Stable Matchings and Preferences of Couples*

Bettina Klaus† Flip Klijn‡

July 2004

Abstract: Couples looking for jobs in the same labor market may cause instabilities. We determine a natural preference domain, the domain of weakly responsive preferences, that guarantees stability. Under a restricted unemployment aversion condition we show that this domain is maximal for the existence of stable matchings. We illustrate how small deviations from (weak) responsiveness, that model the wish of couples to be closer together, cause instability, even when we use a weaker stability notion that excludes myopic blocking. Our remaining results deal with various properties of the set of stable matchings for "responsive couples markets," *viz.*, optimality, filled positions, and manipulation. *Journal of Economic Literature* Classification Numbers: C78; J41.

Key Words: Matching; Couples; Responsiveness; Stability

Running title: Stable Matchings with Couples

1 Introduction

Labor markets are in a continuous process of change. The growing number of couples with the same professional interests is part of this process. Couples seeking positions in the same labor market form a growing part of the demand side. However, they increase the complexity of the matching problem considerably since now, in addition to finding a mutually agreeable solution for both sides of the labor market, one also has to deal with group decision making on the demand side.

*We thank Ahmet Alkan, Jordi Massó, Howard Petith, Alvin Roth, William Thomson, the referee, and the participants of the Stanford Institute for Theoretical Economics (SITE) 2003 summer workshop for helpful comments and suggestions. B. Klaus's and F. Klijn's research is supported by Ramón y Cajal contracts of the Spanish *Ministerio de Ciencia y Tecnología.* The main part of F. Klijn's work was supported by a Marie Curie Fellowship of the European Community programme "Improving Human Research Potential and the Socio-economic Knowledge Base" under contract number HPMF-CT-2001-01232, carried out at the Departament d'Economia i d'Història Econòmica, Universitat Autònoma de Barcelona. The work of the authors is partially supported by Research Grant BEC2002-02130 from the Spanish *Ministerio de Ciencia y Tecnología* and by the Barcelona Economics Program of CREA.

†Corresponding author. Departament d'Economia i d'Història Econòmica, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain; Tel. (34) 93 581 1719; Fax. (34) 93 581 2012; e-mail: `bettina.klaus@uab.es`

‡Institut d'Anàlisi Econòmica (CSIC), Campus UAB, 08193 Bellaterra (Barcelona), Spain; e-mail: `fklijn@pareto.uab.es`

In addition to individual job quality, couples' preferences may capture certain "complementarities" that are induced by the distance between jobs. Loosely speaking, by complementarities we mean that the valuation of one partners job may crucially depend on the other partner's job, that is, the couple may consider job a to be a good job for the husband while the wife holds job b, but unacceptable if the wife holds job c. As in many other economic environments (*e.g.*, multi-object auctions or efficient resource allocation with indivisibilities) the presence of complementarities, or in other words the absence of sufficient substitutability, may imply that "desirable" economic outcomes (*e.g.*, Nash or general equilibria) fail to exist.

In many centralized labor markets, clearinghouses are most often successful if they produce *stable* matchings.[1] In order to explain stability, let us assume for the moment that one side of the market consists only of single workers, and the other side consists of firms each with one position. A matching is then a partition of all workers and firms into pairs (one worker is matched to one firm) and unmatched workers and/or firms. Such a matching is "stable" if (a) each firm and worker has an acceptable match, and (b) no firm and no worker prefer one another to their respective matches. For matching markets with sufficient substitutability instabilities can be ruled out.[2] For one-to-one matching markets considered in this article, Roth (1984) demonstrates the possibility of instability in the presence of couples. In his example, the couples' preferences over pairs of positions (one position for each member of the couple) seem to be somewhat arbitrary (see Table 1). In this article we give some more intuitive examples of instability and aim to obtain a better understanding of what happens when instabilities occur.

First, we show that for a natural preference domain for couples, namely the domain of "(weakly) responsive" preferences, stable matchings exist (Theorem 3.3). A couple's preferences are responsive if the unilateral improvement of one partner's job is considered beneficial for the couple as well. If responsiveness only applies to acceptable positions, then preferences are weakly responsive. Hence, (weakly) responsive preferences may reflect situations where couples search for jobs in the same metropolitan area (if one partner switches to a job he/she prefers and the couple can still live together, then the couple is better off). Since responsiveness essentially excludes complementarities in couples' preferences that may for instance be caused by distance considerations, this result – to some extent – may seem trivial. However, it mirrors other results showing that a sufficient amount of substitutability implies the existence of desirable outcomes for the markets in question (see for instance Roth (1985, many-to-one matching without money), Kelso and Crawford (1982, many-to-one matching with money), Alkan and Gale (2003, many-to-many schedule matching), and Hatfield and Milgrom (2004, two-sided matching with contracts). In contrast to our article, the substitutability conditions in those papers[3] apply to preferences on the supply side (hospitals or firms) over sets of agents (students or workers), while our responsiveness condition applies to preferences on the demand side (couples of students) over ordered pairs of hospitals (not sets!). We show that under a restricted unemployment aversion condition, the domain of weakly responsive preferences is maximal for the existence of stable

[1] Empirical evidence is given in Roth (1984, 1990, 1991) and Roth and Xing (1994).

[2] For one-to-one and many-to-one matching markets without money see Gale and Shapley (1962) and Roth (1985), for many-to-one matching markets with money see Kelso and Crawford (1982), for many-to-one matching with affirmative action constraints see Abdulkadiroğlu (2003), for many-to-many schedule matching see Alkan and Gale (2003), and for two-sided matching with contracts see Hatfield and Milgrom (2004). This list is not exhaustive.

[3] For instance, Alkan and Gale's (2003) "persistence" condition and Hatfield and Milgrom's (2004) "substitution" condition both encompass Roth's (1985) "responsiveness" and Kelso and Crawford's (1982) "gross substitutes" condition on preferences of hospitals/firms over sets of students/workers and corresponding wages.

matchings (Theorem 3.5). This implies that for strictly unemployment averse couples the domain of responsive preferences where all positions are considered to be acceptable is a maximal domain for the existence of stable matchings (Corollary 3.6).

Next, we analyze the existence of stable matchings for couples markets without any unemployment aversion condition. Then, proceeding from our possibility result for responsive preferences, we show that the absence of stable matchings in couples markets is not a theoretical irregularity: a single couple may cause a labor market to be unstable even if its preference list is very consistently based on their individual preferences and the desire to not live too far away from each other. In one of our examples we demonstrate that even a small deviation from responsiveness can cause instability (Example 3.8). Our nonexistence result persists even when we relax the requirement of stability and use a weaker stability notion that excludes myopic blocking (Theorem 3.7). By means of another instructive example (Example 3.9) we demonstrate how couples that do not want to be separated cause instability.

We base our remaining analysis of the set of stable matchings for couples markets on the fact that for responsive preferences one can construct a unique associated singles market with a nonempty set of stable matchings that is included in the set of stable matchings of the original couples market. This might lead one to conclude that, apart from the existence of stable matchings, other desirable properties of stable matchings for singles markets (not generally transferred to general couples markets) would carry over to couples markets with responsive preferences as well. Unfortunately this is not the case. First, we demonstrate that even for the domain of responsive preferences the set of stable matchings no longer needs to be a distributive lattice (Theorem 4.2). More precisely, we strengthen results due to Aldershof and Carducci (1996) by showing that for couples markets with responsive preferences there may be no optimal stable matching for either side of the market. Furthermore, we demonstrate that different stable matchings may assign positions to different applicants and/or have a different number of positions filled (Theorem 4.3). Finally, we show that for the domain of responsive preferences there exists no strategy-proof stable-matching mechanism based on revealed preferences. More precisely, we show that there is no stable-matching mechanism for which stating the true preferences is a dominant strategy for every couple (Theorem 4.5).

The paper is organized as follows. In Section 2, we introduce a simple couples market where the labor market modelled consists of a supply side of four hospitals and a demand side of two couples composed of medical students. In Section 3 we establish the existence of stable matchings for weakly responsive preferences and demonstrate that under restricted strict unemployment aversion the domain of weakly responsive preferences is maximal for the existence of stable matchings. We also demonstrate with two examples how small deviations from (weak) responsiveness that incorporate the distance considerations of a couple may lead to instability. In Section 4 we show that both the lattice structure and the invariable group of matched agents of the set of stable matchings need not carry over from singles markets to couples markets with responsive preferences. Finally, still assuming preferences to be responsive, we show that any stable-matching mechanism is prone to manipulation by couples misrepresenting their preferences. We conclude with Section 5, where we discuss the relation of our results for couples markets to those of Hatfield and Milgrom (2004) for matching markets with contracts.

2 Matching with Couples: The Model

For convenience and without loss of generality, we describe a simple couples market where the labor market modelled consists of a supply side of four hospitals and a demand side of two couples composed of medical students; $H = \{h_1, h_2, h_3, h_4\}$, $S = \{s_1, s_2, s_3, s_4\}$, and $C = \{c_1, c_2\} = \{(s_1, s_2), (s_3, s_4)\}$ are the sets of hospitals, students, and couples, respectively. Each hospital has exactly one position to be filled. All of our results can easily be adapted to more general situations that include other couples as well as single agents and hospitals with multiple positions.[4] Next, we describe preferences of hospitals, students, and couples.

Hospitals' preferences: Each hospital $h \in H$ has a strict, transitive, and complete preference relation $\succeq_h$ over the set of students S and the prospect of having its position unfilled, denoted by $\emptyset$. Hospital h's preferences can be represented by a strict ordering of the elements in $S \cup \{\emptyset\}$; for instance, $P(h) = s_4, s_2, \emptyset, s_1, s_3$ indicates that hospital h prefers student s_4 to s_2, and considers students s_1 and s_3 to be unacceptable. In the remainder of the paper each hospital typically prefers its position filled by some student rather than unfilled. Let $P^H = \{P(h)\}_{h \in H}$.

Students' preferences: Similarly, each student $s \in S$ has an individual strict, transitive, and complete preference relation $\succeq_s$ over the set of hospitals and the prospect of being unemployed, denoted by u. Let $h \in H$. If $h \succ_s u$, then hospital h is *acceptable* to student s; if $u \succ_s h$, then hospital h is *unacceptable* to student s. We assume that these individual preferences are the preferences a student has if he/she is single. Student s's individual preferences can be represented by a strict ordering of the elements in $H \cup \{u\}$; for instance, $P(s) = h_1, h_2, h_3, h_4, u$ indicates that student s prefers h_i to h_{i+1} for $i = 1, 2, 3$ and prefers being employed to being unemployed. Let $P^S = \{P(s)\}_{s \in S}$.

Couples' preferences: Finally, each couple $c \in C$ has a strict, transitive, and complete preference relation $\succeq_c$ over all possible combination of ordered pairs of (different) hospitals and the prospect of being unemployed. Couple c's preferences can be represented by a strict ordering of the elements in $\mathcal{H} := [(H \cup \{u\}) \times (H \cup \{u\})] \backslash \{(h, h) : h \in H\}$. To simplify notation, we denote a generic element of $\mathcal{H}$ by (h_p, h_q), where h_p and h_q indicate a hospital or being unemployed. For instance, $P(c) = (h_4, h_2), (h_3, h_4), (h_4, u)$, *etc.*, indicates that couple $c = (s_1, s_2)$ prefers s_1 and s_2 being matched to h_4 and h_2, respectively, to being matched to h_3 and h_4, respectively, and so on. Let $P^C = \{P(c)\}_{c \in C}$.

Note that when presenting preferences in examples, we often use column notation. Furthermore, whenever we use the strict part $\succ$ of a preference relation, we assume that we compare different elements in $S \cup \{\emptyset\}$, $H \cup \{u\}$, or $\mathcal{H}$.

We use the following restrictions on the couples' preferences in the remainder of the paper.

Unemployment aversion: A couple c is *strongly unemployment averse* if it prefers full employment to the employment of only one partner and the employment of only one partner to the unemployment of both partners. Formally, for all $h_p, h_q, h_r \neq u$, $(h_p, h_q) \succ_c (h_r, u) \succ_c (u, u)$ and $(h_p, h_q) \succ_c (u, h_r) \succ_c (u, u)$.

A couple c is *strictly unemployment averse* if it is worse off if one of its partners looses his/her position. Formally, for all $h_p, h_q \neq u$, $(h_p, h_q) \succ_c (h_p, u) \succ_c (u, u)$ and $(h_p, h_q) \succ_c (u, h_q) \succ_c (u, u)$.

Note that strong unemployment aversion implies strict unemployment aversion.

[4] To be more precise, in order to straightforwardly derive all results for the case of hospitals with multiple positions, we would require that hospitals' preferences are "responsive over sets of students" (see Roth (1985) and Section 5).

Responsive preferences:[5] Note that *a priori* we do not require any relation between students' individual preferences and couples' preferences. In fact, we cannot or do not always wish to specify individual preferences when couples are concerned. However, we do study some situations in which there is a clear relationship. This is the case when the unilateral improvement of one partner's job is considered beneficial for the couple as well. Couple $c = (s_k, s_l)$ has *responsive* preferences if there exist preferences $\succeq_{s_k}$ and $\succeq_{s_l}$ such that for all $h_p, h_q, h_r \in H \cup \{u\}$, $[h_p \succ_{s_k} h_r$ implies $(h_p, h_q) \succ_c (h_r, h_q)]$ and $[h_p \succ_{s_l} h_r$ implies $(h_q, h_p) \succ_c (h_q, h_r)]$. If these *associated individual preferences* $\succeq_{s_k}$ and $\succeq_{s_l}$ exist, then they are unique.[6] Note that if a couple (s_k, s_l) has responsive preferences, then one can easily derive the associated individual preferences $\succeq_{s_k}$ and $\succeq_{s_l}$ (see for instance Klaus *et al.* (2003), Example 2.1).

Leader-follower responsive preferences: A couple $c = (s_k, s_l)$ has *leader-follower responsive* preferences if it has responsive preferences and in addition gives precedence to the job quality for one of its members first (without loss of generality we assume that s_k is the leader and s_l the follower), *i.e.*, for all $h_p, h_x, h_q, h_y \in H \cup \{u\}$, $(h_p, h_x) \succ_c (h_q, h_y)$ implies $(h_p, h'_x) \succ_c (h_q, h'_y)$ for all $h'_x, h'_y \in H \cup \{u\}$.

Singles and couples markets: Now, the standard one-to-one two-sided matching market with single students, or *singles market* for short, is denoted by (P^H, P^S). Since singles markets and some of the classical results for singles markets are well-known, for a detailed description we refer to Roth and Sotomayor (1990) who give an excellent introduction to this model and review all results that are relevant here. For instance, the set of stable matchings is nonempty and coincides with the core. A one-to-one matching market with couples, or a *couples market* for short, is denoted by (P^H, P^C).[7]

Matchings: A *matching* μ for a couples market (P^H, P^C) is an assignment of students and hospitals such that each student is assigned to at most one hospital in H or to u (which can be assigned to multiple students), each hospital in H is assigned to at most one student or to $\emptyset$ (which can be assigned to multiple hospitals), and a student is assigned to a hospital if and only if the hospital is assigned to the student.

By $\mu(S) = \mu(s_1), \mu(s_2), \mu(s_3), \mu(s_4)$ we denote the hospital in H or u matched to students s_1, s_2, s_3, s_4. Thus, s_k is matched to $\mu(s_k)$. Alternatively, by $\mu(H) = \mu(h_1), \mu(h_2), \mu(h_3), \mu(h_4)$ we denote the students in S or $\emptyset$ matched to hospitals h_1, h_2, h_3, h_4. Note that the matching μ associated to (P^H, P^C) can be completely described either by $\mu(S)$ or by $\mu(H)$, but both notations will be useful later.[8]

[5]The concept of responsive preferences was first introduced in Roth (1985), but differs from ours as it deals with hospitals having preferences over (unordered) sets of students.

[6]Note that these derived preferences $\succeq_{s_k}$ and $\succeq_{s_l}$ need not coincide with the students' individual preferences. However, in order to keep notation as simple as possible, we denote the derived preferences the same way as we denote students' individual preferences.

[7]Instead of denoting a couples market by (P^H, P^C), we could add students' individual preferences and consider (P^H, P^S, P^C). Since we do not explicitly use the students' individual preferences, we suppress them in our notation.

