Smith and Rawls Share a Room: Stability and Medians[*]

Bettina Klaus[†]   Flip Klijn[‡]

February 2009

Abstract

We consider one-to-one, one-sided matching (roommate) problems in which agents can either be matched as pairs or remain single. We introduce a so-called bi-choice graph for each pair of stable matchings and characterize its structure. Exploiting this structure we obtain as a corollary the "lone wolf" theorem and a decomposability result. The latter result together with transitivity of blocking leads to an elementary proof of the so-called stable median matching theorem, showing how the often incompatible concepts of stability (represented by the political economist Adam Smith) and fairness (represented by the political philosopher John Rawls) can be reconciled for roommate problems. Finally, we extend our results to two-sided matching problems.

JEL classification: C62, C78.

Keywords: fairness, matching, median, stability.

1 Introduction

Gale and Shapley (1962, Example 3) introduced the so-called roommate problems as follows: "An even number of boys wish to divide up into pairs of roommates." A very

common extension is to allow also for odd numbers of agents and to consider the formation of pairs and singletons (rooms can be occupied either by one or by two agents). Therefore, an outcome for a roommate problem, a matching, is a partition of agents in pairs (of matched agents) and singletons. The class of roommate problems include as special cases the well-known marriage problems (Gale and Shapley, 1962).[1]

In a roommate problem, each agent has preferences over being matched with any of the other agents and remaining unmatched. A key property of a matching is stability: a matching is stable if it satisfies individual rationality and no pair of agents that are not matched with one another prefer to be so. Gale and Shapley (1962) exhibited an unsolvable roommate problem, i.e., a roommate problem in which there is no stable matching. Instead of focusing on necessary and/or sufficient conditions on the preferences for the existence of a stable matching,[2] we directly study the class of solvable roommate problems. For a solvable roommate problem there are typically multiple stable matchings. Our quest is to single out particularly appealing stable matchings. However, before dealing with this selection problem, we introduce for any two stable matchings a so-called bi-choice graph and characterize its structure (Lemma 1). This graphical tool will allow us to provide elementary and illustrative proofs of various well-known results.

A first idea to find appealing stable matchings is to select stable matchings that maximize the number of matched agents. However, Gusfield and Irving (1989) have shown that an agent who is single at any stable matching is also single at all other stable matchings. Using the bi-choice graph structure, we provide an elementary graphical proof of this "lone wolf" theorem (Theorem 1). The bi-choice graph structure can also be used to derive an elementary proof of the so-called decomposability lemma (Lemma 2), which turns out to be crucial for later results.

Then, since no selection can be based on the set of matched students, we try to find a stable matching that will be perceived as fair by the agents. Imagine that we ask each agent to rank all his matches at the stable matchings according to his preferences. Note that since an agent might be matched to the same agent in several stable matchings, this ranking is not strict. Clearly, we cannot always give the best match to every agent, but can we implement fairness by finding a matching that matches each agent to his l-th ranked match for some natural number l? We show that this idea of fairness or compromise is feasible if there is an odd number of stable matchings: the so-called median matching that assigns to each agent his median (ranked) match is well-defined and stable (Theorem 2). In a similar fashion, fairness is "almost" feasible if there is an even number of stable matchings (Corollary 1). Hence, stability (represented by the political economist

---

[1]There is a large literature on the marriage problem; see, for instance, Roth and Sotomayor (1990) and the two-sided matching bibliography on Al Roth's game theory, experimental economics, and market design page. In comparison, relatively few papers and books deal with roommate problems; some of the key references concerning roommate problems are Gusfield and Irving (1989); Tan (1991); Chung (2000); Diamantoudi et al. (2004).

[2]See, for instance, Tan (1991) and Chung (2000).

Adam Smith[3]) and fairness (represented by the political philosopher John Rawls[4]) can be reconciled for solvable roommate problems and "Smith and Rawls (almost) share a room." This result was already stated by Sethuraman and Teo (2001). Here we provide an elementary proof that does not recur to linear programming tools.

In the second part of the paper we turn to two-sided matching problems. A marriage problem is a roommate problem where the set of agents is exogenously partitioned into two sets such that each agent can only be matched with an agent of the other set. A further generalization of marriage problems are college admissions problems (with responsive preferences[5]), where one side of the market has several positions (seats or slots). Unlike roommate problems, marriage and college admissions problems (with responsive preferences) always allow for stable matchings.

Using linear programming tools, Teo and Sethuraman (1998) and Sethuraman et al. (2006) established the existence of natural compromise mechanisms for marriage and college admissions problems, respectively. Specifically, they showed that "generalized median matchings" are well-defined and stable. More formally, if all agents order again their matches at the, say, k stable matchings from best to worst, then the map that assigns to each agent of one side of the market its l-th best match and to each agent of the other side its $(k-l+1)$-st best match constitutes a stable matching. Fleiner (2002) and Klaus and Klijn (2006) independently provided alternative, short proofs for the above mentioned generalized median result for college admissions problems based on the so-called lattice structure of the set of stable matchings. Here we show that a slight adaptation of our elementary proof of the "Smith and Rawls share a room" theorem immediately leads to the stronger result of stable generalized median matchings for the (more specific) class of marriage problems (Theorem 3). Essentially the same proof is also valid for college admissions problems (Theorem 4). Unfortunately, in this case the extended proof can no longer be considered elementary because the decomposability lemma for college admissions problems (Lemma 3) is not an elementary result and in addition we have to establish a (non-elementary) "transitivity of blocking" property for college admissions problems (Lemma 4).

Our paper is organized as follows. Sections 2 and 3 contain the results as discussed in detail above for roommates and two-sided matching problems, respectively. We conclude in Section 4 with two examples of more general models in which a "median outcome" is either not well-defined or not stable.

---

[3]Adam Smith (1723–1790) propagated the view that individuals even though interested only in their own gains will still advance public interest (Smith, 1796).

[4]John Rawls (1921–2002) discussed important aspects of fairness and justice particularly suited for economic applications (Rawls, 1971).

[5]By responsiveness (Roth, 1985), a college's preference relation over sets of students is related to its ranking of single students in the following way: the college always prefers to add an acceptable student to any set of students (provided this does not violate the capacity constraint) and it prefers to replace any student by a better student.

2 Roommate Problems

A *roommate problem* (Gale and Shapley, 1962) is a pair $(N, (\succeq_i)_{i \in N})$ where N is a finite set of agents and, for each $i \in N$, $\succeq_i$ is a complete, transitive, and strict preference relation over N. For each $i \in N$, we interpret $\succeq_i$ as agent i's preferences over sharing a room with any of the agents in $N \setminus \{i\}$ and having a room for himself (or consuming an outside option such as living off-campus). Preferences are strict, i.e., $k \succeq_i j$ and $j \succeq_i k$ if and only if $j = k$. The indifference and strict preference relation associated with $\succeq_i$ are denoted by $\sim_i$ and $\succ_i$, respectively. A solution to a roommate problem, a *matching* μ, is a partition of N in pairs and singletons. Alternatively, we describe a matching by a function $\mu : N \to N$ of order two, i.e., for all $i \in N$, $\mu(\mu(i)) = i$. Agent $\mu(i)$ is agent i's *match*, i.e., the agent with whom he is matched to share a room (possibly himself). If $\mu(i) = i$ then we call i a single.

A *marriage problem* (Gale and Shapley, 1962) is a roommate problem $(N, (\succeq_i)_{i \in N})$ such that N is the union of two disjoint sets M and W (men and women), and each agent in M (respectively W) prefers being alone to being matched with any other agent in M (respectively W).

A matching μ is *blocked by a pair* $\{i, j\} \subseteq N$ (possibly $i = j$) if $j \succ_i \mu(i)$ and $i \succ_j \mu(j)$. If $\{i, j\}$ blocks μ, then $\{i, j\}$ is called a *blocking pair* for μ. A matching is *individually rational* if there is no blocking pair $\{i, j\}$ with $i = j$. A matching is *stable* if there is no blocking pair. A roommate problem is *solvable* if the set of stable matchings is non-empty. Gale and Shapley (1962) showed that all marriage problems are solvable and provided an unsolvable roommate problem (Gale and Shapley, 1962, Example 3).

The following is a simplified version of Gale and Shapley's example with three agents: $2 \succ_1 3 \succ_1 1$, $3 \succ_2 1 \succ_2 2$, and $1 \succ_3 2 \succ_3 3$. Clearly, the matching where all agents are singles is not stable (any two agents can block). So, assume two agents share a room. Then, the single agent is the best roommate for one of these two agents and hence a blocking pair can be formed. Tan (1991) provided a necessary and sufficient condition for the existence of a stable matching.

Note that for marriage problems, an individually rational matching never matches two men or two women, i.e., the partition consists of man-woman pairs and singletons.

