SIAM J. DISCRETE MATH.
Vol. 30, No. 4, pp. 2047–2063

ON THE STABLE MATCHINGS THAT CAN BE REACHED WHEN THE AGENTS GO MARCHING IN ONE BY ONE*

CHRISTINE T. CHENG†

Abstract. The *random order mechanism* (ROM) can be thought of as a sequential version of Gale and Shapley's deferred-acceptance (DA) algorithm, where agents are arriving one at a time, and each newly arrived agent has an opportunity to propose. Like the DA algorithm, ROM can be implemented in polynomial time. Unlike the DA algorithm, it is possible for ROM to output a stable matching that is *different* from the man-optimal and woman-optimal stable matchings. We say that a stable matching μ is *ROM-reachable* if ROM can output μ. In this paper, we investigate computational questions related to ROM-reachability. First, we prove that determining if a particular stable matching is ROM-reachable is NP-complete. However, we show that there is an efficient algorithm for determining if ROM can reach a nontrivial stable matching in the case when every agent has at least two stable partners. We then study two restricted versions of this problem. In the first version, we consider stable matchings that can be reached by ROM in a "direct" manner. We show that they are computationally easy to recognize. In the second version, we restrict the class of stable matchings to what we call extreme stable matchings and prove that the computational complexity of determining if they are ROM-reachable depends on the number of unstable partners of the agents.

Key words. stable matchings, random order mechanism, reachability

AMS subject classifications. 68R05, 68R10, 68Q25

DOI. 10.1137/140996690

1. Introduction.

Since Gale and Shapley's seminal publication [10] on stable matchings, economists, mathematicians and computer scientists alike have flocked into the field. The subject is rich and deep—four books [14, 11, 20, 16] and hundreds, if not thousands, of papers on stable matchings have been published. As a solution concept, it is also widely used in practice—many centralized matching markets such as those for NRMP [19], the Boston Public School Match [2], and the New York City High School Match, etc. [1][1], aim to match agents from two sides of the market in a stable way.

Our initial interest in stable matchings, however, had a more mundane reason. We simply found Gale and Shapley's deferred-acceptance (DA) algorithm to be a lot of fun. Our students share this enthusiasm whenever we have lectured on this topic. The reason it seems is how the agents behave in the algorithm. Their actions more or less capture what most people do in practice. An "active" agent (classically a "man") will initiate offers starting with the person from the other group that he prefers the most. A "passive" agent (classically a "woman"), on the other hand, will just wait for offers but will still act in a self-interested way. The students are taken aback though when they learn that the DA algorithm can only produce two kinds of stable matchings—the man-optimal/woman-pessimal and woman-optimal/man-pessimal stable matchings—even when an instance has an exponential number of stable matchings. In a few of these occasions, they have asked if there are other algorithms where *both* men and

*Received by the editors November 20, 2014; accepted for publication (in revised form) August 5, 2016; published electronically November 1, 2016.
http://www.siam.org/journals/sidma/30-4/99669.html

†Department of Electrical Engineering and Computer Science, University of Wisconsin-Milwaukee, Milwaukee, WI (ccheng@uwm.edu).

[1]See also the references in [17].

women can make proposals, and whether such an algorithm might output a stable matching that is less biased toward one side of the matching. The answer to their question turns out to be "yes," and determining the stable matchings the algorithm can reach is the subject of our investigation.

Ma proposed the *random order mechanism* (ROM) in 1996 [15] as a variant to Roth and Vande Vate's [21] work on random paths to stability. It works as follows: Let π be an ordering of the agents chosen uniformly at random. Think of the agents as arriving in a room (or a market) one at a time. In between arrivals, the room is closed so that a stable matching can be found for the agents in the room. The initial stable matching μ_0 is empty. Let μ_{i-1} denote the stable matching obtained prior to the arrival of $\pi(i)$. When $\pi(i)$ enters the room, μ_{i-1} may or may no longer be a stable matching of the instance consisting of the i agents in the room. If the former is true, μ_i is just μ_{i-1}; if the latter is true, $\pi(i)$ must form a blocking pair with one of the agents in the room. Resolve this in a *best response manner*. That is, among all the agents with which $\pi(i)$ forms a blocking pair, $\pi(i)$ is matched to the agent she prefers the most, say, a. Now if a had a partner in μ_{i-1}, this partner is now unmatched and may create new blocking pairs. Let him resolve in a best response manner again. This process is repeated until a stable matching is obtained. Set μ_i to be this stable matching. The final stable matching, $\mu_{|\pi|}$, formed after all the agents have arrived must then be a stable matching of the original instance.

We note that the step of determining if some agent b is part of some blocking pair of an existing matching μ and then resolving it in a best response manner can be simulated by a procedure that is reminiscent of the DA algorithm: b goes through her preference list and proposes to those who are currently in the market starting with the agent she prefers the most. The first agent to accept her proposal forms a blocking pair with b and is the agent that b prefers the most among all those that form a blocking pair with her. If no agent accepts her proposal, the current matching is stable. We have been referring to $\pi(i)$ as a woman but $\pi(i)$ can be a man too. Hence, we can think of ROM as a sequential version of the DA algorithm where agents from *both* sides of the market have an opportunity to propose. Like the DA algorithm, ROM can also be implemented in polynomial time. Unlike the DA algorithm, it is possible for ROM to output a stable matching *different* from the man-optimal and woman-optimal stable matchings. Ma [15] used an example of Knuth's to show that ROM can produce six out of the ten stable matchings of the instance.

Let us say that a stable matching μ is *reachable by ROM* or *ROM-reachable* if there is an ordering π of the agents so that when ROM processes π, the output is μ. One of the most useful properties of ROM-reachable stable matchings is due to Cechlárová [9] and Blum, Roth, and Rothblum [8]. It states that every ROM-reachable stable matching must have at least one agent matched to their best stable partner. Other properties can be also found in [8] and [7]; nonetheless, a nice characterization of ROM-reachable stable matchings is still missing. In this paper, our goal is to address computational questions about ROM-reachability.

Our results. Let I be an instance with n agents, and let us refer to a stable matching of I as *nontrivial* if it is different from the man-optimal and woman-optimal stable matchings. First, we ask a basic question—given a nontrivial stable matching μ of I, is μ ROM-reachable? We prove that the problem is NP-complete, even in the case when every agent has a preference list of length at most 4. Our result answers an open problem in [16]. In our reduction, the stable matching of interest, μ, has many disjoint submatchings where no agent is matched to their best stable partner. To

reach these submatchings, ROM has to first form submatchings that are not part of μ and then use them as stepping stones to reach the said submatchings. It is this two-step process that makes the problem difficult because the intermediate submatchings are intertwined with each other.