[8]In our model with two couples and four hospitals we have 209 different matchings: $24 = \binom{4}{4}^2 \cdot 4!$ matchings with full employment, $96 = \binom{4}{3}^2 \cdot 3!$ matchings with one unemployed student, $72 = \binom{4}{2}^2 \cdot 2!$ matchings with two unemployed students, $16 = \binom{4}{1}^2 \cdot 1!$ matchings with three unemployed students, and $1 = \binom{4}{0}^2 \cdot 0!$ full unemployment matching.

Stability: Finally, we define stability for couples markets (see Roth and Sotomayor (1990)). First, for a matching to be stable, it should always be better for students (one or both members in a couple) to accept the position(s) offered by the matching instead of voluntarily choosing unemployment and for hospitals it should always be better to accept the student assigned by the matching instead of leaving the position unfilled. A matching μ is *individually rational* if

(i1) for all $c = (s_k, s_l)$, $(\mu(s_k), \mu(s_l)) \succeq_c (\mu(s_k), u)$, $(\mu(s_k), \mu(s_l)) \succeq_c (u, \mu(s_l))$, and $(\mu(s_k), \mu(s_l)) \succeq_c (u, u)$;

(i2) for all $h \in H$, $\mu(h) \succeq_h \emptyset$.

Second, if one partner in a couple can improve the given matching for the couple by switching to another hospital such that this hospital is better off as well, then we would expect this mutually beneficial trade to be carried out, rendering the given matching unstable. A similar statement holds if both students in the couple can improve. For a given matching μ, $(c = (s_k, s_l), (h_p, h_q))$ is a *blocking coalition* if

(b1) $(h_p, h_q) \succ_c (\mu(s_k), \mu(s_l))$;

(b2) $[h_p \in H$ implies $s_k \succeq_{h_p} \mu(h_p)]$ and $[h_q \in H$ implies $s_l \succeq_{h_q} \mu(h_q)]$.

A matching is *stable* if it is individually rational and if there are no blocking coalitions.[9]

Instability in a couples market: Roth (1984, Theorem 10) shows that stable matchings may not exist in the presence of couples. He considers the couples market (P^H, P^C) given by Table 1.[10] We use the following convention for this and future examples. If $\emptyset$ is not listed for hospitals, then all students are acceptable.

By giving a blocking coalition for each of the 24 individually rational full employment matchings, Roth shows that no stable matching exists. Note that neither couple's preferences are responsive. (For couple $c_1 = (s_1, s_2)$ this follows for instance from $(h_1, h_4) \succ_{c_1} (h_1, h_3)$ and $(h_2, h_3) \succ_{c_1} (h_2, h_4)$.)

In the next section, departing from Roth's example, we address one of the open questions and research directions that Roth and Sotomayor (1990, p. 246) indicate, namely to "`find reasonable assumptions about the preferences of married couples that assure the nonemptiness of the core.`" In other words, are there classes of "real-world preferences" for which stable matchings always exist? Given the NP completeness (computational complexity) of determining if a couples market has a stable matching (Ronn (1990)), this question seems even more intricate.

[9] In order to keep notation as simple as possible, we allow some redundancy in the definition of stability with respect to (i1) and (b1).

[10] Roth's (1984) and our later results do not depend on the tails (not specified by Roth (1984)) of the couples' preferences, which only contain unacceptable combinations of positions.

P^H				P^C	
h_1	h_2	h_3	h_4	s_1s_2	s_3s_4
s_4	s_4	s_2	s_2	h_1h_2	h_4h_2
s_2	s_3	s_3	s_4	h_4h_1	h_4h_3
s_1	s_2	s_1	s_1	h_4h_3	h_4h_1
s_3	s_1	s_4	s_3	h_4h_2	h_3h_1
				h_1h_4	h_3h_2
				h_1h_3	h_3h_4
				h_3h_4	h_2h_4
				h_3h_1	h_2h_1
				h_3h_2	h_2h_3
				h_2h_3	h_1h_2
				h_2h_4	h_1h_4
				h_2h_1	h_1h_3
				uu	uu
				$\dots$	$\dots$

Table 1: No Stable Matching in a Couples Market (Roth (1984))

3 Main Results: Existence of (Weakly) Stable Matchings and Domain Maximality

First, we establish an existence result. It is based on the intuition that if there exists no negative externality from one partner's job for the other partner or for the couple, then we can treat the market as if only singles participate. By doing this, we can guarantee the existence of a stable matching (Gale and Shapley (1962)). This would be the case if couples only apply for jobs in one city or metropolitan area so that different regional preferences or travel distance are no longer part of couples' preferences and therefore the preferences are responsive. For our existence result, we slightly extend the domain of responsive preferences. The idea of this extension is that the exact associated preferences that deal with the comparison of unacceptable positions are irrelevant with respect to stability since an agent can always replace any unacceptable position with unemployment.

Weakly responsive preferences: Couple $c = (s_k, s_l)$ has *weakly responsive* preferences if there exist preferences $\succeq_{s_k}$ and $\succeq_{s_l}$ such that

(i) for all $h \in H$,

$(u, h) \succ_c (u, u)$ if and only if $h \succ_{s_l} u$,

$(h, u) \succ_c (u, u)$ if and only if $h \succ_{s_k} u$, and

(ii) for all $h_p, h_q, h_r \in H \cup \{u\}$,

$[h_p \succeq_{s_k} u,\ h_q \succeq_{s_l} u$, and $h_p \succ_{s_k} h_r$ imply $(h_p, h_q) \succ_c (h_r, h_q)]$ and

$[h_p \succeq_{s_l} u,\ h_q \succeq_{s_k} u$, and $h_p \succ_{s_l} h_r$ imply $(h_q, h_p) \succ_c (h_q, h_r)]$.

Remark 3.1 If these *associated individual preferences* $\succeq_{s_k}$ and $\succeq_{s_l}$ exist, then they are only uniquely determined for acceptable positions. In other words, if both $[\succeq_{s_k}$ and $\succeq_{s_l}]$ and $[\succeq'_{s_k}$ and $\succeq'_{s_l}]$ satisfy the two conditions above, then for all $h_p, h_q \in H \cup \{u\}$, $h_p \succ_{s_k} h_q \succeq_{s_k} u$ implies $h_p \succ'_{s_k} h_q \succeq'_{s_k} u$, and $h_p \succ_{s_l} h_q \succeq_{s_l} u$ implies $h_p \succ'_{s_l} h_q \succeq'_{s_l} u$.

Note that responsive preferences are weakly responsive. In the next example we show that not all weakly responsive preferences are responsive.

Example 3.2 ***Weakly Responsive but not Responsive***
Consider couple $c_1 = (s_1, s_2)$'s preferences given by $P(c_1) = (h_1, h_2), (h_1, u), (u, h_2), (h_2, u), (u, u), (h_3, u), \ldots$.

Suppose couple c_1's preferences are responsive. Then the (unique) associated individual preferences are of the form $P(s_1) = h_1, h_2, u, h_3, h_4$ and $P(s_2) = h_2, u, \ldots$. By responsiveness, $(h_3, h_2) \succ_{c_1} (h_3, u)$, a contradiction.

It is easy to see that c_1's preferences are weakly responsive: for any preferences $\succeq_{s_1}$ and $\succeq_{s_2}$ with $P(s_1) = h_1, h_2, u, \ldots$ and $P(s_2) = h_2, u, \ldots$ (tails can be anything) conditions (i) and (ii) of weak responsiveness are satisfied, independently of the couple's preferences after (u, u). $\diamond$

Let (P^H, P^C) be a couples market and assume that couples have weakly responsive preferences. Then, from the couples' weakly responsive preferences we can determine associated individual preferences for all students that are members of a couple. By $(P^H, P^S(P^C))$ we denote an *associated singles market* we obtain by replacing couples and their preferences in (P^H, P^C) by individual students and their (possibly not uniquely determined) associated individual preferences $P^S(P^C)$. It is important to note and easy to see that all associated singles markets have the same set of stable matchings (see Remark 3.1). Notice also that for responsive preferences there exists a unique associated singles market $(P^H, P^S(P^C))$.

Theorem 3.3 ***Stability for Weakly Responsive Preferences***
Let (P^H, P^C) be a couples market where couples have weakly responsive preferences. Then, any matching that is stable for an associated singles market $(P^H, P^S(P^C))$ is also stable for (P^H, P^C). In particular, there exists a stable matching for (P^H, P^C).

Proof. Let μ be a stable matching for $(P^H, P^S(P^C))$ and consider any couple $c = (s_k, s_l)$. Stability of μ in $(P^H, P^S(P^C))$ implies that

$$\mu(s_k) \succeq_{s_k} u \text{ and } \mu(s_l) \succeq_{s_l} u.$$

If $(\mu(s_k), \mu(s_l)) = (u, u)$, then stability condition (i1) is trivially satisfied. If $\mu(s_k) \succ_{s_k} u$ and $\mu(s_l) = u$, then by weak responsiveness (i), $(\mu(s_k), u) \succ_c (u, u)$, which implies (i1). Similarly, $\mu(s_k) = u$ and $\mu(s_l) \succ_{s_l} u$ implies (i1). Finally, assume $\mu(s_k) \succ_{s_k} u$ and $\mu(s_l) \succ_{s_l} u$. Then by weak responsiveness (ii), $(\mu(s_k), \mu(s_l)) \succ_c (\mu(s_k), u) \succ_c (u, u)$. Similarly, $(\mu(s_k), \mu(s_l)) \succ_c (u, \mu(s_l)) \succ_c (u, u)$. Hence, any stable matching μ in $(P^H, P^S(P^C))$ is individually rational for (P^H, P^C) as well.

Suppose now that μ is not stable for (P^H, P^C). Hence, there exists a blocking coalition, for instance $((s_k, s_l), (h_p, h_q))$. Then, (b1) $(h_p, h_q) \succ_c (\mu(s_k), \mu(s_l))$ and (b2) $[h_p \in H$ implies $s_k \succeq_{h_p} \mu(h_p)]$ and $[h_q \in H$ implies $s_l \succeq_{h_q} \mu(h_q)]$.

Assume $h_p \prec_{s_k} u$ and $h_q \prec_{s_l} u$. Then by weak responsiveness (ii), $(u,u) \succ_c (u,h_q) \succ_c (h_p,h_q)$. Using (b1) it follows that $(u,u) \succ_c (\mu(s_k),\mu(s_l))$, contradicting individual rationality of μ in (P^H,P^C). Hence, $h_p \succeq_{s_k} u$ or $h_q \succeq_{s_l} u$.

Assume that $h_p \succeq_{s_k} u$ and $h_q \prec_{s_l} u$. Then by weak responsiveness (ii), $(h_p,u) \succ_c (h_p,h_q)$. Hence, $((s_k,s_l),(h_p,u))$ is a blocking coalition for μ. Similarly, if $h_p \prec_{s_k} u$ and $h_q \succeq_{s_l} u$, then $(u,h_q) \succ_c (h_p,h_q)$ and $((s_k,s_l),(u,h_q))$ is a blocking coalition for μ. Hence, without loss of generality, one can assume that, for blocking coalition $((s_k,s_l),(h_p,h_q))$,

$$h_p \succeq_{s_k} u \text{ and } h_q \succeq_{s_l} u.$$

Suppose that $h_p \succ_{s_k} \mu(s_k)$ or $h_q \succ_{s_l} \mu(s_l)$. Then, according to (b2), (s_k,h_p) or (s_l,h_q) can block μ in $(P^H,P^S(P^C))$. Hence,

$$\mu(s_k) \succeq_{s_k} h_p \text{ and } \mu(s_l) \succeq_{s_l} h_q.$$

But then weak responsiveness (ii) implies $(\mu(s_k),\mu(s_l)) \succeq_c (h_p,\mu(s_l)) \succeq_c (h_p,h_q)$, which contradicts (b1). Hence, μ is also stable for (P^H,P^C). Finally, by Gale and Shapley (1962) a stable matching for $(P^H,P^S(P^C))$ always exists. □

The following example shows that not all stable matchings for (P^H,P^C) may be stable for $(P^H,P^S(P^C))$, even when couples are strongly unemployment averse and have responsive preferences. The intuition is that some matching that would be unstable in a singles market is now stable because a student may not want to block it by taking the position of his/her partner.

We use the following convention for this and future examples. If the unemployment option u is not listed for students, then both couples are strongly unemployment averse.

Example 3.4 ***(P^H,P^C) has More Stable Matchings than (P^H,$P^S(P^C)$)***
Consider the couples market (P^H,P^C) where preferences are given by Table 2 and the students' individual preferences P^S equal $P(s_1)=P(s_3)=h_4,h_1,h_2,h_3,u$ and $P(s_2)=P(s_4)=h_2,h_1,h_4,h_3,u$. It can easily be checked that the couples' preferences can be completed such that they are responsive with respect to the individual preferences (and are in addition identical). There are four stable matchings for the couples market (P^H,P^C) given by $\mu_1(S)=h_1,h_4,h_3,h_2$, $\mu_2(S)=h_4,h_1,h_3,h_2$, $\mu_3(S)=h_3,h_4,h_2,h_1$, and $\mu_4(S)=h_4,h_3,h_2,h_1$ (see Appendix). However, matching μ_2 is the unique stable matching for the associated singles market $(P^H,P^S(P^C))$. ◇

For our next result, a maximal domain result for the existence of stable matchings, we first introduce a weaker notion of strict unemployment aversion by requiring strict unemployment aversion only for "acceptable positions." Since we do not require any type of responsiveness for the couple's preferences, we adapt the definition of an acceptable position as follows. Let $c=(s_k,s_l)$ and $h \in H$. Then, *h is acceptable to student s_k* if $(h,u) \succ_c (u,u)$ and *h is acceptable to student s_l* if $(u,h) \succ_c (u,u)$.

Restricted strict unemployment aversion: Couple c has *restricted strictly unemployment averse preferences* if for any pair of acceptable positions it is worse off if one of its partners looses his/her acceptable position. Formally, for all (h_p,h_q) such that $(h_p,u) \succ_c (u,u)$ and $(u,h_q) \succ_c (u,u)$, $(h_p,h_q) \succ_c (h_p,u)$ and $(h_p,h_q) \succ_c (u,h_q)$.[11]

[11]The assumption of restricted strict unemployment aversion is particularly realistic in entry level labor markets where choosing unemployment, while acceptable jobs are available, may be harmful for future job prospects.

P^H					P^C	
h_1	h_2	h_3	h_4		$s_1 s_2$	$s_3 s_4$
s_4	s_4	s_2	s_2		$h_4 h_2$	$h_4 h_2$
s_2	s_3	s_3	s_4		$h_4 h_1$	$h_4 h_1$
s_1	s_2	s_1	s_1		$h_1 h_2$	$h_1 h_2$
s_3	s_1	s_4	s_3		$h_4 h_3$	$h_4 h_3$
					$h_1 h_4$	$h_1 h_4$
					$h_2 h_1$	$h_2 h_1$
					$h_1 h_3$	$h_1 h_3$
					$h_2 h_4$	$h_2 h_4$
					$h_3 h_2$	$h_3 h_2$
					$h_2 h_3$	$h_2 h_3$
					$h_3 h_1$	$h_3 h_1$
					$h_3 h_4$	$h_3 h_4$

Table 2: (P^H, P^C) has More Stable Matchings than $(P^H, P^S(P^C))$

Next we prove that if couples are restricted strictly unemployment averse, then the domain of weakly responsive preferences is a maximal domain for the existence of stable matchings. In other words, we show that in a couples market with only restricted strictly unemployment averse couples and at least one couple whose preferences are not weakly responsive, we can construct (weakly) responsive preferences for the other couple(s) such that no stable matching exists.

Theorem 3.5 ***Maximal Domain I***
For couples markets with restricted strictly unemployment averse couples, the domain of weakly responsive preferences is a maximal domain for the existence of stable matchings.

