The Exchange-Stable Marriage Problem*

Katarína Cechlárová[1] and David F. Manlove[2,†]

[1] *Institute of Mathematics, P.J. Šafárik University, Faculty of Science, Jesenná 5, 041 54 Košice, Slovakia. Email:* `cechlarova@science.upjs.sk`.

[2] *Department of Computing Science, University of Glasgow, Glasgow G12 8QQ, UK. Email:* `davidm@dcs.gla.ac.uk`.

Abstract

In this paper we consider instances of stable matching problems, namely the classical Stable Marriage (SM) and Stable Roommates (SR) problems and their variants. In such instances we consider a stability criterion that has recently been proposed, that of *exchange-stability*. In particular, we prove that ESM – the problem of deciding, given an SM instance, whether an exchange-stable matching exists – is NP-complete. This result is in marked constrast with Gale and Shapley's classical linear-time algorithm for finding a stable matching in an instance of SM. We also extend the result for ESM to the SR case. Finally, we study some variants of ESM under weaker forms of exchange-stability, presenting both polynomial-time solvability and NP-completeness results for the corresponding existence questions.

Keywords: Stable Marriage problem; Stable Roommates problem; Matching; Coalition-exchange-stable; Man-exchange-stable

Mathematics Subject Classification: 68Q25, 91A06, 91B68

1 Introduction

In their seminal paper [5], Gale and Shapley introduced the well-known Stable Marriage problem (SM). An instance I of SM involves a set U containing n men and a set W containing n women. Each person $q \in U \cup W$ has a *preference list* $P(q)$ in which he/she ranks all the members of the opposite sex in strict order of preference. A *matching* M in I is a one-one correspondence between the men and women. If $(m, w) \in M$ for some man m and woman w, then m is the *mate* of w, denoted by $M(w)$, and vice versa. We say that a (man,woman) pair (m, w) is a *blocking pair* of M if each of m and w prefers the other to his/her mate in M. A matching that admits no blocking pair is said to be *stable*. Gale and Shapley showed that every instance of SM admits at least one stable matching, and gave an $O(n^2)$ algorithm for finding such a matching.

The Stable Marriage problem has been studied extensively [5, 14, 7, 20] and has a number of important practical applications. Probably the largest and best-known of these is concerned with the annual assignment of graduating medical students, or *residents*, to their first hospital appointments [18, 11]. In a number of countries (for example the US, Canada and Scotland), an automated matching scheme constructs stable matchings of

*Both authors were supported by Slovak Agency for Science contract "Combinatorial Structures and Complexity of Algorithms", and Engineering and Physical Sciences Research Council grant GR/R84597/01.

†Supported by Nuffield Foundation award NUF-NAL-02 and Royal Society of Edinburgh/Scottish Executive Personal Research Fellowship.

residents to hospitals based on the preferences of residents over hospitals and vice versa, using extensions of the Gale/Shapley algorithm for SM [7, Section 1.6].

A non-bipartite generalisation of SM that has also received much attention in the literature is the Stable Roommates problem (SR) [5, 10], [7, Chapter 4]. An instance I of SR involves $2n$ participants, and each participant ranks the $2n-1$ others in strict order of preference. In this case a matching M is a set of n unordered pairs of participants in I. If $\{p, q\} \in M$ then p is the *mate* of q, denoted by $M(q)$. A *blocking pair* of M is a pair of participants $\{p, q\}$ such that each of p and q prefers the other to his/her mate in M. M is *stable* if it admits no blocking pair. SR was first studied by Gale and Shapley [5] who showed that an instance need not admit a stable matching. Irving [10] formulated a linear-time algorithm that finds a stable matching, or reports that none exists, given an instance of SR. As the SR problem name suggests, an application is in assigning $2n$ students to share n two-bed rooms, based on their preferences over one another.

Recently, an alternative notion of stability, so-called *exchange-stability*, has been introduced by Alcalde in the context of SR (using the terminology *ξ-stability*) [3]. Given a matching M in an instance I of SR, a sequence of participants $\langle p_0, p_1, \ldots, p_{r-1}\rangle$, for some $r \geq 2$, is an *exchange-blocking coalition* if p_i prefers $M(p_{i+1})$ to $M(p_i)$, where $0 \leq i \leq r-1$ (throughout this paper, all subscripts are taken modulo r when reasoning about exchange-blocking coalitions). M is *coalition-exchange-stable* if M admits no exchange-blocking coalition. The special case of an exchange-blocking coalition with $r = 2$ is also defined as an *exchange-blocking pair* and is also denoted as an unordered pair $\{p_0, p_1\}$. Following terminology in [4], M is *exchange-stable* if M admits no exchange-blocking pair. It follows immediately that a coalition-exchange-stable matching is exchange-stable.

Alcalde showed that exchange-stability and classical stability are independent notions, i.e. neither criterion implies the other. Indeed, he constructed an instance I of SR that admits a stable matching but no exchange-stable matching, and an instance J of SR that admits an exchange-stable matching but no stable matching. (These instances also suffice to illustrate the same properties when 'exchange' is replaced by 'coalition-exchange' in the preceding two sentences.) Alcalde argued that, in situations when participants have "property rights", exchange-stability could be more appropriate than classical stability. For example, in the context of assigning $2n$ students to n two-bed rooms, an individual's property would be the bed that he/she occupies. A blocking pair $\{p, q\}$ in the classical sense might not lead to any disruption of the matching in practice, since there is no extra room for p and q to occupy, and moreover each of the mates of p and q in the matching could exercise their property rights by refusing to give up their bed in order to make a room available. However an exchange-blocking pair $\{p, q\}$ would in practice lead p and q to simply swap beds.

