set of so-called hospital-quasi-stable or student-quasi-stable matchings (Blum, Roth, and Rothblum, 1997, p. 377, and Sotomayor, 1996).²⁵

References

- [1] Y. Blum, A.E. Roth, and U.G. Rothblum (1997). Vacancy Chains and Equilibration in Senior-Level Labor Markets. *Journal of Economic Theory* 76(2): 362–411.
- [2] L.E. Dubins and D.A. Freedman (1981). Machiavelli and the Gale-Shapley Algorithm. *American Mathematical Monthly* 88(7): 485–494.
- [3] D. Gale and L.S. Shapley (1962). College Admissions and the Stability of Marriage. *American Mathematical Monthly* 69(1): 9–15.
- [4] D. Gale and M.A.O. Sotomayor (1985a). Ms Machiavelli and the Stable Matching Problem. *American Mathematical Monthly* 92(4): 261–268.
- [5] D. Gale and M.A.O. Sotomayor (1985b). Some Remarks on the Stable Matching Problem. *Discrete Applied Mathematics* 11(3): 223–232.

²⁴In Example 1, the equilibrium does not consist of dropping strategies, but there we establish that dropping strategies do not generate the full set of equilibrium outcomes.

²⁵A matching μ is hospital-quasi-stable if it is individually rational and for each blocking pair (s, h) , hospital h does not fill its quota at μ , i.e., $|\mu(h)| < q_h$. Student-quasi-stable matchings are defined similarly.

- [6] D. Gusfield and R.W. Irving (1989). *The Stable Marriage Problem: Structure and Algorithms*. Cambridge (Massachusetts): MIT Press.
- [7] P. Jaramillo, Ç. Kayı, and F. Klijn (2012). On the Exhaustiveness of Truncation and Dropping Strategies in Many-to-Many Matching Markets. Barcelona Graduate School of Economics, Working Paper 632.
- [8] F. Klijn and A. Yazıcı (2012). A Many-to-Many ‘Rural Hospital Theorem’. Barcelona Graduate School of Economics, Working Paper 567.
- [9] F. Kojima and P.A. Pathak (2009). Incentives and Stability in Large Two-Sided Matching Markets. *American Economic Review*, 99(3): 608–627.
- [10] J. Ma (2010). The Singleton Core in the Hospital-admissions Problem and its Application to the National Resident Matching Program (NRMP). *Games and Economic Behavior* 69(1): 150–164.
- [11] A.E. Roth (1982). The Economics of Matching: Stability and Incentives. *Mathematics of Operations Research* 7(4): 617–628.
- [12] A.E. Roth (1984a). The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory. *Journal of Political Economy* 92(6): 991–1016.
- [13] A.E. Roth (1984b). Misrepresentation and Stability in the Marriage Problem. *Journal of Economic Theory* 34(2): 383–387.
- [14] A.E. Roth (1985). The Hospital Admission Problem is not Equivalent to the Marriage Problem. *Journal of Economic Theory* 36(2): 277–288.
- [15] A.E. Roth (1986). On the Allocation of Residents to Rural Hospitals: A General Property of Two-Sided Matching Markets. *Econometrica*, 54(2), 425–427.
- [16] A.E. Roth (2008). Deferred Acceptance Algorithms: History, Theory, Practice, and Open Questions. *International Journal of Game Theory* 36(3): 537–569.
- [17] A.E. Roth and M.A.O. Sotomayor (1989). The College Admissions Problem Revisited. *Econometrica* 57(3): 559–570.
- [18] A.E. Roth and M.A.O. Sotomayor (1990). *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*. Econometric Society Monograph Series. New York: Cambridge University Press.
- [19] A.E. Roth and J.H. Vande Vate (1991). Incentives in Two-Sided Matching with Random Stable Mechanisms. *Economic Theory* 1(1): 31–44.

- [20] T. Sönmez (1997). Manipulation via Capacities in Two-Sided Matching Markets. *Journal of Economic Theory* 77(1): 197–204.
- [21] M.A.O. Sotomayor (1996). A Non-Constructive Elementary Proof of the Existence of Stable Marriages. *Games and Economic Behavior* 13(1): 135–137.
- [22] M.A.O. Sotomayor (2008). The Stability of the Equilibrium Outcomes in the Admission Games Induced by Stable Matching Rules. *International Journal of Game Theory* 36(3): 621–640.
- [23] M.A.O. Sotomayor (2012). A Further Note on the College Admission Game. *International Journal of Game Theory* 41(1): 179–193.