Stability and Nash implementation in matching markets with couples

Claus-Jochen Haake · Bettina Klaus

Abstract   We consider two-sided matching markets with couples. First, we extend a result by Klaus and Klijn (J Econ Theory 21: 75–106, 2005, Theorem 3.3) and show that for any weakly responsive couples market, there always exists a "double stable" matching, i.e., a matching that is stable for the couples market and for any associated singles market. Second, we show that for weakly responsive couples markets, the associated stable correspondence is (Maskin) monotonic and Nash implementable. In contrast, the correspondence that assigns all double stable matchings is neither monotonic nor Nash implementable.

Keywords   Matching with couples · (Maskin) Monotonicity · Nash implementation · Stability · Weakly responsive preferences

JEL Classification   C62 · C78 · D78 · J41

1 Introduction

We consider two-sided matching markets consisting of medical students (graduates, workers) on one side and of residencies (jobs, firms) on the other side. In the medical market as well as in many other labor markets, the number of couples with the

C.-J. Haake
Institute of Mathematical Economics, Bielefeld University, P.O. Box 100131, 33501 Bielefeld, Germany
e-mail: CHaake@wiwi.uni-bielefeld.de

B. Klaus (✉)
Harvard Business School, Baker Library, Bloomberg Center 437, Soldier Field, Boston, MA 02163, USA
e-mail: bklaus@hbs.edu

 Springer

same professional interests has been growing. Therefore, we focus on labor markets in which both members of couples seek positions. Examples of such labor markets with couples are medical markets where each year many medical school graduates seek their first employment as residents or interns. In the US, the National Resident Matching Program (NRMP) administers the matching of medical graduates with residencies and internships (see Roth 1984a, Roth and Sotomayor 1990). One of the reasons for centralizing this market, and later for reorganizing it (Roth and Peranson 1999) was that market outcomes did not seem to be "stable" as indicated by unraveling (pre NRMP) or a significant reduction of voluntary participation in the NRMP (pre reorganized NRMP). In particular, couples searching for residencies together were dropping out of the NRMP before it was reorganized—in its current version couples can file joint NRMP applications. For simplicity, we continue to refer to one side of the market as students and to the other side of the market as hospitals, even though we do not exclusively restrict our attention to centralized markets such as the NRMP.

Loosely speaking, an outcome or matching is stable if there are no students/ couples and hospitals that are not matched with each other, but in fact would prefer to be. Thus, it is easy to see that whenever contracts are not enforceable, stability is a minimal requirement for the persistence of an outcome for any market (centralized or decentralized). Gale and Shapley (1962) provided an algorithm, the famous deferred acceptance (DA) algorithm, which always finds a stable matching in markets with only single students. Thus, as long as we restrict attention to singles markets, in these markets—centralized or not—stability is possible. Unfortunately, once couples emerge on one side of the market, stability is in danger: stable matchings may not exist (Roth 1984a) and it may be very difficult (in fact NP-hard) to decide if stable matchings exist for a given couples market (Ronn 1990).

It is well known that for matching markets with sufficient substitutability, instabilities can be ruled out: 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-many schedule matching see Alkan and Gale (2003), and for matching with contracts see Roth (1984b) and Hatfield and Milgrom (2005). For couples markets, a similar result holds: Klaus and Klijn (2005) (Klaus et al. 2008) show that if couples have "(weakly) responsive" preferences (i.e., couples' preferences show some degree of substitutability), then stable matchings always exist. In other words, if couples' preferences are weakly responsive, then the stable correspondence that assigns to each couples market its set of stable matchings is well defined.

To achieve stable matchings in a couples market, information on preferences is necessary. In general, a centralized clearinghouse would not have this information and so a solution may be vulnerable to strategic manipulation of preferences. An appealing property for solutions which requires that no agent can ever benefit from misrepresenting his/her preferences is strategy-proofness, i.e., truth-telling is a weakly dominant strategy for all agents. Unfortunately, there exists no stable (single-valued) solution for which stating the true preferences is a dominant strategy for every couple (see Klaus and Klijn 2005, Theorem 4.5). A weaker requirement than obtaining stable matchings through "dominant truth telling" equilibria in the direct revelation game is to obtain them through Nash equilibria in an associated game (see

Jackson 2001, Maskin and Sjöström 2002, for comprehensive surveys on implementation theory). Then, to *Nash implement* the stable correspondence means that the set of matchings induced by all Nash equilibria of the underlying game coincides with the set of stable matchings.

Maskin (1977, 1999) provided a key condition for Nash implementation: *(Maskin) monotonicity* is a necessary (but not sufficient) condition for Nash implementability. To find an equivalent condition for Nash implementability, one can either strengthen the monotonicity property or pair it with additional requirements. Both tasks have been successfully accomplished. Yamato (1992) shows that the *essential monotonicity* property is equivalent to Nash implementability, and Moore and Repullo (1990) package Maskin's monotonicity and two further (weak) conditions into one property, called *condition* μ.

Several positive Nash implementation results for matching markets have been derived in the past decade. Kara and Sönmez (1996, 1997) show that the stable correspondence in marriage and college admission markets is Nash implementable. Sönmez (1996) obtains a corresponding Nash implementability result for the so-called generalized matching markets: a class of one-sided matching markets that include marriage and roommate markets (Gale and Shapley 1962), as well as housing markets (Shapley and Scarf 1974). Another generalization of the classical marriage and college admission markets are the matching with contracts markets of Roth (1984b) and Hatfield and Milgrom (2005)—in difference to previous models a match between two agents also can specify a variety of contract terms. In a recent article, Haake and Klaus (2008) demonstrate that the previously mentioned Nash implementability of the stable correspondence also extends to matching markets with contracts.

In this article, we focus on stability and Nash implementability in matching markets with couples. Interestingly, even if it is possible to associate a matching market without couples (an associated singles market) to any given couples market in a reasonable way (e.g., if couples have (weakly) responsive preferences), we cannot make use of the implementation results mentioned above (see discussion in Sect. 4). Hence, as already observed by Klaus and Klijn (2005), many results for matching markets without couples can be extended to matching markets with couples, but they are not simple extensions and require different proof techniques.

We introduce matching markets with couples, stability, and weakly responsive preferences in Sect. 2. We extend a result by Klaus and Klijn (2005, Theorem 3.3) and show that for any weakly responsive couples market, there always exists a "double stable" matching, i.e., a matching that is stable for the couples market and for any associated singles market (Proposition 1). Section 3 is devoted to monotonicity and Nash implementability. First, we show that the stable correspondence is monotonic whenever it is well defined (Theorem 2). Our main result in Sect. 3 is the Nash implementability of the stable correspondence for matching markets with couples (Theorem 3). To obtain this result, we verify Moore and Repullo's (1990) condition and thereby apply an alternative approach to Sönmez (1996) and Kara and Sönmez (1996, 1997) who rely on Yamato's (1992) characterization of Nash implementability. Finally, for couples markets with weakly responsive preferences, the correspondence of stable matchings is Nash implementable (Corollary 1), but the subcorrespondence of double stable matchings is not (Corollary 2). In the last section (Sect. 4), we discuss the difference and

difficulties that occur when trying to extend results from matching markets without couples to matching markets with couples.

2 Matching with couples, stability, and responsive preferences

For convenience and without loss of generality, we describe a couples market where the labor market modeled consists of a supply side of hospitals and a demand side of couples composed of medical students; $H = \{h_1, h_2, \ldots, h_m\}$ and $C = \{c_1, c_2, \ldots, c_n\} = \{(s_1, s_2), (s_3, s_4), \ldots, (s_{2n-1}, s_{2n})\}$ are the sets of hospitals and couples. We denote the set of students by $S = \{s_1, s_2, \ldots, s_{2n}\}$. Each hospital has exactly one position to be filled. Our results can easily be adapted to more general situations that could include mutual externalities between more than two individuals (for instance between members of a family or among friends), single students, or hospitals with multiple positions.[1] Next, we describe preferences of hospitals and couples.