The concept of an exchange-blocking coalition may also be defined in the SM case (the members of such a coalition are necessarily of the same sex) and hence the definition of an exchange-blocking pair also follows in this context.

In practice it is more likely that exchange-blocking pairs would arise, relative to a given matching, due to the greater difficulty that the members of a large exchange-blocking coalition may have in determining its existence. Hence in many situations it may suffice to find a matching that is exchange-stable. In the majority of this paper we consider the complexity of finding exchange-stable matchings in various problem contexts.

Further motivation for considering exchange-stability arises from a concrete practical application where in reality an exchange-blocking pair has arisen [12]. In a previous run of the *Scottish PRHO Allocation scheme* (SPA), the matching scheme for allocating graduating medical students to hospital posts in Scotland [11], two participating students discovered that, if they could have exchanged their assigned hospitals with each other, then they would each have ended up with a more favourable assignment. Naturally the hospitals to which the students were matched would not have permitted the exchange (for if they were to have agreed, it would have implied that the original matching contained a blocking pair with respect to classical stability, whereas the primary consideration of SPA is to produce a matching that is stable in the classical sense). Nevertheless, such a situation can lead participants to lose confidence in the optimality criterion involved in the matching scheme.

It is known that the problem of deciding whether an SRII instance I admits an exchange-stable matching is NP-complete [4], where SRII is the variant of SR in which the preference lists in I are permitted to be *incomplete* and *inconsistent*. Incomplete preference lists arise when a given participant need not find all other participants acceptable. If a participant p finds a participant q unacceptable, then q does not appear in p's preference list. Also, inconsistent preference lists arise when there are two participants p and q such that p finds q acceptable and q finds p unacceptable. In this context, a *matching* must not only satisfy the property that each participant p has a unique participant q as his/her mate, but moreover each of p and q must find the other acceptable. The definition of exchange-stability is analogous to that for the complete lists case (however, it should be noted that the definition of an exchange-blocking pair $\{p, q\}$ as given in [4] does not assume that p and q are acceptable to $M(q)$ and $M(p)$ respectively).

We now turn to the case of complete preference lists in SM and SR. Let ESM (respectively ESR) denote the problem of deciding, given an instance I of SM (respectively SR), whether I admits an exchange-stable matching. Cechlárová [4] gave an example instance of SM involving two men and two women that does not admit a (coalition-)exchange-stable matching. For completeness, we repeat the example here, representing a participant q's preference list $P(q)$ in order, starting with the most-preferred mate:

$$\begin{array}{ll} P(m_1) = w_1, w_2 & \qquad P(w_1) = m_2, m_1 \\ P(m_2) = w_2, w_1 & \qquad P(w_2) = m_1, m_2 \end{array}$$

It is easy to see that the matching $\{(m_1, w_1), (m_2, w_2)\}$ admits the exchange-blocking pair $\{w_1, w_2\}$, whilst the matching $\{(m_1, w_2), (m_2, w_1)\}$ admits the exchange-blocking pair $\{m_1, m_2\}$. Cechlárová [4] left as open problems the complexities of both ESM and ESR.

In this paper we resolve the complexities of both of these problems. Firstly, we demonstrate that, somewhat surprisingly, ESM is NP-complete. This result, proved in Section 2, is an interesting departure from what is usually regarded by the community as the "expected" algorithmic behaviour of stable matching problems in general, i.e. polynomial-time solvability. The NP-completeness of ESR then follows as a corollary, since as we shall demonstrate in Section 3, ESM is a special case of ESR. Note that in the literature, there are very few examples of stable matching problems that are NP-complete for strictly ordered and complete preference lists. To the best of our knowledge, apart from the results in this paper, the only examples of such problems are the Hospitals / Residents problem with couples [16], the 3-Person Stable Assignment problem and the 3-Gender Stable Marriage problem [15].

We also consider in Section 4 a weaker form of exchange-stability, given instances of SM and its variants, in which exchange-blocking coalitions are permitted to contain only men (or analogously, only women). Define a matching in an SM instance I to be *man(woman)-coalition-exchange-stable* if there is no exchange-blocking coalition involving only men (women). As before we may define a *man(woman)-exchange-stable* matching. As in the SR case, one can extend these definitions to the variant of SM in which preference lists may be incomplete, which we denote by SMI. The two cases of SMI in which preference lists are consistent and inconsistent are denoted by SMIC and SMII respectively. We study man-(coalition-)exchange-stability in these variants of SM, which may be grouped according to two cases: (i) the preferences of the women are essentially irrelevant (this case includes SM and SMIC), and (ii) the preferences of the women are relevant insofar as they are permitted to exclude potential mates in a matching (this case includes SMII). For Case (i), firstly we show that an instance of SM admits at least one man-coalition-exchange-stable matching, and such a matching may be found in linear time. Secondly, if we are given an instance I of SMIC, we show that the problem of finding a man-coalition-exchange-stable matching in I, or reporting that none exists, is solvable in $O(\sqrt{n}L)$ time, where L is the total length of the preference lists in I. These observations exploit a strong connection between man-coalition-exchange-stable matchings and *Pareto optimal* matchings in instances of house allocation problems [9, 1, 2] and housing markets [21, 19, 17]. By contrast, for Case (ii) we show that problem of deciding whether a man-exchange-stable matching exists, given an instance of SMII, is NP-complete.