Smith and Rawls: Stability and Fairness

The starting point of our analysis is a solvable roommate problem. Typically there are multiple stable matchings and with choice comes the opportunity to select a particularly appealing stable matching. Before we explore this choice in order to address fairness in addition to stability, we introduce a so-called bi-choice graph for each pair of stable matchings and characterize its structure.[6]

---

[6]It is not necessary to restrict the bi-choice graph to pairs of stable matchings. One easily verifies that if μ, μ', $\mu \neq \mu'$, are individually rational matching such that (a) there is no blocking pair $\{i', j'\}$ for

2.1 Bi-Choice Graphs

Let μ and μ' be stable matchings. We consider the following bi-choice graph $G(\mu, \mu') = (V, E)$. The set of vertices is the set of agents, i.e., $V = N$. The set E consists of three types of edges. Let $i, j \in N$. Then,

E1. there is a continuous directed edge from i to j, denoted by $i \bullet\!\!-\!\!\!\longrightarrow\!\!\bullet j$ if $j = \mu(i) \succ_i \mu'(i)$, i.e., agent i strictly prefers his match $j = \mu(i)$ under μ to his match under μ';

E2. there is a discontinuous directed edge from i to j, denoted by $i \bullet\!-\!-\!\blacktriangleright\!\bullet j$ if $j = \mu'(i) \succ_i \mu(i)$, i.e., agent i strictly prefers his match $j = \mu'(i)$ under μ' to his match under μ;

E3. there is a (continuous) undirected edge between i and j, denoted by $i \bullet\!\!-\!\!\!\longrightarrow\!\!\bullet j$ if $j = \mu(i) \sim_i \mu'(i)$, i.e., agent i is indifferent between his match $j = \mu(i)$ under μ and his match under μ'. Note that for $j = i = \mu(i) \sim_i \mu'(i)$ we allow for an undirected edge from i to himself; we call such an edge a loop: $i \bullet\!\!\!\bigcirc$.

We make the following observations about the bi-choice graph $G(\mu, \mu')$.

O1. An edge $i \bullet\!\!-\!\!\!\longrightarrow\!\!\bullet j$ implies $j \neq \mu'(i)$. An edge $i \bullet\!-\!-\!\blacktriangleright\!\bullet j$ implies $j \neq \mu(i)$.

O2. An edge $i \bullet\!\!-\!\!\!\longrightarrow\!\!\bullet j$ implies that $\mu'(i) = \mu(i) = j$, and therefore that there is no other edge adjacent to i or j.

(Follows from strictness of preferences.)

O3. For each agent i, there is either (i) a unique undirected edge to which i is adjacent or (ii) a unique outgoing directed edge from i.

(If there is an undirected edge to which i is adjacent, then by $O2$ this edge is unique. Suppose now there is no undirected edge to which i is adjacent. By completeness of preferences there is an outgoing directed edge ($E1$ or $E2$) from i to either $\mu(i)$ or $\mu'(i)$. In particular, the outgoing directed edge is unique.)

O4. An edge $i \bullet\!\!-\!\!\!\longrightarrow\!\!\bullet j$ or $i \bullet\!-\!-\!\blacktriangleright\!\bullet j$ implies that $j \neq i$.

(W.l.o.g. consider the first case, i.e., $i \bullet\!\!-\!\!\!\longrightarrow\!\!\bullet j$. If $j = i$, then $i = \mu(i) \succ_i \mu'(i)$. So, μ' is not individually rational; contradicting stability of μ'.)

O5. An edge $i \bullet\!\!-\!\!\!\longrightarrow\!\!\bullet j$ or $i \bullet\!-\!-\!\blacktriangleright\!\bullet j$ implies that there is no directed edge from j to i.

(W.l.o.g. consider the first case, i.e., $i \bullet\!\!-\!\!\!\longrightarrow\!\!\bullet j$. By $O4$, $j \neq i$. An edge $j \bullet\!\!-\!\!\!\longrightarrow\!\!\bullet i$ would imply that $\{i, j\}$ is a blocking pair for μ'; contradicting stability of μ'. An edge $j \bullet\!-\!-\!\blacktriangleright\!\bullet i$ would imply $i = \mu'(j)$ contradicting $j \neq \mu'(i)$ from $O1$.)

μ with $\mu'(i') = j'$ and (b) there is no blocking pair $\{i, j\}$ for μ' with $\mu(i) = j$, then Observations $O1$–$O7$ still apply; only the proofs of $O4$ and $O5$ need a slight adaptation. As a direct consequence, for these pairs Lemma 1 and Theorem 1 still hold. Since stability is the main requirement we consider here, we do not deal with these two slightly generalized results.

O6. An edge $i \,\bullet\!\!\longrightarrow\!\!\bullet\, j$ implies $j \,\bullet\!\!\dashrightarrow\!\!\bullet\, k$ for some $k \neq i, j$.
Similarly, an edge $i \,\bullet\!\!\dashrightarrow\!\!\bullet\, j$ implies $j \,\bullet\!\!\longrightarrow\!\!\bullet\, k$ for some $k \neq i, j$.

(W.l.o.g. consider the first case, i.e., $i \,\bullet\!\!\longrightarrow\!\!\bullet\, j$. By $O4$, $j \neq i$. By $O2$ and $O3$, there is an outgoing directed edge from j, say to k. By $O4$, $k \neq j$. By $O5$, $k \neq i$. Finally, suppose the edge from j to k is continuous, i.e., $j \,\bullet\!\!\longrightarrow\!\!\bullet\, k$. Then, the two continuous edges imply $j = \mu(i)$ and $k = \mu(j)$, contradicting $k \neq i$.)

***O7.*** Each vertex has at most 1 incoming directed edge.

(Suppose k has (at least) 2 incoming directed edges, say from i and j, $i \neq j$. By $O4$, $k \neq i, j$. Note $[\mu(i) = k$ and $\mu'(j) = k]$ or $[\mu'(i) = k$ and $\mu(j) = k]$. In either case, by $O2$ and $O3$ there is a directed edge from k to either i or j, contradicting $O5$.)

Lemma 1 (Bi-choice graph components).
Let μ and μ' be stable matchings. Let $i \in N$. Then, agent i's component of $G(\mu, \mu')$ either

(a) *equals $i \,\bullet\!\!\longrightarrow\!\!\bullet\, j$ for some agent j (i.e., $i \,\bigcirc\,$ if $j = i$), or*

(b) *is a directed even cycle, i.e., there is a directed path starting from i that induces a closed cycle $c_i = (i_1, i_2, i_3, \ldots, i_p)$ consisting of an even number $p \geq 4$ of agents (with $i \in \{i_1, \ldots, i_p\}$ and $i_r \neq i_s$ for all $r \neq s$) where continuous and discontinuous edges alternate.*

An example of such a cyclical component is
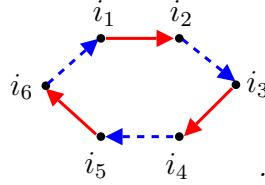
Proof. By $O3$, there is either (i) a unique undirected edge to which i is adjacent or (ii) a unique outgoing directed edge from i.

In Case (i), it follows from $O2$ that agent i's component is as described in (a). We complete the proof by showing that in Case (ii) agent i's component is as described in (b). By $O6$, there is an "alternating" continuation of the unique outgoing directed edge from i. By $O2$, $O3$, and $O7$, each agent on this path has exactly one incoming and one outgoing directed edge. By the finiteness of the set of agents, the path then has to close to a cycle at some point. We can label the set of agents involved in the cycle by $\{i = i_1, i_2, i_3, \ldots, i_p\}$. By the alternation pattern induced by $O6$, p is an even number. $\square$

An example of a bi-choice-graph is
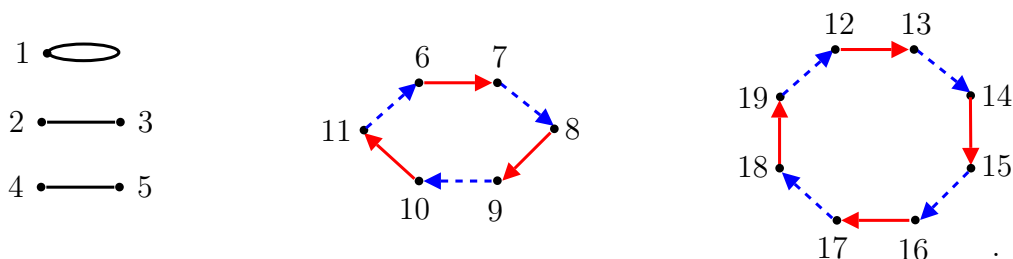
2.2   Lone Wolves, Medians, and Compromise

We now return to our quest of choosing a particularly appealing stable matching for a solvable roommate problem with multiple stable matchings. A first idea is to select a stable matching that maximizes the number of matched pairs. It turns out that no such selection is possible because an agent who is single at any stable matching is also single at all other stable matchings (Gusfield and Irving, 1989, Theorem 4.5.2). According to Roth and Sotomayor (1990, p. 50), the first statement of this theorem for the (sub)class of marriage problems can be found in McVitie and Wilson (1970) for the case when all men and women are mutually acceptable and in Roth (1984) for college admissions problems. Here we provide a new and elementary proof based on the structure of bi-choice graphs.