Proof. We prove the theorem by constructing a counter example for each possible violation of weak responsiveness. Assume that couple $c_1 = (s_1, s_2)$'s preferences are restricted strictly unemployment averse, but not weakly responsive. Consider $\succeq_{s_1}$ and $\succeq_{s_2}$ satisfying weak responsiveness condition (i). Since couple c_1's preferences are not weakly responsive, $\succeq_{s_1}$ and $\succeq_{s_2}$ satisfy weak responsiveness condition (ii). It follows that there exist $h_q, \overline{h}_q \in H \cup \{u\}$, $h_q \neq \overline{h}_q$ such that for some $h_p, h_r \in H \cup \{u\}$, $h_p \neq h_r$ we have

$[h_q, \overline{h}_q \succeq_{s_1} u,\ h_p \succeq_{s_2} u,\ (h_q, h_p) \succ_{c_1} (h_q, h_r)]$ and $[h_q, \overline{h}_q \succeq_{s_1} u,\ h_r \succeq_{s_2} u,\ (\overline{h}_q, h_r) \succ_{c_1} (\overline{h}_q, h_p)]$ or

$[h_p \succeq_{s_1} u,\ h_q, \overline{h}_q \succeq_{s_2} u,\ (h_p, h_q) \succ_{c_1} (h_r, h_q))]$ and $[h_r \succeq_{s_1} u,\ h_q, \overline{h}_q \succeq_{s_2} u,\ (h_r, \overline{h}_q) \succ_{c_1} (h_p, \overline{h}_q)]$.

Thus, with a slight abuse of notation[12] and without loss of generality,[13] there exist $h_1, h_2, h_3, h_4 \in H \cup \{u\}$ such that $h_1 \neq h_2$, $h_3 \neq h_4$, and

$$h_1, h_2 \succeq_{s_1} u,\ h_3, h_4 \succeq_{s_2} u,\ (h_1, h_3) \succ_{c_1} (h_1, h_4), \text{ and } (h_2, h_4) \succ_{c_1} (h_2, h_3).$$

[12]The objects $h_1, h_2, h_3, h_4 \in H \cup \{u\}$ may not be the four hospitals introduced in Section 2.
[13]The role of s_1 and s_2 can be switched.

Since preferences are complete, either (a) $(h_1, h_3) \succ_{c_1} (h_2, h_4)$ or (b) $(h_2, h_4) \succ_{c_1} (h_1, h_3)$. We construct a contradiction for Case (a) (Case (b) is analogous). By (a) and transitivity, $(h_1, h_3) \succ_{c_1} (h_2, h_3)$.

Thus, if $h_1 = u$, then $(u, h_3) \succ_{c_1} (h_2, h_3)$. Since $h_1 \neq h_2$, $h_2 \succ_{s_1} u$. If $h_3 = u$, then $h_2 \succ_{s_1} u$ and weak responsiveness condition (i) implies $(h_2, u) \succ_{c_1} (u, u)$. If $h_3 \neq u$, then $h_3 \succ_{s_2} u$ and weak responsiveness condition (i) implies $(u, h_3) \succ_{c_1} (u, u)$. Thus, by restricted strict unemployment aversion, $(h_2, h_3) \succ_{c_1} (u, h_3)$. Both cases contradict $(u, h_3) \succ_{c_1} (h_2, h_3)$. Hence, $h_1 \in H$. Similarly, weak responsiveness condition (i), restricted strict unemployment aversion, and $(h_1, h_3) \succ_{c_1} (h_1, h_4)$ imply $h_3 \in H$ and weak responsiveness condition (i), restricted strict unemployment aversion, and $(h_2, h_4) \succ_{c_1} (h_2, h_3)$ imply $h_4 \in H$. Now, for $h_1, h_3, h_4 \in H$ we specify

$P(h_1) = s_3, s_1, \emptyset, \ldots$,

$P(h_3) = s_2, s_3, \emptyset, \ldots$, and

$P(h_4) = s_2, \emptyset, \ldots$.

Couple $c_2 = (s_3, s_4)$ has restricted strictly unemployment averse responsive preferences based on $P(s_3) = h_3, h_1, u, \ldots$ and $P(s_4) = u, \ldots$.

Case 1: $h_2 \in H$. Let $P(h_2) = s_1, \emptyset, \ldots$. Note that for any individually rational matching μ, $\mu(c_2) \in \{(h_3, u), (h_1, u), (u, u)\}$. Assume that μ is stable.

If $\mu(c_2) = (u, u)$, then $\mu(c_1) = (h_1, h_3)$. Hence, μ is blocked by $(c_2, (h_1, u))$. If $\mu(c_2) = (h_1, u)$, then $\mu(c_1) = (h_2, h_4)$. Hence, μ is blocked by $(c_2, (h_3, u))$. If $\mu(c_2) = (h_3, u)$, then $\mu(c_1) = (h_1, h_4)$ or $\mu(c_1) = (h_2, h_4)$. Hence, μ is blocked by $(c_1, (h_1, h_3))$. Thus, all candidates for a stable matching are blocked.

Case 2: $h_2 = u$. Note that for any individually rational matching μ, $\mu(c_2) \in \{(h_3, u), (h_1, u), (u, u)\}$. Assume that μ is stable.

If $\mu(c_2) = (u, u)$, then $\mu(c_1) = (h_1, h_3)$. Hence, μ is blocked by $(c_2, (h_1, u))$. If $\mu(c_2) = (h_1, u)$, then $\mu(c_1) = (u, h_4)$. Hence, μ is blocked by $(c_2, (h_3, u))$. If $\mu(c_2) = (h_3, u)$, then $\mu(c_1) = (h_1, h_4)$. Hence, μ is blocked by $(c_1, (h_1, h_3))$. Thus, all candidates for a stable matching are blocked. □

It is easy to find examples that demonstrate that the domain of weakly responsive preferences is no longer maximal once restricted strict unemployment aversion is dropped. For instance a couple c with $P(c) = (h, u), (u, h'), (u, u), \ldots$ will never cause instability, no matter how the remaining preferences are specified.[14]

[14] We sketch the proof of this argument. Consider the following preference domain: any couple c's preferences are weakly responsive or are $P(c) = (h, u), (u, h'), (u, u), \ldots$, for some $h, h' \in H$. Obviously, the new domain strictly includes the domain of weakly responsive preferences. We construct a stable matching for any profile of preferences in the new domain as follows. First, construct associated individual preferences for couples with weakly responsive preferences. Second, for any couple $c = (s_k, s_l)$ with $P(c) = (h, u), (u, h'), (u, u), \ldots$, define associated individual preferences by $P(s_k) = h, u, \ldots$ and $P(s_l) = h', u, \ldots$. Now apply the student-optimal deferred acceptance algorithm (Gale and Shapley (1962)) to the associated singles market to obtain a tentative matching. If this matching is individually rational for all couples, then it is stable in the original couples market. Since individual rationality is automatically satisfied for weakly responsive couples, if individual rationality is violated for a couple $c = (s_k, s_l)$, then there are $h, h' \in H$ such that $P(c) = (h, u), (u, h'), (u, u), \ldots$ and $\mu(c) = (h, h')$. Redefine associated individual preferences by $P(s_k) = h, u, \ldots$ and $P(s_l) = u, \ldots$. The student-optimal deferred acceptance algorithm applied to the adjusted associated singles market gives another tentative matching where students are weakly better off. If this matching is individually rational for all couples, then it is stable in the original couples market. If individual rationality is violated for any couple, then redefine associated individual preferences, and so on. This procedure will finally produce a stable matching for the original couples market.

Note that the only weakly responsive preferences for a couple c that satisfy the stronger condition of strict unemployment aversion are responsive preferences where all hospitals are acceptable, that is, a strictly unemployment averse couple $c = (s_k, s_l)$ with weakly responsive preferences must in fact have responsive preferences with unique associated individual preferences that can be represented by $P(s_k) = \ldots, u$ and $P(s_l) = \ldots, u$.

Corollary 3.6 ***Maximal Domain II***
For couples markets with strictly unemployment averse couples, the domain of responsive preferences where all hospitals are considered acceptable in the associated individual preferences is a maximal domain for the existence of stable matchings.

Next, we drop the condition of (restricted) strict unemployment aversion and address the question whether or not one can enlarge the domain of (weakly) responsive preferences while still guaranteeing the existence of stable matchings. In fact, we start with a somewhat less ambitious task. First we relax the requirement of stability by excluding myopic behavior of blocking coalitions and ask for which reasonable preference domains "weakly stable" matchings always exist (see Klijn and Massó (2003) for weak stability in singles markets).

To model non-myopic behavior we assume that if the assignment of hospitals to students and students to hospitals that a blocking coalition proposes for themselves is not likely to be their final "match," then the blocking will not take place. Let μ be a matching and $((t_1, t_2), (l_1, l_2))$ be a blocking coalition. We model two cases in which a blocking coalition's match most likely will not be their final match:

- the couple (t_1, t_2) that participates in the blocking coalition $((t_1, t_2), (l_1, l_2))$ can do better for themselves in another blocking coalition $((t_1, t_2), (k_1, k_2))$ such that the other agents (one or both hospitals) that are participating in both blocking coalitions are not worse off.

  So, if couple (t_1, t_2) also blocks μ together with hospitals (k_1, k_2), then (t_1, t_2) prefers (k_1, k_2) to (l_1, l_2), which it would be matched with in the other blocking coalition, *i.e.*, (d1) $(k_1, k_2) \succ_{(t_1,t_2)} (l_1, l_2)$.

  If a hospital participates in both blocking coalitions, then it is not worse off, *i.e.*, if for some $i, j = 1, 2$, $k_i = l_j$, then (d2) $t_i \succeq_{k_i} t_j$.

- a hospital l_p that participates in the blocking coalition $((t_1, t_2), (l_1, l_2))$ can do better for itself in another blocking coalition $((z_1, z_2), (k_1, k_2))$ such that the other agents (the other hospital or the couple) participating in both blocking coalitions are not worse off.

  Let $l_p = k_r$, t_p be the student that is assigned to hospital l_p in blocking coalition $((t_1, t_2), (l_1, l_2))$, and z_r be the student that is assigned to hospital $k_r = l_p$ in blocking coalition $((z_1, z_2), (k_1, k_2))$.

  So, if hospital $k_r = l_p$ blocks μ together with hospital k_s ($\{k_r, k_s\} = \{k_1, k_2\}$) and couple (z_1, z_2), then it obtains a better student, *i.e.*, (d2) $z_r \succ_{l_p} t_p$.

  If the other hospital l_q participates in both blocking coalitions (*i.e.*, $k_s = l_q$), then it is not worse off, *i.e.*, (d2) $z_s \succeq_{l_q} t_q$.

  If the new blocking coalition is formed with the same couple, then it is not worse off, *i.e.*, (d1) $(k_1, k_2) \succeq_{(t_1,t_2)} (l_1, l_2)$.

We now give the formal definition. Let μ be a matching. We say that a blocking coalition $((t_1,t_2),(l_1,l_2))$ is *dominated* by another blocking coalition $((z_1,z_2),(k_1,k_2)) \neq ((t_1,t_2),(l_1,l_2))$, if

(d1) if $(z_1,z_2)=(t_1,t_2)$, then $(k_1,k_2) \succeq_{(z_1,z_2)} (l_1,l_2)$;

(d2) for all $i,j=1,2$, if $k_i = l_j \in H$, then $z_i \succeq_{k_i} t_j$;

(d3) $(z_1,z_2)=(t_1,t_2)$ or $k_i=l_j \in H$ for some $i,j=1,2$.[15]

A matching μ is *weakly stable* if it is individually rational and all blocking coalitions are dominated. Clearly, a stable matching is weakly stable. Note also that a matching with a single blocking coalition cannot be weakly stable. In some contexts it is natural to focus only on weakly stable matching with full employment (for instance when couples are strongly unemployment averse). For Roth's example (Table 1) there are three weakly stable matchings with full employment (see Appendix).

Now one might wonder whether with this weaker concept of stability we may extend the existence result in Theorem 3.3 to a larger class of preferences. For singles markets Klijn and Massó (2003) show that the set of weakly stable matchings contains Zhou's (1994) bargaining set. Hence, Zhou's (1994) result that in general the bargaining set is nonempty indicates that studying weak stability might be a fruitful approach. The next theorem, however, crushes any hope for this approach.

Theorem 3.7 ***No Weak Stability***
In couples markets the set of weakly stable matchings may be empty.

The following example proves Theorem 3.7. In the example, we construct a couples market where couples have leader-follower responsive preferences. Then, we create a new market by switching two pairs of hospitals in one couple's preference relation. However similar the two markets may seem, there is no weakly stable matching for the new market. In particular, there is no stable matching for the new market.

Example 3.8 ***Wanting to be a Little Bit Closer may create Instability***
Consider a couples market where preferences are given by Table 3 and the students' individual preferences equal $P(s_1)=h_3,h_4,h_1,h_2,u$, $P(s_2)=h_1,h_2,h_3,h_4,u$, and $P(s_3)=P(s_4)=h_2,h_1,h_3,h_4,u$. Differences in the students' individual preferences can be easily explained by "regional preferences": even though there may exist a unanimous ranking of hospitals according to prestige or salary, students may rank certain hospitals differently because they prefer to live in a certain region, for instance, they prefer to live on the West Coast instead of on the East Coast, or *vice versa*. Note that both couples are strongly unemployment averse and the first couple's preferences are leader-follower responsive. The second couple's preferences are obtained by first constructing leader-follower responsive preferences and then switching the last and second from last entries (in fact, only two hospitals for agent s_4 are switched – the switch is denoted in boldface in Table 3). This switch can be easily justified by assuming that hospital h_3 is closer than hospital h_2 to hospital h_4. In the worst case scenario where leader s_3 is assigned to h_4, the

[15] By (d3) we ensure that we only compare conflicting blocking coalitions in the sense that there exists at least one agent that is present in both blocking coalitions. Otherwise, domination is not possible.

couple's wish to be closer together may overrule any preference for follower s_4. Note also that the hospitals have identical preferences over students, which can be easily justified if hospitals rank students according to final grades or other test scores. It is tedious but not difficult to check that no weakly stable matching with full employment, and therefore by individual rationality no weakly stable matching, exists (see Appendix). ◇

P^H					P^C	
h_1	h_2	h_3	h_4		s_1s_2	s_3s_4
s_4	s_4	s_4	s_4		h_3h_1	h_2h_1
s_1	s_1	s_1	s_1		h_3h_2	h_2h_3
s_2	s_2	s_2	s_2		h_3h_4	h_2h_4
s_3	s_3	s_3	s_3		h_4h_1	h_1h_2
					h_4h_2	h_1h_3
					h_4h_3	h_1h_4
					h_1h_2	h_3h_2
					h_1h_3	h_3h_1
					h_1h_4	h_3h_4
					h_2h_1	$h_4\mathbf{h_3}$
					h_2h_3	h_4h_1
					h_2h_4	$h_4\mathbf{h_2}$

Table 3: Wanting to be a Little Bit Closer may create Instability

Example 3.8 exhibits almost responsive preferences, except for a single switch that can easily be explained by the desire of couple (s_3, s_4) to be closer together if the leader is assigned to hospital h_4, his/her worst option. Therefore, this example brings us closer to answering Roth and Sotomayor's (1990) question in the negative in the following sense. If we extend the domain of (weakly) responsive preferences to allow for non-responsive switches that could be due to distance considerations (which is the very reason that couples may have different preferences than if they were singles), then stable matchings may not exist.

The next example is another one without a stable matching that is based on simple preferences that can be explained intuitively. Note that in the previous example students have different regional preferences (see explanation in Example 3.8), which create different individual preferences. The following example deals with preferences that are based on identical individual preferences of students (no differences because of regional preferences). But, in addition, we assume that if positions are too far away, the unemployment of one partner may be preferred to being separated, that is, we drop the assumption of strong unemployment aversion. This example also illustrates how students' individual preferences may differ from the students' associated preferences as derived from the couples' preferences.[16]

[16]Cantala (2004) also studies the existence of stable matchings in relation to distance aspects. He shows nonexistence of stable matchings for some restricted preference domain, for instance he assumes that "preferences of couples satisfy the strong regional lexicographic conditions and that couples face the same geographical constraint."