Earlier in this section we outlined a number of practical applications involving the computation of stable matchings. We also gave motivation for finding matchings that avoid exchange-blocking pairs and exchange-blocking coalitions in such situations. Hence it is natural to consider the problem of finding a stable matching that is (coalition-)exchange-stable – we consider this problem in Section 5.

Finally in Section 6, we present some concluding remarks.

2 ESM is NP-complete

In this section we establish the NP-completeness of ESM. To justify our construction we shall need the following modification of a result of Alcalde [3].

Lemma 2.1. *Let I be an instance of SM in which U and W are the sets of men and women respectively. Suppose that the men and women can be labelled $m_1, m_2, \ldots, m_n$ and $w_1, w_2, \ldots, w_n$ respectively, in such a way that each $m \in U$ has the same preference list $P(m) = w_1, w_2, \ldots, w_n$, and each $w \in W$ has the same preference list $P(w) = m_1, m_2, \ldots, m_n$. Then any matching in I is coalition-exchange-stable.*

Proof. Let M be a matching in I. Suppose that $(m_i, w_k) \in M$ and $(m_j, w_l) \in M$, and without loss of generality assume that $i < j$. Then w_k prefers $M(w_k)$ to $M(w_l)$. Now assume that $k < l$ (the argument is similar if $l < k$). Then m_i prefers $M(m_i)$ to $M(m_j)$. Hence M admits no exchange-blocking coalition. □

Hence, for each n, there is an SM instance with n men and n women that admits $n!$ coalition-exchange-stable matchings. (We remark as an aside that, by constrast, it is easy to show that an SM instance with n men and n women could never admit as many stable matchings, for each $n > 2$.)

ESM clearly belongs to the class NP, since when an exchange-stable matching is given, checking all pairs of men and all pairs of women verifies this property. We now give a polynomial transformation to ESM from the following problem:

> **Restricted 3-satisfiability (R3SAT).**
> *Instance:* A Boolean formula B in Conjunctive Normal Form, with each clause containing at most three literals and each variable x occurring at most twice as literal x and at most twice as literal $\overline{x}$.
> *Question:* Is there a satisying truth assignment for B?

The NP-completeness of R3SAT was proved in [7, p.210]; see also [6, p.259]. We also suppose, without loss of generality, that no clause contains both a literal and its negation, since such a clause would be satisfied in any truth assignment.

Let B be a Boolean formula given as an instance of R3SAT, containing n variables $x_1, x_2, \ldots, x_n$ and m clauses $C_1, C_2, \ldots, C_m$. We construct an ESM instance I.

In I, we shall have 5 types of participants, distinguished by letters representing them: x- and g-participants are female, whilst u-, y- and h-participants are male. There are $4n$ of each of the x-, g- and y-participants. The number of u-participants is m, and the number of h-participants is $4n - m$ (we remark that $4n > 3n \geq m$).

To each clause C_i one u-participant is assigned, denoted by u_i. For each variable x_j there will be 12 *variable* participants, forming a *family.* Participants x_j^1, y_j^1 and g_j^1 correspond to the first occurence of literal x_j in the formula, x_j^2, y_j^2 and g_j^2 to its second; $\overline{x}_j^1$, $\overline{y}_j^1$ and $\overline{g}_j^1$ correspond to the first occurrence of literal $\overline{x}_j$ and participants $\overline{x}_j^2$, $\overline{y}_j^2$ and $\overline{g}_j^2$ to its second. They are defined even if the corresponding literals are not actually present in the formula. For simplicity, we shall adopt the following notation: the u-participant corresponding to the clause containing the literal corresponding to a particular x-participant will be denoted by $u(x)$, and conversely, *any* x-participant corresponding to a literal contained in the clause corresponding to a particular u-participant will be denoted by $x(u)$. (If a given literal appears in the formula only once or not at all, the function $u(x)$ for the missing x-participant is not defined.) Furthermore, two x- and y-participants with the same set of indices and bars will be denoted by $x(y)$ and $y(x)$ respectively (they will be called *twins*); similarly for the two y- and g-participants with the same set of indices and bars we shall use the notation $y(g)$ and $g(y)$ respectively and call them each other's *son* and *mother* respectively.

Also, the 6 participants without bars from a family (i.e. $x_j^1, x_j^2, y_j^1, y_j^2, g_j^1, g_j^2$) will be said to form its *true side* and the 6 participants with the bar its *false side.* Moreover, for a y-participant, $\overline{x}^1(y)$ and $\overline{x}^2(y)$ are the two x-participants from the other side of his family.

We shall suppose that for each type of participant there exists a linear ordering that enables us to list them as $g_1, \ldots, g_{4n}$, $h_1, \ldots, h_{4n-m}$, $u_1, \ldots, u_m$, $x_1, \ldots, x_{4n}$ and $y_1, \ldots, y_{4n}$. In what follows, each participant p has a "proper" part of his/her preference list, the rightmost extent of which is indicated by a vertical line in p's preference list. If Q denotes the group of participants listed in p's preference list to the right of this vertical line, it should be noted that implicitly we omit from Q any participants who have already appeared in the proper part of p's list.