Hospitals' preferences: Each hospital $h \in H$ has a strict, transitive, and complete preference relation R_h over the set of students S and the prospect of having its position unfilled, denoted by $\emptyset$. The asymmetric part of R_h is referred to as P_h. Given $s, s' \in S \cup \{\emptyset\}$, $s \, P_h \, s'$ means that hospital h strictly prefers student s to student s'; $s \, R_h \, s'$ implies $s \, P_h \, s'$ or $s = s'$. We denote the set of all possible preferences for hospital h by $\mathcal{R}^h$. Let $R^H = \{R_h\}_{h \in H} \in \prod_{h \in H} \mathcal{R}^h$.

Couples' preferences: Each couple $c = (s_k, s_l) \in C$ has a strict, transitive, and complete preference relation R_c over all possible combinations of ordered pairs of (different) hospitals and the prospect of being unemployed $\mathcal{H} := [(H \cup \{u\}) \times (H \cup \{u\})] \setminus \{(h, h) : h \in H\}$. 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. Again, P_c denotes the asymmetric part of R_c. Given $(h_p, h_q), (h'_p, h'_q) \in \mathcal{H}$, $(h_p, h_q) \, P_c \, (h'_p, h'_q)$ means that couple c strictly prefers s_k to be matched to h_p and s_l to be matched to h_q to s_k being matched to h'_p and s_l being matched to h'_q; $(h_p, h_q) \, R_c \, (h'_p, h'_q)$ implies $(h_p, h_q) \, P_c \, (h'_p, h'_q)$ or $(h_p, h_q) = (h'_p, h'_q)$. We denote the set of all possible preferences for couple c by $\mathcal{R}^c$. Let $R^C = \{R_c\}_{c \in C} \in \prod_{c \in C} \mathcal{R}^c$.

Since the set of agents (hospitals and couples) remains fixed throughout this study, a matching market with couples can completely be described by the agents' preferences.

Couples markets: A one-to-one matching market with couples, or a *couples market* for short, is denoted by a preference profile (R^H, R^C). We denote the *set of all matching markets* (i.e., the set of all preference profiles) by $\mathcal{R} \equiv \prod_{h \in H} \mathcal{R}^h \times \prod_{c \in C} \mathcal{R}^c$.

A *matching* μ for a couples market (R^H, R^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

---

[1] Our results would remain the same if the setting would be extended to triples or any other constellation of ordered tuples of students. However, the couples incidence is frequently observed in real-life matching markets, and by restricting the analysis to couples, we can keep notation, without loss of generality, simple. Furthermore, our model already can include single students if we model them as a couple where the student is married to a "fictitious partner" that does not want to work at all (see Klaus and Klijn 2007, Remark 3.6). Finally, in order to straightforwardly derive all the results for the case of hospitals with multiple positions, we would require that hospitals' preferences are "responsive over sets of students" (*cf.* Kara and Sönmez 1997).

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. Formally, a matching is a mapping $\mu : H \cup S \longrightarrow (S \cup \{\emptyset\}) \cup (H \cup \{u\})$ such that for all $h \in H$ and $s \in S$, $\mu(h) \in S \cup \{\emptyset\}$, $\mu(s) \in H \cup \{u\}$, and $[\mu(s) = h$ if and only if $\mu(h) = s]$.

A matching μ is completely described by the list $(\mu(s_1), \mu(s_2), \ldots, \mu(s_{2n}))$ of hospitals or u assigned to students $s_1, s_2, \ldots, s_{2n}$. Equivalently, a matching μ can be completely described by the list $(\mu(h_1), \mu(h_2), \ldots, \mu(h_m))$ of students or $\emptyset$ assigned to hospitals $h_1, h_2, \ldots, h_m$. For any couple $c = (s_k, s_l)$, we also use the notation $\mu(c) = (\mu(s_k), \mu(s_l))$.

We denote the set of matchings by $\mathcal{M}$. Clearly, all preference relations R_i ($i \in H \cup C$) induce *weak preferences over matchings* in a natural way. Using the same notation, for all agents $i \in H \cup C$ and matchings $\mu, \mu' \in \mathcal{M}$, we let $\mu \, R_i \, \mu'$ if and only if $\mu(i) \, R_i \, \mu'(i)$.

Let $\mathcal{R}_0 \subseteq \mathcal{R}$ be a class of couples markets. A *solution* φ (on $\mathcal{R}_0$) is a correspondence $\varphi : \mathcal{R}_0 \rightrightarrows \mathcal{M}$ that assigns to each couples market $(R^H, R^C) \in \mathcal{R}_0$ a non-empty set of matchings $\varphi(R^H, R^C)$. Next, we discuss two basic properties for solutions: Pareto efficiency and individual rationality.

Pareto efficiency: A matching $\mu \in \mathcal{M}$ is *Pareto efficient* for couples market $(R^H, R^C) \in \mathcal{R}$ if there is no other matching $\mu' \in \mathcal{M}$ such that for all $i \in H \cup C$, $\mu' \, R_i \, \mu$ and for some $j \in H \cup C$, $\mu' \, P_j \, \mu$. A solution φ is *Pareto efficient* if it only assigns sets of Pareto efficient matchings.

Individual rationality: Since matches are based on voluntary participation, 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* for couples market $(R^H, R^C) \in \mathcal{R}$ if

(i1) for all $c = (s_k, s_l)$,
$(\mu(s_k), \mu(s_l)) \, R_c \, (\mu(s_k), u)$, $(\mu(s_k), \mu(s_l)) \, R_c \, (u, \mu(s_l))$, and $(\mu(s_k), \mu(s_l)) \, R_c \, (u, u)$;
(i2) for all $h \in H$, $\mu(h) \, R_h \, \emptyset$.

As described in the Introduction (Sect. 1), an important criterion for a matching to be accepted as final outcome in a two-sided matching market (with couples) is stability. First, since the matching markets we consider here are based on voluntary participation of students (couples) and hospitals, a necessary condition for a matching μ to be stable is individual rationality. 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, $(c, (h_p, h_q)) = ((s_k, s_l), (h_p, h_q))$ is a *blocking coalition* if

(b1) $(h_p, h_q) \, P_c \, (\mu(s_k), \mu(s_l))$;
(b2) $[h_p \in H$ implies $s_k \, R_{h_p} \, \mu(h_p)]$ and $[h_q \in H$ implies $s_l \, R_{h_q} \, \mu(h_q)]$.

A matching is *stable* for couples market $(R^H, R^C) \in \mathcal{R}$ if it is individually rational and if there are no blocking coalitions. By $S(R^H, R^C)$ we denote the set of stable matchings for couples market (R^H, R^C). We denote the solution (on $\mathcal{R}_0$) that assigns the set of stable matchings to any couples market $(R^H, R^C) \in \mathcal{R}_0$ by S and refer to it as the *stable correspondence*.

Roth (1984a, Theorem 10) showed that stable matchings may not exist in the presence of couples. Even worse, Ronn (1990) showed that determining if a couples market has a stable matching is an NP complete problem.

Next, we discuss a preference domain for couples that guarantees the existence of stable matchings: the domain of weakly responsive preferences (Klaus and Klijn 2005, Klaus et al. 2008). This restriction of preferences 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, which in turn guarantees 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 distances are no longer part of couples' preferences and therefore the preferences are responsive. Klaus and Klijn (2005) (Klaus et al. 2008) demonstrated that for the existence of stable matchings, one can easily 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. Because the class of weakly responsive preferences ensures the presence of stable matchings, it is a natural starting point for the study of decentralized decision making in couples markets.