Example 3.9 ***No Stability because Separation is out of the Question***
Consider the couples market (P^H, P^C) where preferences are given by Table 4 and the students' individual preferences P^S for $s \in S$ equal $P(s) = h_1, h_2, h_3, h_4, u$. Both couples have the same preference relation. Note that as singles all students like hospital h_1 best. However, assume that hospitals h_2, h_3, and h_4 are close together, while hospital h_1 is very far away. Now, instead of being separated, the partner of a student who is matched to hospital h_1 would not accept his/her position because unemployment is preferable to separation. When ranking matchings consisting of two positions, each couple uses lexicographic preferences with respect to the quality of the position. Note that if we focus only on individually rational matchings with full employment, then the agents' preferences are responsive. In this case, a student's derived associated individual preference over hospitals (excluding u) equals h_2, h_3, h_4, h_1. Comparing this to the student's individual preferences, we see that hospital h_1 moved from being the best position for the single student to being the worst position for the member of a couple, because working at h_1 either means separation from or unemployment of the partner.

It is easy to prove that no stable matching exists. Moreover, there is no weakly stable matching with full employment (this follows easily since any such matching is not individually rational). However, one can show for instance that the matching given by $\mu(S) = u, h_1, h_4, h_3$ is weakly stable. We prove these statements in the Appendix. $\diamond$

P^H				P^C	
h_1	h_2	h_3	h_4	s_1s_2	s_3s_4
s_1	s_1	s_1	s_1	h_2h_3	h_2h_3
s_3	s_3	s_3	s_3	h_2h_4	h_2h_4
s_4	s_4	s_4	s_4	h_3h_2	h_3h_2
s_2	s_2	s_2	s_2	h_3h_4	h_3h_4
				h_4h_2	h_4h_2
				h_4h_3	h_4h_3
				h_1u	h_1u
				h_2u	h_2u
				h_3u	h_3u
				h_4u	h_4u
				uh_1	uh_1
				uh_2	uh_2
				uh_3	uh_3
				uh_4	uh_4
				h_1h_2	h_1h_2
				h_1h_3	h_1h_3
				h_1h_4	h_1h_4
				h_2h_1	h_2h_1
				h_3h_1	h_3h_1
				h_4h_1	h_4h_1
				uu	uu

Table 4: Separation is out of the Question

4 Further Results for Stable Matchings when Preferences are Responsive: Optimality, Filled Positions, and Manipulation

Recall that when preferences are responsive one can construct a unique associated singles market with a nonempty set of stable matchings that is included in the set of stable matchings of the original couples market. In this section we analyze properties of the set of stable matchings for couples markets when preferences are responsive.

Apart from the fact that stable matchings always exist in the absence of couples, singles markets have other interesting features. If preferences are strict, the set of stable matchings has the structure of a (distributive) lattice, which we explain next.

Let (P^H, P^S) be a singles market and μ and μ' two of its matchings. We define a function $\lambda \equiv \mu \vee_S \mu'$ that assigns to each student his/her more preferred match from μ and μ'. Formally, let $\lambda = \mu \vee_S \mu'$ be defined for all $s \in S$ by $\lambda(s) := \mu(s)$ if $\mu(s) \succ_s \mu'(s)$ and $\lambda(s) := \mu'(s)$ otherwise. In a similar way we define the function $\mu \wedge_S \mu'$, which gives each student his/her less preferred match. In a dual way we define a function $\tilde{\lambda} \equiv \mu \vee_H \mu'$ that assigns to each hospital its more preferred match from μ and μ'. Formally, let $\tilde{\lambda} = \mu \vee_H \mu'$ be defined for all $h \in H$ by $\tilde{\lambda}(h) := \mu(h)$ if $\mu(h) \succ_h \mu'(h)$ and $\tilde{\lambda}(h) := \mu'(h)$ otherwise. In a similar way we define the function $\mu \wedge_H \mu'$, which gives each hospital its less preferred match.

For singles markets, Knuth (1976) published the following theorem, but it is attributed to John Conway. One of the implications of the theorem is that there is a polarization of interests between the two sides of the market along the set of stable matchings.

Theorem 4.1 ***Conway's Lattice Theorem for Singles Markets***
Let (P^H, P^S) be a singles market and μ and μ' be two stable matchings. Then, $\mu \vee_S \mu' = \mu \wedge_H \mu'$ and $\mu \wedge_S \mu' = \mu \vee_H \mu'$ are stable matchings. Furthermore, since the "sup operator" $\vee_S$ and the "inf operator" $\wedge_S$ satisfy the law of distributivity, the set of stable matchings for singles markets form a distributive lattice.

Conway's Lattice Theorem implies that there exists a unique best stable matching μ_S (called the *student-optimal matching*) favored by the students, which is the worst stable matching for the hospitals, and *vice versa* there exists a unique best stable matching μ_H (called the *hospital-optimal matching*) favored by the hospitals, which is the worst stable matching for the students. In fact, Gale and Shapley (1962) already proved the existence of μ_S and μ_H, and provided an algorithm, called the Deferred Acceptance algorithm, to calculate these matchings.

In the next theorem we demonstrate that for responsive preferences P^C the lattice structure of the set of stable matchings in $(P^H, P^S(P^C))$ need not carry over to (P^H, P^C). It strengthens the negative result (on the general domain of couples preferences) by Aldershof and Carducci (1996) that there may be no optimal matching for either side of the market. We first introduce some more notation.

Let μ be a matching for couples market (P^H, P^C). Then, for couple $c = (s_k, s_l)$, we define $\mu(c) := (\mu(s_k), \mu(s_l))$. For any two matchings μ and μ', we define a function $\bar{\lambda} \equiv \mu \vee_C \mu'$ that assigns to each couple its more preferred match from μ and μ'. Formally, let $\bar{\lambda} = \mu \vee_C \mu'$ be defined for all $c \in C$ by $\bar{\lambda}(c) := \mu(c)$ if $\mu(c) \succ_c \mu'(c)$ and $\bar{\lambda}(c) := \mu'(c)$ otherwise. In a similar way we define the function $\mu \wedge_C \mu'$, which gives each couple its less preferred match. The definition of functions $\mu \vee_H \mu'$ and $\mu \wedge_H \mu'$ is the same as before. The function $\bar{\lambda} = \mu \vee_C \mu'$ induces in a natural way a matching if $\bar{\lambda}(s) \neq \bar{\lambda}(t)$ for all students $s, t \in S$, $s \neq t$. Similar statements hold for $\mu \wedge_C \mu'$, $\mu \vee_H \mu'$, and $\mu \wedge_H \mu'$.

A stable matching $\overline{\mu}_C$ is the *couples maximal matching* if no other stable matching μ gives to any couple c a pair of positions $\mu(c)$ that the couple (weakly) prefers to $\overline{\mu}_C(c)$. A stable matching $\underline{\mu}_C$ is the *couples minimal matching* if no other stable matching μ gives to any couple c a pair of positions $\mu(c)$ that the couple likes (weakly) less than $\underline{\mu}_C(c)$. Similarly, a stable matching $\overline{\mu}_H$ is the *hospitals maximal matching* if no other stable matching μ gives to any hospital h a match $\mu(h)$ that the hospital (weakly) prefers to $\overline{\mu}_H(h)$. Finally, a stable matching $\underline{\mu}_H$ is the *hospitals minimal matching* if no other stable matching μ gives to any hospital h a match $\mu(h)$ that the hospital likes (weakly) less than $\underline{\mu}_H(h)$.

Theorem 4.2 ***Loss of Lattice Structure***
Let (P^H, P^C) be a couples market where couples have responsive preferences. Let μ and μ' be two stable matchings.

(i) *Functions $\mu \vee_C \mu'$, $\mu \wedge_C \mu'$, $\mu \vee_H \mu'$, and $\mu \wedge_H \mu'$ may not be matchings. Furthermore, the duality for singles markets between $\vee_S$ and $\wedge_H$ ($\wedge_S$ and $\vee_H$ respectively) need not carry over; that is, possibly $\mu \vee_C \mu' \neq \mu \wedge_H \mu'$ and $\mu \wedge_C \mu' \neq \mu \vee_H \mu'$.*

(ii) *The optimal matchings $\overline{\mu}_C$, $\underline{\mu}_C$, $\overline{\mu}_H$, and $\underline{\mu}_H$ may not exist.*

Proof.
(i) If we take $\mu = \mu_2$ and $\mu' = \mu_3$ ($\mu = \mu_2$ and $\mu' = \mu_4$) in Example 3.4, then $\mu \vee_C \mu'$ and $\mu \wedge_C \mu'$ ($\mu \vee_H \mu'$ and $\mu \wedge_H \mu'$) are not matchings.
(ii) It can be checked easily that none of the four stable matchings in Example 3.4 satisfies the definition of $\overline{\mu}_C$, $\underline{\mu}_C$, $\overline{\mu}_H$, or $\underline{\mu}_H$. □

Since in general there is more than one stable matching, a criterion one might want to employ to select a subset of stable matchings is (the maximization of) the number of matched agents. However, for singles markets the set of matched agents does not vary from one stable matching to another. In other words, for singles markets the set of unmatched agents is always the same for all stable matchings (McVitie and Wilson (1970), Roth (1982)).[17] In contrast, for couples markets Aldershof and Carducci (1996) show that on the general domain of couples preferences different stable matchings may have a different number of positions filled. We strengthen this result by showing that on the restricted domain of responsive preferences the number of positions that are filled at different stable matchings may vary as well.

Theorem 4.3 ***Different Number of Filled Positions across Stable Matchings***
Let (P^H, P^C) be a couples market where couples have responsive preferences. Then there may be stable matchings that leave different numbers of positions unfilled.

The following example, which is a slight variation of the example used by Aldershof and Carducci (1996), proves Theorem 4.3.

[17] Martínez *et al.* (2000) study this question for many-to-one matching markets without money. They show that in case hospitals have "substitutable" preferences the number of matched agents may vary from one stable matching to another. On the positive side, however, they establish that if preferences profiles satisfy certain axioms then the set of unmatched agents is the same under every stable matching, among other desirable properties.

Example 4.4 Consider a couples market where preferences are given by Table 5 and the students' individual preferences equal $P(s_1) = h_3, h_2, u, h_4, h_1$, $P(s_2) = h_3, h_2, h_4, u, h_1$, $P(s_3) = h_2, h_1, h_4, u, h_3$, and $P(s_4) = h_1, h_3, u, h_2, h_4$. It can easily be checked that the couples' preferences can be completed such that they are responsive with respect to the individual preferences. There are two stable matchings given by $\mu_1(H) = s_4, s_2, s_1, s_3$ and $\mu_2(H) = s_4, s_3, s_2, \emptyset$ (see Appendix), which obviously leave different numbers of positions unfilled.

For later use we note that one can easily check that for any associated singles market $(P^H, P^S(P^C))$ the unique stable matching is μ_2, which hence is the outcome of the Deferred Acceptance algorithm by Gale and Shapley (1962), *i.e.*, both the student-optimal and hospital-optimal matching. $\diamond$

P^H				P^C	
h_1	h_2	h_3	h_4	s_1s_2	s_3s_4
s_4	s_2	s_2	s_2	h_3h_2	h_2h_1
s_3	s_3	s_4	s_3	h_2h_3	h_2h_3
$\emptyset$	s_1	s_1	$\emptyset$	h_3h_4	h_4h_1
$\cdots$	$\emptyset$	$\emptyset$	$\cdots$	h_2h_4	h_1h_3
	s_4	s_3		uh_3	h_4h_3
				uh_2	h_2u
				uh_4	h_1u
				h_3u	h_4u
				h_2u	uh_1
				uu	uh_3
				$\cdots$	uu
					$\cdots$

Table 5: Different Number of Filled Positions across Stable Matchings

Note that we could also use Example 4.4 to prove Theorem 4.2. However, in Example 4.4 it is essential that some students are not acceptable for some hospitals and that some couples find certain positions unacceptable. Example 3.4 demonstrates that the negative results in Theorem 4.2 remain true on the smaller domain of responsive preferences where all hospitals and all students are mutually acceptable.

Before stating our next result, we define a *matching mechanism* as a function that assigns a matching to each couples market. A *stable-matching mechanism* is a matching mechanism that assigns a stable matching to a couples market whenever it has a nonempty set of stable matchings. A stable-matching mechanism is *strategy-proof* if no couple and no hospital can ever benefit from misrepresenting its preferences. In other words, a stable-matching mechanism is strategy-proof if truth telling is a dominant strategy.

Our final result on the set of stable matchings for couples markets with responsive preferences is that there exists no strategy-proof stable-matching mechanism based on revealed preferences. More precisely, we show that there is no stable-matching mechanism for which stating the true preferences is a dominant strategy for *every couple*.

Theorem 4.5 ***No Strategy-Proof Stable-Matching Mechanism***
There is no stable-matching mechanism for couples markets with responsive preferences for which stating the true preferences is a dominant strategy for every couple. In other words, for any stable-matching mechanism there exists a couples market with responsive preferences such that at least one couple can profit from misrepresenting its preferences.

Proof. To prove the theorem we consider the couples market (P^H, P^C) in Example 4.4 (where couples' preferences are responsive). We show that every stable-matching mechanism makes it possible for some couple to profit by misrepresenting its preferences.

Suppose the mechanism chooses matching μ_1. If couple c_2 changes its preferences from $P(c_2)$ to responsive preferences $Q(c_2) = (h_2, h_3), (h_2, h_1), (h_1, h_3), (h_4, h_3), (h_4, h_1), (h_2, u), (h_1, u), (h_4, u), (u, h_3), (u, h_1), (u, u), \ldots$ while everyone else states their true preferences, then μ_2 is the only stable matching with respect to the stated preferences $(P^H, \{P(c_1), Q(c_2)\})$, see Appendix. So, any stable-matching mechanism must select μ_2 when the stated preferences are $(P^H, \{P(c_1), Q(c_2)\})$. Since $\mu_2(c_2) = (h_2, h_1) \succ_{c_2} (h_4, h_1) = \mu_1(c_2)$, it is not a dominant strategy for couple c_2 to state its true preferences.

Suppose the mechanism chooses matching μ_2. If couple c_1 changes its preferences from $P(c_1)$ to responsive preferences $Q(c_1) = (h_3, h_2), (h_3, u), (u, h_2), (u, u), \ldots$ while everyone else states their true preferences, then μ_1 is the only stable matching with respect to the stated preferences $(P^H, \{Q(c_1), P(c_2)\})$, see Appendix. So, any stable-matching mechanism must select μ_1 when the stated preferences are $(P^H, \{Q(c_1), P(c_2)\})$. Since $\mu_1(c_1) = (h_3, h_2) \succ_{c_1} (u, h_3) = \mu_2(c_1)$, it is not a dominant strategy for couple c_1 to state its true preferences. $\square$

In fact, if in Example 4.4 the stable-matching mechanism chooses matching μ_1, then also hospital h_3 can profit by changing its preferences from $P(h_3)$ to $Q(h_3) = s_2, s_4, \emptyset, \ldots$, since the unique stable matching for $(\{P(h_1), P(h_2), Q(h_3), P(h_4)\}, P^C)$ is μ_2 and $\mu_2(h_3) = s_2 \succ_{h_3} s_1 = \mu_1(h_3)$ (see Appendix).

Note that it is not surprising that none of the hospitals in Example 4.4 can profitably misstate its preferences when the matching mechanism chooses μ_2. The reason for this is that μ_2 is the hospital-optimal matching in (any of) the associated singles market(s) (see remark in Example 4.4). By a result due to Dubins and Freedman (1981) and Roth (1982) it is a dominant strategy for each hospital to state its true preferences in the associated singles market(s) if the hospital-optimal matching is always picked. In other words, a misrepresentation of some hospital h's preferences in a couples market will always give rise to a stable matching that is weakly worse for h compared with the hospital-optimal matching (of the associated singles market(s)) when stating its true preferences. The following possibility theorem, which also holds on the domain of weakly responsive preferences, is an immediate consequence of this observation.

Theorem 4.6 ***No Profitable Misrepresentation by Individual Hospitals***
The stable-matching mechanism that yields the hospital-optimal matching (in the associated singles market(s)) makes it a dominant strategy for each hospital to state its true preferences.