Now we are ready to define the preference lists in I. The preference lists of u-participants are as follows:

$$P(u_i) = t_i^1, w_i^1, t_i^2, w_i^2, t_i^3, w_i^3, z_i^1, z_i^2, z_i^3 \mid g_1, \ldots, g_{4n}, x_1, \ldots, x_{4n}$$

where the z-participants denote x-participants representing the first, second and third literal of clause C_i. Participants t_i^k and w_i^k are the two mothers from the other side of the family of z_i^k.

For example, suppose that the second clause C_2 in the formula has the form

$$C_2 = x_3 \vee \overline{x}_2 \vee x_1,$$

where for the literals $\overline{x}_2, x_1$ this is their first ocurrence and for x_3 its second, then

$$P(u_2) = \overline{g}_3^1, \overline{g}_3^2, g_2^1, g_2^2, \overline{g}_1^1, \overline{g}_1^2, x_3^2, \overline{x}_2^1, x_1^1 \mid g_1, \ldots, g_{4n}, x_1, \ldots, x_{4n}.$$

If the clause C_i contains only two literals, then z_i^3, t_i^3 and w_i^3 are simply not defined and they will not appear in the preference list of u_i.

Preferences of other types of participants are as follows. (Note that, as mentioned above, if a given literal appears in the formula only once or not at all, the function $u(x)$

for the missing x-participant is not defined and the corresponding entry does not appear in her preference list.)

$$\begin{array}{rcl} P(y) & = & x(y), \overline{x}^1(y), \overline{x}^2(y), g(y) \mid x_1, \ldots, x_{4n}, g_1, \ldots, g_{4n} \\ P(h) & = & g_1, \ldots, g_{4n} \mid x_1, \ldots, x_{4n} \\ P(x) & = & u(x), y(x) \mid h_1, \ldots, h_{4n-m}, y_1, \ldots, y_{4n}, u_1, \ldots, u_m \\ P(g) & = & y(g), h_1, \ldots, h_{4n-m} \mid y_1, \ldots, y_{4n}, u_1, \ldots, u_m \end{array}$$

Suppose that there exists a satisfying truth assigment f for the given formula B. We shall define on its basis a matching M of the set of participants and then show that it is exchange-stable in I. First, match each clause participant u_i with the x-participant corresponding to the first true literal in clause C_i under f. It can happen, that in one family, either both x-participants from one side will be matched with the u-participants, or only one of them, or none. In any case, match the remaining x-participants to their twins. The remaining y-participants will be matched to their mothers.

There remains to be matched a subset G of the set of g-participants of cardinality $4n - m$. If we apply to G the ordering induced by the original ordering of g-participants, we can use Lemma 2.1 to obtain an exchange-stable matching of G and the set of all h-participants.

Lemma 2.2. *The matching M is exchange-stable in I.*

Proof. We shall successively consider all types of participants.

Woman x – she is matched either to $u(x)$ (who is her first choice) or to $y(x)$, to whom she prefers only $u(x)$. But in the latter case, $u(x)$ is matched to some x' and is her first choice, so x' will not take part in any exchange-blocking pair.

Woman g – if she is matched to her son $y(g)$, no exchange-blocking pair involving g is possible. Otherwise she is matched to some h_k, to whom she prefers:

- $y(g)$ – denote him by y'. But then y' is matched to $x(y')$, and $x(y')$ prefers y' to h_k.
- Another h_l. But h-participants are matched to g-participants according to some exchange-stable matching by Lemma 2.1, so no exchange-blocking pair involving g occurs.

Man y – he is matched either to his twin (in that case there is no possibility for y to improve) or to his mother $g(y)$. Let us now consider the participants preferred by y to $g(y)$.

- Twin x. But x must be matched to $u(x)$, and the u-participants prefer to those x only the g-participants who correspond to negations of literals in the corresponding clause. However, we supposed that no clause contains a literal and its negation, so $u(x)$ will not co-operate with y in an exchange-blocking pair.
- An x' from the other side of his family. But x' must be matched to her twin $y(x')$, and since x' is the first choice of $y(x')$, no exchange-blocking pair involving x will arise.

Man h – he is matched to some woman g and can prefer to his mate only another woman g'. If g' is matched to some h', then no exchange-blocking pair involving h can occur by Lemma 2.1. If g' is matched to $y(g')$ then no exchange-blocking pair involving $y(g')$ can arise, since $y(g')$ prefers g' to g.

Man u – since no other types of participants can take part in any exchange-blocking pair, we only need to consider pairs comprising two u-participants. But these could exchange only x-participants corresponding to occurrences of literals contained in the respective clauses. As the sets of those literal occurrences are mutually disjoint, no exchange-blocking pair involving u-participants is possible. ■

Conversely, suppose that M is an exchange-stable matching in I. We shall prove that on the basis of M, it is possible to define a satisfying truth assigment for B. For that we firstly need to derive some properties of exchange-stable matchings for the defined preference lists in I.

Lemma 2.3. *No u-participant can be matched in M to an x-participant not corresponding to a literal contained in the respective clause.*

Proof. Suppose that some u, let us call him u_{i_1}, is matched to a woman x_0, who does not correspond to a literal contained in C_{i_1}. Now choose any $x(u_{i_1})$. If $x(u_{i_1})$ is matched to some y or some h, then clearly x_0 prefers the mate of $x(u_{i_1})$ to u_{i_1}, and since u_{i_1} is the first choice of $x(u_{i_1})$, we have an exchange-blocking pair $\{x_0, x(u_{i_1})\}$.