But suppose we simplify the problem and just ask if ROM can reach a nontrivial stable matching of I. We show that the problem can be answered in polynomial time provided every agent of I has at least two stable partners. All we have to do is run at most n permutations on ROM. If ROM can reach a nontrivial stable matching, then one of these runs will also output a nontrivial stable matching of I.

Next, we say that a stable matching μ is *strongly ROM-reachable* if there is some input permutation π of I's agents so that the output of ROM is μ, and $\mu_1 \subseteq \mu_2 \subseteq \cdots \subseteq \mu_{|\pi|} = \mu$. That is, once a pair of agents is part of a μ_i, it is part of the remaining stable matchings until ROM ends. Anecdotally, many of the ROM-reachable stable matchings we have found are also strongly ROM-reachable. In the third part of our paper, we characterize the strongly ROM-reachable stable matchings using directed graphs. We then present an efficient algorithm that recognizes these kinds of stable matchings. Our characterization makes use of subgraphs of *jealousy graphs* defined recently by Hoffman, Moeller, and Paturi [12] to obtain more refined bounds for the convergence time of random better response dynamics. It is interesting that jealousy graphs are also relevant to ROM.

Finally, we consider a class of stable matchings we call *extreme stable matchings.* They are the stable matchings where every pair has one agent matched to his/her best stable partner and the other to his/her worst stable partner. Unlike the stable matching of interest in our first NP-completeness reduction, these stable matchings do not have submatchings that lie in the "middle." We show that when each agent has at most one unstable partner in I (i.e., the agent and the unstable partner are never matched in a stable matching of I), every extreme stable matching of I is strongly ROM-reachable. However, there is an instance where some agents have two unstable partners, and this instance has an extreme stable matching that is not reachable by ROM. Using this instance as a gadget, we then prove that when agents have two or more unstable partners, determining if an extreme stable matching of I is ROM-reachable is NP-complete. These results are highly unusual in that we know of no computational problems on stable matchings where the unstable pairs of the instance determine the complexity of the problem.

Related work. We have presented ROM as a sequential version of the DA algorithm where the starting matching is the empty matching and agents from both sides of the matching are allowed to propose. Ma [15] also allowed ROM to start at an arbitrary matching. He based ROM on Roth and Vande Vate's proof [21] that the random better response dynamics converges to a stable matching with probability 1. Interestingly, ROM is quite different from the random better response dynamics in at least two ways. First, starting with an empty matching, the latter can reach *every* stable matching of an instance [21]. Such a property does not hold for ROM. Second, there are instances where the random better response dynamics can take exponential time to converge to a stable matching [3]. In contrast, ROM always reaches a stable matching in a polynomial number of steps.

In [8], Blum, Roth, and Rothblum sought to model the dynamics of senior-level labor markets (e.g., head football coaches of US college teams, etc.). Assume a stable matching already exists for the firms and workers in the market. Then some workers retire and some firms open up new positions. Blocking pairs involving unmatched firms can now exist. The paper studied how the market can restabilize itself using the

DA algorithm. Their results describe how stable matchings change from one iteration of ROM to the next. Biro, Cechlárová, and Fleiner [7] extended their work to the stable roommates setting.

Other probabilistic mechanisms for generating stable matchings have been studied in the past (see [3, 16] and references therein). Perhaps the one that is most similar to ROM is *employment by lotto* (EBL) by Aldershof, Carducci, and Lorenc [4], which is just the *random serial dictatorship* (RSD) applied to the stable matchings setting. Like ROM, the input of EBL is a random permutation π of the agents. Initially, S_0 consists of all the stable matchings of the instance. In the ith iteration, S_i is reduced to the set of stable matchings where $\pi(i)$ is matched to his/her best stable partner in S_{i-1}. The algorithm ends when all the agents in π have been processed and outputs $S_{|\pi|}$. The recent work by Aziz, Brandt, and Brill [5] on RSD in the one-sided matching setting imply that there is an efficient algorithm for determining if a particular stable matching can be reached by EBL. What is computationally hard is determining the probabilities induced by EBL on the stable matchings.

Last, many researchers have proposed different notions of "fair" stable matchings. Klaus and Klijn [13] argued that while both ROM and EBL do not guarantee end-state fairness (i.e., they may not output the stable matchings of the instance with equal probability), they are *procedurally fair* because "the sequence of moves for the agents is drawn uniformly at random."

2. Preliminaries. *Stable marriage with incomplete lists* (SMI) instances model two-sided matching markets. One side consists of "men," the other of "women." Each agent has a preference list that ranks members from the opposite group the agent has deemed acceptable in a linear order. A pair (m, w) is *acceptable* if m and w appear in each other's preference lists. A *matching* μ is a set of acceptable man-woman pairs so that every agent is part of at most one pair. The matching has a *blocking pair* (m, w) if (i) m is unmatched or m prefers w to his partner in μ and (ii) w is unmatched or w prefers m to her partner in μ. A goal in two-sided matching markets is to find *stable matchings*, which are matchings with no blocking pairs, because the agents are less likely to break their assignments.

Throughout this paper, we will assume that in every SMI instance I, an agent a is in another agent b's preference list if and only if b is in a's preference list. The two agents are *stable partners* if they are matched to each other in some stable matching of I; otherwise, they are *unstable partners*. Additionally, b is a's *best (worst) stable partner* if among all of his/her stable partners b is his/her most (least) preferred one. Gale and Shapley [10] showed that when it is the men who propose in their algorithm, the result is the *man-optimal/woman-pessimal stable matching*—that is, every man is matched to his best stable partner and simultaneously every woman is matched to her worst stable partner. On the other hand, when the women are the ones who propose in their algorithm, the output is the *woman-optimal/man-pessimal stable matching* which is defined similarly. A simple corollary of this result is that when b is a's best stable partner, a is b's worst stable partner.

Since the number of men and the number of women in I need not be the same, some agents may be unmatched in a stable matching of I. The *rural hospitals theorem* [18] states that when an agent is unmatched in one stable matching of I, the agent will be unmatched in *all* stable matchings of I. Thus, the set of matched agents is the same for all stable matchings of I. The set can be easily determined by computing the man-optimal stable matching of I.