Remark 4.7 ***Discussion of Possible Implications of Theorems 4.3, 4.5, and 4.6.***
One of the main results of this article is the existence of stable matchings if couples have (weakly) responsive preferences. If a labor market the couples apply to is regional and/or the positions' duration is short, which for example is the case of some U.K. entry level labor markets for

physicians and surgeons (see Roth (1991)), then it seems likely that couples have responsive preferences. Thus, given such a situation, one could derive the (unique) associated individual preferences from the couples' preferences and apply the Deferred Acceptance algorithm by Gale and Shapley (1962) to obtain a stable matching. However, in view of Theorem 4.3 it is not clear whether this is desirable regarding the number of matched agents. For instance, consider Example 4.4 where the Deferred Acceptance algorithm picks stable matching μ_2 which does not maximize the number of matched agents and leaves one agent unemployed. On the other hand, Theorem 4.6 shows that if we choose μ_2, which is the hospital-optimal matching resulting from the Deferred Acceptance algorithm, then at least hospitals have no incentives to misrepresent their preferences. However, no matter which matching the stable-matching mechanism chooses, by Theorem 4.5 stating their true preferences is not a dominant strategy for every couple.

5 Responsiveness for Couples Markets and Previous Notions of Substitutability for Singles Markets

As already discussed in the Introduction, our existence results of stable matchings when couples have (weakly) responsive preferences to some extent mirrors other results that demonstrate that a sufficient amount of substitutability implies the existence of stable matchings for the matching market in question; see for instance Roth (1985, many-to-one matching without money, also called the college admissions model), Kelso and Crawford (1982, many-to-one matching with money), Abdulkadiroğlu (2003, college admissions with affirmative action), and Hatfield and Milgrom (2004, two-sided matching with contracts). In contrast to our notion of (weak) responsiveness, all substitutability conditions in these papers apply to the preferences of the supply side (hospitals or firms) over sets of agents (students or workers), while our responsiveness condition applies to preferences of the demand side (couples of students) over ordered pairs of hospitals (not sets!). Alkan and Gale's (2003, many-to-many schedule matching) substitutability condition of "persistence" in fact applies to both, the demand and the supply side, but still does not apply to ordered pairs as in our model. It is interesting to note that both, Alkan and Gale's "persistence" as well as Hatfield and Milgrom's (2004) "substitution" condition encompass Roth's (1985) "responsiveness" and Kelso and Crawford's (1982) "gross substitutes" condition. Here, in order to compare our results with previous results for "matching markets with substitutability" in a comprehensive way, we focus on Hatfield and Milgrom's (2004) results for matching markets with contracts.

Hatfield and Milgrom (2004) present a new model of matching with contracts that encompasses some of the previous classical models such as Gale and Shapley's (1962) and Roth's (1985) college admissions problem or the Kelso-Crawford's (1982) *tâtonnement* model of wage determination in labor markets. In one of their main results Hatfield and Milgrom (2004, Theorem 5) identify a maximal set of preferences over contracts (the domain of substitutable preferences) for which a stable matching exists. The proof of this result inspired our proof of Theorem 3.5 (Maximal Domain I result), but in addition to our substitutability requirement of responsiveness, we had to add restricted unemployment aversion (the fact that we consider ordered pairs of positions for couples changes the formulation and several parts of the proof). Once we drop the unemployment aversion requirement, we were not able to obtain a similar maximal domain result. Instead, we demonstrated with two instructive examples (Examples 3.8 and 3.9) how a single couple may cause a labor market to be unstable even though its preferences may be almost responsive. Example 3.8 also proves that once preferences are not responsive, even weakly stable matchings may not exist.

In addition to the maximal domain result, Hatfield and Milgrom (2004, Theorem 3) demonstrate that under their substitution condition, a stable matching can be obtained by applying a generalization of Gale and Shapley's (1962) Deferred Acceptance algorithm. In contrast to this approach, we show that if preferences are (weakly) responsive one can construct a singles market such that any stable matching in the singles market is also stable in the original couples market (Theorem 3.3). In particular, this means that we can construct a stable matching for any (weakly) responsive couples market using the original Gale and Shapley's (1962) Deferred Acceptance algorithm. In fact, a generalization of the Deferred Acceptance mechanism à la Hatfield and Milgrom (2004) would not work for responsive couples markets since the set of stable matchings may not form a lattice (see Theorem 4.2 versus Hatfield and Milgrom (2004), Theorem 4).

Next, Theorem 4.3 demonstrates that even for responsive couples markets different numbers of positions may be filled across stable matchings. Hatfield and Milgrom (2004) also confirm this violation of the so-called rural hospital theorem (its original version is due to Roth (1986)) for matching markets with contracts under the assumption of substitutability. However, by additionally requiring that preferences on the supply side also satisfy the "law of aggregate demand," they are able to restore the rural hospital theorem (Hatfield and Milgrom (2004), Theorem 8). Since the definition of the law of aggregate demand depends on the cardinality of sets of students chosen by the hospitals, no corresponding requirement exists for couples markets where we compare ordered pairs of positions and not the sets of positions a couple consumes.

Finally, we prove that there is no stable-matching mechanism for couples markets with responsive preferences for which stating the true preferences is a dominant strategy for every couple (Theorem 4.5). A similar result has been obtained already for the college admissions problem: Roth (1985) proves that even though colleges have responsive preferences over sets of students, no stable-matching mechanism exists that makes it a dominant strategy for all colleges to state their true preferences. Since Hatfield and Milgrom's (2004) model encompasses Roth's (1985) formulation of the college admissions problem with responsive preferences, Roth's counterexample also holds in the matching with contracts context.

6 Appendix: Remaining Proofs

Proof of statement in Example 3.4: In Table 6 we list all 24 individually rational (full employment) matchings for the couples market with preferences given by Table 2. For each of the 20 unstable matchings we provide a blocking coalition. Note that the stable matchings detected in Table 6 correspond to the stable matchings listed in Example 3.4 as follows: matching no. 6 $\sim \mu_1$, no. 12 $\sim \mu_2$, no. 23 $\sim \mu_3$, and no. 24 $\sim \mu_4$. □

Proof of existence of three weakly stable matching with full employment for Roth's (1984) example: We show that for the couples market with preferences given by Table 1 there are at least three weakly stable matchings with full employment. It can easily be checked that none of the other 21 individually rational (full employment) matchings is weakly stable.[18] In Table 7 we list the three weakly stable matchings along with *all* blocking coalitions.

[18]We only want to point out that even if there are no stable matchings, there may be weakly stable matchings. In fact, as this example shows, the set of weakly stable matchings may contain more than one matching. For this

For each matching and for each blocking coalition we provide another, dominating blocking coalition. □

Proof of nonexistence of weakly stable matchings in Example 3.8: We still have to check that for the couples market with preferences given by Table 3 none of the 24 individually rational (full employment) matchings is weakly stable. We do this below by providing in Tables 8 and 9 at least one undominated blocking coalition for each full employment matching. □

Proof of statements in Example 3.9: To show that for the couples market defined by Table 4 no stable matching exists, let $\mathcal{H}^*$ be the seven most preferred hospital combinations depicted in Table 4, *i.e.*, $\mathcal{H}^* = \{(h_2,h_3),(h_2,h_4),(h_3,h_2),(h_3,h_4),(h_4,h_2),(h_4,h_3),(h_1,u)\}$.

Let μ be a stable matching. Suppose that $(\mu(s_1),\mu(s_2)) \notin \mathcal{H}^*$. Then, $((s_1,s_2),(h_1,\emptyset))$ is a blocking coalition. Hence, $(\mu(s_1),\mu(s_2)) \in \mathcal{H}^*$.

Suppose that $(\mu(s_3),\mu(s_4)) \notin \mathcal{H}^*$. If $(\mu(s_1),\mu(s_2)) = (h_1,u)$, then $((s_3,s_4),(h_2,h_3))$ or $((s_3,s_4),(h_2,h_4))$ is a blocking coalition. If $(\mu(s_1),\mu(s_2)) \neq (h_1,u)$, then $((s_3,s_4),(h_1,\emptyset))$ is a blocking coalition. Hence, $(\mu(s_3),\mu(s_4)) \in \mathcal{H}^*$.

So, μ is one of the 12 matchings depicted in Table 10. However, for each of these matchings a blocking coalition exists: a contradiction. Hence, there is no stable matching.

It remains to be proven that the (individually rational) matching μ given by $\mu(S) = u,h_1,h_4,h_3$ is weakly stable. In Table 11 we list all blocking coalitions for this matching, along with the blocking coalitions they are dominated by. Since each blocking coalition is dominated by some other blocking coalition it follows that μ is weakly stable. □

Proof of statement in Example 4.4: To show that for the couples market defined by Table 5 the only two stable matchings are given by $\mu_1(H) = s_4,s_2,s_1,s_3$ and $\mu_2(H) = s_4,s_3,s_2,\emptyset$, we first consider all 69 individually rational matchings. If we delete all the matchings that leave a student and a hospital unmatched, while in fact they are mutually acceptable, then only the 13 matchings in Table 12 remain. In the table we give for each matching a blocking coalition whenever possible. Note that the stable matchings detected in Table 12 correspond to the stable matchings listed in Example 4.4 as follows: matching no. 11 $\sim \mu_1$ and no. 13 $\sim \mu_2$. □

Proof of statements in Theorem 4.5: First we prove that μ_2 is the only stable matching in couples market $(P^H,\{P(c_1),Q(c_2)\})$. We consider all 69 individually rational matchings. If we delete all the matchings that leave a student and a hospital unmatched, while in fact they are mutually acceptable, then only the 13 matchings in Table 13 remain. In the table we give for each matching a blocking coalition whenever possible. Note that the only stable matching detected in Table 13 is μ_2.

It remains to prove that μ_1 is the only stable matching in couples market $(P^H,\{Q(c_1),P(c_2)\})$. We consider all 31 individually rational matchings. If we delete all the matchings that leave a student and a hospital unmatched, while in fact they are mutually acceptable, then only the 6 matchings in Table 14 remain. In the table we give for each matching a blocking coalition whenever possible. Note that the only stable matching detected in Table 14 is μ_1. □

reason, and also to save space, we do not elaborate the proof that there are no other weakly stable matchings, which can be obtained upon request.

Proof of statement right after Theorem 4.5: We have to prove that μ_2 is the unique stable matching for $(\{P(h_1), P(h_2), Q(h_3), P(h_4)\}, P^C)$. We consider all 52 individually rational matchings. If we delete all the matchings that leave a student and a hospital unmatched, while in fact they are mutually acceptable, then only the 11 matchings in Table 15 remain. In the table we give for each matching a blocking coalition whenever possible. Note that the only stable matching detected in Table 15 is μ_2. $\square$

	Hospitals				Blocking coalitions?	
no.	h_1	h_2	h_3	h_4	Students	Hospitals
1	s_1	s_2	s_3	s_4	(s_3, s_4)	(h_2, h_1)
2	s_1	s_2	s_4	s_3	(s_1, s_2)	(h_4, h_2)
3	s_1	s_3	s_2	s_4	(s_1, s_2)	(h_1, h_4)
4	s_1	s_3	s_4	s_2	(s_3, s_4)	(h_2, h_1)
5	s_1	s_4	s_2	s_3	(s_1, s_2)	(h_4, h_1)
6	s_1	s_4	s_3	s_2	−−	−−
7	s_2	s_1	s_3	s_4	(s_3, s_4)	(h_2, h_1)
8	s_2	s_1	s_4	s_3	(s_1, s_2)	(h_4, h_2)
9	s_2	s_3	s_1	s_4	(s_3, s_4)	(h_2, h_1)
10	s_2	s_3	s_4	s_1	(s_3, s_4)	(h_2, h_1)
11	s_2	s_4	s_1	s_3	(s_1, s_2)	(h_4, h_1)
12	s_2	s_4	s_3	s_1	−−	−−
13	s_3	s_1	s_2	s_4	(s_1, s_2)	(h_1, h_2)
14	s_3	s_1	s_4	s_2	(s_1, s_2)	(h_1, h_2)
15	s_3	s_2	s_1	s_4	(s_1, s_2)	(h_1, h_2)
16	s_3	s_2	s_4	s_1	(s_3, s_4)	(h_1, h_2)
17	s_3	s_4	s_1	s_2	(s_1, s_2)	(h_1, h_4)
18	s_3	s_4	s_2	s_1	(s_1, s_2)	(h_4, h_1)
19	s_4	s_1	s_2	s_3	(s_1, s_2)	(h_4, h_2)
20	s_4	s_1	s_3	s_2	(s_3, s_4)	(h_2, h_1)
21	s_4	s_2	s_1	s_3	(s_1, s_2)	(h_4, h_2)
22	s_4	s_2	s_3	s_1	(s_3, s_4)	(h_2, h_1)
23	s_4	s_3	s_1	s_2	−−	−−
24	s_4	s_3	s_2	s_1	−−	−−

Table 6: Example 3.4 / Table 2, all individually rational matchings (with a blocking coalition when possible)

	Hospitals				Blocking coalitions		Dominated by	
no.	h_1	h_2	h_3	h_4	Students	Hospitals		
1	s_3	s_1	s_2	s_4	(s_1, s_2)	(h_1, h_2)	(s_3, s_4)	(h_2, h_4)
					(s_1, s_2)	(h_1, h_4)	(s_1, s_2)	(h_1, h_2)
					(s_1, s_2)	(h_1, h_3)	(s_1, s_2)	(h_1, h_2)
					(s_3, s_4)	(h_2, h_4)	(s_1, s_2)	(h_1, h_4)
					(s_3, s_4)	(h_2, h_1)	(s_3, s_4)	(h_2, h_4)
					(s_3, s_4)	(h_1, h_2)	(s_1, s_2)	(h_1, h_3)
2	s_3	s_1	s_4	s_2	(s_1, s_2)	(h_1, h_2)	(s_3, s_4)	(h_3, h_1)
					(s_1, s_2)	(h_1, h_4)	(s_1, s_2)	(h_1, h_2)
					(s_1, s_2)	(h_1, h_3)	(s_1, s_2)	(h_1, h_2)
					(s_1, s_2)	(h_3, h_4)	(s_1, s_2)	(h_1, h_2)
					(s_1, s_2)	(h_3, h_1)	(s_1, s_2)	(h_3, h_4)
					(s_1, s_2)	(h_3, h_2)	(s_1, s_2)	(h_3, h_4)
					(s_1, s_2)	(h_2, h_3)	(s_1, s_2)	(h_1, h_4)
					(s_3, s_4)	(h_3, h_1)	(s_1, s_2)	(h_2, h_3)
					(s_3, s_4)	(h_3, h_2)	(s_3, s_4)	(h_3, h_1)
					(s_3, s_4)	(h_2, h_1)	(s_3, s_4)	(h_3, h_1)
					(s_3, s_4)	(h_2, h_3)	(s_3, s_4)	(h_3, h_1)
					(s_3, s_4)	(h_1, h_2)	(s_3, s_4)	(h_3, h_2)
3	s_3	s_2	s_4	s_1	(s_1, s_2)	(h_1, h_2)	(s_3, s_4)	(h_2, h_4)
					(s_1, s_2)	(h_4, h_1)	(s_3, s_4)	(h_2, h_1)
					(s_1, s_2)	(h_4, h_3)	(s_1, s_2)	(h_1, h_2)
					(s_3, s_4)	(h_3, h_1)	(s_1, s_2)	(h_4, h_3)
					(s_3, s_4)	(h_3, h_2)	(s_3, s_4)	(h_3, h_1)
					(s_3, s_4)	(h_3, h_4)	(s_3, s_4)	(h_3, h_1)
					(s_3, s_4)	(h_2, h_4)	(s_3, s_4)	(h_3, h_1)
					(s_3, s_4)	(h_2, h_1)	(s_3, s_4)	(h_3, h_1)
					(s_3, s_4)	(h_2, h_3)	(s_3, s_4)	(h_3, h_1)
					(s_3, s_4)	(h_1, h_2)	(s_3, s_4)	(h_3, h_2)
					(s_3, s_4)	(h_1, h_4)	(s_3, s_4)	(h_1, h_2)