Therefore let $x(u_{i_1})$ be matched to some u_{i_2}. If $i_2 < i_1$, then x_0 prefers u_{i_2} to u_{i_1}, hence again we have an exchange-blocking pair $\{x_0, x(u_{i_1})\}$. Therefore $i_2 > i_1$. Now choose any $x(u_{i_2})$. Similarly as before, if $x(u_{i_2})$ is matched to some y, h or u_{i_3} with $i_3 < i_2$, women $\{x(u_{i_1}), x(u_{i_2})\}$ form an exchange-blocking pair. Hence $i_3 > i_2$, and in this way we construct a sequence $u_{i_1}, u_{i_2}, u_{i_3}, \ldots$ with $i_1 < i_2 < i_3 < \ldots$. As this sequence cannot continue indefinitely, eventually we must obtain an exchange-blocking pair. □

Lemma 2.4. *No u-participant can be matched in M to a g-participant.*

Proof. Suppose that $(u, g) \in M$. Let us take an arbitrary $x(u)$. For this participant, u is now unavailable, so Lemma 2.3 implies that $x(u)$ must be matched to some y or some h. Hence women $\{g, x(u)\}$ form an exchange-blocking pair. □

Lemma 2.5. *No g-participant can be matched in M to a y-participant other than $y(g)$.*

Proof. Suppose that y_{i_1} is matched to some g_0 other than $g(y_{i_1})$. Then we consider $g(y_{i_1})$. Clearly, $g(y_{i_1})$ prefers y_{i_1} to her mate. On the other hand, Lemma 2.4 implies that $g(y_{i_1})$ cannot be matched to a u-participant, and if she is matched to any h-participant or to some y_{i_2} with $i_2 < i_1$, then we have an exchange-blocking pair $\{g_0, g(y_{i_1})\}$. Therefore $i_2 > i_1$. Now consider $g(y_{i_2})$. Similarly as before, $g(y_{i_2})$ must be matched to some y_{i_3} with $i_3 > i_2$. If we continue this argument, we get a sequence $y_{i_1}, y_{i_2}, y_{i_3}, \ldots$ with $i_1 < i_2 < i_3 < \ldots$. As this sequence cannot continue indefinitely, eventually an exchange-blocking pair must occur. □

Lemma 2.6. *No y-participant can be matched in M to an x-participant other than $x(y)$.*

Proof. Let y be matched to some x who is not his twin. Then his mother $g(y)$ must be matched to some h (by Lemmas 2.4 and 2.5) and so x prefers this h to y, and $g(y)$ prefers y to her mate too, so $\{x, g(y)\}$ form an exchange-blocking pair. □

Lemmas 2.3-2.6 now imply:

Corollary 2.7. *Each u-participant is matched in M to an x-participant, corresponding to a literal contained in the respective clause.*

Corollary 2.8. *For each x-participant matched in M to $u(x)$, participant $y(x)$ is matched in M to his mother.*

Lemma 2.9. *For a given family, it is impossible to have two x-participants from opposite sides matched in M to u-participants.*

Proof. Suppose $(u(x_j^i), x_j^i) \in M$ and $(u(\overline{x}_j^i), \overline{x}_j^i) \in M$ (the case $(u(\overline{x}_j^{3-i}), \overline{x}_j^{3-i}) \in M$ can be treated similarly). Then we also have $(y_j^i, g_j^i) \in M$ by Corollary 2.8. Hence $u(\overline{x}_j^i)$ prefers g_j^i to his mate, and y_j^i prefers $\overline{x}_j^i$ to his mate, so that $\{u(\overline{x}_j^i), y_j^i\}$ form an exchange-blocking pair. □

Therefore we may define on the basis of the exchange-stable matching M the Boolean values of the variables in the formula B in the following way: x_j will be true if x_j^1 or x_j^2 are matched to $u(x_j^1)$ or to $u(x_j^2)$ respectively, and x_j will be false if $\overline{x}_j^1$ or $\overline{x}_j^2$ are matched to $u(\overline{x}_j^1)$ or to $u(\overline{x}_j^2)$ respectively. (If some variable is not assigned a Boolean value in this way, its value can be arbitrary.) Lemma 2.9 ensures that these Boolean values are defined consistently. Moreover, Corollary 2.7 ensures that under this truth assignment all clauses are satisfied.

Hence the Boolean formula B is satisfiable if and only if the constructed ESM instance I admits an exchange-stable matching. Therefore, together with our earlier observation that ESM belongs to the class NP, we have proved the following result:

Theorem 2.10. *ESM is NP-complete.*

3 ESR is NP-complete

In this section we consider ESR, the non-bipartite generalisation of ESM. It is known that SM is a special case of SR, i.e. given an instance I of SM, we may construct in polynomial time an instance J of SR such that the set of stable matchings in I is equal to the set of stable matchings in J [7, Lemma 4.1.1]. Analogously, ESM can be viewed as a special case of ESR, as is shown by the following lemma (the proof is similar to that of Lemma 4.1.1 of [7]).