Below is the pseudocode for Ma's ROM [15]. Let I consist of n agents. When S is a subset of I's agents, we use $I_{|S}$ to denote the SMI instance obtained by restricting I to the agents in S. We say that a is a *blocking agent* of a matching μ_i of $I_{|S}$ if it is part of a blocking pair of μ_i. We also say that b is the *best blocking partner* of a if b is the one that a prefers the most among all agents that form a blocking pair with a. Let π be a permutation of I's agents chosen uniformly at random.

```
ROM(π, I)
S ← ∅, μ_0 ← ∅
for i = 1 to n
    a_i ← π(i), S ← S ∪ {a_i}
    μ_i ← μ_{i−1}
    while a_i is a blocking agent of μ_i with respect to instance I_|S
        let b_j be the best blocking partner of a_i
        a_z ← a_i
        if b_j is matched in μ_i
            let a_i now denote the partner of b_j
            μ_i ← μ_i − {(a_i, b_j)}
        μ_i ← μ_i ∪ {(a_z, b_j)}
return μ_n
```

FACT 1 (see [21, 16]). *ROM(π, I) always terminates and outputs a stable matching of I.*

The while loop runs in $O(|I|)$ time, where $|I|$ is the size of instance I, so ROM(π, I) runs in $O(n|I|)$ time. Let us call a stable matching μ of I *reachable by ROM* or *ROM-reachable* if there is some permutation π of its agents so that μ is the output of ROM(π, I). Ma noted that if π consists of all the women first followed by all the men, the output of ROM is the man-optimal stable matching of I; similarly, if π consists of all the men first followed by all the women, the output of ROM is the woman-optimal stable matching of I. Unlike Gale and Shapley's algorithm, however, ROM can in some cases output stable matchings different from the man-optimal and woman-optimal stable matchings. For example, Ma presented an SMI instance that had 10 stable matchings, six of which are reachable by ROM. Several researchers have derived necessary conditions for a stable matching to be ROM-reachable. Here are some of them.

FACT 2 (Cechlárová [9], Blum, Roth, and Rothblum [8]). *Suppose SMI instance I has n agents, and π is an ordering of I's agents.*

(i) *Let $\pi(n) = a$. Then a is matched to his/her best stable partner in μ_n, the output of ROM(π, I).*

(ii) *Let $b \neq a$. If the partner of b in μ_{n-1} is one of his/her stable partners in I, then b will remain matched to this partner in μ_n. Otherwise, b is matched to his/her best stable partner in μ_n if b has the same gender as a and to his/her worst stable partner in μ_n if b has the opposite gender as a.*

3. ROM-reachability is NP-complete. In this section, we prove the NP-completeness of *ROM-reachability: Given a stable matching μ of instance I, is there a permutation π of I's agents so that when ROM processes π, the output is μ?* Consider the SMI instance I^* below. The men are m_i and a_{i1}, a_{i2}, $i = 1, 2, 3$ while the women are w_i and b_{i1}, b_{i2}, $i = 1, 2, 3$.

m_1:	w_1	w_2	b_{11}	w_3	w_1:	m_2	m_3	m_1
m_2:	w_2	w_3	b_{21}	w_1	w_2:	m_3	m_1	m_2
m_3:	w_3	w_1	b_{31}	w_2	w_3:	m_1	m_2	m_3
a_{11}:	b_{11}	b_{12}			b_{11}:	a_{12}	a_{11}	m_1
a_{12}:	b_{12}	b_{11}			b_{12}:	a_{11}	a_{12}	
a_{21}:	b_{21}	b_{22}			b_{21}:	a_{22}	a_{21}	m_2
a_{22}:	b_{22}	b_{21}			b_{22}:	a_{11}	a_{12}	
a_{31}:	b_{31}	b_{32}			b_{31}:	a_{32}	a_{31}	m_3
a_{32}:	b_{32}	b_{31}			b_{32}:	a_{31}	a_{32}	

Let $\alpha_1 = \{(m_1, w_1), (m_2, w_2), (m_3, w_3)\}, \alpha_2 = \{(m_1, w_2), (m_2, w_3), (m_3, w_1)\}, \alpha_3 = \{(m_1, w_3),\ (m_2, w_1), (m_3, w_2)\}$. Notice that $\alpha_1, \alpha_2, \alpha_3$ are exactly the stable matchings of the subinstance of I^* when restricted to the agents m_i's and w_i's. Both α_1 and α_3 are ROM-reachable stable matchings of this subinstance because they are the man-optimal and woman-optimal stable matchings, respectively. However, α_2 is not ROM-reachable for this subinstance since none of the m_i's nor the w_i's are matched to their best stable partners in the subinstance.

For $j = 1, 2, 3$, let $\beta_{j1} = \{(a_{j1}, b_{j1}), (a_{j2}, b_{j2})\}$ and $\beta_{j2} = \{(a_{j1}, b_{j2}), (a_{j2}, b_{j1})\}$. Notice also that β_{j1} and β_{j2} are exactly the stable matchings of the subinstance of I^* when restricted to $a_{j1}, a_{j2}, b_{j1}, b_{j2}$. It is easy to check that the stable matchings of I^* are exactly of the form $\alpha_i \cup \beta_{1k_1} \cup \beta_{2k_2} \cup \beta_{3k_3}$, where $i \in \{1, 2, 3\}$ and $k_1, k_2, k_3 \in \{1, 2\}$. Of interest to us is the stable matching $\mu^* = \alpha_2 \cup \beta_{12} \cup \beta_{22} \cup \beta_{32}$.

PROPOSITION 1. *When* $\pi = m_1, m_2, m_3, w_1, w_2, w_3, b_{11}, a_{11}, a_{12}, b_{12}, b_{21}, a_{21}, a_{22}, b_{22}, b_{31}, a_{31}, a_{32}, b_{32}$, *ROM*$(\pi, I)$ *outputs* μ^*.

LEMMA 1. *Let* π *be a permutation of the participants of* I^*. *Suppose ROM*(π, I^*) *outputs* μ^*. *Then the following must be true:*

(i) *For* $j = 1, 2, 3$, *among* a_{j1}, a_{j2}, b_{j1}, *and* b_{j2}, *the last agent to appear in* π *is* b_{j1} *or* b_{j2}.

(ii) *There must be some* $j \in \{1, 2, 3\}$ *so that* b_{j1} *appears first and* b_{j2} *appears last in the ordering of* a_{j1}, a_{j2}, b_{j1}, *and* b_{j2} *in* π.

Proof. To prove (i), among a_{j1}, a_{j2}, b_{j1}, and b_{j2}, let c be the last agent to appear in π. Let μ' be the stable matching prior to ROM processing c. Suppose $c = a_{j1}$. If b_{j2} is unmatched in μ', $\{a_{j2}, b_{j2}\}$ is a blocking pair of μ'. Since the only person b_{j2} can be matched to is a_{j2}, (a_{j2}, b_{j2}) must belong to μ'. Consequently, either $(m_j, b_{j1}) \in \mu'$ or b_{j1} is unmatched in μ'. When ROM finally processes a_{j1}, he will get matched to b_{j1} so that β_{j1} is part of the stable matching at the end of this iteration. Since none of the remaining agents after a_{j1} in π can form a blocking pair with a_{j1}, a_{j2}, b_{j1}, and b_{j2} when the latter are matched according to β_{j1}, β_{j1} will be a submatching of the output of ROM. This contradicts our assumption that ROM(π, I^*)'s output is μ^*. Thus, $c \neq a_{j1}$. The same reasoning applies as to why $c \neq a_{j2}$ so $c = b_{j1}$ or b_{j2}.