Theorem 1 (The lone wolf theorem).
Let μ and μ' be stable matchings. Then, μ and μ' have the same set of single agents, i.e., $\mu(i) = i$ if and only if $\mu'(i) = i$.

Proof. Suppose w.l.o.g. $\mu(i) = i$ but $\mu'(i) \neq i$. Consider $G(\mu, \mu')$. Since preferences are strict and μ' is individually rational, $\mu'(i) \succ_i i = \mu(i)$. Let $j = \mu'(i)$. Thus, by $E2$, $i \dashrightarrow j$ is a directed edge in $G(\mu, \mu')$ and by Lemma 1 (b), agent i is part of a (directed) cycle, where his predecessor and successor are his two different matches under μ and μ'. In particular, $\mu(i) \neq i$, contradicting the initial assumption. $\qquad \square$

Since no selection can be based on the set of matched agents, we next try to find a stable matching that will be perceived as fair by the agents. Imagine that we ask each agent to rank all stable matchings according to his preferences. We extend agents' preferences over the set of agents to matchings as follows. For any agent $i \in N$ and any two (stable) matchings μ and μ', $\mu \succeq_i \mu'$ if and only if $\mu(i) \succeq_i \mu'(i)$. Note that since an agent might be matched to the same agent at different matchings, this ranking is not strict. The indifference and strict preference relation (over matchings) associated with $\succeq_i$ are denoted by $\sim_i$ and $\succ_i$, respectively.

Clearly, we cannot always give the best match to every agent, but can we implement fairness by finding a matching that matches each agent with his l-th ranked match for some natural number l? We explain this idea in our next example and show that this idea of fairness or compromise is not always feasible.

Example 1 (An even number of stable matchings and no compromise).
Consider the following roommate problem with 8 agents and preferences as listed in the table below. The first column, for instance, means that $8 \succ_1 2 \succ_1 7 \succ_1 4 \succ_1 1 \succ_1 5 \succ_1 3 \succ_1 6$.

$\succ_1$	$\succ_2$	$\succ_3$	$\succ_4$	$\succ_5$	$\succ_6$	$\succ_7$	$\succ_8$
8	3	4	2	6	7	8	5
2	5	6	1	3	5	6	7
7	1	2	3	8	4	1	1
4	4	5	5	2	2	2	2
1	6	1	6	1	6	3	8
5	8	7	8	5	1	4	6
3	7	3	7	7	3	5	4
6	2	8	4	4	8	7	3

There are 4 stable matchings:

$$\mu_1 = \{\{1,2\}, \{3,4\}, \{5,6\}, \{7,8\}\}$$
$$\mu_2 = \{\{1,2\}, \{3,4\}, \{5,8\}, \{6,7\}\}$$
$$\mu_3 = \{\{1,4\}, \{2,3\}, \{5,6\}, \{7,8\}\}$$
$$\mu_4 = \{\{1,4\}, \{2,3\}, \{5,8\}, \{6,7\}\}$$

The following (weak) rankings on matchings are induced:

$$\mu_1 \sim_1 \mu_2 \succ_1 \mu_3 \sim_1 \mu_4$$
$$\mu_3 \sim_2 \mu_4 \succ_2 \mu_1 \sim_2 \mu_2$$
$$\mu_1 \sim_3 \mu_2 \succ_3 \mu_3 \sim_3 \mu_4$$
$$\mu_3 \sim_4 \mu_4 \succ_4 \mu_1 \sim_4 \mu_2$$
$$\mu_1 \sim_5 \mu_3 \succ_5 \mu_2 \sim_5 \mu_4$$
$$\mu_2 \sim_6 \mu_4 \succ_6 \mu_1 \sim_6 \mu_3$$
$$\mu_1 \sim_7 \mu_3 \succ_7 \mu_2 \sim_7 \mu_4$$
$$\mu_2 \sim_8 \mu_4 \succ_8 \mu_1 \sim_8 \mu_3$$

Agents 1, 2, and 4 order their matches at the four stable matchings as $2, 2, 4, 4$; $3, 3, 1, 1$; and $1, 1, 3, 3$ respectively. If agent 1 is given his first or second choice, then he is matched with agent 2 who then receives his third or fourth choice. If agent 1 is given his third or fourth choice, then he is matched with agent 4 who then receives his first or second choice. It follows that matching all agents with a l-th ranked match is not possible. $\diamond$

The impossibility result exhibited in the example is due to the fact that there is an even number of stable matchings. Next, we show that for roommate problems with an odd number of stable matchings a compromise matching where each agent is matched to a match of the same rank is possible. In fact, we prove that for any odd number of stable matchings, a stable matching at which each agent is matched to his "median" match always exists. Thus, for roommate problems with an odd number of stable matchings Adam Smith (who stands for stability) and John Rawls (who stands for fairness) represent compatible criteria and hence "can share a room."[7]

The next lemma, which appeared in Gusfield and Irving (1989, Lemma 4.3.9),[8] facilitates the proof of our main result.

Lemma 2 (Decomposability).
Let μ and μ' be stable matchings. Let $i \in N$. Suppose $\mu'(i) \neq \mu(i) = j$ for some $j \in N$. Then, $j, \mu'(i) \neq i$. Moreover,

(a) $\mu(i) \succ_i \mu'(i)$ *implies* $\mu'(j) \succ_j \mu(j)$;

(b) $\mu'(i) \succ_i \mu(i)$ *implies* $\mu(j) \succ_j \mu'(j)$.

Proof. By Theorem 1, $j, \mu'(i) \neq i$. To prove (a), assume $j = \mu(i) \succ_i \mu'(i)$. In terms of the bi-choice graph $G(\mu, \mu')$, $i \:\bullet\!\!\longrightarrow\!\!\bullet\: j$. By $O6$, $j \:\bullet\!\!-\!\!-\!\!\blacktriangleright\!\bullet\: k$ where $k = \mu'(j) \neq i$. Hence, $\mu'(j) \succ_j \mu(j)$. To prove (b), assume $\mu'(i) \succ_i \mu(i)$. In terms of the bi-choice graph $G(\mu, \mu')$, $i \:\bullet\!\!-\!\!-\!\!\blacktriangleright\!\bullet\: \mu'(i)$. By Lemma 1 (b), j is i's predecessor in $G(\mu, \mu')$ and there is a continuous directed edge from j to i. Hence, $\mu(j) = i \succ_j \mu'(j)$. $\square$

Our main result, Theorem 2, extends Theorem 4.3.5 in Gusfield and Irving (1989) from three to any odd number of stable matchings.

Let $\mu_1, \ldots, \mu_{2k+1}$ be an odd number of (possibly non-distinct) stable matchings. Let each agent rank these matchings according to his preferences as explained before (Example 1). We define agent i's median match as its $(k+1)$-st ranked match, and denote it by $\mathrm{med}\{\mu_1(i), \ldots, \mu_{2k+1}(i)\}$.

Theorem 2 (The median matching theorem or Smith and Rawls share a room).
Let $\mu_1, \ldots, \mu_{2k+1}$ be an odd number of (possibly non-distinct) stable matchings. Then, $\mu^* : N \to N$ defined by

$$\mu^*(i) := \mathrm{med}\{\mu_1(i), \ldots, \mu_{2k+1}(i)\} \text{ for all } i \in N$$

is a well-defined stable matching. We call μ^* the median matching of $\mu_1, \ldots, \mu_{2k+1}$.

---

[7]Example 1 is an example where the equilibrium and the fairness principles represented by Adam Smith and John Rawls are incompatible: there they will never be able to share a room.

[8]According to Roth and Sotomayor (1990, p. 50), the first statement of this lemma for the (sub)class of marriage problems can be found in Knuth (1976).

In the context of the linear programming approach to so-called bistable matching problems, Sethuraman and Teo (2001, Theorem 3.2) mentioned this result (without proof) as an interesting structural property of stable roommate matchings. Here we provide an elementary proof. In the first part of the proof we use the decomposability result (Lemma 2) to show that the median matching is well-defined. In the second part of the proof we use "transitivity of blocking" to show that the median matching is stable.

For any agent $i \in N$ we define partial preferences over sets of matchings as follows. Let U and V be two sets of matchings. If for all $\mu^U \in U$ and $\mu^V \in V$, $\mu^U(i) \succ_i \mu^V(i)$, then $U \succ_i V$. Furthermore, if for all $\mu^U \in U$ and $\mu^V \in V$, $\mu^U(i) \sim_i \mu^V(i)$, then $U \sim_i V$.