Lemma 3.1. *Given an instance I of SM, we may construct in polynomial time an instance J of SR such that the set of exchange-stable matchings in I is equal to the set of exchange-stable matchings in J.*

Proof. Suppose that there are n men and n women in I. The set of participants in J is equal to the $2n$ men and women occurring in I. Each participant q's preference list in J is obtained from q's list $P(q)$ in I by appending all other members of the same sex, in arbitrary order, to $P(q)$.

Let M be an exchange-stable matching in I. Suppose $\{x, y\}$ is an exchange-blocking pair of M in J. Suppose, without loss of generality, that x is a man and y is a woman. Then x cannot prefer y's mate (a man) to his own mate (a woman), a contradiction. Hence x, y are of the same sex, so that $\{x, y\}$ is an exchange-blocking pair of M in I, a contradiction. Hence M is exchange-stable in J.

Conversely suppose that M is an exchange-stable matching in J. Suppose that $\{x, y\} \in M$ where x, y are both men. Then there exist two women x', y' such that $\{x', y'\} \in M$. Hence $\{x, x'\}$ is an exchange-blocking pair of M in J, a contradiction. Therefore M is a matching in I. Suppose that $\{x, y\}$ is an exchange-blocking pair of M in I. Then $\{x, y\}$ is also an exchange-blocking pair of M in J, a contradiction. Hence M is exchange-stable in I. □

The following corollary is an immediate consequence of Lemma 3.1 and Theorem 2.10.

Corollary 3.2. *ESR is NP-complete.*

4 Man-coalition-exchange-stability in SM

In this section we consider man-coalition-exchange-stability, given an instance I of SMI. A matching M in I must satisfy the property that, if $(m, w) \in M$ then each of m and w finds the other acceptable. The remainder of this section is split into two cases according to whether the preference lists in I are consistent. In Section 4.1 we consider instances of SM and SMIC, whilst in Section 4.2 we consider instances of SMII. Henceforth we denote by L the total length of the preference lists in I.

4.1 Consistent preference lists

We begin by assuming that the preference lists in I are consistent – in this context, from the point of view of finding a man-coalition-exchange-stable matching, essentially the preferences of the women are irrelevant. The results in this section are closely related to results already known in the literature for the *house allocation problem* [9, 1] and for *housing markets* [21, 19, 17]. An instance J of the house allocation problem may be obtained from I by deleting the women's preferences, and by interpreting the men as *agents* and the women as *houses*, following terminology used in the literature.

We firstly consider the special case that all preference lists in I are complete (so that I is an instance of SM). It turns out that, in contrast to the exchange-stability case, I is bound to admit a man-coalition-exchange-stable matching, and moreover such a matching may be found in linear time using a greedy algorithm known as the *serial dictatorship mechanism* [1] applied to J. This algorithm assumes an ordering $m_1, \ldots, m_n$ of the agents in J. For each i in turn from $1, \ldots, n$, we match m_i to the most-preferred unmatched house on m_i's preference list in J. Denote by M the resultant matching. Abdulkadiroğlu and Sönmez [1] prove that M is *Pareto optimal*, i.e. there is no other matching M' in J such that (i) some agent m_i prefers $M'(m_i)$ to $M(m_i)$ and (ii) no agent m_i prefers $M(m_i)$ to $M'(m_i)$. A Pareto optimal matching in J is man-coalition-exchange-stable in I [2]. Putting these observations together gives the following result.

Theorem 4.1. *Every instance of SM admits a man-coalition-exchange-stable matching, and such a matching may be found in $O(n^2)$ time using the serial dictatorship mechanism.*

The serial dictatorship mechanism is also outlined by Roth and Sotomayor [20, Example 4.3], who remark that it is strategy proof (though does not in general lead to a stable matching) and that a very similar procedure is used by the United States Naval Academy in order to match graduating students to their first posts as Naval Officers.

We now consider the general case where preference lists in I may be incomplete. We firstly define the *underlying graph* $G = (V, E)$ of I. This contains a vertex corresponding to each man m_i and woman w_j, and an edge $\{m_i, w_j\}$ if and only if m_i finds w_j acceptable. Clearly if G does not admit a perfect matching then I does not admit a man-coalition-exchange-stable matching. The problem of deciding whether G admits a perfect matching can be solved in $O(\sqrt{n}L)$ time [8]. We therefore assume that G admits a perfect matching M – in general M may admit an exchange-blocking coalition consisting of men.

In order to eliminate such coalitions, we proceed as follows. If we take the house allocation instance J and let the *initial endowment* be M (i.e. each agent m_i owns the house $M(m_i)$ initially) we obtain an instance K of a *housing market* [21, 19, 17]. Gale's Top Trading Cycles algorithm [21] may be applied to K in order to produce the unique matching M' that belongs to the *core* of the housing market K [21, 19]. That is, there is no other matching M'' and set of agents A in I such that:

1. $M''(m_i) \in \{w_j \in W : w_j = M(m_k) \text{ for some } m_k \in A\}$ for all $m_i \in A$;

2. some agent $m_i \in A$ prefers $M''(m_i)$ to $M'(m_i)$;

3. no agent $m_i \in A$ prefers $M'(m_i)$ to $M''(m_i)$.

Therefore M' is Pareto optimal in K, and hence by [2], M is man-coalition-exchange-stable in I. The Top Trading Cycles algorithm may be implemented to run in $O(L)$ time [2]. Hence we may summarise these discussions as follows.