Couple $c = (s_k, s_l)$ has *weakly responsive* preferences if there exist strict, transitive, and complete preference relations R_{s_k} and R_{s_l} over $H \times \{u\}$, i.e., the set of hospitals and the prospect of being unemployed u such that

(a) for all $h \in H$,
$(u, h) \, P_c \, (u, u)$ if and only if $h \, P_{s_l} \, u$,
$(h, u) \, P_c \, (u, u)$ if and only if $h \, P_{s_k} \, u$;

(b) for all $h_p, h_q, h_r \in H \cup \{u\}$,
$[h_p \, R_{s_k} \, u, \, h_q \, R_{s_l} \, u$, and $h_p \, P_{s_k} \, h_r$ imply $(h_p, h_q) \, P_c \, (h_r, h_q)]$,
$[h_p \, R_{s_l} \, u, \, h_q \, R_{s_k} \, u$, and $h_p \, P_{s_l} \, h_r$ imply $(h_q, h_p) \, P_c \, (h_q, h_r)]$;

(c) for all $h', h'' \in H, \, h' \neq h''$, $u \, R_{s_k} \, h'$ and $u \, R_{s_l} \, h''$ imply $(u, u) \, P_c \, (h', h'')$.

We denote the set of all possible weakly responsive preferences for couple c by $\mathcal{RR}^c$. A couples market (R^H, R^C) is *weakly responsive* if all couples have weakly responsive preferences. We denote the *set of all weakly responsive couples markets* by $\mathcal{RR} \equiv \prod_{h \in H} \mathcal{R}^h \times \prod_{c \in C} \mathcal{RR}^c$.

Let $(R^H, R^C) \in \mathcal{RR}$. Then, for all $c = (s_k, s_l)$, the *associated individual preferences* R_{s_k} and R_{s_l} are only uniquely determined for acceptable positions. In other words, if both $[R_{s_k}$ and $R_{s_l}]$ and $[R'_{s_k}$ and $R'_{s_l}]$ satisfy the three weak responsiveness conditions, then for all $h_p, h_q \in H \cup \{u\}$, $h_p \, P_{s_k} \, h_q \, R_{s_k} \, u$ implies $h_p \, P'_{s_k} \, h_q \, R'_{s_k} \, u$, and $h_p \, P_{s_l} \, h_q \, R_{s_l} \, u$ implies $h_p \, P'_{s_l} \, h_q \, R'_{s_l} \, u$.

Given associated individual preferences for all students that are members of a couple, we obtain an *associated singles market* $(R^H, R^S(R^C))$ by replacing couples and

their preferences in (R^H, R^C) by individual students and their (possibly not uniquely determined) associated individual preferences $R^S(R^C)$. In what follows, we further explore the relation between stability in a couples market and stability in associated singles markets.

Remark 1 Associated singles markets and stability
Consider a weakly responsive couples market. Then, for each associate singles market, the set of acceptable hospitals is the same for each student. Moreover, students' preferences over acceptable hospitals are the same in each associated singles market. Hence, all associated singles markets have the same set of stable matchings.

With a slight abuse of notation and in view of Remark 1, we use $(R^H, R^S(R^C))$ to denote some singles market associated with the couples market (R^H, R^C), although preferences in $R^S(R^C)$ may differ for unacceptable hospitals. The following existence theorem is due to Klaus and Klijn (2005, Theorem 3.3) and Klaus et al. (2008).

Theorem 1 Stability for weakly responsive preferences
Let $(R^H, R^C) \in \mathcal{RR}$. Then, any matching that is stable for an associated singles market $(R^H, R^S(R^C))$ induces[2] a stable matching for the original couples market (R^H, R^C). In particular, there exists a stable matching for (R^H, R^C).

Apart from the stable correspondence S, we are interested in the following solutions on $\mathcal{RR}$. Let $\bar{S}$ be the correspondence that assigns to each couples market $(R^H, R^C) \in \mathcal{RR}$ the set of matchings that are stable for each associated singles market. By S^*, we denote the solution that assigns to each couples market $(R^H, R^C) \in \mathcal{RR}$ the *set of double stable matchings*, i.e., the set of matchings that are stable for couples market (R^H, R^C) and that are stable for all the associated single markets $(R^H, R^S(R^C))$—we shall prove in Proposition 1 that S^* is well defined. Hence, $S^* = S \cap \bar{S}$ and, in particular, $S^* \subseteq S$ and $S^* \subseteq \bar{S}$.[3]

Two examples demonstrate that the two correspondences S and $\bar{S}$ are logically unrelated, i.e., neither correspondence includes the other. Klaus et al. (2008, Example 1.2) show that not every matching that is stable for an associated singles market $(P^H, P^S(P^C))$ is also stable for the original couples market $(R^H, R^C) \in \mathcal{RR}$. Hence $\bar{S} \nsubseteq S$. Vice versa, Klaus and Klijn (2005, Example 3.4) show that not every matching that is stable for a couples market $(R^H, R^C) \in \mathcal{RR}$ is also stable for any of the associated singles market $(R^H, R^S(R^C))$. 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. Therefore, $S \nsubseteq \bar{S}$.

Proposition 1 "Double stability" for weakly responsive preferences
Let $(R^H, R^C) \in \mathcal{RR}$. Then, there always exists a matching that is stable for the couples market $(R^H, R^C) \in \mathcal{RR}$ and any associated singles market $(R^H, R^S(R^C))$, i.e., $S^*(R^H, R^C) \neq \emptyset$.

---

[2] That is, any matching μ that is stable for $(P^H, P^S(P^C))$ either is also stable for (P^H, P^C) or the only blocking coalitions are of the form $((s_k, s_l), (\mu(s_l), \mu(s_k)))$. However, then we can easily obtain a stable matching μ' from μ by satisfying all these blocking coalitions (see Klaus et al., 2008, for details).
[3] That is, for all $(R^H, R^C) \in \mathcal{RR}$, $S^*(R^H, R^C) = S(R^H, R^C) \cap \bar{S}(R^H, R^C)$, $S^*(R^H, R^C) \subseteq S(R^H, R^C)$, and $S^*(R^H, R^C) \subseteq \bar{S}(R^H, R^C)$.

Note that Proposition 1 implies Theorem 1 and is therefore a strengthening of Klaus and Klijn (2005) (Klaus et al. 2008).

Proof Let $(R^H, R^C) \in \mathcal{RR}$ and $\mu \in S(R^H, R^S(R^C))$ (note that $S(R^H, R^S(R^C)) \neq \emptyset$). If $\mu \in S(R^H, R^C)$, then $S^*(R^H, R^C) \neq \emptyset$. Thus, assume that $\mu \notin S(R^H, R^C)$.

First, we show that μ is individually rational for couples market (P^H, P^C). Let $c = (s_k, s_l) \in C$. Since $\mu \in S(R^H, R^S(R^C))$, $\mu(s_k) R_{s_k} u$ and $\mu(s_l) R_{s_l} u$. If $(\mu(s_k), \mu(s_l)) = (u, u)$, then individual rationality condition (i1) for (R^H, R^C) is trivially satisfied. If $\mu(s_k) P_{s_k} u$ and $\mu(s_l) = u$, then by weak responsiveness condition (a), $(\mu(s_k), u) P_c (u, u)$, which implies individual rationality condition (i1) for (R^H, R^C). Similarly, $\mu(s_k) = u$ and $\mu(s_l) P_{s_l} u$ implies individual rationality condition (i1) for (R^H, R^C). Finally, assume $\mu(s_k) P_{s_k} u$ and $\mu(s_l) P_{s_l} u$. Then, by weak responsiveness condition (b), $(\mu(s_k), \mu(s_l)) P_c (\mu(s_k), u) P_c (u, u)$. Similarly, $(\mu(s_k), \mu(s_l)) P_c (u, \mu(s_l)) P_c (u, u)$. Since $\mu \in S(R^H, R^S(R^C))$, for all $h \in H$, $\mu(h) R_h \emptyset$. Thus, individual rationality condition (i2) is also satisfied for (R^H, R^C).