Proof. First, we show that μ^* is a well-defined matching, i.e., μ^* is of order two. Let $i \in N$ with $\mu^*(i) \neq i$. Let $j := \mu^*(i)$. We have to prove that $\mu^*(j) = i$. W.l.o.g. $\mu_1(i) \succeq_i \mu_2(i) \succeq_i \cdots \succeq_i \mu_{2k}(i) \succeq_i \mu_{2k+1}(i)$. Then, $\mu^*(i) = \mu_{k+1}(i) = j$ and $\mu_{k+1}(j) = i$. Let $\Sigma_1 := \{\mu_1, \ldots, \mu_k\}$ and $\Sigma_2 := \{\mu_{k+2}, \ldots, \mu_{2k+1}\}$. Define

$$
\begin{aligned}
SA_i &:= \{\mu \in \Sigma_1 \mid \mu \succ_i \mu_{k+1}\} && \text{(matchings that are "strictly above" } \mu_{k+1}\text{),} \\
IA_i &:= \{\mu \in \Sigma_1 \mid \mu \sim_i \mu_{k+1}\} && \text{(matchings that are "indifferent above" } \mu_{k+1}\text{),} \\
SB_i &:= \{\mu \in \Sigma_2 \mid \mu \prec_i \mu_{k+1}\} && \text{(matchings that are "strictly below" } \mu_{k+1}\text{), and} \\
IB_i &:= \{\mu \in \Sigma_2 \mid \mu \sim_i \mu_{k+1}\} && \text{(matchings that are "indifferent below" } \mu_{k+1}\text{).}
\end{aligned}
$$

For notational convenience we denote the singleton set $\{\mu_{k+1}\}$ by μ_{k+1}. Then,

$$SA_i \succ_i IA_i \sim_i \mu_{k+1} \sim_i IB_i \succ_i SB_i.$$

Note that

$$\mu \in IB_i \cup IA_i \Rightarrow \mu(i) = \mu_{k+1}(i) = j. \tag{1}$$

By decomposability (Lemma 2, a) and $\mu_{k+1} \succ_i SB_i$, $SB_i \succ_j \mu_{k+1}$. By decomposability (Lemma 2, b) and $SA_i \succ_i \mu_{k+1}$, $\mu_{k+1} \succ_j SA_i$. By (1), $IB_i \sim_j \mu_{k+1} \sim_j IA_i$. Summarizing,

$$SB_i \succ_j IB_i \sim_j \mu_{k+1} \sim_j IA_i \succ_j SA_i.$$

Since $SA_i \cup IA_i = \Sigma_1$ and $SB_i \cup IB_i = \Sigma_2$, we have $|SA_i \cup IA_i| = |IB_i \cup SB_i| = k$ and therefore, $\mu^*(j) = \mu_{k+1}(j) = i$. Hence, μ^* is a well-defined matching.

We now prove that μ^* is stable. By definition, μ^* is individually rational. Suppose there is a blocking pair $\{i, j\}$ with $i \neq j$ for μ^*, i.e., $j \succ_i \mu^*(i)$ and $i \succ_j \mu^*(j)$. Then, i prefers j to his match under at least $k + 1$ stable matchings in $\Sigma := \{\mu_1, \ldots, \mu_{2k+1}\}$. Similarly, j prefers i to his match under at least $k + 1$ stable matchings in Σ. Since Σ contains only $2k + 1$ matchings, for at least one matching $\mu \in \Sigma$, both $j \succ_i \mu(i)$ and $i \succ_j \mu(j)$ (it is here that we apply transitivity of blocking[9]). Hence, $\{i, j\}$ is a blocking pair for μ. This however contradicts stability of μ (the set Σ only contains stable matchings). Therefore, there is no blocking pair for μ^*. Hence, μ^* is stable. $\qquad\square$

---

[9]Note that $j \succ_i \mu^*(i) \succeq_i \ldots \succeq_i \mu(i) \succeq_i \ldots$ and $i \succ_j \mu^*(j) \succeq_j \ldots \succeq_j \mu(j) \succeq_j \ldots$ together with transitivity implies $j \succ_i \mu(i)$ and $i \succ_j \mu(j)$ (for the pairwise blocking notion we consider here this is immediate, but for more general blocking notions transitivity of blocking might not hold).

We can easily extend the result of Theorem 2 to an even number of stable matchings (Sethuraman and Teo, 2001, Theorem 3.3).

Corollary 1 (Smith and Rawls (almost) share a room).
Let $\mu_1, \ldots, \mu_{2k}$ be an even number of (possibly non-distinct) stable matchings. Then, there exists a stable matching at which each agent is assigned a match of rank k or $k+1$.

Proof. By Theorem 2, the median matching of the first $2k-1$ stable matchings, i.e., $\mu^* = \text{med}\{\mu_1, \ldots, \mu_{2k-1}\}$, is a well-defined and stable matching. Then, for any $i \in N$,

agent i is matched at μ^* with his $\begin{cases} k\text{-th ranked match} & \text{if } \mu^*(i) \succeq_i \mu_{2k}(i); \\ (k+1)\text{-st ranked match} & \text{if } \mu^*(i) \preceq_i \mu_{2k}(i). \end{cases}$ $\qquad \square$

In the next example we calculate the median matching for three stable matchings and demonstrate that the "median operator" is not closed, i.e., that the resulting median matching is not always one of the stable matchings that were used to calculate it.

Example 2 (An odd number of stable matchings and median matchings). [10]
Consider the following roommate problem with 8 agents and preferences as listed in the table below.

$\succ_1$	$\succ_2$	$\succ_3$	$\succ_4$	$\succ_5$	$\succ_6$	$\succ_7$	$\succ_8$
6	7	2	1	2	4	6	7
2	6	4	2	8	2	4	2
3	4	6	8	3	3	1	3
5	8	7	5	6	8	5	1
8	1	8	6	4	5	3	6
7	3	5	3	1	1	2	4
4	5	1	7	7	7	8	5
1	2	3	4	5	6	7	8

---

[10]Note that for this particular roommate problem the set of stable matchings induces a partition of the set of agents $N = M \cup W$ into 'men' $M = \{1, 2, 3, 4\}$ and 'women' $W = \{5, 6, 7, 8\}$ in a natural way. More precisely, at each stable matching, neither is any 'man' matched to another 'man,' nor is any 'woman' matched to another 'woman.' This suggests that this roommate problem might essentially be a marriage problem. Formally, one can associate with this roommate problem a marriage problem $(M \cup W, \succeq') = (M \cup W, (\succeq'_i)_{i \in N})$ where 'man' m's preferences $\succeq'_m$ are obtained from $\succeq_m$ by making the other 'men' unacceptable. Similarly, 'woman' w's preferences $\succeq'_w$ are obtained from $\succeq_w$ by making the other 'women' unacceptable. Note that in the associated marriage problem $(M \cup W, \succeq')$ there are fewer possible blocking pairs than in $(N, \succeq)$. Hence, any matching that is stable in the roommate problem is also stable in the associated marriage problem. However, the associated marriage problem has an additional stable matching given by

$$\nu = \{\{1, 7\}, \{2, 8\}, \{3, 5\}, \{4, 6\}\}.$$

Note that ν is not stable in the original roommate problem since it is blocked by $\{2, 4\}$ (two 'men'). Therefore, the roommate problem is not, as suspected, essentially equal to a marriage problem.

11

There are 5 stable matchings:

$$\mu_1 = \{\{1,5\}, \{2,6\}, \{3,7\}, \{4,8\}\}$$
$$\mu_2 = \{\{1,5\}, \{2,7\}, \{3,6\}, \{4,8\}\}$$
$$\mu_3 = \{\{1,7\}, \{2,6\}, \{3,8\}, \{4,5\}\}$$
$$\mu_4 = \{\{1,8\}, \{2,6\}, \{3,7\}, \{4,5\}\}$$
$$\mu_5 = \{\{1,8\}, \{2,7\}, \{3,6\}, \{4,5\}\}$$

Notice that μ_4 is not only the median matching of $\{\mu_1, \mu_2, \mu_3, \mu_4, \mu_5\}$ but also of $\{\mu_1, \mu_3, \mu_5\}$. However, $\mu_4 \notin \{\mu_1, \mu_3, \mu_5\}$. $\qquad\qquad\qquad\qquad\diamond$

3 Two-Sided Matching and Generalized Medians

3.1 Marriage Problems

We now turn to the subclass of marriage problems. These two-sided roommate problems exhibit some additional structure and therefore we can strengthen the median matching theorem. Let $\mu_1, \ldots, \mu_k$ be k (possibly non-distinct) stable matchings for a marriage problem and assume that each agent ranks these matchings according to his preferences. Using linear programming tools, Teo and Sethuraman (1998, Theorem 2) showed that the map that assigns to each man his l-th (weakly) best match and to each woman her $(k - l + 1)$-st (weakly) best match determines a well-defined and stable matching. Klaus and Klijn (2006) introduced and discussed (generalized) medians as compromise solutions for two-sided matching problems. Next, we show that we can adapt the proof of Theorem 2 to prove this stronger result for marriage problems. In contrast to previous proofs that use the so-called lattice structure of the set of stable matchings (see Fleiner, 2002, Theorem 5.5, Klaus and Klijn, 2006, Theorem 3.2, and Yenmez and Schwarz, 2008, Theorem 1)[11] our proof is elementary.