Theorem 4.2. *Given an instance of SMIC, we may find a man-coalition-exchange-stable matching, or report that no such matching exists, in* $O(\sqrt{n}L)$ *time.*

4.2 Inconsistent preference lists

We now consider instances of SMII. We denote by MESMII the problem of deciding, given an instance of SMII, whether a man-exchange-stable matching exists. In this context it is not necessary for either of the mates w_1, w_2 of the men involved in an exchange-blocking pair to find their new mates acceptable. Indeed, if the MESMII problem definition were to include the assumption that w_1, w_2 should find their new mate acceptable, then we may delete inconsistent entries from the preference lists to obtain an instance of SMIC, and hence the previous subsection would apply.

It turns out that MESMII is NP-complete, as we now prove.

Theorem 4.3. *MESMII is NP-complete.*

Proof. We consider a similar transformation to the one used in order to prove Theorem 2.10. As before, we are given a Boolean formula as an instance of R3SAT. However in this case we create an instance I of SMII rather than SM.

The set of participants is exactly the same as before – the only modification to the constructed instance of ESM is that each participant's preference list comprises only the proper part as defined previously (i.e. the participants to the left of the vertical line, in the same order).

Given a satisfying truth assignment for the given formula, we construct an assignment M of men to women as before. It is immediate that M is a matching in I, since every participant was given a mate from the proper part of his/her list in the proof of Theorem 2.10. Furthermore, M is certainly man-exchange-stable (and is in fact exchange-stable), since an exchange-blocking pair in I would be an exchange-blocking pair in the original instance of ESM, contradicting Lemma 2.2.

Conversely suppose that M is a man-exchange-stable matching in I. Then each u-participant must be matched with some $x(u)$ (since g-participants do not find u-participants acceptable in I). It remains to show that, for a given family, no two x-participants from opposite sides can be matched in M to u-participants. In order to demonstrate this, the proof of Lemma 2.9 can be adapted for I. For, suppose that $(u(x_j^i), x_j^i) \in M$ and $(u(\overline{x}_j^i), \overline{x}_j^i) \in M$ (the case $(u(\overline{x}_j^{3-i}), \overline{x}_j^{3-i}) \in M$ can be treated similarly). Since $(y_j^i, x_j^i) \notin M$, the only other possible mate for y_j^i in M is g_j^i (since neither $\overline{x}_j^1$ nor $\overline{x}_j^2$ finds y_j^i acceptable). Hence the men $\{u(\overline{x}_j^i), y_j^i\}$ form an exchange-blocking pair, a contradiction. Therefore, the assignment of Boolean values to variables may be carried out consistently as in the argument following the proof of Lemma 2.9, and it constitutes a satisfying truth assignment for the given Boolean formula. □

5 Stable matchings that are (coalition-)exchange-stable

In this section we consider the problem of finding a stable matching that is also (coalition-) exchange-stable, given an instance of SM or SR. We begin with the SM case. Certainly it is possible that an instance of SM need not admit a stable matching that is even man-exchange-stable [13].

In SM, let us recall that, for each stable matching M other than the man-optimal stable matching, another stable matching M' and a rotation ρ exist, such that M is obtained from M' by eliminating ρ [7, Section 2.5.1]. Since each man m_i in the rotation ρ prefers $M'(m_i)$ to $M(m_i)$, it is easy to see that the men involved in ρ form an exchange-blocking coalition for M. Hence, a necessary condition for the existence of a stable matching M that is man-coalition-exchange-stable is that M is the man-optimal stable matching, and similarly for a woman-coalition-exchange-stable matching. Consequently, a necessary and sufficient condition for the existence of a stable matching that is coalition-exchange-stable is that there exists a unique stable matching (i.e. the man-optimal and woman-optimal stable matchings coincide) and it is coalition-exchange-stable.

Testing for the coalition-exchange-stability of a given matching M may be carried out using the *envy graph* relative to M. This is a digraph $G(M)$ with a vertex corresponding to each man m_i and each woman w_j, and an arc from the vertex corresponding to participant p to the vertex corresponding to participant q if and only if p prefers $M(q)$ to $M(p)$. It is easy to see that M is coalition-exchange-stable if and only if $G(M)$ is acyclic. Similarly M is man-coalition-exchange-stable if and only if $G(M)$ admits no cycle containing only vertices corresponding to men. We have therefore proved the following.

Theorem 5.1. *Given an instance I of SM, we may find a stable matching that is coalition-exchange-stable, or report that no such matching exists, in $O(n^2)$ time. The same result also holds if the matching is required to be both stable and man(woman)-coalition-exchange-stable, and/or if I is an instance of SMIC.*

By contrast, the problem of deciding whether an instance of SM admits a stable matching that is man-exchange-stable is NP-complete [13]. The same result also holds in the case that the stable matching is required to be exchange-stable [13].