Since μ is individually rational for couples market (P^H, P^C), $\mu \notin S(R^H, R^C)$ implies that there exists a blocking coalition $((s_k, s_l), (h_p, h_q))$ with (b1) $(h_p, h_q) P_c (\mu(s_k), \mu(s_l))$ and (b2) $[h_p \in H$ implies $s_k R_{h_p} \mu(h_p)]$ and $[h_q \in H$ implies $s_l R_{h_q} \mu(h_q)]$.

Assume $u P_{s_k} h_p$ and $u P_{s_l} h_q$. Then, by weak responsiveness condition (c), $(u, u) P_c (h_p, h_q)$. Using (b1), it follows that $(u, u) P_c (\mu(s_k), \mu(s_l))$, contradicting individual rationality of μ for (P^H, P^C). Hence, $h_p R_{s_k} u$ or $h_q R_{s_l} u$.

Assume that $u P_{s_k} h_p$ and $h_q R_{s_l} u$. Then, by weak responsiveness condition (b), $(u, h_q) P_c (h_p, h_q)$. Hence, $((s_k, s_l), (u, h_q))$ is a blocking coalition for μ. Similarly, if $h_p R_{s_k} u$ and $u P_{s_l} h_q$, then $(h_p, u) P_c (h_p, h_q)$ and $((s_k, s_l), (h_p, u))$ is a blocking coalition for μ. Hence, it is without loss of generality to assume that for the blocking coalition $((s_k, s_l), (h_p, h_q))$, $h_p R_{s_k} u$ and $h_q R_{s_l} u$. Suppose that $h_p P_{s_k} \mu(s_k)$ or $h_q P_{s_l} \mu(s_l)$. Then, according to (b2), (s_k, h_p) or (s_l, h_q) can block μ in $(P^H, P^S(P^C))$. Hence, $\mu(s_k) R_{s_k} h_p$ and $\mu(s_l) R_{s_l} h_q$.

If $h_p \neq \mu(s_l)$, then weak responsiveness (b) implies $(\mu(s_k), \mu(s_l)) R_c (h_p, \mu(s_l)) R_c (h_p, h_q)$, which contradicts (b1). If $h_q \neq \mu(s_k)$, then similarly, we obtain a contradiction to (b2). In both cases, it follows that $\mu \in S(R^H, R^C)$; a contradiction. Note that the only case for which the last argument does not apply is the case where $h_p = \mu(s_l)$ and $h_q = \mu(s_k)$. In other words, all blocking coalitions are of the form $((s_k, s_l), (\mu(s_l), \mu(s_k)))$.

Then, we can easily obtain a stable matching $\mu' \in S(R^H, R^C)$ from μ by satisfying all these blocking coalitions, i.e., for all couples (s_k, s_l) involved in a blocking coalition, $\mu'(s_k) = \mu(s_l)$ and $\mu'(s_l) = \mu(s_k)$. (Note that up to this point we have proven Theorem 1.)

We complete the proof by showing that $\mu' \in S(R^H, R^S(R^C))$. Assume that $\mu' \notin S(R^H, R^S(R^C))$. Then, there exists a blocking coalition (s, h) such that (b1) $h P_s \mu'(s)$ and (b2) $s P_h \mu'(h)$.

Suppose that $\mu'(s) = \mu(s)$ and observe that by construction of μ', $\mu'(h) R_h \mu(h)$ for all $h \in H$. Then, $h P_s \mu(s)$ and $s P_h \mu(h)$; contradicting $\mu \in S(R^H, R^S(R^C))$. Thus, assume that $s \in \{s_k, s_l\}$, where couple $c = (s_k, s_l)$ was involved in a blocking

coalition and $\mu'(s_k) = \mu(s_l)$ and $\mu'(s_l) = \mu(s_k)$. Without loss of generality, assume that $s = s_k$.

Then, $\mu \in S(R^H, R^S(R^C))$ implies (i) $\mu(s_k) \, P_{s_k} \, h$ and (ii) $\mu(s_l) \, P_{s_l} \, \mu'(s_l)$. By (b1) and (i), $\mu(s_k) \, P_{s_k} \, h \, P_{s_k} \, \mu'(s_k)$. This and the individual rationality of μ implies $h \in H \setminus \{\mu(s_k), \mu(s_l)\}$. Recall that $\mu'(s_k) = \mu(s_l)$ and $\mu'(s_l) = \mu(s_k)$. Then, by weak responsiveness condition (b), $(\mu(s_k), \mu(s_l)) \, P_c \, (h, \mu(s_l)) \, P_c \, (h, \mu(s_k)) \, P_c \, (\mu(s_l), \mu(s_k)) = (\mu'(s_k), \mu'(s_l))$; contradicting that couple $c = (s_k, s_l)$ was involved in a blocking coalition.

To summarize, for any weakly responsive couples market (R^H, R^C), either (i) there exists $\mu \in S(R^H, R^S(R^C))$ such that $\mu \in S(R^H, R^C)$ or (ii) there exists $\mu' \in S(R^H, R^C)$ such that $\mu' \in S(R^H, R^S(R^C))$. Hence, $S^*(R^H, R^C) \neq \emptyset$. □

The proof of Proposition 1 in particular shows that the correspondence S^* collects all stable matchings in the couples market that are induced by some stable matching in an associated singles market. As mentioned before, not any matching μ that is stable for a singles market needs to be stable for the couples market. In that case, blocking coalitions consist of a couple and their employer(s) so that switching jobs within the couple makes everyone better off. The matching μ' that is obtained by such job switches is stable in the couples market as well as in the singles markets.

Finally, Klaus and Klijn (2005, Theorem 3.5) (Klaus et al. 2008) showed that under a restricted unemployment aversion condition, the domain of weakly responsive preferences is maximal for the existence of stable matchings. In view of this result, the class of weakly responsive preferences is an important preference domain for the study of stability in couples markets.

In what follows, we are interested in the Nash implementability of correspondences S, $\bar{S}$, and S^* for weakly responsive couples markets. We use the concept of *full implementation*, meaning that *any* stable matching is achievable (in Nash equilibrium) by letting agents act strategically. Given weakly responsive preferences, one straightforward idea to approach the problem is to first decompose any couples market into an associated singles market and then to apply previously established implementation results for singles market (e.g., Kara and Sönmez 1996). However, as argued above, the stable correspondence of the associated singles markets and the stable correspondence for couples markets are not logically related. So, implementing the stable correspondence for associated singles markets is not the same as implementing the stable correspondence for couples markets.[4] In the next sections, we prove that the stable correspondence is Nash implementable, but that neither the correspondence S^* nor the correspondence $\bar{S}$ is Nash implementable.

3 Monotonicity and Nash implementation

So far, we have described all ingredients for the implementation environment, which is given by the set of agents $H \cup C$, the set of alternatives (matchings) $\mathcal{M}$, and the set

---

[4] Also note that when considering associated singles markets, the players of the underlying games are hospitals and students, whereas in the original problem of couples markets, the players of the underlying games are hospitals and couples.

of preference profiles (couples markets) $\mathcal{R}$. Before focusing on Nash implementability of the stable correspondence, we consider a necessary condition for this endeavor (Maskin 1999): (Maskin) monotonicity.

In order to introduce (Maskin) monotonicity, we need some standard terms and notation. For any agent $i \in H \cup C$, preference relation $R_i \in \mathcal{R}^i$, and matching $\mu \in \mathcal{M}$, the *lower contour set of R_i at μ* is $L_i(R_i, \mu) := \{\mu' \in \mathcal{M} \mid \mu \, R_i \, \mu'\}$.

Next, we define (Maskin) monotonic transformations. Loosely speaking, for any preference profile R and any matching μ, if at a preference profile R' all agents $i \in H \cup C$ consider their match $\mu(i)$ to be (weakly) better, then R' is a (Maskin) monotonic transformation of R at μ. Formally, for preference profiles $R, R' \in \mathcal{R}$ and matching $\mu \in \mathcal{M}$, R' is a *(Maskin) monotonic transformation of R at μ* if for all $i \in H \cup C$, $L_i(R_i, \mu) \subseteq L_i(R_i', \mu)$. By $MT(R, \mu)$, we denote the *set of all monotonic transformations of R at μ*. For agent $i \in H \cup C$, preference relation $R_i' \in \mathcal{R}^i$, and preference profile $R \in \mathcal{R}$, we obtain preference profile (R_i', R_{-i}) by replacing R_i at R by R_i'.