Table 7: Roth's (1984) example / Table 1, three weakly stable matchings

	Hospitals				Blocking coalitions		
no.	h_1	h_2	h_3	h_4	Students	Hospitals	Undominated?
1	s_1	s_2	s_3	s_4	(s_1, s_2)	(h_3, h_2)	x
					(s_3, s_4)	(h_3, h_2)	x
					(s_3, s_4)	(h_3, h_1)	
2	s_1	s_2	s_4	s_3	(s_1, s_2)	(h_4, h_2)	x
3	s_1	s_3	s_2	s_4	(s_1, s_2)	(h_3, h_2)	x
					(s_1, s_2)	(h_1, h_2)	
					(s_3, s_4)	(h_2, h_1)	
					(s_3, s_4)	(h_2, h_3)	
4	s_1	s_3	s_4	s_2	(s_1, s_2)	(h_4, h_2)	x
					(s_1, s_2)	(h_1, h_2)	
					(s_3, s_4)	(h_2, h_1)	
5	s_1	s_4	s_2	s_3	(s_1, s_2)	(h_3, h_4)	x
					(s_1, s_2)	(h_4, h_3)	x
					(s_3, s_4)	(h_4, h_3)	x
					(s_3, s_4)	(h_4, h_1)	
6	s_1	s_4	s_3	s_2	(s_1, s_2)	(h_3, h_4)	x
					(s_1, s_2)	(h_4, h_3)	x
					(s_1, s_2)	(h_1, h_3)	
7	s_2	s_1	s_3	s_4	(s_1, s_2)	(h_3, h_1)	x
					(s_1, s_2)	(h_1, h_3)	x
					(s_3, s_4)	(h_3, h_2)	
					(s_3, s_4)	(h_3, h_1)	
8	s_2	s_1	s_4	s_3	(s_1, s_2)	(h_4, h_1)	x
					(s_1, s_2)	(h_1, h_4)	x
9	s_2	s_3	s_1	s_4	(s_3, s_4)	(h_2, h_1)	x
					(s_3, s_4)	(h_2, h_3)	
10	s_2	s_3	s_4	s_1	(s_3, s_4)	(h_2, h_1)	x
11	s_2	s_4	s_1	s_3	(s_3, s_4)	(h_4, h_3)	x
					(s_3, s_4)	(h_4, h_1)	
12	s_2	s_4	s_3	s_1	(s_1, s_2)	(h_3, h_1)	x
13	s_3	s_1	s_2	s_4	(s_1, s_2)	(h_3, h_1)	x
					(s_1, s_2)	(h_1, h_3)	x
					(s_1, s_2)	(h_2, h_1)	
					(s_3, s_4)	(h_1, h_2)	
					(s_3, s_4)	(h_1, h_3)	

Table 8: Example 3.8 / Table 3, matchings 1-13 not weakly stable

no.	Hospitals h_1	h_2	h_3	h_4	Blocking coalitions Students	Hospitals	Undominated?
14	s_3	s_1	s_4	s_2	(s_1, s_2)	(h_4, h_1)	x
					(s_1, s_2)	(h_1, h_4)	x
					(s_1, s_2)	(h_2, h_1)	
					(s_3, s_4)	(h_1, h_2)	
15	s_3	s_2	s_1	s_4	(s_1, s_2)	(h_3, h_1)	x
					(s_3, s_4)	(h_1, h_2)	
					(s_3, s_4)	(h_1, h_3)	
16	s_3	s_2	s_4	s_1	(s_1, s_2)	(h_4, h_1)	x
					(s_3, s_4)	(h_1, h_2)	
17	s_3	s_4	s_1	s_2	(s_1, s_2)	(h_3, h_1)	x
18	s_3	s_4	s_2	s_1	(s_1, s_2)	(h_3, h_1)	x
					(s_1, s_2)	(h_4, h_1)	
19	s_4	s_1	s_2	s_3	(s_1, s_2)	(h_3, h_4)	x
					(s_1, s_2)	(h_4, h_3)	x
					(s_3, s_4)	(h_4, h_3)	x
20	s_4	s_1	s_3	s_2	(s_1, s_2)	(h_3, h_4)	x
					(s_1, s_2)	(h_4, h_3)	x
					(s_1, s_2)	(h_2, h_3)	
					(s_3, s_4)	(h_3, h_2)	
21	s_4	s_2	s_1	s_3	(s_3, s_4)	(h_4, h_3)	x
22	s_4	s_2	s_3	s_1	(s_1, s_2)	(h_3, h_2)	x
					(s_3, s_4)	(h_3, h_2)	x
23	s_4	s_3	s_1	s_2	(s_1, s_2)	(h_3, h_2)	x
24	s_4	s_3	s_2	s_1	(s_1, s_2)	(h_3, h_2)	x
					(s_1, s_2)	(h_4, h_2)	

Table 9: Example 3.8 / Table 3, matchings 14-24 not weakly stable

	Students				A blocking coalition	
no.	s_1	s_2	s_3	s_4	Students	Hospitals
1	h_1	u	h_2	h_3	(s_1, s_2)	(h_2, h_4)
2	h_1	u	h_2	h_4	(s_1, s_2)	(h_2, h_3)
3	h_1	u	h_3	h_2	(s_1, s_2)	(h_2, h_4)
4	h_1	u	h_3	h_4	(s_1, s_2)	(h_3, h_2)
5	h_1	u	h_4	h_2	(s_1, s_2)	(h_2, h_3)
6	h_1	u	h_4	h_3	(s_1, s_2)	(h_3, h_2)
7	h_2	h_3	h_1	u	(s_3, s_4)	(h_3, h_4)
8	h_2	h_4	h_1	u	(s_3, s_4)	(h_4, h_3)
9	h_3	h_2	h_1	u	(s_3, s_4)	(h_2, h_4)
10	h_3	h_4	h_1	u	(s_3, s_4)	(h_4, h_2)
11	h_4	h_2	h_1	u	(s_3, s_4)	(h_2, h_3)
12	h_4	h_3	h_1	u	(s_3, s_4)	(h_3, h_2)

Table 10: Example 3.9 / Table 4, no stable matchings

	Blocking coalitions		
no.	Students	Hospitals	Dominated by no.
1	(s_1, s_2)	(h_3, h_2)	9
2	(s_1, s_2)	(h_4, h_2)	1
3	(s_1, s_2)	(h_1, u)	2
4	(s_1, s_2)	(h_2, u)	3
5	(s_1, s_2)	(h_3, u)	4
6	(s_1, s_2)	(h_4, u)	5
7	(s_3, s_4)	(h_2, h_3)	5
8	(s_3, s_4)	(h_3, h_2)	5
9	(s_3, s_4)	(h_4, h_2)	8

Table 11: Example 3.9 / Table 4, a weakly stable matching: $\mu(S) = u, h_1, h_4, h_3$

	Hospitals				Blocking coalitions?	
no.	h_1	h_2	h_3	h_4	Students	Hospitals
1	$\emptyset$	s_2	s_4	s_3	(s_3, s_4)	(h_4, h_1)
2	$\emptyset$	s_3	s_4	s_2	(s_3, s_4)	(h_2, h_1)
3	$\emptyset$	s_1	s_4	s_3	(s_3, s_4)	(h_4, h_1)
4	s_3	s_1	s_2	$\emptyset$	(s_3, s_4)	(h_2, h_1)
5	s_3	s_1	s_4	s_2	(s_3, s_4)	(h_2, h_3)
6	s_3	s_2	s_1	$\emptyset$	(s_3, s_4)	(h_1, h_3)
7	s_3	s_2	s_4	$\emptyset$	(s_1, s_2)	(u, h_3)
8	s_3	$\emptyset$	s_1	s_2	(s_1, s_2)	(h_3, h_2)
9	s_4	s_1	s_2	s_3	(s_3, s_4)	(h_2, h_1)
10	s_4	s_1	$\emptyset$	s_2	(s_3, s_4)	(h_2, h_1)
11	s_4	s_2	s_1	s_3	$--$	$--$
12	s_4	s_3	s_1	s_2	(s_1, s_2)	(h_3, h_2)
13	s_4	s_3	s_2	$\emptyset$	$--$	$--$

Table 12: Example 4.4 / Table 5, 13 matchings (with a blocking coalition when possible)

	Hospitals				Blocking coalitions?	
no.	h_1	h_2	h_3	h_4	Students	Hospitals
1	$\emptyset$	s_2	s_4	s_3	(s_3, s_4)	(h_1, h_3)
2	$\emptyset$	s_3	s_4	s_2	(s_1, s_2)	(u, h_3)
3	$\emptyset$	s_1	s_4	s_3	(s_1, s_2)	(h_2, h_3)
4	s_3	s_1	s_2	$\emptyset$	(s_3, s_4)	(h_2, h_1)
5	s_3	s_1	s_4	s_2	(s_1, s_2)	(h_2, h_3)
6	s_3	s_2	s_1	$\emptyset$	(s_3, s_4)	(h_1, h_3)
7	s_3	s_2	s_4	$\emptyset$	(s_1, s_2)	(u, h_3)
8	s_3	$\emptyset$	s_1	s_2	(s_1, s_2)	(h_3, h_2)
9	s_4	s_1	s_2	s_3	(s_3, s_4)	(h_2, h_1)
10	s_4	s_1	$\emptyset$	s_2	(s_1, s_2)	(h_3, h_2)
11	s_4	s_2	s_1	s_3	(s_3, s_4)	(h_4, h_3)
12	s_4	s_3	s_1	s_2	(s_1, s_2)	(h_3, h_2)
13	s_4	s_3	s_2	$\emptyset$	$--$	$--$

Table 13: Theorem 4.5 / Manipulation by c_2, 13 matchings (a blocking coalition when possible)

	Hospitals				Blocking coalitions?	
no.	h_1	h_2	h_3	h_4	Students	Hospitals
1	$\emptyset$	s_2	s_4	s_3	(s_3, s_4)	(h_4, h_1)
2	s_3	s_2	s_1	$\emptyset$	(s_3, s_4)	(h_4, h_1)
3	s_3	s_2	s_4	$\emptyset$	(s_3, s_4)	(h_4, h_1)
4	s_4	s_3	s_1	$\emptyset$	(s_1, s_2)	(h_3, h_2)
5	$\emptyset$	s_3	s_4	$\emptyset$	(s_1, s_2)	(u, h_2)
6	s_4	s_2	s_1	s_3	$--$	$--$

Table 14: Theorem 4.5 / Manipulation by c_1, 6 matchings (a blocking coalition when possible)

	Hospitals				Blocking coalitions?	
no.	h_1	h_2	h_3	h_4	Students	Hospitals
1	$\emptyset$	s_1	s_4	s_3	(s_1, s_2)	(h_2, h_3)
2	$\emptyset$	s_2	s_4	s_3	(s_1, s_2)	(u, h_3)
3	$\emptyset$	s_3	s_4	s_2	(s_1, s_2)	(u, h_3)
4	s_3	s_1	s_2	$\emptyset$	(s_3, s_4)	(h_2, h_1)
5	s_3	s_1	s_4	s_2	(s_1, s_2)	(h_2, h_3)
6	s_3	s_2	s_4	$\emptyset$	(s_1, s_2)	(u, h_3)
7	s_4	s_1	$\emptyset$	s_2	(s_1, s_2)	(h_2, h_3)
8	s_4	s_1	s_2	$\emptyset$	(s_3, s_4)	(h_2, h_1)
9	s_4	s_2	$\emptyset$	s_3	(s_1, s_2)	(u, h_3)
10	s_4	s_3	$\emptyset$	s_2	(s_1, s_2)	(u, h_3)
11	s_4	s_3	s_2	$\emptyset$	$--$	$--$

Table 15: Comment after Theorem 4.5 / Manipulation by h_3, 11 matchings (a blocking coalition when possible)

References

[1] A. Abdulkadiroğlu, College admissions with affirmative action, Department of Economics, Columbia University, Working Paper, May 2003.

[2] B. Aldershof and O.M. Carducci, Stable matchings with couples, *Discrete Applied Mathematics* **68** (1996), 203-207.

[3] A. Alkan and D. Gale, Stable schedule matching under revealed preferences, *J. Econ. Theory* **112** (2003), 289-306.

[4] D. Cantala, Matching markets: the particular case of couples, Centro de Estudios Económicos, El Colegio de México, Working Paper, April 2004.

[5] L.E. Dubins and D.A. Freedman, Machiavelli and the Gale-Shapley algorithm, *American Mathematical Monthly* **88** (1981), 485-494.

[6] D. Gale and L.S. Shapley, College admissions and the stability of marriage, *American Mathematical Monthly* **69** (1962), 9-15.

[7] J.W. Hatfield and P. Milgrom, Auctions, matching and the law of aggregate demand, Department of Economics, Stanford University, Working Paper 04-003, February 2004.

[8] A.S. Kelso and V.P. Crawford, Job matching, coalition formation, and gross substitutes, *Econometrica* **50** (1982), 1483-1504.

[9] B. Klaus, F. Klijn, and J. Massó, Some things couples always wanted to know about stable matchings (but were afraid to ask), Departament d'Economia i d'Història Econòmica, Universitat Autònoma de Barcelona, Working Paper, September 2003.

[10] F. Klijn and J. Massó, Weak stability and a bargaining set for the marriage model, *Games Econ. Behav.* **42** (2003), 91-100.

[11] D.E. Knuth, "Marriages Stables," Les Presses de l'Université, Montreal, 1976.

[12] R. Martínez, J. Massó, A. Neme, and J. Oviedo, Single agents and the set of many-to-one stable matchings, *J. Econ. Theory* **91** (2000), 91-105.

[13] D.G. McVitie and L.B. Wilson, Stable marriage assignments for unequal sets, *BIT* **10** (1970), 295-309.

[14] E. Ronn, NP-Complete stable matching problems, *Journal of Algorithms* **11** (1990), 285-304.

[15] A.E. Roth, The economics of matching: stability and incentives, *Mathematics of Operations Research* **7** (1982), 617-628.

[16] A.E. Roth, The evolution of the labor market for medical interns and residents: a case study in game theory, *J. Polit. Economy* **92** (1984), 991-1016.

[17] A.E. Roth, The college admissions problem is not equivalent to the marriage problem, *J. Econ. Theory* **36** (1985), 277-288.

[18] A.E. Roth, On the allocation of residents to rural hospitals: a general property of two-sided matching markets, *Econometrica* **54** (1986), 425-427.

[19] A.E. Roth, New physicians: a natural experiment in market organization, *Science* **250** (1990), 1524-1528.

[20] A.E. Roth, A natural experiment in the organization of entry-level labor markets: regional markets for new physicians and surgeons in the United Kingdom, *Amer. Econ. Rev.* **81** (1991), 415-440.

[21] A.E. Roth and M.A.O. Sotomayor, "Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis," Econometric Society Monograph Series, Cambridge University Press, New York, 1990.

[22] A.E. Roth and X. Xing, Jumping the gun: imperfections and institutions related to the timing of market transactions, *Amer. Econ. Rev.* **84** (1994), 992-1044.

[23] L. Zhou, A new bargaining set of an N-person game and endogenous coalition formation, *Games Econ. Behav.* **6** (1994), 512-526.

Some things couples always wanted to know about stable matchings (but were afraid to ask)*

Bettina Klaus[†] Flip Klijn[‡] Jordi Massó[§]

November 2004

Abstract: In this note we study the National Resident Matching Program (NRMP) algorithm in the US market for physicians. We report on two problems that concern the presence of couples, a feature explicitly incorporated in the new NRMP algorithm (*cf.* Roth and Peranson, 1999). First, we show that the new NRMP algorithm may not find an existing stable matching, even when couples' preferences are 'responsive,' *i.e.*, when Gale and Shapley's (1962) deferred acceptance algorithm (on which the old NRMP algorithm is based) is applicable. Second, we demonstrate that the new NRMP algorithm may also be manipulated by couples acting as singles.

Keywords: Matching; Stability; Couples
JEL classification: C78; D78; J41

1 Introduction

In many countries, the proportion of women attending college has steadily been increasing during the last decades. Thus, it is not surprising that the number of couples searching jointly for a job in the same labor market has been increasing as well. In this paper we deal with a specific US labor market that, because of its history and development, is a bench mark for several other matching markets in the US and Canada. Each year thousands of medical school graduates seek their first employment through a centralized matching process, the National Resident Matching Program (NRMP).[1] This clearinghouse was initiated in the 1950's in response to persistent failures to

*We thank Howard Petith, Alvin Roth, and William Thomson for helpful comments and suggestions. B. Klaus's and F. Klijn's research is supported by Ramón y Cajal contracts of the Spanish *Ministerio de Ciencia y Tecnología*. The main part of F. Klijn's work was supported by a Marie Curie Fellowship of the European Community programme "Improving Human Research Potential and the Socio-economic Knowledge Base" under contract number HPMF-CT-2001-01232, carried out at the Departament d'Economia i d'Història Econòmica, Universitat Autònoma de Barcelona. J. Massó's research is partially supported by Research Grant 2001SGR-00162 from the Department d'Universitats, Recerca i Societat de la Informació (Generalitat de Catalunya). The work of the authors is partially supported by Research Grant BEC2002-02130 from the Spanish *Ministerio de Ciencia y Tecnología* and by the Barcelona Economics Program of CREA.