To prove (ii), first we note that during the execution of ROM(π, I^*) there must be a step where some m_j is temporarily matched to b_{j1} even though the two may not be matched to each other in the stable matching for that iteration. Otherwise, let π' be the permutation obtained from π by removing the a_{jk}'s and b_{jk}'s. Let I' be the instance obtained from I^* by removing the same set of agents. Then it must be the case that ROM(π', I') can simulate how ROM(π, I^*) matched the agents in I' so that the output of ROM(π', I') is a submatching of ROM(π, I^*). But the former will only output α_1 or α_3 and therefore μ^* cannot be the output of ROM(π, I^*). Since this is

a contradiction, some m_j must be temporarily matched to b_{j1} during the execution of ROM(π, I^*).

Now, consider an arbitrary b_{j1}. If it appears second or third in the ordering of a_{j1}, a_{j2}, b_{j1}, and b_{j2} in π, b_{j2} must appear last in the ordering because of (i). Thus, either a_{j1} or a_{j2} appears first. When ROM begins to process b_{j1}, at least one of these men will be matched to b_{j1} immediately so b_{j1} will never be temporarily matched to m_j in this iteration. In the later iterations, b_{j1} may change her partner from a_{j1} to a_{j2} but other than this change b_{j1} will never be matched to any one else.

Suppose b_{j1} appears last in the ordering of a_{j1}, a_{j2}, b_{j1}, and b_{j2} in π. Using the same reasoning in the first paragraph, (a_{j1}, b_{j2}) will be part of the stable matching just before ROM processes b_{j1} while a_{j2} is unmatched. Thus, when ROM processes b_{j1}, she will be immediately matched to a_{j2} and β_{j2} will be the resulting submatching until the end of ROM. Throughout the execution of ROM, b_{j1} will never be matched to anyone else.

Thus, the only way for b_{j1} to be temporarily matched to m_j is for her to appear first in the ordering of a_{j1}, a_{j2}, b_{j1}, b_{j2} in π and consequently for b_{j2} to appear last because of (i). Since this must be true for some b_{j1}, (ii) follows. ∎

Now consider an arbitrary 3-SAT instance Φ with n variables $x_1, x_2, \ldots, x_n$ and q clauses $C_1, C_2, \ldots, C_q$. Our goal is to construct an SMI instance I_Φ so that a particular stable matching of I_Φ is reachable by ROM if and only if Φ is satisfiable. For each variable x_j, create the subinstance

$$\begin{array}{llll} a_{j1}\colon & b_{j1} & b_{j2} & \\ a_{j2}\colon & b_{j2} & b_{j1} & \end{array} \qquad \begin{array}{llll} b_{j1}\colon & a_{j2} & a_{j1} & \cdots \\ b_{j2}\colon & a_{j1} & a_{j2} & \cdots \end{array}$$

where the dots indicate that b_{j1} and b_{j2} may have more agents in their preference lists. Let us call b_{j1} and b_{j2} *twins* and a_{j1} and a_{j2} the *counterparts* of b_{j1} and b_{j2}. For each clause C_i, create the subinstance

$$\begin{array}{lllll} m_{i1}\colon & w_{i1} & w_{i2} & z_{i1} & w_{i3} \\ m_{i2}\colon & w_{i2} & w_{i3} & z_{i2} & w_{i1} \\ m_{i3}\colon & w_{i3} & w_{i1} & z_{i3} & w_{i2} \end{array} \qquad \begin{array}{llll} w_{i1}\colon & m_{i2} & m_{i3} & m_{i1} \\ w_{i2}\colon & m_{i3} & m_{i1} & m_{i2} \\ w_{i3}\colon & m_{i1} & m_{i2} & m_{i3} \end{array}$$

where z_{ik}, $k = 1, 2, 3$, is based on the kth literal in C_j. If this literal is x_j, set z_{ik} to b_{j1} and add m_{ik} to the end of b_{j1}'s preference list; otherwise, if the literal is $\overline{x_j}$, set z_{ik} to b_{j2} and add m_{ik} to the end of b_{j2}'s preference list. Thus, when we restrict I_Φ to the participants associated with C_j and its variables, the subinstance looks just like the instance I^* we considered with the exception that b_{j2} may sometimes play the role of b_{j1} and vice versa. For $i = 1, \ldots, q$, let $\alpha_{i1} = \{(m_{i1}, w_{i1}), (m_{i2}, w_{i2}), (m_{i3}, w_{i3})\}$, $\alpha_{i2} = \{(m_{i1}, w_{i2}), (m_{i2}, w_{i3}), (m_{i3}, w_{i1})\}$, $\alpha_{i3} = \{(m_{i1}, w_{i3}), (m_{i2}, w_{i1}), (m_{i3}, w_{i2})\}$. Define β_{j1} and β_{j2} as before for $j = 1, \ldots, n$. Again, it is straightforward to verify that the stable matchings of I_Φ are exactly of the form $\alpha_{1g_1} \cup \alpha_{2g_2} \cup \cdots \cup \alpha_{qg_q} \cup \beta_{1k_1} \cup \beta_{2k_2} \cup \cdots \cup \beta_{nk_n}$, where each $g_i \in \{1, 2, 3\}$ and each $k_j \in \{1, 2\}$.

Let $\mu^{**} = \alpha_{12} \cup \alpha_{22} \cup \cdots \cup \alpha_{q2} \cup \beta_{12} \cup \beta_{22} \cup \cdots \cup \beta_{n2}$. When we restrict μ^{**} to the agents associated with clause C_j and its literals, μ^{**} is just like μ^*. It is, however, much trickier for ROM to reach μ^{**} because b_{j1} and b_{j2} can be part of other subinstances. The two women cannot simultaneously help obtain their subinstances' "middle" stable matchings since one of them has to appear last in the ordering of $a_{j1}, a_{j2}, b_{j1}, b_{j2}$. We are now ready to prove our main result.