A solution φ is (Maskin) monotonic if a matching μ that is chosen at preference profile R is also chosen at a preference profile R' where μ is considered (weakly) better by all agents. Formally, a solution $\varphi : \mathcal{R}_0 \rightrightarrows \mathcal{M}$ is *(Maskin) monotonic* if for all preference profiles $R, R' \in \mathcal{R}_0 \subseteq \mathcal{R}$, $\mu \in \varphi(R)$ and $R' \in MT(R, \mu)$ imply $\mu \in \varphi(R')$.

Theorem 2 *The stable correspondence S is monotonic on any preference domain $\mathcal{R}_0$ for which S is well defined, i.e., for all $R \in \mathcal{R}_0$, $S(R) \neq \emptyset$.*

Proof Assume that couples' preferences are such that the stable correspondence is well defined (e.g., couples' preferences are weakly responsive). Consider a couples market $R \in \mathcal{R}_0$ and a stable matching $\mu \in S(R)$. Let $R' \in \mathcal{R}_0$ be a monotonic transformation of R at μ, i.e., $R' \in MT(R, \mu)$. In order to show that the stable correspondence is monotonic, we need to show that $\mu \in S(R')$.

Individual rationality. Matching μ is individually rational at couples market R'. Let $c = (s_k, s_l) \in C$ [$h \in H$]. Since $\mu \in S(R)$, by individual rationality (i1) [(i2)], $(\mu(s_k), u), (u, \mu(s_l)), (u, u) \in L_c(R_c, \mu)$ [$\emptyset \in L_h(R_h, \mu)$]. Since $R' \in MT(R, \mu)$, $(\mu(s_k), u), (u, \mu(s_l)), (u, u) \in L_c(R_c', \mu)$ [$\emptyset \in L_h(R_h', \mu)$]. Hence, μ being individually rational at R implies that μ is individually rational at R'.

No blocking. No blocking coalition exists for matching μ at couples market R'. Suppose, by contradiction, that a blocking coalition $(c, (h_p, h_q))$ with $c = (s_k, s_l)$ exists. Hence, (b1) $(h_p, h_q) \, P_c' \, (\mu(s_k), \mu(s_l))$ and (b2) [$h_p \in H$ implies $s_k \, R_{h_p}' \, \mu(h_p)$] and [$h_q \in H$ implies $s_l \, R_{h_q}' \, \mu(h_q)$].

Suppose $(\mu(s_k), \mu(s_l)) \, R_c \, (h_p, h_q)$. Then, $R' \in MT(R, \mu)$ implies $(\mu(s_k), \mu(s_l)) \, R_c' \, (h_p, h_q)$, a contradiction to (b1). Hence, (b1) $(h_p, h_q) \, P_c \, (\mu(s_k), \mu(s_l))$.

Since μ is individually rational at R', $(\mu(s_k), \mu(s_l)) R_c'(u, u)$, and therefore, $h_p \in H$ or $h_q \in H$. Suppose $h_p \in H$. If $\mu(h_p) P_{h_p} s_k$, then $R' \in MT(R, \mu)$ implies $\mu(h_p) P_{h_p}' s_k$, a contradiction to (b2). Hence, $s_k \, R_{h_p} \, \mu(h_p)$. Similarly, it follows that $h_q \in H$ implies $s_l \, R_{h_q} \, \mu(h_q)$. Thus, (b2) is satisfied for couples market R.

Since (b1) and (b2) are satisfied for couples market R, $(c, (h_p, h_q)) = ((s_k, s_l), (h_p, h_q))$ is a blocking coalition for matching μ at couples market R. This contradicts $\mu \in S(R)$.

We have shown that μ is individually rational at R' and that no blocking coalition exists for matching μ at couples market R'. Hence, $\mu \in S(R')$. □

Note that, in particular, Theorem 2 is valid for weakly responsive couples markets. One may be tempted to think that Theorem 2 for the weakly responsive preferences domain can be proved by considering an associated singles market, applying previous results on monotonicity of the stable correspondence (e.g., Kara and Sönmez 1996) in such markets, and then transferring the result back to the original couples market. However, there are two main problems with this "proof strategy."

First, as discussed earlier, the correspondence of matchings that are stable for the associated singles markets, $\bar{S}$ is not related to the correspondence S of stable matchings for couples markets. Second, a monotonic transformation of a couple's preference relation does not necessarily involve two monotonic transformations of the corresponding singles' preference relations in an associated singles market. Hence, the monotonic transformations used in the proof of Theorem 2 may not have corresponding monotonic transformations in an associated singles market.

The next proposition shows that neither of the solutions $\bar{S}$ or S^* is monotonic.

Proposition 2 *For weakly responsive couples markets, the correspondences of double stable matchings S^* and of stable matchings in singles markets $\bar{S}$ are not monotonic.*

Proof Consider the following example with two hospitals and one couple in the market. We define weakly responsive preferences R and $\bar{R}$ as follows.

R^H		R^C	$R^S(R^C)$		$\bar{R}^H$		$\bar{R}^C$	$R^S(\bar{R}^C)$	
h_1	h_2	(s_1, s_2)	s_1	s_2	h_1	h_2	(s_1, s_2)	s_1	s_2
$\boxed{s_1}$	s_1	h_2, h_1	$\boxed{h_1}$	h_1	$\boxed{s_1}$	s_1	h_2, h_1	h_2	h_1
s_2	$\boxed{s_2}$	$\boxed{h_1, h_2}$	h_2	$\boxed{h_2}$	s_2	$\boxed{s_2}$	$\boxed{h_1, h_2}$	$\boxed{h_1}$	$\boxed{h_2}$
$\emptyset$	$\emptyset$	h_1, u	u	u	$\emptyset$	$\emptyset$	h_2, u	u	u
		h_2, u					h_1, u		
		u, h_1					u, h_1		
		u, h_2					u, h_2		
		u, u					u, u		

Note that at both couples markets R and $\bar{R}$ complete unemployment (u, u) is the worst option. Hence, the associated singles preferences $R^S(R^C)$ and $R^S(\bar{R}^C)$ listed above are unique. Thus,

$$S(R) = S(\bar{R}) = \{(h_1, h_2), (h_2, h_1)\}, \ \bar{S}(R) = S^*(R) = \{(h_1, h_2)\},$$

$$\bar{S}(\bar{R}) = S^*(\bar{R}) = \{(h_2, h_1)\}.$$

Now, define the matching μ by $\mu(c) = (h_1, h_2)$. Observe that the couple's preferences $\bar{R}^C$ are obtained from R^C by rearranging the lower contour set at $\mu(c)$. Since the

hospitals' preferences remain unaltered, $\bar{R} \in MT(R, \mu)$. However, $\mu \in S^*(R)$ and $\mu \notin S^*(\bar{R})$ imply that S^* is not monotonic. Indeed, in the associated singles market $(\bar{R}^H, R^S(\bar{R}^C))$, μ can be blocked by the pair (h_2, s_1). Since S^* and $\bar{S}$ coincide on R and $\bar{R}$, we get the same contradiction for $\bar{S}$. Observe that the monotonic transformation of the couple's preferences at μ triggered a non-monotonic transformation of the associated singles preferences, rendering the matching μ unstable in the singles market. □

Next, given that stable matchings exist, an important question to ask is if a stable matching indeed will be reached through decentralized market interactions. It is well observed that strategic behavior of agents often leads to undesirable (e.g., inefficient) economic outcomes. We now analyze if it is possible to achieve stability in matching markets with couples through strategic interaction. We use the concept of *full implementation*, meaning that *any* stable matching is achievable (in Nash equilibrium) by letting agents act strategically. Next, we introduce the (implementation-) theoretical framework for this analysis.