Consider a marriage problem $(M \cup W, (\succeq_i)_{i \in M \cup W})$. Let $\mu_1, \ldots, \mu_k$ be k (possibly non-distinct) stable matchings. Let each agent rank these matchings according to his preferences as explained before. Formally, for each $i \in M \cup W$ there is a sequence of matchings $(\mu_1^i, \ldots, \mu_k^i)$ such that $\{\mu_1^i, \ldots, \mu_k^i\} = \{\mu_1, \ldots, \mu_k\}$ and for any $l \in \{1, \ldots, k - 1\}$, $\mu_l^i(i) \succeq_i \mu_{l+1}^i(i)$. Thus, for any $l \in \{1, \ldots, k\}$, under μ_l^i agent i is assigned to his l-th (weakly) best match (among the k stable matchings).

For any $l \in \{1, \ldots, k\}$, we define the *generalized median matching* α_l as the function $\alpha_l : M \cup W \to M \cup W$ such that

$$\alpha_l(i) := \begin{cases} \mu_l^i(i) & \text{if } i \in M; \\ \mu_{(k-l+1)}^i(i) & \text{if } i \in W. \end{cases}$$

---

[11] Yenmez and Schwarz (2008) extend the analysis of stable median matchings from the class of marriage problems to two-sided (one-to-one) matching problems with side payments.

Theorem 3 (Marriage and compromise – generalized medians).
Let $\mu_1, \ldots, \mu_k$ be k (possibly non-distinct) stable matchings for a marriage problem. Then, for any $l \in \{1, \ldots, k\}$, α_l is a well-defined stable matching.

Proof. Let $l \in \{1, \ldots, k\}$. First, we show that α_l is a well-defined matching, i.e., α_l is of order two. W.l.o.g. let $m \in M$ with $\alpha_l(m) = w \in W$. We have to prove that $\alpha_l(w) = m$. W.l.o.g. $\mu_1(m) \succeq_m \mu_2(m) \succeq_m \cdots \succeq_m \mu_l(m) \succeq_m \cdots \succeq_m \mu_{k-1}(m) \succeq_m \mu_k(m)$ and $\alpha_l(m) = \mu_l(m)$. Then, $\mu_l(m) = \alpha_l(m) = w$ and $\mu_l(w) = m$. Using decomposability and arguments similar to those in the first part of the proof of Theorem 2,[12]

$$\mu_1 \succeq_m \cdots \succeq_m \mu_{l-1} \succeq_m \quad \mu_l \quad \succeq_m \mu_{l+1} \succeq_m \cdots \succeq_m \mu_k$$
$$\{\mu_{l+1}, \ldots, \mu_k\} \succeq_w \quad \mu_l \quad \succeq_w \{\mu_1, \ldots, \mu_{l-1}\}.$$

In particular, $\mu_l(w)$ is the $(k-l+1)$-st ranked match for woman w and therefore $\alpha_l(w) = \mu_l(w) = m$. Therefore, α_l is a well-defined matching.

We now prove that α_l is stable. By definition, α_l is individually rational. Suppose there is a blocking pair $\{m, w\}$ with $m \in M$ and $w \in W$ for α_l, i.e., $w \succ_m \alpha_l(m)$ and $m \succ_w \alpha_l(w)$. Then, m prefers w to his match under at least $k - l + 1$ stable matchings in $\Sigma := \{\mu_1, \ldots, \mu_k\}$. Similarly, w prefers m to her match under at least l stable matchings in Σ. Since Σ contains only $k < (k-l+1)+l$ matchings, for at least one matching $\mu \in \Sigma$, both $w \succ_m \mu(m)$ and $m \succ_w \mu(w)$ (here we apply transitivity of blocking). Hence, $\{m, w\}$ is a blocking pair for μ. This however contradicts stability of μ (the set Σ only contains stable matchings). Therefore, there is no blocking pair for α_l. Hence, α_l is stable. $\square$

3.2 College Admissions Problems

Fleiner (2002, Theorem 5.5), Klaus and Klijn (2006, Theorem 3.2), and Sethuraman et al. (2006, Theorem 9) generalized Theorem 3 to college admissions problems (Gale and Shapley, 1962) with so-called responsive preferences in which students have to be matched to colleges based on the students' and the colleges' preferences over the other side of the market and colleges' capacity constraints. Hence, like the roommate model, the college admissions model is a generalization of the marriage model. (However, the college admissions model is not a generalization of the roommate model, nor vice versa.) Next, we show that the proof of Theorem 3 is essentially valid for the college admissions model. More precisely, we can extend Theorem 3 to college admissions problems by using the same steps as in the proof of Theorem 3 (or, in fact, Theorem 2). Unfortunately, the extended proof will no longer be elementary in the sense that in order to be able to apply the key steps we need to recur to well-known but non-trivial results for college admissions problems.

---

[12]Recall that in the first part of the proof of Theorem 2 only the relative ranking of the stable matchings with respect to μ_{k+1} mattered – which exact rank μ_{k+1} had did not play any role in that part of the proof.

We first extend the marriage model to the college admissions model by introducing the following notation. There are two finite and disjoint sets of agents: a set $S = \{s_1, \ldots, s_m\}$ of students and a set $\mathcal{C} = \{C_1, \ldots, C_n\}$ of colleges. We denote a generic student by s and a generic college by C. For each college C, there is a fixed quota q_C that represents the number of positions it offers.[13]

Each student has a complete, transitive, and strict preference relation $\succeq_s$ over the colleges and the prospect of being *unmatched*. Hence, student s's preferences are defined over the elements in $\mathcal{C} \cup \{s\}$. If $C \in \mathcal{C}$ such that $C \succ_s s$, then we call C an *acceptable college for student s*.

A set of students $S' \subseteq S$ *is feasible for college C* if $|S'| \leq q_C$. Each college has a complete and transitive preference relation $\succeq_C$ over feasible sets of students. Hence, college C's preferences are defined over the elements in $\mathcal{P}(S, q_C) := \{S' \subseteq S : |S'| \leq q_C\}$. We make two assumptions on the preferences of a college C.[14]

First, C's preferences over singleton sets of students, or equivalently over individual students, are strict. For notational convenience we denote a singleton set $\{s\}$ by s. The second assumption describes comparisons of feasible sets of students when a single student is added or replaced. If $s \in S$ is such that $s \succ_C \emptyset$, then we call s an *acceptable student for college C*. If $s, s' \in S$ are such that $s \succ_C s'$, then we call student s a *better student than student s' for college C*. We assume that each college C's preferences over feasible sets of students are based on preferences over individual students such that C always prefers to add an acceptable student and it also prefers to replace any student by a better student. More formally, we assume that C's preferences are *responsive*, i.e., for all $S' \in \mathcal{P}(S, q_C)$,

- if $s \notin S'$ and $|S'| < q_C$, then $(S' \cup s) \succ_C S'$ if and only if $s \succ_C \emptyset$ and

- if $s \notin S'$ and $t \in S'$, then $((S' \backslash t) \cup s) \succ_C S'$ if and only if $s \succ_C t$.

A *college admissions problem* is a triple $(S, \mathcal{C}, (\succeq_i)_{i \in S \cup \mathcal{C}})$. A *matching* for college admissions problem $(S, \mathcal{C}, (\succeq_i)_{i \in S \cup \mathcal{C}})$ is a function μ on the set $S \cup \mathcal{C}$ such that

- each student is either matched to exactly one college or unmatched, i.e., for all $s \in S$, either $\mu(s) \in \mathcal{C}$ or $\mu(s) = s$,

- each college is matched to a feasible set of students, i.e., for all $C \in \mathcal{C}$, $\mu(C) \in \mathcal{P}(S, q_C)$,   and

- a student is matched to a college if and only if the college is matched to the student, i.e., for all $s \in S$ and $C \in \mathcal{C}$, $\mu(s) = C$ if and only if $s \in \mu(C)$.

Given matching μ, we call $\mu(s)$ *student s's match* and $\mu(C)$ *college C's match*.