[†]Departament d'Economia i d'Història Econòmica, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain; e-mail: `bettina.klaus@uab.es`

[‡]Corresponding author. Institut d'Anàlisi Econòmica (CSIC), Campus UAB, 08193 Bellaterra (Barcelona), Spain; Tel. (34) 93 580 6612; Fax. (34) 93 580 1452; e-mail: `fklijn@pareto.uab.es`

[§]CODE and Departament d'Economia i d'Història Econòmica, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain; e-mail: `jordi.masso@uab.es`

[1]See Roth (1984), Roth and Sotomayor (1990), and Roth and Peranson (1999).

organize the market in a timely and orderly way by decentralized means. Roth (1984) would later explain its success by showing that the clearinghouse in fact employed the hospital-optimal variant of the so-called deferred acceptance algorithm, developed by Gale and Shapley (1962) who at the time were not aware of the relation with the NRMP.

Around the mid 1970's voluntary and orderly participation in the NRMP started to drop. What happened then was that a growing number of couples in need of two positions in the same vicinity left the centralized market and started to negotiate directly with hospitals (see Checker, 1973). As a consequence, the labor market became, just as before the 1950's, prone to chaos and dissatisfaction on all sides. A hypothesis offered by Roth (1984) is that the chaotic conditions reflect the instability of the matching procedure. If couples and hospitals find it profitable to make their own arrangements outside of the matching program it must be that the matching procedure is unstable with respect to couples. This indeed turned out to be the case. In the mid 1990's a crisis of confidence[2] in the matching procedure on the applicants side finally induced the NRMP Board of Directors to design a new algorithm. Apart from recovering students' confidence by favoring their side, the algorithm was also meant to deal with couples in an appropriate manner.[3] The first match with the new algorithm was carried out in 1998. Roth and Peranson (1999) describe how the new algorithm was designed. Furthermore, using computational simulations and analyzing previous data, they show that the new algorithm is to be expected to perform well in practice. Roth (2002) gives a more recent review of the redesign of the NRMP algorithm in the context of analyzing the 'engineering aspects' of economic design. A nice overview of how the new algorithm was designed to address the problems that occur in the presence of couples is given in Roth (2002, Section 2.4.1).

In this note, we address two issues that are of importance for couples that participate in the new NRMP algorithm. The first issue concerns the stability of the matching obtained by the NRMP algorithm. A matching is 'stable' if (a) each couple and each hospital have acceptable matches, and (b) no couple and no pair of hospitals prefer a mutual match to the ones that have been assigned. Roth (1984) and Sotomayor (unpublished note) independently demonstrate the possibility of instability in the presence of couples. However, Klaus and Klijn (2004) show that for a natural preference domain for couples, namely the domain of 'responsive' preferences, there is always a stable matching that moreover can be found by applying Gale and Shapley's (1962) deferred acceptance algorithm. In Example 3.1 we construct a matching market in which couples have responsive preferences and where there is a unique stable matching, which hence can be easily obtained by Gale and Shapley's (1962) deferred acceptance algorithm. Surprisingly, we still find that the new NRMP algorithm may cycle, that is, not find the stable matching.

Next, couples may wonder if they should apply to the NRMP as a couple or as two seemingly independent applicants. With our second couples market (Example 3.2) we show that the new NRMP algorithm may indeed be prone to strategic manipulation by the members of a couple pretending to be singles, *i.e.*, a couple may be better off by applying as two separate applicants.

Both our couples markets are realistic in the sense that (a) couples' preferences can be easily explained in terms of individual preferences and distances between hospitals, and (b) hospitals have the same preferences over students (which can be justified by final grades or test scores).

[2]Many students believed that the matching was not conducted in their best interest and that possibilities for strategic manipulations existed; see Roth and Peranson (1999).

[3]In fact, three other issues that are not relevant in the present discussion were addressed as well. See Roth (2002, p. 1355) for details.

2 Matching with couples: the model

We describe a model with 4 hospitals and 2 pairs of students; $H = \{h_1, h_2, h_3, h_4\}$, $S = \{s_1, s_2, s_3, s_4\}$, and $C = \{c_1, c_2\} = \{(s_1, s_2), (s_3, s_4)\}$ are the sets of hospitals, students, and couples, respectively. Each hospital has exactly one position to be filled. Our definitions and results can easily be adapted to more general situations that include other couples as well as single agents and hospitals with multiple positions.

Hospitals' preferences: Each hospital $h \in H$ has a strict, transitive, and complete preference relation $\succeq_h$ over the set of students S and the prospect of having its position unfilled, denoted by $\emptyset$; for instance, $P(h) = s_4, s_2, \emptyset, s_1, s_3$ indicates that hospital h prefers student s_4 to s_2, and considers students s_1 and s_3 to be unacceptable. Let $P^H = \{P(h)\}_{h\in H}$.

Students' preferences: Similarly, each student $s \in S$ has an individual strict, transitive, and complete preference relation $\succeq_s$ over the set of hospitals and the prospect of being unemployed, denoted by u. We assume that these individual preferences are the preferences a student has if he is single. For instance, $P(s) = h_1, h_2, h_3, h_4, u$ indicates that student s prefers h_i to h_{i+1} for $i = 1, 2, 3$ and prefers being employed to being unemployed. Let $P^S = \{P(s)\}_{s\in S}$.

Couples' preferences: Each couple $c \in C$ has a strict, transitive, and complete preference relation $\succeq_c$ over all possible combination of ordered pairs of (different) hospitals and the prospect of being unemployed. We denote a generic ordered pair of hospitals by (h_p, h_q), where h_p and h_q indicate either a hospital or being unemployed. For instance, $P(c) = (h_4, h_2), (h_3, h_4), (h_4, u)$, *etc.*, indicates that couple $c = (s_1, s_2)$ prefers s_1 and s_2 being matched to h_4 and h_2, respectively, to being matched to h_3 and h_4, respectively, and so on. Let $P^C = \{P(c)\}_{c\in C}$.

Singles and couples markets: Now, the standard one-to-one two-sided matching market with single students, or *singles market* for short, is denoted by (P^H, P^S). Since singles markets and some of the classical results for singles markets are well-known, for a detailed description we refer to Roth and Sotomayor (1990). We define a one-to-one matching market with couples, or a *couples market* for short, by (P^H, P^C).

Matchings: A *matching* μ for a couples market (P^H, P^C) is an assignment of students and hospitals such that each student is assigned to at most one hospital in H or to u (which can be assigned to multiple students), each hospital in H is assigned to at most one student or to $\emptyset$ (which can be assigned to multiple hospitals), and a student is assigned to a hospital if and only if the hospital is assigned to the student. A matching μ is completely described by the list $\mu(H) = \mu(h_1), \mu(h_2), \mu(h_3), \mu(h_4)$ of students in S or $\emptyset$ matched to hospitals h_1, h_2, h_3, h_4.

Stability: Finally, we define stability for couples markets (see Roth and Sotomayor, 1990). First, for a matching to be stable, it should always be better for students (one or both members in a couple) to accept the position(s) offered by the matching instead of voluntarily choosing unemployment and for hospitals it should always be better to accept the student assigned by the matching instead of leaving the position unfilled. A matching μ is *individually rational* if

(i1) for all $c = (s_k, s_l)$, $(\mu(s_k), \mu(s_l)) \succeq_c (\mu(s_k), u)$, $(\mu(s_k), \mu(s_l)) \succeq_c (u, \mu(s_l))$, and $(\mu(s_k), \mu(s_l)) \succeq_c (u, u)$;

(i2) for all $h \in H$, $\mu(h) \succeq_h \emptyset$.

Second, if one partner in a couple can improve the given matching for the couple by switching to another hospital such that this hospital is better off as well, then we would expect this mutually beneficial trade to be carried out, rendering the given matching unstable. A similar statement holds if both students in the couple can improve. For a given matching μ, $(c = (s_k, s_l), (h_p, h_q))$ is a *blocking coalition* if

(b1) $(h_p, h_q) \succ_c (\mu(s_k), \mu(s_l))$;

(b2) $[h_p \in H$ implies $s_k \succeq_{h_p} \mu(h_p)]$ and $[h_q \in H$ implies $s_l \succeq_{h_q} \mu(h_q)]$.

A matching is *stable* if it is individually rational and if there are no blocking coalitions.[4]

Roth (1984) and Sotomayor (unpublished note) demonstrate that stable matchings may not exist in the presence of couples. Klaus and Klijn (2004) prove existence for couples market where couples' preferences are 'responsive' by applying Gale and Shapley's (1962) deferred acceptance algorithm to the 'associated individual preferences.' A couple's preferences are responsive if the unilateral improvement of one partner's job is considered beneficial for the couple as well. For instance, responsive preferences may reflect situations where couples search for jobs in the same metropolitan area (if one partner switches to a job he/she prefers and the couple can still live together, then the couple is better off).[5]

Responsive preferences: Couple $c = (s_k, s_l)$ has *responsive* preferences if there exist *associated individual preferences* $\succeq_{s_k}$ and $\succeq_{s_l}$ such that for all $h_p, h_q, h_r \in H \cup \{u\}$, $[h_p \succ_{s_k} h_r$ implies $(h_p, h_q) \succ_c (h_r, h_q)]$ and $[h_p \succ_{s_l} h_r$ implies $(h_q, h_p) \succ_c (h_q, h_r)]$. If these preferences $\succeq_{s_k}$ and $\succeq_{s_l}$ exist, then they are unique.

3 Couples and the new NRMP algorithm

The formal and complete description of the new algorithm used by the NRMP is outside the scope of this note. Essentially, the new algorithm adapts (from the original one-to-one model without couples) Roth and Vande Vate's (1990) dynamic process to find stable matchings. Flowchart 1 in the Appendix describes concisely the parts of the new Applicant Proposing Couples Algorithm (APCA) used by the NRMP that are relevant for applying the algorithm to Examples 3.1 and 3.2 below. (The complete flowchart describing the APCA algorithm that we used to derive Flowchart 1 was kindly made available to us by Alvin Roth.[6])

First, we demonstrate that even for responsive preferences the algorithm might cycle without selecting a stable matching. This example is particularly interesting because the unique stable matching could be easily found using the deferred acceptance algorithm.

Example 3.1 ***The New NRMP Algorithm may cycle for Responsive Preferences***
Consider the couples market (P^H, P^C) given in Table 1. Students' individual preferences P^S equal $P(s_1) = P(s_2) = h_1, h_2, h_3, h_4, u$, $P(s_3) = h_2, h_1, h_3, h_4, u$, and $P(s_4) = h_3, h_4, h_2, h_1, u$. Note that the couples' preferences are responsive: any unilateral improvement of one partner's job is considered beneficial for the couple as well. Moreover, hospitals have identical preferences over students, which can be easily justified if hospitals rank students according to final grades or other test scores. The unique stable matching for (P^H, P^C) is $\mu(H) = s_2, s_3, s_1, s_4$ (Appendix, Table 2), which, because of responsiveness, is the outcome of the deferred acceptance algorithm.

[4] In order to keep notation as simple as possible, we allow some redundancy in the definition of stability with respect to (i1) and (b1).

[5] For a discussion of results related to responsiveness in couples markets we refer to Klaus and Klijn (2004).

[6] We contacted the NRMP and asked them to 'run' the algorithm with our examples below. Since the NRMP is involved in an antitrust lawsuit they were advised by their legal counsel not to accommodate our request.

Next, we apply the Applicant Proposing Couples Algorithm to this couples market. Suppose that couple (s_3, s_4) is at the top of the stack (a symmetric process occurs if instead couple (s_1, s_2) is at the top of the stack). The algorithm starts with the empty matching $\mu^0(H) = \emptyset, \emptyset, \emptyset, \emptyset$ and cycles over the unstable matchings $\mu^I(H) = \emptyset, s_3, s_4, \emptyset$, $\mu^{II}(H) = s_1, s_2, \emptyset, \emptyset$, $\mu^{III}(H) = s_3, \emptyset, s_4, \emptyset$, $\mu^{IV}(H) = s_2, s_1, \emptyset, \emptyset$, and finally back to $\mu^I(H)$. $\diamond$

P^H				P^C	
h_1	h_2	h_3	h_4	s_1s_2	s_3s_4
s_2	s_2	s_2	s_2	h_1h_2	h_2h_3
s_3	s_3	s_3	s_3	h_1h_3	h_2h_4
s_1	s_1	s_1	s_1	h_1h_4	h_2h_1
s_4	s_4	s_4	s_4	h_2h_1	h_1h_3
$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$	h_2h_3	h_1h_4
				h_2h_4	h_1h_2
				h_3h_1	h_3h_4
				h_3h_2	h_3h_2
				h_3h_4	h_3h_1
				h_4h_1	h_4h_3
				h_4h_2	h_4h_2
				h_4h_3	h_4h_1
				$\vdots$	$\vdots$

Table 1: Responsive preferences for which the new NRMP algorithm cycles

Finally, we illustrate the possibility that, if the APCA is used, a couple may obtain a better pair of positions by registering as single students rather than as a couple.

Example 3.2 ***Pretending to be Single may be Beneficial***
Consider the couples market (P^H, P^C) where hospitals h_1 and h_2 are located in one city and hospitals h_3 and h_4 in some other, distant city. Students and couples have the same preferences over hospitals: $P(s) = h_1, h_2, h_3, h_4, u$ for each $s \in S$ and $P(s_1, s_2) = P(s_3, s_4) = (h_1, h_2), (h_2, h_1), (h_3, h_4), (h_4, h_3), \ldots$ (the tail can be anything). In other words, a couple would look for two positions in another market before accepting two positions located in different cities. The hospitals' preferences over students are $P(h) = s_1, s_3, s_4, s_2, \emptyset$ for every $h \in H$.

Assume first that the four students register as couples, and couple (s_1, s_2) is at the top of the stack. Then, the APCA yields the matching $\tilde{\mu}(H) = s_1, s_2, s_3, s_4$.

However, if s_3 and s_4 register as single students and, as a consequence, the order in the stack changes to $s_3, s_4, (s_1, s_2)$, then the algorithm yields the matching $\hat{\mu}(H) = s_3, s_4, s_1, s_2$. At this matching couple (s_3, s_4) is strictly better off than at matching $\tilde{\mu}$.[7] $\diamond$

[7] If instead the order in the stack changes to $s_4, s_3, (s_1, s_2)$, then the algorithm produces the matching $\bar{\mu}(H) = s_4, s_3, s_1, s_2$, in which couple (s_3, s_4) is also strictly better off than at matching $\tilde{\mu}$.

References

Checker, A. (1973) "The National Intern and Resident Matching Program, 1966-1972," *Journal of Medical Education*, **48**, 107-109.

Gale, D. and Shapley, L.S. (1962) "College Admissions and the Stability of Marriage," *American Mathematical Monthly*, **69**, 9-15.

Klaus, B. and Klijn, F. (2004) "Stable Matchings and Preferences of Couples," forthcoming in *Journal of Economic Theory*.

Roth, A.E. (1984) "The Evolution of the Labor Market for Medical Interns and Residents: a Case Study in Game Theory," *Journal of Political Economy*, **92**, 991-1016.

Roth, A.E. (2002) "The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics," *Econometrica*, **70**, 1341-1378.

Roth, A.E. and Peranson, E. (1999) "The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design," *American Economic Review*, **89**, 748-780.