A *mechanism* is a pair (M, g), where $M := \prod_{i \in N} M_i$ denotes *a set of messages* or *strategy profiles* and a function $g : M \longrightarrow \mathcal{A}$ called *outcome function*. The outcome function assigns to each strategy profile a matching in $\mathcal{M}$. Since g contains all the relevant information, we identify a mechanism with its outcome function. A mechanism g together with a preference profile $R \in \mathcal{R}$ induces a non-cooperative game in strategic form, denoted by $\Gamma(g, R)$, as follows: Each strategy profile $m \in M$ is mapped to a matching $g(m) \in \mathcal{M}$. These outcomes of the game are then evaluated using the agents' preferences at R. Note that the fact that preferences in our context are ordinal does not limit the game theoretical analysis: either use ordinal preferences at R to compare outcomes or choose a representing utility function $u^R : \mathcal{M} \longrightarrow \mathbb{R}^l$ ($l = |H \cup C|$) of agents' preferences at R, and define the payoffs for $\Gamma(g, R)$ by the composition $u^R \circ g : M \longrightarrow \mathbb{R}^l$. Note that the player set in $\Gamma(g, R)$ consists of hospitals and couples. Students, who form a couple, therefore have to coordinate on their strategies in the game.

Mechanism g *Nash implements solution* φ (on $\mathcal{R}_0$), if for all $R \in \mathcal{R}_0$, we obtain $g(NE(\Gamma(g, R))) = \varphi(R)$, where $NE(\cdot)$ denotes the Nash equilibrium correspondence. Hence, for a given preference profile R and any matching $\mu \in \varphi(R)$ there is a Nash equilibrium of the induced game $\Gamma(g, R)$ with outcome μ. Conversely, the outcome of any Nash equilibrium of $\Gamma(g, R)$ belongs to $\varphi(R)$. We say that a solution φ is *Nash implementable*, if there exists a mechanism that Nash implements it.

Loosely speaking, a mechanism g describes a list of rules for a game in strategic form. These rules are independent of the true preference profile. Then, for any possible set of agents, represented by their preference profile R, any (desired) matching in $\varphi(R)$ can be achieved by strategic interaction in equilibrium, and thus can be obtained in a non-cooperative fashion.

Maskin (1977, 1999) showed that a Nash implementable solution necessarily has to be monotonic. However, monotonicity alone is not a sufficient condition for Nash implementability, unless it is paired with some other condition(s), e.g., *no veto power*. Moore and Repullo (1990) provide a necessary and sufficient condition (condition μ) for Nash implementability, which is a combination of monotonicity and a weaker form

of no veto power. The stable correspondence does not satisfy no veto power, but we show next that it does satisfy Moore and Repullo's (1990) condition μ and hence can be implemented by a version of Maskin's (1999) mechanism (see Moore and Repullo 1990, Appendix).

For any agent $i \in H \cup C$, couples market $R \in \mathcal{R}$, and subset $\mathcal{M}' \subseteq \mathcal{M}$ of matchings, let $B_i(R, \mathcal{M}')$ denote the set of "best matchings for agent i in $\mathcal{M}'$ with respect to preference relation R_i", i.e., $B_i(R, \mathcal{M}') := \{\mu \in \mathcal{M}' \mid \mu(i) \ R_i \ \mu'(i) \text{ for all } \mu' \in \mathcal{M}'\}$. Note that if $\mu \in B_i(R, \mathcal{M}')$ and $\mu' \in \mathcal{M}'$ is such that $\mu(i) = \mu'(i)$, then $\mu' \in B_i(R, \mathcal{M}')$. That means that all best matchings in $B_i(R, \mathcal{M}')$ specify the same match for agent i.

Definition 1 *Condition μ in Moore and Repullo (1990)*
Let $\mathcal{R}_0 \subseteq \mathcal{R}$. A solution $\varphi : \mathcal{R}_0 \rightrightarrows \mathcal{M}$ satisfies *condition μ* if there exists a set $\mathcal{L} \subseteq \mathcal{M}$ and for each $i \in H \cup C$, $R \in \mathcal{R}_0$ and $\mu \in \varphi(R)$ there is a set $T_i(R, \mu) \subseteq \mathcal{L}$ with $\mu \in B_i(R, T_i(R, \mu))$ such that for each $\bar{R} \in \mathcal{R}_0$ and $j \in H \cup C$, the following three conditions are satisfied:

$(\mu 1)$ $\mu \in \bigcap_{i \in H \cup C} B_i(\bar{R}, T_i(R, \mu))$ *implies* $\mu \in \varphi(\bar{R})$;

$(\mu 2)$ $\mu^* \in B_j(\bar{R}, T_j(R, \mu)) \cap \bigcap_{i \in H \cup C \setminus \{j\}} B_i(\bar{R}, \mathcal{L})$ *implies* $\mu^* \in \varphi(\bar{R})$;

$(\mu 3)$ $\mu^* \in \bigcap_{i \in H \cup C} B_i(\bar{R}, \mathcal{L})$ *implies* $\mu^* \in \varphi(\bar{R})$.

Moore and Repullo (1990, Theorem 1) show that in the presence of three or more agents, a solution φ is Nash implementable if and only if it satisfies condition μ.

Theorem 3 *If $|H \cup C| \geq 3$, then the stable correspondence S is Nash implementable.*

Before we prove Theorem 3, we pause for an observation. As Moore and Repullo (1990) pointed out, the sets $\mathcal{L}$ and $T_i(R, \mu)$ have an intuitive interpretation. The set $\mathcal{L}$ describes all the outcomes (matchings) that at most result from strategic interaction. Take the implementing mechanism g, a preference profile $R \in \mathcal{R}_0$, and a Nash equilibrium in $\Gamma(g, R)$, with stable matching μ as the outcome. Then, the set $T_i(R, \mu)$ contains all possible outcomes (matchings) that agent i might enforce when deviating from his equilibrium strategy. Due to the voluntary participation requirement (individual rationality) that is built into the stability concept, an agent can always enforce to stay unmatched. Suppose a couple is matched to two hospitals, then there are three possibilities for the couple to enforce their individual rationality constraints: either of the students may wish to resign from the job while the other keeps his or her match, or both of them resign at the same time. We introduce some notation to model "deviations to unemployment" as described above.

Consider a matching $\mu' \in \mathcal{M}$ and a couple $c = (s_k, s_l) \in C$. Define the matching $\mu'^{(c,1)}$ obtained from matching μ' when the first student of the couple switches to unemployment by

- $\mu'^{(c,1)}(c) = (u, \mu'(s_l))$,
- if $\mu'(s_k) \in H$, then $\mu'^{(c,1)}(\mu'(s_k)) = \emptyset$, and

- for all $i \in (H \cup C) \setminus \{c, \mu'(s_k)\}$, $\mu'^{(c,1)}(i) = \mu'(i)$.

Define the matching $\mu'^{(c,2)}$ obtained from matching μ' when the second student of the couple switches to unemployment by

- $\mu'^{(c,2)}(c) = (\mu'(s_k), u)$,
- if $\mu'(s_l) \in H$, then $\mu'^{(c,2)}(\mu'(s_l)) = \emptyset$, and
- for all $i \in (H \cup C) \setminus \{c, \mu'(s_l)\}$, $\mu'^{(c,2)}(i) = \mu'(i)$.

Finally, define the matching $\mu'^{(c,12)}$ obtained from matching μ' when both students of the couple switch to unemployment by

- $\mu'^{(c,12)}(c) = (u, u)$,
- if $\mu'(s_k) \in H$, then $\mu'^{(c,12)}(\mu'(s_k)) = \emptyset$,
- if $\mu'(s_l) \in H$, then $\mu'^{(c,12)}(\mu'(s_l)) = \emptyset$, and
- for all $i \in (H \cup C) \setminus \{c, \mu'(s_k), \mu'(s_l)\}$, $\mu'^{(c,12)}(i) = \mu'(i)$.