Similar to the roommate and marriage model, a key property of matchings in the college admissions model is *stability*. First, we impose a voluntary participation condition.

---

[13]The marriage model is the special case where for all $C \in \mathcal{C}$, $q_C = 1$.

[14]See Roth and Sotomayor (1989) for a discussion of these assumptions.

A matching μ is *individually rational* if neither a student nor a college would be better off by breaking a current match, i.e., if $\mu(s) = C$, then $C \succ_s s$ and $\mu(C) \succ_C (\mu(C)\backslash s)$. By responsiveness of $\succeq_C$, the latter requirement can be replaced by $s \succ_C \emptyset$. Thus alternatively, a matching μ is individually rational if any student and any college that are matched to one another are mutually acceptable. Second, if a student s and a college C are not matched to one another at a matching μ but the student would prefer to be matched to the college and the college would prefer to either add the student or replace another student by student s, then we would expect this mutually beneficial adjustment to be carried out. Formally, a pair (s, C) *blocks* $(\mu(s), \mu(C))$ if $C \succ_s \mu(s)$ and

B1. $[\,|\mu(C)| < q_C$ and $s \succ_C \emptyset\,]$ or

B2. $[$there exists $t \in \mu(C)$ such that $s \succ_C t\,]$.[15]

A matching μ is *stable* if it is individually rational and there is no pair (s, C) that blocks $(\mu(s), \mu(C))$.[16] Gale and Shapley (1962) proved that each college admissions problem has at least one stable matching.

Consider a college admissions problem $(S, \mathcal{C}, (\succeq_i)_{i \in S \cup \mathcal{C}})$. Let $\mu_1, \ldots, \mu_k$ be k (possibly non-distinct) stable matchings. Let each student rank these matchings according to his preferences. Formally, for each $s \in S$ there is a sequence of matchings $(\mu_1^s, \ldots, \mu_k^s)$ such that $\{\mu_1^s, \ldots, \mu_k^s\} = \{\mu_1, \ldots, \mu_k\}$ and for any $l \in \{1, \ldots, k-1\}$, $\mu_l^s(s) \succeq_s \mu_{l+1}^s(s)$. Thus, for any $l \in \{1, \ldots, k\}$, under μ_l^s student s is assigned to his l-th (weakly) best match (among the k stable matchings). For any $l \in \{1, \ldots, k\}$, define the function α_l^S on the set S such that for all $s \in S$, $\alpha_l^S(s) := \mu_l^s(s)$.

By Roth and Sotomayor (1989, Theorem 3), each college can proceed similarly.[17] Formally, for each $C \in \mathcal{C}$ there is a sequence of matchings $(\mu_1^C, \ldots, \mu_k^C)$ such that $\{\mu_1^C, \ldots, \mu_k^C\} = \{\mu_1, \ldots, \mu_k\}$ and for any $l \in \{1, \ldots, k-1\}$, either $\mu_l^C(C) \succ_C \mu_{l+1}^C(C)$ or $\mu_l^C(C) = \mu_{l+1}^C(C)$. Thus, for any $l \in \{1, \ldots, k\}$, under μ_l^C college C is assigned to its l-th (weakly) best match (among the k stable matchings). For any $l \in \{1, \ldots, k\}$, define the function $\alpha_l^{\mathcal{C}}$ on the set $\mathcal{C}$ such that for all $C \in \mathcal{C}$, $\alpha_l^{\mathcal{C}}(C) := \mu_l^C(C)$.

For any $l \in \{1, \ldots, k\}$, we define the *generalized median matching* α_l by

$$\alpha_l(i) := \begin{cases} \alpha_l^S(s) & \text{if } i \in S; \\ \alpha_{k-l+1}^{\mathcal{C}}(C) & \text{if } i \in \mathcal{C}. \end{cases}$$

Theorem 4 (College admissions and compromise – generalized medians).
Let $\mu_1, \ldots, \mu_k$ be k (possibly non-distinct) stable matchings for a college admissions problem. Then, for any $l \in \{1, \ldots, k\}$, α_l is a well-defined stable matching.

---

[15]Recall that by responsiveness $B1$ implies $(\mu(C) \cup s) \succ_C \mu(C)$ and $B2$ implies $((\mu(C)\backslash t) \cup s) \succ_C \mu(C)$.

[16]Roth and Sotomayor (1989, Proposition 1) showed that this is the "correct" concept of stability in the sense that there is no blocking *coalition* if and only if there is no blocking pair. Also note that in the special case of $q_C = 1$ (for all $C \in \mathcal{C}$) the concept coincides with the previously introduced stability concept for marriage problems.

[17]Roth and Sotomayor (1989, Theorem 3) stated that for all stable matchings μ and μ' and all $C \in \mathcal{C}$, either $\mu(C) \succ_C \mu'(C)$, $\mu'(C) \succ_C \mu(C)$, or $\mu(C) = \mu'(C)$.

The proof is similar to that of Theorem 3. It is based on the two properties that were already key in the proofs of Theorems 2 and 3: (weak) decomposability and transitivity of blocking.

Lemma 3 (**Weak decomposability**, Roth and Sotomayor, 1990, Theorem 5.33)**.**
Let μ and μ' be stable matchings. Let $C \in \mathcal{C}$, $s \in S$, and $s \in \mu(C) \cup \mu'(C)$. Then,

(a) $\mu(C) \succ_C \mu'(C)$ *implies* $\mu'(s) \succeq_s \mu(s)$;

(b) $\mu(s) \succ_s \mu'(s)$ *implies* $\mu'(C) \succeq_C \mu(C)$.

Note that matchings μ and μ' play a symmetric role in Lemma 3.

Next we extend the transitivity of blocking property (see Footnote 9) used in the proofs of Theorems 2 and 3 to college admissions problems.

Lemma 4 (**Transitivity of blocking for college admissions**)**.**
Let μ and μ' be matchings, $C \in \mathcal{C}$, and $s \in S$. Suppose (s, C) blocks $(\mu(s), \mu(C))$. Suppose also that C is assigned groups of students $\mu(C)$ and $\mu'(C)$ under some stable *matchings.*[18] *If $\mu(s) \succeq_s \mu'(s)$ and $\mu(C) \succeq_C \mu'(C)$, then (s, C) also blocks $(\mu'(s), \mu'(C))$.*[19]

Proof. Suppose $\mu(s) \succeq_s \mu'(s)$ and $\mu(C) \succeq_C \mu'(C)$. We prove that (s, C) blocks $(\mu'(s), \mu'(C))$. Since $C \succ_s \mu(s) \succeq_s \mu'(s)$ we only have to check that

B1'. $\left[|\mu'(C)| < q_C \text{ and } s \succ_C \emptyset \right]$ or
B2'. $\left[\text{there exists } t' \in \mu'(C) \text{ such that } s \succ_C t' \right]$.

Since (s, C) blocks $(\mu(s), \mu(C))$, we have that blocking condition $B1$ or $B2$ is satisfied. Note that by Roth (1984, Theorem 9), $|\mu'(C)| = |\mu(C)|$.

Suppose $B1$ is satisfied. Then, $|\mu(C)| < q_C$ and $s \succ_C \emptyset$. So, $|\mu'(C)| = |\mu(C)| < q_C$. Hence, $B1'$ is satisfied.

Now suppose $B2$ is satisfied. Then, there exists $t \in \mu(C)$ such that $s \succ_C t$. If $t \in \mu'(C)$, then $B2'$ holds trivially with $t' = t$. If $|\mu'(C)| < q_C$, then by the Rural Hospital Theorem (Roth, 1986), $\mu'(C) = \mu(C)$, $t \in \mu'(C)$, and therefore $B2'$. So, suppose $|\mu'(C)| = q_C$ and $t \notin \mu'(C)$. Hence, $|\mu(C)| = |\mu'(C)| = q_C$ and $\mu'(C) \neq \mu(C)$. But then, by Roth and Sotomayor (1989, Theorem 3) and $\mu(C) \succeq_C \mu'(C)$, $\mu(C) \succ_C \mu'(C)$. Now Roth and Sotomayor (1989, Theorem 4) and $|\mu(C)| = |\mu'(C)| = q_C$ imply that there is a student $t' \in \mu'(C) \backslash \mu(C)$ with $t \succ_C t'$. So, $s \succ_C t \succ_C t'$ and $t' \in \mu'(C)$. Hence, $B2'$ is satisfied. □

We are now ready to prove Theorem 4. In addition to the fact that (weak) decomposability and transitivity of blocking (Lemmas 3 and 4) are not elementary results as in the case of roommate (marriage) problems, the proof also has to be adapted because now we can only use *weak* decomposability.

---

[18]That is, there are stable matchings $\bar{\mu}$ and $\bar{\mu}'$ such that $\bar{\mu}(C) = \mu(C)$ and $\bar{\mu}'(C) = \mu'(C)$.