THEOREM 1. *The 3-SAT instance Φ has a satisfying assignment if and only if there is a permutation π of I_Φ's agents so that the output of ROM(π, I_Φ) is μ^{**}.*

Proof. Let f be a satisfying assignment of Φ. We now order the agents of I_Φ based on f. Initially set π_f to the empty sequence. For $i = 1$ to q, add $m_{i1}, m_{i2}, m_{i3}, w_{i1}, w_{i2}, w_{i3}$ to the end of π_f. We call this the first part of the sequence. Next, for $j = 1$ to n, if $f(x_j) = 1$, add b_{j1} to the end of π_f; otherwise add b_{j2} to the end of π_f. We call this the second part of the sequence. Finally, for $j = 1$ to n, if $f(x_j) = 1$, add a_{j1}, a_{j2}, b_{j2} to the end of π_f; otherwise, add a_{j1}, a_{j2}, b_{j1} to the end of π_f. We call this the third part of the sequence.

Now consider what happens in ROM(π_f, I_Φ). After ROM processes the first part of π_f, the resulting stable matching is $\alpha_{13} \cup \alpha_{23} \cup \cdots \cup \alpha_{q3}$. Next, ROM processes the second part of π_f. Suppose b_{j1} is one of the women in this sequence. Let $C_{i_1}, C_{i_2}, \ldots, C_{i_r}$ be the clauses that have x_j as a literal. Without loss of generality, assume that x_j is their first literal so that $m_{i_1 1}, m_{i_2 1}, \ldots, m_{i_r 1}$ appear after a_{j2} and a_{j1} in b_{j1}'s preference list. Consider the beginning of the iteration that processes b_{j1}. There are two possible cases:

(1) $\alpha_{i_1 3}$ *is part of the current stable matching.* Then $\{m_{i_1 1}, b_{j1}\}$ is a blocking pair of $\alpha_{i_1 3}$. As a result, $m_{i_1 1}$ rejects $w_{i_1 3}$, and $w_{i_1 3}$ will in turn propose to her second choice who then accepts her proposal. Blocking pairs will continue to get resolved until $\alpha_{i_1 2}$ replaces $\alpha_{i_1 3}$ as a submatching of the current stable matching and b_{j1} is unmatched.

(2) $\alpha_{i_1 3}$ *is not part of the current stable matching.* This implies that in a prior iteration, case (1) happened (via the agent associated with the second or third literal of C_{i_1}) and $\alpha_{i_1 2}$ already replaced $\alpha_{i_1 3}$ as a submatching. Since none of the women in the second part of π_f can form a blocking pair with the agents in $\alpha_{i_1 2}$, $\alpha_{i_1 2}$ is still a submatching of the current stable matching. Furthermore, b_{j1} will just skip over $m_{i_1 1}$ and remain unmatched.

Thus, after b_{j1} considers $m_{i_1 1}$, $\alpha_{i_1 2}$ is a submatching of the current stable matching and b_{j1} is unmatched. Similar results apply as b_{j1} considers $m_{i_2 1}, \ldots, m_{i_r 1}$ so that at the end of the iteration that processes b_{j1}, α_{i2} has replaced α_{i3} for all clauses C_i that have x_j as a literal. Additionally, b_{j1} is unmatched.

When b_{j2} instead of b_{j1} is in the second part of the sequence, α_{i2} will replace α_{i3} for all clauses C_i that have $\overline{x_j}$ as a literal at the end of the iteration that processes b_{j2}. Additionally, b_{j2} is unmatched. Hence, once ROM processes the second part of π_f, the resulting stable matching is $\alpha_{12} \cup \alpha_{22} \cup \cdots \cup \alpha_{m2}$ because f is a satisfying assignment of Φ. All of the women that appears in the second part of π_f are unmatched.

Finally, after ROM processes the third part of π_f, it is easy to see that $\beta_{12} \cup \cdots \beta_{n2}$ becomes part of the output. Moreover, $\alpha_{12} \cup \alpha_{22} \cup \cdots \cup \alpha_{q2}$ remains unchanged because none of the agents in the third part of π_f forms a blocking pair with the agents of this submatching. We have shown that μ^{**} is the output of ROM(π_f, I_Φ).

We prove the converse next. Suppose ROM(π, I_Φ) outputs μ^{**}. When we restrict I_Φ to the agents associated with C_i and its variables, the instance is just like our first example I. Similarly, when we restrict μ^{**} to the same set of agents, the stable matching looks just like μ^* of I. Thus, a result like Lemma 1 should apply to the ordering of the agents in π. We restate part (ii) as follows:

(ii′) For $i = 1, \ldots, q$, there is some $k \in \{1, 2, 3\}$ so that z_{ik} appears first and her twin appears last in the ordering of z_{ik}, her twin, and their counterparts in π.

We now construct a truth assignment f_π as follows: for $j = 1, \ldots, n$, set $f_\pi(x_j)$ to 0 if b_{j1} is the last agent to appear in π among $a_{j1}, a_{j2}, b_{j1}, b_{j2}$; otherwise set $f_\pi(x_j)$ to 1. We know from (i) that when $f_\pi(x_j)$ is set to 1, b_{j2} is the last agent to appear in π among $a_{j1}, a_{j2}, b_{j1}, b_{j2}$. Additionally, from (ii′), we know that f_π has set one of the literals in C_i to 1 for $i = 1, \ldots, q$. Thus, f_π is a satisfying assignment for Φ. ∎

COROLLARY 1. ROM-reachability *is NP-complete even in the restricted case when all agents have a preference list of length at most* 4.

Proof. Given 3-SAT instance Φ with n variables and q clauses, we created an instance I_Φ that has $4n + 6q$ agents so that Φ is satisfiable if and only if a particular stable matching of I_Φ is reachable by ROM. Thus, 3-SAT is polynomially reducible to ROM-reachability. Additionally, it is easy to verify that ROM-reachability is in NP. It follows that ROM-reachability is NP-complete.

But it is also known that 3-SAT is NP-complete even in the special case when each literal appears twice among the clauses—i.e., there are two clauses that contain x_i and two other clauses that contain $\overline{x_i}$ for $i = 1, \ldots, n$ [6]. When Φ is such an instance, then in I_Φ both b_{j1} and b_{j2} have exactly two unstable partners in their preference lists for $j = 1, \ldots, n$. The restriction on ROM-reachability follows. ∎

In our hardness result above, every agent in the SM instance has at least two stable partners. We show next that for these kinds of instances determining if ROM can reach a nontrivial stable matching can be answered in polynomial time. For each agent a, let π_a denote a permutation that consists first of an ordering of the agents of the same gender as a except for a, followed by an ordering of the agents of the opposite gender as a, and then ending with a. Let $\mu(a)$ refer to the partner of a in the stable matching μ.