Now, let μ be a stable matching at couples market $R \in \mathcal{R}$. Then, for $i \in H \cup C$, we define sets $T_i(R, \mu)$ as follows:

$$\text{if } i = h \in H, \quad \text{then } T_h(R, \mu) := L_h(R_h, \mu) \tag{1}$$

$$\text{if } i = c \in C, \quad \text{then } T_c(R, \mu) := \left\{ \mu' \in L_c(R_c, \mu) \mid \mu'^{(c,1)}, \mu'^{(c,2)} \in L_c(R_c, \mu) \right\} \tag{2}$$

Thus, for any couple c, the set $T_c(R, \mu)$ is obtained from the couple's lower contour set $L_c(R_c, \mu)$ by removing all matchings that admit a better match than $\mu(c)$ for c when one of the students in the couple resigns from his/her job at μ. For instance, if $\tilde{\mu} \in L_c(R_c, \mu)$ with $\tilde{\mu}(c) = (h_p, h_q)$ is such that $(h_p, u) \, P_c \, \mu(c) \, P_c \, (h_p, h_q)$, then $\tilde{\mu}^{(c,2)} \notin L_c(R_c, \mu)$ and $\tilde{\mu} \notin T_c(R, \mu)$. Since μ is stable for couples market R, any matching that assigns (u, u) to couple c is in the lower contour set $L_c(R_c, \mu)$ and belongs to $T_c(R, \mu)$. Note that $\mu' \in T_c(R, \mu)$ implies $\mu'^{(c,1)}, \mu'^{(c,2)}, \mu'^{(c,12)} \in T_c(R, \mu)$. Furthermore, as $\mu \in T_c(R, \mu)$ (μ is stable) and $T_c(R, \mu) \subseteq L_c(R_c, \mu)$, we conclude $\mu \in B_c(R, T_c(R, \mu))$ as required in condition μ.

Since μ is stable, all matchings that do not assign any student to hospital h are contained in the lower contour set $L_h(R_h, \mu)$. Hence, for any hospital, the definition of $T_h(R, \mu)$ as the hospital's lower contour set $L_h(R_h, \mu)$ can be interpreted to be in the same spirit as the definition of $T_c(R, \mu)$ for couple c.

Proof of Theorem 3 Let the stable correspondence S be well defined on $\mathcal{R}_0 \subseteq \mathcal{R}$. We show that S (given $R \in \mathcal{R}_0$ and $\mu \in S(R)$) satisfies condition μ (Definition 1) with $\mathcal{L} = \mathcal{M}$ and sets $T_i(R, \mu)$ as defined in (1) and (2).

Let $R, \bar{R} \in \mathcal{R}_0$, $\mu \in S(R)$, and $j \in H \cup C$. Then, condition μ reads as follows:

$(\mu 1)$ $\mu \in \bigcap_{i \in H \cup C} B_i(\bar{R}, T_i(R, \mu))$ implies $\mu \in S(\bar{R})$;

$(\mu 2)$ $\mu^* \in B_j(\bar{R}, T_j(R, \mu)) \cap \bigcap_{i \in H \cup C \setminus \{j\}} B_i(\bar{R}, \mathcal{M})$ implies $\mu^* \in S(\bar{R})$;

(μ3) $\mu^* \in \bigcap_{i \in H \cup C} B_i(\bar{R}, \mathcal{M})$ implies $\mu^* \in S(\bar{R})$.

Step 0 Individual rationality at $\bar{R}$.

Let $\mu \in S(R)$, $\mu' \in \mathcal{M}$, $i \in H \cup C$, and $\mu' \in B_i(\bar{R}, T_i(R, \mu))$. We prove that μ' is individually rational for i at couples market $\bar{R}$.

First, consider a hospital $h \in H$. It is easy to see that any matching $\mu' \in B_h(\bar{R}, T_h(R, \mu)) = B_h(\bar{R}, L_h(R_h, \mu))$ satisfies $\mu'(h) \, \bar{R}_h \, \emptyset$, i.e., μ' is individually rational for h at couples market $\bar{R}$.

Second, consider a couple $c = (s_k, s_l) \in C$. Since $\mu' \in B_c(\bar{R}, T_c(R, \mu))$, in particular, $\mu' \in T_c(R, \mu)$. Hence, by the construction of the set $T_c(R, \mu)$, $\mu'^{(c,1)}$, $\mu'^{(c,2)}$, $\mu'^{(c,12)} \in T_c(R, \mu)$. However, μ' is the best matching in $T_c(R, \mu)$ for c according to $\bar{R}_c$, which means $\mu'(c) \, \bar{R}_c \, \mu'^{(c,1)}(c) = (u, \mu'(s_l))$, $\mu'(c) \, \bar{R}_c \, \mu'^{(c,2)}(c) = (\mu'(s_k), u)$, and $\mu'(c) \, \bar{R}_c \, \mu'^{(c,12)}(c) = (u, u)$, i.e., μ' is individually rational for c at couples market $\bar{R}$.

Finally, if $\mu' \in \bigcap_{i \in H \cup C} B_i(\bar{R}, T_i(R, \mu))$, then (i1) and (i2) are satisfied and μ' is individually rational for couples market $\bar{R}$.

Step 1 S satisfies condition (μ1).

Note that by definition of the sets $T_c(R, \mu)$ for couples $c \in C$, we cannot straightforwardly apply monotonicity (Theorem 2) to prove condition (μ1).[5] Even though Moore and Repullo (1990, p. 1089) argue that condition (μ1) is equivalent to monotonicity, for the sake of completeness, we provide a direct proof of (μ1).

The assumption $\mu \in \bigcap_{i \in H \cup C} B_i(\bar{R}, T_i(R, \mu))$ together with Step 0 implies individual rationality of μ at $\bar{R}$. Assume that there exists a blocking coalition $(c, (h_p, h_q))$ with $c = (s_k, s_l)$ for μ at $\bar{R}$ and denote by μ^B a matching that matches the blocking agents accordingly. Thus,

(b1) $(h_p, h_q) \, \bar{P}_c \, (\mu(s_k), \mu(s_l))$ and

(b2) $[h_p \in H$ implies $s_k \, \bar{R}_{h_p} \, \mu(h_p)]$ and $[h_q \in H$ implies $s_l \, \bar{R}_{h_q} \, \mu(h_q)]$.

Note that by individual rationality of μ at $\bar{R}$, $h_p \in H$ or $h_q \in H$.

Recall that for all $h \in H$, $\mu \in B_h(\bar{R}, L_h(R_h, \mu))$. Thus, μ is among the best matchings in $L_h(R_h, \mu)$ according to $\bar{R}_h$. Hence, $L_h(R_h, \mu) \subseteq L_h(\bar{R}_h, \mu)$ and (b2) above implies

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

If $(h_p, h_q) \, P_c \, (\mu(s_k), \mu(s_l))$, then $(c, (h_p, h_q))$ is a blocking coalition for μ at R, a contradiction to $\mu \in S(R)$. Hence, $(\mu(s_k), \mu(s_l)) \, P_c \, (h_p, h_q)$. Therefore, $\mu^B \in L_c(R_c, \mu)$. Recall that $\mu \in B_c(\bar{R}, T_c(R, \mu))$ and (b1) $(h_p, h_q) \, \bar{P}_c \, (\mu(s_k), \mu(s_l))$. Thus, $\mu^B \in L_c(R_c, \mu) \setminus T_c(R, \mu)$. Then, by the definition of $T_c(R, \mu)$,

(b1)' $(h_p, u) \, P_c \, (\mu(s_k), \mu(s_l))$ or $(u, h_q) \, P_c \, (\mu(s_k), \mu(s_l))$.