We now consider the SR case. It turns out that there is a strong necessary condition for the existence of a stable matching that is coalition-exchange-stable. Consider an execution of Irving's algorithm [10] as applied to a given SR instance I that admits a stable matching. Suppose that Phase 2 of the algorithm is executed, terminating with stable matching M. This phase involves the elimination of one or more rotations – let ρ be the final rotation to be eliminated. Then $\rho = (x_0, y_0), (x_1, y_1), \ldots, (x_{r-1}, y_{r-1})$ for some $r \geq 2$, where $\{x_i, y_{i+1}\} \in M$ and x_i prefers y_i to y_{i+1} ($0 \leq i \leq r-1$, and $+$ is taken modulo r). Hence $\langle x_{r-1}, \ldots, x_1, x_0 \rangle$ is an exchange-blocking coalition of M. Thus a necessary condition for I to admit a stable matching that is coalition-exchange-stable is that Phase 1 of Irving's algorithm terminates with a stable matching M (which is therefore unique). To check that M is indeed coalition-exchange-stable, an analogous construction to the envy graph defined for the SM case can be used. We therefore have:

Theorem 5.2. *Given an instance of SR, we may find a stable matching that is coalition-exchange-stable, or report that no such matching exists, in $O(n^2)$ time.*

It follows by [13], [7, Lemma 4.1.1] and Lemma 3.1 that the problem of deciding whether an SR instance admits a stable matching that is exchange-stable is NP-complete.

6 Concluding remarks

The results of this paper indicate that the algorithmic behaviour of matching problems involving exchange-stability can be in marked contrast to that of matching problems involving classical stability. The NP-completeness results for ESM and ESR reflect the inherent difficulty of satisfying the exchange-stability criterion, given an instance of either of these problems. However we have seen that polynomial-time algorithms are possible for weaker forms of exchange-stability. In particular, the consideration of man-coalition-exchange-stability in instances of SM and SMIC reveals a strong connection with existing results relating to house allocation problems and housing markets.

We conclude with an open problem. Let CESM denote the problem of deciding whether a coalition-exchange-stable matching exists, given an instance of SM. The algorithmic complexity of CESM is open, thought we conjecture that the problem is NP-complete (by considering the envy graph it can be seen that CESM belongs to NP).

Acknowledgements

We would like to thank David Abraham, José Alcalde and Rob Irving for useful discussions, and also the anonymous referees for their comments. Part of this work was carried out whilst both authors visited the University of Alicante.

References

[1] A. Abdulkadiroğlu and T. Sönmez. Random serial dictatorship and the core from random endowments in house allocation problems. *Econometrica*, 66(3):689–701, 1998.

[2] D.J. Abraham, K. Cechlárová, D.F. Manlove, and K. Mehlhorn. Pareto optimality in house allocation problems. In *Proceedings of ISAAC 2004: the 15th Annual International Symposium on Algorithms and Computation*, volume 3341 of *Lecture Notes in Computer Science*, pages 3–15. Springer-Verlag, 2004.

[3] J. Alcalde. Exchance-proofness or divorce-proofness? Stability in one-sided matching markets. *Economic Design*, 1:275–287, 1995.

[4] K. Cechlárová. On the complexity of exchange-stable roommates. *Discrete Applied Mathematics*, 116(3):279–287, 2002.

[5] D. Gale and L.S. Shapley. College admissions and the stability of marriage. *American Mathematical Monthly*, 69:9–15, 1962.

[6] M.R. Garey and D.S. Johnson. *Computers and Intractability.* Freeman, San Francisco, CA., 1979.

[7] D. Gusfield and R.W. Irving. *The Stable Marriage Problem: Structure and Algorithms.* MIT Press, 1989.

[8] J.E. Hopcroft and R.M. Karp. A $n^{5/2}$ algorithm for maximum matchings in bipartite graphs. *SIAM Journal on Computing*, 2:225–231, 1973.

[9] A. Hylland and R. Zeckhauser. The efficient allocation of individuals to positions. *Journal of Political Economy*, 87(2):293–314, 1979.

[10] R.W. Irving. An efficient algorithm for the "stable roommates" problem. *Journal of Algorithms*, 6:577–595, 1985.

[11] R.W. Irving. Matching medical students to pairs of hospitals: a new variation on a well-known theme. In *Proceedings of ESA '98: the Sixth European Symposium*

on Algorithms, volume 1461 of *Lecture Notes in Computer Science*, pages 381–392. Springer-Verlag, 1998.

[12] R.W. Irving. Personal communication, 2002.

[13] R.W. Irving. The Man-Exchange Stable Marriage problem. Technical Report TR-2004-177, University of Glasgow, Department of Computing Science, 2004.

[14] D.E. Knuth. *Stable Marriage and its Relation to Other Combinatorial Problems*, volume 10 of *CRM Proceedings and Lecture Notes*. American Mathematical Society, 1997. English translation of *Mariages Stables*, Les Presses de L'Université de Montréal, 1976.

[15] C. Ng and D.S. Hirschberg. Three-dimensional stable matching problems. *SIAM Journal on Discrete Mathematics*, 4:245–252, 1991.

[16] E. Ronn. NP-complete stable matching problems. *Journal of Algorithms*, 11:285–304, 1990.

[17] A.E. Roth. Incentive compatibility in a market with indivisible goods. *Economics Letters*, 9:127–132, 1982.

[18] A.E. Roth. The evolution of the labor market for medical interns and residents: a case study in game theory. *Journal of Political Economy*, 92(6):991–1016, 1984.

[19] A.E. Roth and A. Postlewaite. Weak versus strong domination in a market with indivisible goods. *Journal of Mathematical Economics*, 4:131–137, 1977.

[20] A.E. Roth and M.A.O. Sotomayor. *Two-sided matching: a study in game-theoretic modeling and analysis*, volume 18 of *Econometric Society Monographs*. Cambridge University Press, 1990.

[21] L. Shapley and H. Scarf. On cores and indivisibility. *Journal of Mathematical Economics*, 1:23–37, 1974.