LEMMA 2. *Let I be an SM instance where each agent has at least two stable partners. Suppose there is a permutation π of I so that ROM(π, I) outputs a nontrivial stable matching of I. Let a be the last agent in π. Then ROM(π_a, I) will also output a nontrivial stable matching of I.*

Proof. Denote the output of ROM(π, I) as μ. Without loss of generality, assume a is a man, and let μ_M be the man-optimal stable matching of I. Then $\mu(a) = \mu_M(a)$ according to Fact 2(i). But since $\mu \neq \mu_M$, there must be some woman b so that $\mu(b) \neq \mu_M(b)$. In particular, b prefers $\mu(b)$ over $\mu_M(b)$ since $\mu_M(b)$ is her worst stable partner in I. Additionally, $\mu(b)$ has to be a stable partner of b in I_{-a}, the instance obtained from I by removing agent a; otherwise, according to Fact 2(ii), $\mu(b) = \mu_M(b)$.

Now consider what happens in ROM(π_a, I). At the end of iteration $n - 1$, the stable matching is the woman-optimal stable matching of I_{-a}. Hence b is matched to $\mu(b)$ or somebody she prefers over $\mu(b)$ since $\mu(b)$ is one of her stable partners in I_{-a}. At iteration n, agent a arrives and a sequence of proposals are made by the men until a stable matching is obtained. Since b is already matched, she will accept a new offer only if it came from men she prefers over her current partner. In other words, b will always be matched throughout iteration n and her partner will stay the same or get better and better. Thus, at the end of iteration n, b has to be matched to $\mu(b)$ or somebody she prefers over $\mu(b)$; that is, her partner cannot be $\mu_M(b)$. On the other hand, since a is the last agent to arrive, a has to be matched to $\mu_M(a)$, which we know is different from $\mu_W(a)$ since a has at least two stable partners. It follows that the outcome of ROM(π_a, I) is neither the man-optimal nor woman-optimal stable matchings. ∎

We emphasize that the lemma does not say that ROM(π_a, I) and ROM(π, I) have the same outputs; rather, if ROM(π, I) outputs a nontrivial stable matching, then so will ROM(π_a, I).

THEOREM 2. *Suppose I has n agents each of which has at least two stable partners. Then checking if ROM can reach a nontrivial stable matching of I takes $O(n^2|I|)$ time.*

Proof. First, determine the man-optimal and woman-optimal stable matchings of I. Then for each agent a, construct a permutation π_a and run ROM(π_a, I). If one run outputs a nontrivial stable matching of I, return "yes"; otherwise if the outputs of all the runs are just the trivial stable matchings of I, return "no." The correctness follows from the previous lemma. Since there are n permutations to consider and ROM can be implemented in $O(n|I|)$ time, the whole procedure takes $O(n^2|I|)$ time. ∎

4. Strongly ROM-reachable stable matchings. Recall that a stable matching μ of I is strongly ROM-reachable if there is a permutation π of the agents of I so that (i) ROM(π, I) outputs μ and (ii) $\mu_1 \subseteq \mu_2 \subseteq \cdots \subseteq \mu_{|\pi|} = \mu$, where μ_i is the stable matching at the end of iteration i of ROM(π, I). Call π a permutation *associated with* μ. Notice that the definition implies that once an agent is matched in some μ_i, his or her partner must be the same one as in μ and remains so until the end of ROM. Intuitively, it is easier to determine if a stable matching is strongly ROM-reachable because ROM can build it one pair at a time.

PROPOSITION 2. *Let μ be a strongly ROM-reachable stable matching of I, and let π be a permutation associated with μ. Suppose $(m, w) \in \mu$, $\pi(k) = m$, $\pi(k') = w$, and $k < k'$. Then for $i = k, k+1, \ldots, k'-1$, m is unmatched in μ_i while for $i = k', k'+1, \ldots |A|$, $(m, w) \in \mu_i$.*

In [12], Hoffman, Moeller, and Paturi defined the *jealousy graph* of a stable matching μ, $J(\mu)$ as follows: The vertices of $J(\mu)$ are the pairs in μ, and there is a directed edge from the pair (m, w) to another pair (m', w') whenever m' prefers w to w' or w' prefers m to m'. In this section, we consider a "labeled" version of $J(\mu)$. Let L be a labeling that assigns each agent of I as *lucky* or *unlucky*. We say that L *respects* μ if for every pair in μ one agent is labeled lucky while the other is labeled unlucky. For such a labeling, denote as $J_L(\mu)$ the graph whose vertices are the pairs in μ such that there is a directed edge from (m, w) to (m', w') if w is an unlucky agent and m' prefers w to w' or m is an unlucky agent and w' prefers m to m'. Thus, $J_L(\mu)$ is a subgraph of $J(\mu)$ and keeps only the edges "caused" by the agents labeled unlucky by L. Here now is our characterization of strongly ROM-reachable stable matchings based on labeled jealousy graphs.

THEOREM 3. *A stable matching μ of I is strongly ROM-reachable if and only if there is a labeling L that respects μ such that $J_L(\mu)$ is acyclic.*

Proof. Suppose μ is a stongly ROM-reachable stable matching of I. Let π be a permutation that is associated with μ. For each pair $(m, w) \in \mu$, label the agent that appears first in π as unlucky and the agent that appears later in π as lucky. For the unmatched agents, arbitrarily label them as lucky or unlucky. Call the labeling L^*. We argue that $J_{L^*}(\mu)$ is acyclic next.

Order the pairs of μ based on when the pairs' lucky agents appeared in π. Denote the ordering as $p_1, p_2, \ldots, p_{|\mu|}$. Thus, among all pairs in μ, p_1's lucky agent appeared in π first, followed by that of p_2, etc. Now, consider an edge (p_j, p_k) in $J_{L^*}(\mu)$. Assume $k < j$. Let i be the iteration when ROM processes the lucky agent in p_j. Then at the end of iteration $i-1$, the unlucky agent of p_j is unmatched while p_k is part of the stable matching μ_{i-1}. But the edge from p_j to p_k implies that μ_{i-1} has a blocking pair, a contradiction. Thus, it must be the case that $k > j$. Since we

chose the edge (p_j, p_k) arbitrarily, $p_1, p_2, \ldots, p_{|\mu|}$ must be a topological ordering of the vertices of $J_{L^*}(\mu)$; i.e., $J_{L^*}(\mu)$ is acyclic.

Let us now prove the converse. Suppose $J_L(\mu)$ is acyclic. To prove that μ is strongly ROM-reachable, we need to show that there is a permutation that is associated with μ. Let $p_1, p_2, \ldots, p_{|\mu|}$ be a topological ordering of $J_L(\mu)$. Construct π by making its $(2j-1)$st agent be the unlucky agent in p_j and the $(2j)$th agent be the lucky agent in p_j for $j = 1, \ldots, |\mu|$. Then add any unmatched agents in μ to the end of the sequence. Consider ROM(π, I) next.