Then, $h_p \in H$ or $h_q \in H$ in combination with (b1)' and (b2)' implies that $(c, (h_p, u))$ or $(c, (u, h_q))$ is a blocking coalition for μ at R, a contradiction to $\mu \in S(R)$.

---

[5] Note that for all $h \in H$, $L_h(R_h, \mu) \subseteq L_h(\bar{R}_h, \mu)$ (we explain this step in the proof of Step 1). Thus, if for all $c \in C$, $L_c(R_c, \mu) \subseteq L_c(\bar{R}_c, \mu)$, then $\bar{R} \in MT(R, \mu)$ and by Theorem 2, $\mu \in S(\bar{R})$. However, it is possible that $T_c(R, \mu) \subsetneq L_c(R_c, \mu)$ and $L_c(R_c, \mu) \nsubseteq L_c(\bar{R}_c, \mu)$, in which case $\bar{R} \notin MT(R, \mu)$.

Step 2 S satisfies condition ($\mu 2$).

Since all agents, possibly except agent j, consider μ^* as their best matching at $\bar{R}$, there is no blocking coalition. Using Step 0, μ^* is also individually rational; hence $\mu^* \in S(\bar{R})$.

Step 3 S satisfies condition ($\mu 3$).

Since all hospitals and all couples obtain their best match in μ^*, there is no blocking coalition and individual rationality at $\bar{R}$ is trivially satisfied; hence $\mu^* \in S(\bar{R})$. ☐

Observe that we did not impose any restrictions on preferences in the proof of Theorem 3. Therefore, the result is valid on any domain on which the existence of stable matchings is guaranteed. In view of Theorem 1, we therefore immediately obtain the following corollary.

Corollary 1 *Suppose* $|H \cup C| \geq 3$. *The stable correspondence S is Nash implementable for weakly responsive couples markets.*

Finally, by Proposition 2, neither of the solutions $\bar{S}$ or S^* is Nash implementable, as neither of them is even monotonic.

Corollary 2 *For weakly responsive couples markets, the correspondences of double stable matchings S^* and of stable matchings in singles markets $\bar{S}$ are not Nash implementable.*

4 Conclusion

We close our analysis with some remarks on why the Nash implementation of the stable correspondence in couples markets is different from the Nash implementation of the stable correspondence in singles markets.

We consider "full implementation," which requires that any outcome in the solution, i.e., any stable matching, is reachable via an equilibrium and any equilibrium yields a stable matching.[6] As discussed earlier, for each couples markets with (weakly) responsive preferences, there exists at least one associated singles markets. Any stable matching in an associated singles market induces a stable matching in the couples market. However, some couples markets have stable matchings that are not stable for any of the associated singles markets (e.g., Klaus and Klijn 2005, Example 3.4). Therefore, any implementation result for singles markets could at best show that some stable matchings can be reached through strategic behavior, but not all.

Therefore, one could think of using previous results for singles markets to *weakly implement* the stable correspondence on the domain of couples markets. This means that any Nash equilibrium outcome in the underlying games $\Gamma(g, R)$ is a stable matching for the couples market R, but that not all stable matchings have to be reached through Nash equilibria.[7] However, as not all stable matchings in associated singles

---

[6] For a survey on different implementation concepts see Thomson (1996).

[7] Formally, the difference between full and weak implementation is to require $g(NE(\Gamma(g, R)) \subseteq \varphi(R)$ instead of $g(NE(\Gamma(g, R)) = \varphi(R)$.

markets are also stable for the couples market (e.g., Klaus et al. 2008, Example 1.2), we cannot guarantee that the outcome of strategic interaction is stable.

The reason for this fundamental difference between couples and singles markets is that multiple weak responsive preferences for a couple can be associated with the same pair of students' preferences. This causes a crucial difference in the formulations of monotonicity for singles and for couples markets. For example, a monotonic transformation of a couple's preference relation does not have to correspond to monotonic transformations of both students' preferences. Consider for instance a couple $c = (s_1, s_2)$ and preferences R_c and R_c', where R_c' is a monotonic transformation of R_c at a matching μ. Thus, at R_c', couple c finds its match $(\mu(s_1), \mu(s_2))$ better than at R_c. This improvement can be caused by a transformation of the students associated singles preferences where student s_1 likes her match $\mu(s_1)$ at R_c' a lot better than at R_c while her partner likes his match $\mu(s_2)$ at R_c' less than at R_c. Hence, R_c' can be a monotonic transformation of R_c at μ for c, while the associated singles preference relation R_{s_2}' is not a monotonic transformation of the associated singles preference relation R_{s_2} at μ (see, e.g., the example in the proof of Proposition 2). So, on the one hand, requiring that a stable matching is still stable after a monotonic transformation of preferences in a couples market is a more demanding restriction than monotonicity for the associated singles markets. On the other hand, there are possibly "more" stable matchings in a couples market, which makes it "easier" for a matching to remain stable after a preference transformation.

References

Alkan, A., & Gale, D. (2003). Stable schedule matching under revealed preferences. *Journal of Economic Theory, 112*, 289–306.

Gale, D., & Shapley, L. (1962). College admissions and the stability of marriage. *American Mathematical Monthly, 69*, 9–15.

Haake, C.-J., & Klaus, B. (2008). Monotonicity and Nash implementation in matching markets with contracts. *Economic Theory* (in press). doi:10.1007/s00199-008-0399-8.

Hatfield, J. W., & Milgrom, P. (2005). Matching with contracts. *American Economic Review, 95*(4), 913–935.

Jackson, M. O. (2001). A crash course in implementation theory. *Social Choice and Welfare, 18*(3), 655–708.

Kara, T., & Sönmez, T. (1996). Nash implementation of matching rules. *Journal of Economic Theory, 68*, 425–439.

Kara, T., & Sönmez, T. (1997). Implementation of college admission rules. *Economic Theory, 9*, 197–218.

Kelso, A. S., & Crawford, V. P. (1982). Job matching, coalition formation, and gross substitutes. *Econometrica, 6*, 1483–1504.

Klaus, B., & Klijn, F. (2005). Stable matchings and preferences of couples. *Journal of Economic Theory, 121*, 75–106.

Klaus, B., & Klijn, F. (2007). Paths to stability for matching markets with couples. *Games and Economic Behavior, 58*, 154–171.

Klaus, B., Klijn, F., & Nakamura, T. (2008). Corrigendum: Stable matchings and preferences of couples. *Journal of Economic Theory* (Forthcoming)

Maskin, E. (1977). *Nash equilibrium and welfare optimality.* MIT working paper.

Maskin, E. (1999). Nash equilibrium and welfare optimality. *Review of Economic Studies, 66*, 23–38.

Maskin, E., & Sjöström, T. (2002). Implementation theory. In *Handbook of social choice and welfare* (*Vol. 1*). Amsterdam: North-Holland.

Moore, J., & Repullo, R. (1990). Nash implementation: A full characterization. *Econometrica, 58*(5), 1083–1099.

Ronn, E. (1990). NP-complete stable matching problems. *Journal of Algorithms, 11*, 285–304.

Roth, A. E. (1984a). 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. (1984b). Stability and polarization of interests in job matching. *Econometrica, 52*, 47–57.

Roth, A. E. (1985). The college admissions problem is not equivalent to the marriage problem. *Journal of Economic Theory, 36*, 277–288.

Roth, A. E., & 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., & Sotomayor, M. A. (1990). *Two-sided matching: A study in game-theoretic modeling and analysis*. Econometric Society Monographs. Cambridge: Cambridge University Press.

Shapley, L. S., & Scarf, H. (1974). On cores and indivisibility. *Journal of Mathematical Economics, 1*, 23–28.

Sönmez, T. (1996). Implementation in generalized matching problems. *Journal of Mathematical Economics, 26*, 429–439.

Thomson, W. (1996). Concepts of implementation. *Japanese Economic Review, 47*, 133–143.

Yamato, T. (1992). On Nash implementation of social choice correspondences. *Games and Economic Behavior, 4*, 484–492.