[19]By "(s, C) also blocks $(\mu'(s), \mu'(C))$" we mean that $C \succ_s \mu'(s)$ and *B1'* $\left[|\mu'(C)| < q_C \text{ and } s \succ_C \emptyset \right]$ or *B2'* $\left[\text{there exists } t' \in \mu'(C) \text{ such that } s \succ_C t' \right]$.

Proof of Theorem 4. Define $\Sigma := \{\mu_1, \ldots, \mu_k\}$. Let $l \in \{1, \ldots, k\}$. First, we show that α_l is a well-defined matching, i.e., α_l^S and $\alpha_{k-l+1}^{\mathcal{C}}$ together constitute a well-defined matching.

Let $s \in S$ with $\alpha_l^S(s) = C \in \mathcal{C}$. We have to prove that $s \in \alpha_{k-l+1}^{\mathcal{C}}(C)$. Recall that stable matchings $\mu_1^s, \ldots, \mu_k^s$ are such that $\{\mu_1^s, \ldots, \mu_k^s\} = \Sigma$ and $\mu_1^s(s) \succeq_s \mu_2^s(s) \succeq_s \cdots \succeq_s \mu_l^s(s) \succeq_s \cdots \succeq_s \mu_{k-1}^s(s) \succeq_s \mu_k^s(s)$. Hence, $\mu_l^s(s) = \alpha_l^S(s) = C$ and $s \in \mu_l^s(C)$. Let $\Sigma_s^l := \{\mu^s \in \Sigma \mid \mu^s(s) \sim_s \mu_l^s(s)\} = \{\mu^s \in \Sigma \mid \mu^s(s) = \mu_l^s(s)\}$. Note that $\Sigma_s^l = \{\mu_p^s, \mu_{p+1}^s, \ldots, \mu_l^s, \ldots, \mu_{q-1}^s, \mu_q^s\}$ for some $p \leq l \leq q$. W.l.o.g. we may assume that

$$\mu_q^s(C) \succeq_C \cdots \succeq_C \mu_{l+1}^s(C) \succeq_C \mu_l^s(C) \succeq_C \mu_{l-1}^s(C) \succeq_C \cdots \succeq_C \mu_p^s(C). \tag{2}$$

By weak decomposability (Lemma 3, b) and (2),

$$\mu_1^s \succeq_s \cdots \succeq_s \mu_{l-1}^s \succeq_s \quad \mu_l^s \quad \succeq_s \mu_{l+1}^s \succeq_s \cdots \succeq_s \mu_k^s$$
$$\{\mu_k^s, \ldots, \mu_{l+1}^s\} \succeq_C \quad \mu_l^s \quad \succeq_C \{\mu_{l-1}^s, \ldots, \mu_1^s\}.^{20}$$

In particular, $\mu_l^s(C)$ is the $(k-l+1)$-st ranked match for college C and therefore $s \in \mu_l^s(C) = \alpha_{k-l+1}^{\mathcal{C}}(C)$.

Let $C \in \mathcal{C}$. Let $s \in \alpha_{k-l+1}^{\mathcal{C}}(C)$. We have to prove that $\alpha_l^S(s) = C$. Recall that stable matchings $\mu_1^C, \ldots, \mu_k^C$ are such that $\{\mu_1^C, \ldots, \mu_k^C\} = \Sigma$ and $\mu_1^C(C) \succeq_C \mu_2^C(C) \succeq_C \cdots \succeq_C \mu_{k-l+1}^C(C) \succeq_C \cdots \succeq_C \mu_{k-1}^C(C) \succeq_C \mu_k^C(C)$. Hence, $\mu_{k-l+1}^C(C) = \alpha_{k-l+1}^{\mathcal{C}}(C) \ni s$ and $\mu_{k-l+1}^C(s) = C$. Let $\Sigma_C^{k-l+1} := \{\mu^C \in \Sigma \mid \mu^C(C) \sim_C \mu_{k-l+1}^C(C)\}$. Note that $\Sigma_C^{k-l+1} = \{\mu_x^C, \mu_{x+1}^C, \ldots, \mu_{k-l+1}^C, \ldots, \mu_{y-1}^C, \mu_y^C\}$ for some $x \leq k-l+1 \leq y$. By Roth and Sotomayor (1989, Theorem 3), $\Sigma_C^{k-l+1} = \{\mu^C \in \Sigma \mid \mu^C(C) = \mu_{k-l+1}^C(C)\}$. Hence, $s \in \mu^C(C)$ for all $\mu^C \in \Sigma_C^{k-l+1}$. So,

$$\mu_y^C(s) \sim_s \cdots \sim_s \mu_{k-l+2}^C(s) \sim_s \mu_{k-l+1}^C(s) \sim_s \mu_{k-l}^C(s) \sim_s \cdots \sim_s \mu_x^C(s). \tag{3}$$

By weak decomposability (Lemma 3, a) and (3),

$$\mu_1^C \succeq_C \cdots \succeq_C \mu_{k-l}^C \succeq_C \quad \mu_{k-l+1}^C \quad \succeq_C \mu_{k-l+2}^C \succeq_C \cdots \succeq_C \mu_k^C$$
$$\{\mu_k^C, \ldots, \mu_{k-l+2}^C\} \succeq_s \quad \mu_{k-l+1}^C \quad \succeq_s \{\mu_{k-l}^C, \ldots, \mu_1^C\}.$$

In particular, $\mu_{k-l+1}^C(s)$ is the l-th ranked match for student s and therefore $C = \mu_{k-l+1}^C(s) = \alpha_l^S(s)$.

Hence, we have proven that α_l^S and $\alpha_{k-l+1}^{\mathcal{C}}$ together constitute a well-defined matching. Thus, α_l is a well-defined matching.

We now prove that α_l is stable. By definition, α_l is individually rational. Suppose there is a pair (s, C) that blocks $(\alpha_l(s), \alpha_l(C))$. Note that $\alpha_l(s) \succeq_s \mu(s)$ for at least $k-l+1$ matchings μ in $\Sigma = \{\mu_1, \ldots, \mu_k\}$. Similarly, $\alpha_l(C) \succeq_C \mu(C)$ for at least l matchings μ in Σ. Since Σ contains only $k < (k-l+1)+l$ matchings, for at least one matching $\mu' \in \Sigma$, both $\alpha_l(s) \succeq_s \mu'(s)$ and $\alpha_l(C) \succeq_C \mu'(C)$.

By construction there exist stable matchings under which C is assigned $\alpha_l(C)$ and $\mu'(C)$. By transitivity of blocking for college admissions (Lemma 4), (s, C) blocks $(\mu'(s), \mu'(C))$. This however contradicts stability of μ' (the set Σ only contains stable matchings). Therefore, α_l is stable. $\square$

4   Conclusion

For three different matching models (roommate, marriage, and college admissions) we have shown that certain compromise stable matchings exist in the form of so-called (generalized) median matchings. For roommate problems, we prove the existence and stability of the median matching of an odd number of stable matchings (Theorem 2). For the two-sided (marriage and college admissions) matching problems we prove the existence and stability of generalized median matchings for any number of stable matchings (Theorems 3 and 4). In all proofs we use a decomposability property to show that (generalized) median matchings are well-defined and a transitivity of blocking property to show that they are stable. The following examples demonstrate that our results can break down if decomposability or transitivity of blocking is violated.

The following example shows how a lack of decomposability causes the median matching to not be well-defined. The example also demonstrates that the "lone wolf" property is not sufficient for median matchings to be well-defined. It is well-known that for college admissions problems with responsive preferences the set of stable matchings is still non-empty and has the "lone wolf" property, i.e., the set of single students does not vary from one stable matching to another (see, for instance, Roth and Sotomayor, 1990, Lemma 5.6 and Theorem 5.12). Martínez et al. (2000) introduced the domain of q-separable and substitutable preferences for colleges that contains the domain of responsive preferences. They prove that even for this larger domain the set of stable matchings is non-empty and that the "lone wolf" property holds.

Example 3. *No compromise for q-separable and substitutable preferences*
Consider Martínez et al.'s (2000, Example 2) college admissions problem with four students s_1, s_2, s_3, s_4, two colleges C_1 and C_2 with two seats each, and preferences as listed in the table below. The colleges' preferences are q-separable and substitutable (for details see Martínez et al., 2000).