Claim. During the execution of ROM(π, I), $\mu_{2j-1} = \mu_{2j-2}$ while $\mu_{2j} = \mu_{2j-2} \cup \{p_j\}$ for $j = 1, \ldots, |\mu|$.

Proof of claim. It is clear that μ_1 is an empty matching while $\mu_2 = \{p_1\}$. So assume that the claim is true for $j = 1, \ldots, t'$. Without loss of generality, let m be an unlucky agent in $p_{t'+1} = (m, w)$. When ROM processes m, m can propose to the women on his list that are also in $p_1, \ldots, p_{t'}$. These may include women that he prefers less over w. If such a woman prefers m to her current partner, it would mean that there is a directed edge from $p_{t'+1}$ to p_k for some $k < t' + 1$ in $J_L(\mu)$, a contradiction. Thus, every woman that m proposes to rejects him. It follows that at the end of iteration $2t' + 1$, $\mu_{2t'+1} = \mu_{2t'}$ and m is unmatched.

When ROM processes w, w will begin by proposing to the men on her list that she prefers over m and who are also in $p_1, \ldots, p_{t'}$. But every such man m' must prefer his current partner over w because not doing so will mean that (m', w) is a blocking pair of μ. Since this cannot be the case, every man that w proposes to before m rejects her. Thus, w will propose to m and he will accept because he is unmatched. At the end of iteration $2t' + 2$, $\mu_{2t'+2} = \mu_{2t'} \cup \{p_{t'+1}\}$. By induction, we have shown that the claim is true. ∎

Finally, when ROM processes an unmatched agent a in μ, none of the agents a proposes to will accept the proposal since it would mean that μ has a blocking pair. Thus, after ROM has processed all the agents in $p_1, \ldots, p_{|\mu|}$, the stable matching is μ and will remain so until the end of ROM. We have shown that π is a permutation that accompanies μ so μ is a strongly ROM-reachable stable matching of I. ∎

So how do we take advantage of Theorem 3 to determine if a stable matching μ is strongly ROM-reachable? There are at least $2^{|\mu|}$ labelings of the agents of I that respect μ so the brute force method of checking if one of the labelings L yields an acyclic $J_L(\mu)$ is infeasible. First, we note that strongly ROM-reachable stable matchings are made up of strongly ROM-reachable stable submatchings.

LEMMA 3. *Suppose μ is a strongly ROM-reachable stable matching of I. Let $\mu' \subseteq \mu$. Then μ' is also a strongly ROM-reachable stable matching of $I_{|\mu'}$, the instance obtained by restricting I to the agents in μ'.*

Proof. Since μ is a strongly ROM-reachable stable matching, by Theorem 3 there is a labeling L that respects μ such that $J_L(\mu)$ is acyclic. Additionally, μ' has to be a stable matching of $I_{|\mu'}$; otherwise, if it has a blocking pair, then so will μ. Restrict L to the agents in μ' and call it L'. Clearly, L' respects μ' and $J_{L'}(\mu')$ is acyclic since it is a subgraph of $J_L(\mu)$. It follows that μ' is a strongly ROM-reachable stable matching of $I_{|\mu'}$. ∎

Second, we define the notion of a *sink agent* whose name is meant to suggest that it behaves like the sink node of a directed acyclic graph. We shall say that a stable matching τ of I (not necessarily strongly ROM-reachable) has a sink agent if

(i) there is a man m so that $\tau(m) = w$ is his best stable partner in I and for any other pair $(m', w') \in \tau$, m' does not prefer w to his partner w' or
(ii) there is a woman w so that $\tau(w) = m$ is her best stable partner in I and for any other pair $(m', w') \in \tau$, w' does not prefer m to her partner m'.

In (i), we say m is a sink agent of τ while in (ii) w is a sink agent of τ.

LEMMA 4. *Every strongly ROM-reachable stable matching μ of I has a sink agent.*

Proof. Let π be a permutation that is associated with μ. Consider the very last matched agent that appears in π. Without loss of generality, let this agent be m who is matched to w in μ, and let $\pi(i) = m$. This means that $\mu_i = \mu$ but that w is unmatched in μ_{i-1}. The latter implies that for every other pair (m', w') in μ, m' preferred w' to w. Hence, m is a sink agent of μ. If the last matched agent that appears in π is a woman w, a similar proof will show that w is a sink agent of μ. ∎

Our algorithm for determining if a stable matching is strongly ROM-reachable is patterned after the standard algorithm for topologically sorting a directed acyclic graph.

CheckDirectROM(τ, I)
Set $i = 1$, $I_1 = I$, and $\tau_1 = \tau$.
While τ_i has a sink agent a
let p_i be the pair that consists of a and $\tau_i(a)$
label a as lucky and $\tau_i(a)$ as unlucky
$\tau_{i+1} \leftarrow \tau_i - \{p_i\}$ and $I_{i+1} \leftarrow I_{|\tau_{i+1}}$
$i \leftarrow i + 1$
If τ_i is empty return ("yes"; $p_1, p_2, \ldots, p_{|\tau|}$)
Else return ("no"; τ_i)

THEOREM 4. CheckDirectROM *correctly determines if a stable matching τ of I is strongly ROM-reachable in $O(|\tau| \times |I|)$ time. In each case, the algorithm also returns a certificate that can be used to verify that the algorithm's answer is correct.*

Proof. Since every τ_i is a subset of τ, if τ is a strongly-ROM reachable stable matching of I, then every τ_i is a strongly ROM-reachable stable matching of $I_{|\tau_i}$ according to Lemma 3. By Lemma 4, every τ_i has a sink agent. Thus, the algorithm is correct in concluding that when some τ_i has no sink agent, the input τ is not a strongly ROM-reachable stable matching. The lack of a sink agent in τ_i is evidence that τ is not a strongly ROM-reachable stable matching.

Let $|\tau| = n$. Now suppose that $\tau_1, \ldots, \tau_n$ have sink agents. Let L be the labeling that assigns each sink agent in p_i as lucky and the partner as unlucky; the unmatched agents are labeled arbitrarily. By the definition of sink agents, $p_n, p_{n-1}, \ldots, p_1$ is a topological ordering of $J_L(\mu)$ because p_i will not have any edges to $p_{i+1}, \ldots, p_n$ in $J_L(\mu)$. Thus, μ is a strongly ROM-reachable stable matching of I, and the permutation π based on $p_n, p_{n-1}, \ldots, p_1$ as described in the proof of Theorem 3 can be used to verify that ROM(π, I) outputs τ.