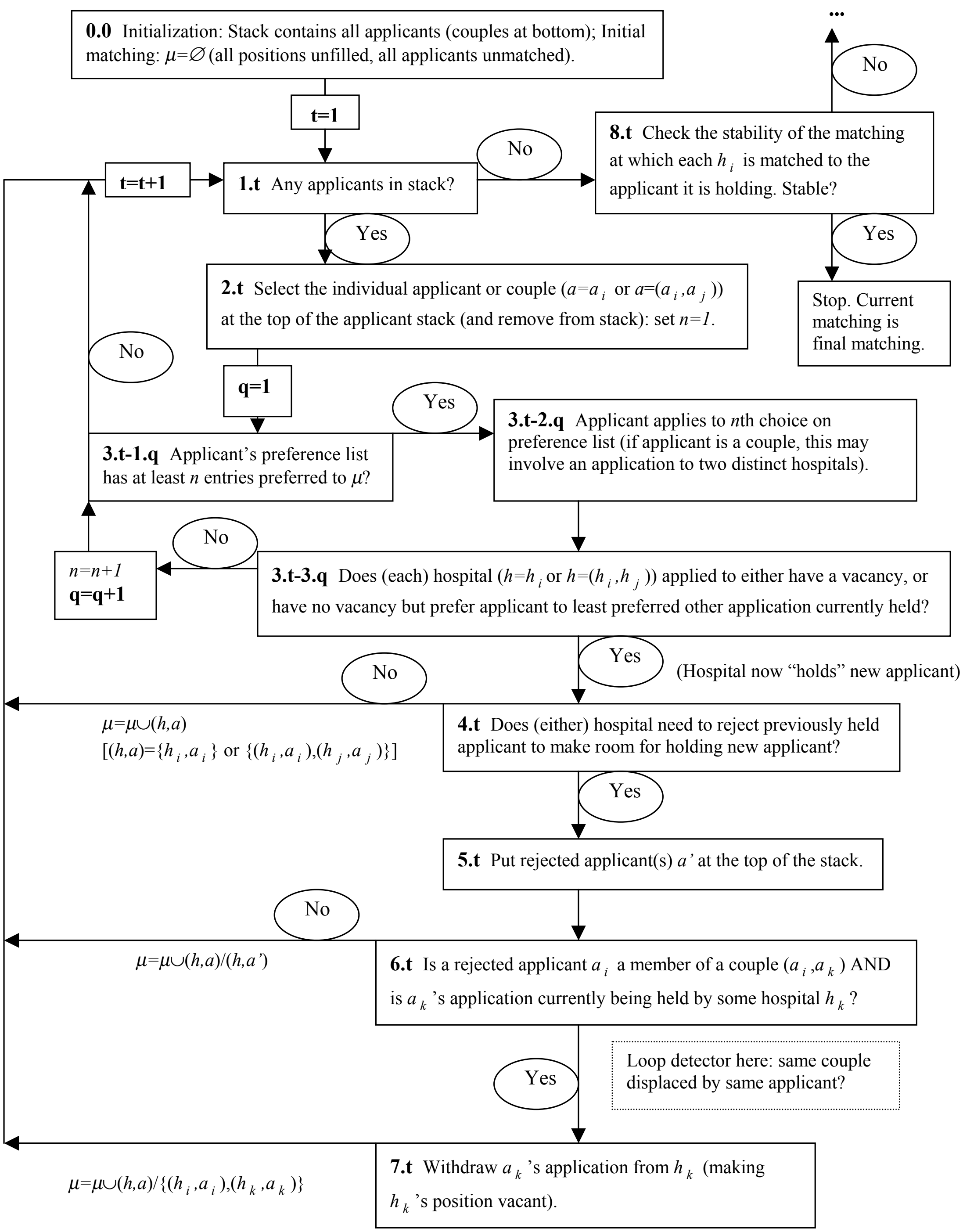
Roth, A.E. and Sotomayor, M.A.O. (1990) Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis. Econometric Society Monograph Series. New York: Cambridge University Press.

Roth, A.E. and Vande Vate, J.H. (1990) "Random Paths to Stability in Two-Sided Matching," *Econometrica*, **58**, 1475-1480.

4 Appendix

no.	Hospitals				Blocking coalitions?	
	h_1	h_2	h_3	h_4	Students	Hospitals
1	s_1	s_2	s_3	s_4	(s_3, s_4)	(h_1, h_4)
2	s_1	s_2	s_4	s_3	(s_3, s_4)	(h_1, h_3)
3	s_1	s_3	s_2	s_4	(s_1, s_2)	(h_1, h_2)
4	s_1	s_3	s_4	s_2	(s_1, s_2)	(h_1, h_2)
5	s_1	s_4	s_2	s_3	(s_1, s_2)	(h_1, h_2)
6	s_1	s_4	s_3	s_2	(s_1, s_2)	(h_1, h_2)
7	s_2	s_1	s_3	s_4	(s_3, s_4)	(h_2, h_4)
8	s_2	s_1	s_4	s_3	(s_3, s_4)	(h_2, h_3)
9	s_2	s_3	s_1	s_4	$--$	$--$
10	s_2	s_3	s_4	s_1	(s_1, s_2)	(h_3, h_1)
11	s_2	s_4	s_1	s_3	(s_1, s_2)	(h_2, h_1)
12	s_2	s_4	s_3	s_1	(s_1, s_2)	(h_2, h_1)
13	s_3	s_1	s_2	s_4	(s_3, s_4)	(h_2, h_4)
14	s_3	s_1	s_4	s_2	(s_1, s_2)	(h_2, h_1)
15	s_3	s_2	s_1	s_4	(s_1, s_2)	(h_3, h_1)
16	s_3	s_2	s_4	s_1	(s_1, s_2)	(h_3, h_1)
17	s_3	s_4	s_1	s_2	(s_1, s_2)	(h_2, h_1)
18	s_3	s_4	s_2	s_1	(s_1, s_2)	(h_2, h_1)
19	s_4	s_1	s_2	s_3	(s_1, s_2)	(h_1, h_2)
20	s_4	s_1	s_3	s_2	(s_1, s_2)	(h_1, h_2)
21	s_4	s_2	s_1	s_3	(s_1, s_2)	(h_1, h_2)
22	s_4	s_2	s_3	s_1	(s_1, s_2)	(h_1, h_2)
23	s_4	s_3	s_1	s_2	(s_1, s_2)	(h_1, h_2)
24	s_4	s_3	s_2	s_1	(s_1, s_2)	(h_1, h_2)

Table 2: Example 3.1, all individually rational matchings (with blocking coalition when possible)



Flowchart 1: The analyzed part of the Applicant Proposing Couples Algorithm (APCA)

The APCA applied to Example 3.1:

0.0 $\mu^0(H) = \emptyset, \emptyset, \emptyset, \emptyset$.

1.1 Yes.
2.1 (s_3, s_4) is selected and set $n = 1$.
3.1-1.1 Yes, (s_3, s_4) has more than one entry preferred to $\mu^0(s_3, s_4) = (u, u)$.
3.1-2.1 (s_3, s_4) applies to (h_2, h_3).
3.1-3.1 Yes, h_2 "holds" s_3 and h_3 "holds" s_4.
4.1 No, no rejection is needed; $\mu^I(H) = \emptyset, s_3, s_4, \emptyset$.

1.2 Yes.
2.2 (s_1, s_2) is selected and set $n = 1$.
3.2-1.1 Yes, (s_1, s_2) has more than one entry preferred to $\mu^I(s_1, s_2) = (u, u)$.
3.2-2.1 (s_1, s_2) applies to (h_1, h_2).
3.2-3.1 Yes, h_1 "holds" s_1 and h_2 "holds" s_2.
4.2 Yes, h_2 rejects s_3; $\mu^{II}(H) = s_1, s_2, \emptyset, \emptyset$.

5.2 (s_3, s_4) is at the top of the stack.
6.2 s_3 is rejected and s_4 is currently being held by h_3.
7.2 s_4 is withdrawn from h_3; $\mu^{II}(H) = s_1, s_2, \emptyset, \emptyset$.

1.3 Yes.
2.3 (s_3, s_4) is selected and set $n = 1$.
3.3-1.1 Yes, (s_3, s_4) has more than one entry preferred to $\mu^{II}(s_3, s_4) = (u, u)$.
3.3-2.1 (s_3, s_4) applies to (h_2, h_3).
3.3-3.1 No, h_2 prefers s_2 to s_3. Set $n = 2$.
3.3-1.2 Yes, (s_3, s_4) has more than two entries preferred to $\mu^{II}(s_3, s_4) = (u, u)$.
3.3-2.2 (s_3, s_4) applies to (h_2, h_4).
3.3-3.2 No, h_2 prefers s_2 to s_3. Set $n = 3$.
3.3-1.3 Yes, (s_3, s_4) has more than three entries preferred to $\mu^{II}(s_3, s_4) = (u, u)$.
3.3-2.3 (s_3, s_4) applies to (h_2, h_1).
3.3-3.3 No, h_2 prefers s_2 to s_3 and h_1 prefers s_1 to s_4. Set $n = 4$.
3.3-1.4 Yes, (s_3, s_4) has more than four entries preferred to $\mu^{II}(s_3, s_4) = (u, u)$.
3.3-2.4 (s_3, s_4) applies to (h_1, h_3).
3.3-3.4 Yes, h_3 has a vacancy and h_1 prefers s_3 to s_1.
4.3 Yes, h_1 rejects s_1; $\mu^{4.3}(H) = s_3, s_2, s_4, \emptyset$.
5.3 (s_1, s_2) is at the top of the stack.
6.3 s_1 is rejected and s_2 is currently being held by h_2.
7.3 s_2 is withdrawn from h_2; $\mu^{III}(H) = s_3, \emptyset, s_4, \emptyset$.

1.4 Yes.
2.4 (s_1, s_2) is selected and set $n = 1$.
3.4-1-1 Yes, (s_1, s_2) has more than one entry preferred to $\mu^{III}(s_1, s_2) = (u, u)$.
3.4-2.1 (s_1, s_2) applies to (h_1, h_2).
3.4-3.1 No, h_1 prefers s_3 to s_1. Set $n = 2$.
3.4-1.2 Yes, (s_1, s_2) has more than two entries preferred to $\mu^{III}(s_1, s_2) = (u, u)$.
3.4-2.2 (s_1, s_2) applies to (h_1, h_3).
3.4-3.2 No, h_1 prefers s_3 to s_1. Set $n = 3$.
3.4-1.3 Yes, (s_1, s_2) has more than three entries preferred to $\mu^{III}(s_1, s_2) = (u, u)$.
3.4-2.3 (s_1, s_2) applies to (h_1, h_4).
3.4-3.3 No, h_1 prefers s_3 to s_1. Set $n = 4$.

3.4-1.4 Yes, (s_1, s_2) has more than four entries preferred to $\mu^{III}(s_1, s_2) = (u, u)$.
3.4-2.4 (s_1, s_2) applies to (h_2, h_1).
3.4-3.4 Yes, h_2 has a vacancy and h_1 prefers s_2 to s_3.
4.4 Yes, h_1 rejects s_3; $\mu^{4.4}(H) = s_2, s_1, s_4, \emptyset$.
5.4 (s_3, s_4) is at the top of the stack.
6.4 s_3 is rejected and s_4 is currently being held by h_3.
7.4 s_4 is withdrawn from h_3; $\mu^{IV}(H) = s_2, s_1, \emptyset, \emptyset$.

1.5 Yes.
2.5 (s_3, s_4) is selected and set $n = 1$.
3.5-1.1 Yes, (s_3, s_4) has more than one entry preferred to $\mu^{IV}(s_3, s_4) = (u, u)$.
3.5-2.1 (s_3, s_4) applies to (h_2, h_3).
3.5-3.1 Yes, h_3 has a vacancy and h_2 prefers s_3 to s_1.
4.5 Yes, h_2 rejects s_1; $\mu^{4.5}(H) = s_2, s_3, s_4, \emptyset$.
5.5 (s_1, s_2) is at the top of the stack.
6.5 s_1 is rejected and s_2 is currently being held by h_1.
7.5 s_1 is withdrawn from h_2; $\mu^{V}(H) = \mu^{I}(H) = \emptyset, s_3, s_4, \emptyset$.

The APCA cycles and stops at one of the unstable matchings $\mu^I, \mu^{II}, \mu^{III}, \mu^{IV}$, or μ^V.

The APCA applied to Example 3.2 with $(s_1, s_2), (s_3, s_4)$ in the stack:

0.0 $\mu^0(H) = \emptyset, \emptyset, \emptyset, \emptyset$.

1.1 Yes.
2.1 (s_1, s_2) is selected and set $n = 1$.
3.1-1.1 Yes, (s_1, s_2) has more than one entry preferred to $\mu^0(s_1, s_2) = (u, u)$.
3.1-2.1 (s_1, s_2) applies to (h_1, h_2).
3.1-3.1 Yes, h_1 "holds" s_1 and h_2 "holds" s_2.
4.1 No, no rejection is needed; $\mu^I(H) = s_1, s_2, \emptyset, \emptyset$.

1.2 Yes.
2.2 (s_3, s_4) is selected and set $n = 1$.
3.2-1.1 Yes, (s_3, s_4) has more than one entry preferred to $\mu^I(s_3, s_4) = (u, u)$.
3.2-2.1 (s_3, s_4) applies to (h_1, h_2).
3.2-3.1 No, h_1 prefers s_1 to s_3. Set $n = 2$.
3.2-1.2 Yes, (s_3, s_4) has more than two entries preferred to $\mu^I(s_3, s_4) = (u, u)$.
3.2-2.2 (s_3, s_4) applies to (h_2, h_1).
3.2-3.2 No, h_1 prefers s_1 to s_4. Set $n = 3$.
3.2-1.3 Yes, (s_3, s_4) has more than three entries preferred to $\mu^I(s_3, s_4) = (u, u)$.
3.2-2.3 (s_3, s_4) applies to (h_3, h_4).
3.2-3.3 Yes, h_3 "holds" s_3 and h_4 "holds" s_4.
4.2 No, no rejection is needed; $\mu^{II}(H) = s_1, s_2, s_3, s_4$.

1.3 No, the stack is empty.
8.3 Yes, $\tilde{\mu}(H) \equiv \mu^{II}(H) = s_1, s_2, s_3, s_4$ is stable and it is the outcome of the algorithm.

The APCA applied to Example 3.2 with $s_3, s_4, (s_1, s_2)$ in the stack:

0.0 $\mu^0(H) = \emptyset, \emptyset, \emptyset, \emptyset$.

1.1 Yes.
2.1 s_3 is selected and set $n = 1$.
3.1-1.1 Yes, s_3 has more than one entry preferred to $\mu^0(s_3) = u$.
3.1-2.1 s_3 applies to h_1.
3.1-3.1 Yes, h_1 "holds" s_3.
4.1 No, no rejection is needed; $\mu^I(H) = s_3, \emptyset, \emptyset, \emptyset$.

1.2 Yes.
2.2 s_4 is selected and set $n = 1$.
3.2-1.1 Yes, s_4 has more than one entry preferred to $\mu^I(s_4) = u$.
3.2-2.1 s_4 applies to h_1.
3.2-3.1 No, h_1 prefers s_3 to s_4. Set $n = 2$.
3.2-1.2 Yes, s_4 has more than two entries preferred to $\mu^I(s_4) = u$.
3.2-2.2 s_4 applies to h_2.
3.2-3.2 Yes, h_2 "holds" s_4.
4.2 No, no rejection is needed; $\mu^{II}(H) = s_3, s_4, \emptyset, \emptyset$.

1.3 Yes.
2.3 (s_1, s_2) is selected and set $n = 1$.
3.3-1.1 Yes, (s_1, s_2) has more than one entry preferred to $\mu^{II}(s_1, s_2) = (u, u)$.
3.3-2.1 (s_1, s_2) applies to (h_1, h_2).
3.3-3.1 No, h_2 prefers s_4 to s_2. Set $n = 2$.
3.3-1.2 Yes, (s_1, s_2) has more than two entries preferred to $\mu^{II}(s_1, s_2) = (u, u)$.
3.3-2.2 (s_1, s_2) applies to (h_2, h_1).
3.3-3.2 No, h_1 prefers s_3 to s_2. Set $n = 3$.
3.3-1.3 Yes, (s_1, s_2) has more than three entries preferred to $\mu^{II}(s_1, s_2) = (u, u)$.
3.3-2.3 (s_1, s_2) applies to (h_3, h_4).
3.3-3.3 Yes, h_3 "holds" s_1 and h_4 "holds" s_2.
4.3 No, no rejection is needed; $\mu^{III}(H) = s_3, s_4, s_1, s_2$.
1.4 No, the stack is empty.
8.4 Yes, $\hat{\mu}(H) \equiv \mu^{III}(H) = s_3, s_4, s_1, s_2$ is stable and it is the outcome of the algorithm.