$\succ_{C_1}$	$\succ_{C_2}$	$\succ_{s_1}$	$\succ_{s_2}$	$\succ_{s_3}$	$\succ_{s_4}$
$\{s_1, s_2\}$	$\{s_3, s_4\}$	C_2	C_2	C_1	C_1
$\{s_1, s_3\}$	$\{s_2, s_4\}$	C_1	C_1	C_2	C_2
$\{s_2, s_4\}$	$\{s_1, s_3\}$				
$\{s_3, s_4\}$	$\{s_1, s_2\}$				
$\{s_1, s_4\}$	$\{s_1, s_4\}$				
$\{s_2, s_3\}$	$\{s_2, s_3\}$				
$\{s_1\}$	$\{s_1\}$				
$\{s_2\}$	$\{s_2\}$				
$\{s_3\}$	$\{s_3\}$				
$\{s_4\}$	$\{s_4\}$				

There are four stable matchings:

$$\mu_1 = \{\{C_1, s_1, s_2\}, \{C_2, s_3, s_4\}\}$$
$$\mu_2 = \{\{C_1, s_1, s_3\}, \{C_2, s_2, s_4\}\}$$
$$\mu_3 = \{\{C_1, s_2, s_4\}, \{C_2, s_1, s_3\}\}$$
$$\mu_4 = \{\{C_1, s_3, s_4\}, \{C_2, s_1, s_2\}\}$$

Considering the first three matchings, one straightforwardly checks that matching each agent with its median match is not a matching: C_1 would be matched with $\{s_1, s_3\}$ but at the same time s_3 would be matched with C_2. Hence, the median matching is not well-defined.

We have the following violation of weak decomposability. Consider s_3 and C_1. Then, $s_3 \in \mu_2(C_1)$, $\mu_2(s_3) \succ_{s_3} \mu_3(s_3)$, and $\mu_2(C_1) \succ_{C_1} \mu_3(C_1)$; a violation of weak decomposability (Lemma 3). $\diamond$

The following example demonstrates that in a more general setting a median matching may exist but fail to be stable.

Example 4. *An unstable compromise for a network formation problem*
Consider the network formation problem (see Jackson and Watts, 2002) with three agents labeled 1,2, and 3, who have preferences over all possible networks as listed in the table below. Loosely speaking, agents like to be connected but the complete network is too "costly."

$\succ_1$	$\succ_2$	$\succ_3$
1—2—3	2—3—1	3—1—2
3—1—2	1—2—3	2—3—1
2—3—1	3—1—2	1—2—3
triangle (1,2,3)	triangle (1,2,3)	triangle (1,2,3)
1—2 3	2—3 1	1—3 2
1—3 2	1—2 3	2—3 1
2—3 1	1—3 2	1—2 3
1 2 3	1 2 3	1 2 3

19

We extend the notion of blocking for the matching problems considered so far to this network formation problem in a natural way as follows (we use Jackson and Watts's, 2002, definition of pairwise blocking). Two agents can block a given network by adding a link if and only if this is beneficial for both agents. Furthermore, a single agent can block a given network by destroying a link if that is beneficial for him.

There are three stable networks μ_1, μ_2, and μ_3 which are given by

$$\underset{1}{\bullet}\!\!-\!\!\underset{2}{\bullet}\!\!-\!\!\underset{3}{\bullet} \; , \quad \underset{3}{\bullet}\!\!-\!\!\underset{1}{\bullet}\!\!-\!\!\underset{2}{\bullet} \; , \text{ and } \quad \underset{2}{\bullet}\!\!-\!\!\underset{3}{\bullet}\!\!-\!\!\underset{1}{\bullet} \; ,$$

respectively. Let each agent choose the median of the three sets of links with which he can be associated. Then, each agent chooses to connect with both of the other agents. Hence, the resulting median network is the (well-defined!) complete network
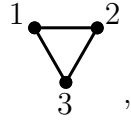
which is not stable. Hence, the median network is well-defined but not stable.

We now demonstrate that transitivity of blocking is violated. To do so we first extend the definition to the network formation model. Note that agents now care about the structure of the whole network and not only about their direct links. Thus, in contrast to all other models we considered so far this is a model with externalities. For any network μ and any agent k, we denote by $\mu(k)$ agent k's links at network μ.

Transitivity of blocking for network formation

Let μ and μ' be networks and i, j agents such that $\{i, j\}$ (possibly $i = j$) blocks network μ. Suppose also that for all $k = i, j$, agent k is assigned the set of links $\mu(k)$ and $\mu'(k)$ under some *stable* network.[21] If $\mu \succeq_i \mu'$ and $\mu \succeq_j \mu'$, then $\{i, j\}$ also blocks μ'.

To see that transitivity of blocking is violated, consider

$$\mu = \underset{3}{\bigvee\limits^{1 \quad 2}} \quad \text{and } \mu' = \underset{1}{\bullet}\!\!-\!\!\underset{3}{\bullet} \quad \underset{2}{\bullet} \; .$$

Note that $\mu_2 = \underset{3}{\bullet}\!\!-\!\!\underset{1}{\bullet}\!\!-\!\!\underset{2}{\bullet}$ and $\mu_3 = \underset{2}{\bullet}\!\!-\!\!\underset{3}{\bullet}\!\!-\!\!\underset{1}{\bullet}$ are stable networks with $\mu_2(1) = \mu(1)$ and $\mu_3(1) = \mu'(1)$. Agent 1 blocks network μ by breaking either the link with agent 2 or the link with agent 3. Furthermore, $\mu \succeq_1 \mu'$. However, in contradiction to transitivity of blocking, agent 1 cannot block network μ'.[22]   ◇

---

[21]That is, there are stable networks $\bar{\mu}$ and $\bar{\mu}'$ such that $\bar{\mu}(k) = \mu(k)$ and $\bar{\mu}'(k) = \mu'(k)$.

[22]Note that weak decomposability is violated as well: $\mu_2 \succ_1 \mu_3$ and agent 1 is linked with agent 3 (at μ_2), but nevertheless $\mu_2 \succ_3 \mu_3$. So far, we did not succeed in constructing an example where the median outcome is well-defined but unstable, (weak) decomposability is satisfied, and transitivity of blocking is violated.

References

Chung, K.-S. (2000): "On the Existence of Stable Roommate Matchings." *Games and Economic Behavior*, 33: 206–230.

Diamantoudi, E., Miyagawa, E., and Xue, L. (2004): "Random Paths to Stability in the Roommate Problem." *Games and Economic Behavior*, 48: 18–28.

Fleiner, T. (2002): "Some Results on Stable Matchings and Fixed Points." Technical Report TR-2002-08, Egerváry Research Group, Budapest. http://www.cs.elte.hu/egres.

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

Gusfield, D. and Irving, R. W. (1989): *The Stable Marriage Problem: Structure and Algorithms.* The MIT Press, Cambridge.

Jackson, M. O. and Watts, A. (2002): "The Evolution of Social and Economic Networks." *Journal of Economic Theory*, 106: 265–295.

Klaus, B. and Klijn, F. (2006): "Median Stable Matching for College Admissions." *International Journal of Game Theory*, 34: 1–11.

Knuth, D. E. (1976): *Marriages Stables.* Les Presses de l'Université de Montreal, Montreal.

Martínez, R., Massó, J., Neme, A., and Oviedo, J. (2000): "Single Agents and the Set of Many-to-One Stable Matchings." *Journal of Economic Theory*, 91: 91–105.

McVitie, D. G. and Wilson, L. B. (1970): "Stable Marriage Assignments for Unequal Sets." *BIT*, 10: 295–309.

Rawls, J. (1971): *A Theory of Justice.* Harvard University Press, Cambridge.

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. (1985): "The College Admissions Problem is not Equivalent to the Marriage Problem." *Journal of Economic Theory*, 36: 277–288.

Roth, A. E. (1986): "On the Allocation of Residents to Rural Hospitals: A General Property of Two-Sided Matching Markets." *Econometrica*, 54: 425–428.

Roth, A. E. and Sotomayor, M. A. O. (1989): "The College Admissions Problem Revisited." *Econometrica*, 57: 559–570.

Roth, A. E. and Sotomayor, M. A. O. (1990): *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis.* Cambridge University Press, Cambridge.

Sethuraman, J. and Teo, C.-P. (2001): "A Polynomial-Time Algorithm for the Bistable Roommates Problem." *Journal of Computer and Systems Sciences*, 63: 486–497.

Sethuraman, J., Teo, C.-P., and Qian, L. (2006): "Many-to-One Stable Matching: Geometry and Fairness." *Mathematics of Operations Research*, 31: 581–596.

Smith, A. (1796): *An Inquiry into the Nature and Causes of the Wealth of Nations.* Oxford University Press, Oxford.

Tan, J. (1991): "A Necessary and Sufficient Condition for the Existence of a Complete Stable Matching." *Journal of Algorithms*, 12: 154–178.

Teo, C.-P. and Sethuraman, J. (1998): "The Geometry of Fractional Stable Matchings and its Applications." *Mathematics of Operations Research*, 23: 874–891.

Yenmez, M. B. and Schwarz, M. (2008): "Median Stable Matching." Mimeo. http://ssrn.com/abstract=1031277.