Finally, to determine if τ_i has a sink agent, we run the Gale–Shapley algorithm to identify every agent's best stable partner. The algorithm can be implemented in $O(|I|)$ time. Next, for each agent a matched to their best stable partner, we check if there is a person who prefers $\tau_i(a)$ over their current partner in τ_i. We can do this by going through the preference list of $\tau_i(a)$, and, for each person b that appears in this list, we compare the rank of $\tau_i(b)$ and $\tau_i(a)$ in b's preference list. Using the appropriate data structure so that rank-checking can be done in $O(1)$ time, this step

can again be implemented in $O(|I|)$ time. But there can be n τ_i's so implementing CheckDirectROM takes $O(|\tau| \times |I|)$ time. ∎

5. Extreme stable matchings. A stable matching is an *extreme stable matching* if for every pair in the stable matching, either the man or the woman is matched to his/her best stable partner (and consequently the other person is matched to his/her worst stable partner). These stable matchings are interesting because they do not have the "middle" submatchings like the α_{i2}'s that μ^{**} had when we proved the NP-completeness of ROM-reachability. Are all extreme stable matchings ROM-reachable? We show that the answer to this question depends on the number of unstable partners of the agents.

THEOREM 5. *Let I be an SMI instance where every agent has at most one unstable partner. Then every extreme stable matching μ of I is strongly ROM-reachable.*

Proof. Let μ be an extreme stable matching of I. For each pair $(m, w) \in \mu$, label the agent matched to his/her worst stable partner as unlucky and the other agent as lucky.[2] Call the labeling L. Clearly, L respects μ. To prove the theorem, we will show that $J_L(\mu)$ is acyclic.

Suppose this is not the case and the pairs $(a_1, b_1), (a_2, b_2), \ldots, (a_k, b_k)$ form a directed cycle in $J_L(\mu)$. Without loss of generality, we also assume that a_1 is the unlucky agent in (a_1, b_1). We will now establish that a_i is the unlucky agent in (a_i, b_i) for $i = 2, \ldots, k$.

When $k = 2$ (i.e., the directed cycle has length 2), there is an edge from (a_1, b_1) to (a_2, b_2) and from (a_2, b_2) to (a_1, b_1). The first edge implies that b_2 prefers a_1 over a_2. If additionally b_2 is the unlucky agent in (a_2, b_2), then a_1 prefers b_2 over b_1. This makes $\{a_1, b_2\}$ a blocking pair of μ, a contradiction. Thus, a_2 must be the unlucky agent in (a_2, b_2).

When $k \geq 3$, there is an edge from (a_1, b_1) to (a_2, b_2) and from (a_2, b_2) to (a_3, b_3). Again, the first edge implies that b_2 prefers a_1 over a_2. In order for $\{a_1, b_2\}$ not to form a blocking pair, a_1 must prefer b_1 over b_2. But since a_1 is the unlucky agent in (a_1, b_1), b_1 is a_1's worst stable partner so a_1 and b_2 cannot be stable partners in I. If b_2 is the unlucky agent in (a_2, b_2), then the edge from (a_2, b_2) to (a_3, b_3) will also imply that b_2 and a_3 are unstable partners in I using a similar reasoning. In other words, b_2 will have two unstable partners, contradicting our assumption about I. So a_2 must be the unlucky agent in (a_2, b_2). Applying the same reasoning around the directed cycle, we conclude that a_i is the unlucky agent in (a_i, b_i) for $i = 2, \ldots, k$.

The above observation implies that each b_i prefers a_{i-1} over a_i. Since a_i is already b_i's best stable partner, b_i and a_{i-1} must be unstable partners, and the preference list of b_i must consist of a_{i-1} first followed by a_i and the rest of her stable partners. On the other hand, consider a_i. We already know that a_i is the first person in b_{i+1}'s list. In order for $\{a_i, b_{i+1}\}$ not be a blocking pair of μ, a_i must prefer b_i over b_{i+1}. Thus, the preference list of a_i consists of all his stable partners including b_i, his worst stable partner, and is then followed by b_{i+1}.

Let μ' be the matching obtained by removing the pairs $(a_1, b_1), (a_2, b_2), \ldots, (a_k, b_k)$ from μ and replacing them with $(a_1, b_2), (a_2, b_3), \ldots, (a_k, b_1)$. We argue next that μ' has to be a stable matching of I too. If μ' is not a stable matching, then it has a blocking pair. Clearly, one of the agents in the blocking pair must be from the set

[2] If (m, w) is a fixed pair—i.e., they are matched to each other in all of the stable matchings of I—then they are each other's best and worst stable partners. Arbitrarily label one as lucky and the other as unlucky.

$\{a_1, \ldots, a_k, b_1, \ldots, b_k\}$; otherwise, the same pair will be blocking μ as well. Additionally, none of $b_1, \ldots, b_k$ are part of the blocking pair since each one is matched to her first choice. So suppose the blocking pair is (a_i, b'). Now, the partner of b' in μ and μ' are the same but a_i prefers b_i, his partner in μ, over his partner b_{i+1}, his partner in μ'. Furthermore, b_i and b_{i+1} are next to each other in a_i's preference list. Thus, if (a_i, b') is a blocking pair of μ', then b' is ahead of b_i in a_i's preference list and has to be a blocking pair of μ too. Since this is a contradiction, μ' has no blocking pairs and must be a stable matching of I.

But we already noted that a_i and b_{i+1} are unstable partners in I. It must be the case then that $J_L(\mu)$ is acyclic and, consequently, μ is strongly ROM-reachable. ∎

In the next lemma, we show that when we relax the condition on Theorem 5 and allow agents to have two unstable partners in I, an extreme stable matching of I may no longer be ROM-reachable.

LEMMA 5. *There is an SMI instance whose agents have at most two unstable partners, and this instance has an extreme stable matching that is not ROM-reachable.*

Proof. Consider the following SMI instance I:

m_1:	w_1	w_2			w_1:	m_6	m_2	m_1	m_3
m_2:	w_2	w_1			w_2:	m_5	m_1	m_2	m_4
m_3:	w_1	w_3	w_4	w_8	w_3:	m_4	m_3		
m_4:	w_2	w_4	w_3	w_7	w_4:	m_3	m_4		
m_5:	w_7	w_5	w_6	w_2	w_5:	m_6	m_5		
m_6:	w_8	w_6	w_5	w_1	w_6:	m_5	m_6		
m_7:	w_7	w_8			w_7:	m_4	m_8	m_7	m_5
m_8:	w_8	w_7			w_8:	m_3	m_7	m_8	m_6

For $i = 1, 2, 3, 4$, let $\alpha_{i1} = \{(m_{2i-1}, w_{2i-1}), (m_{2i}, w_{2i})\}$ and $\alpha_{i2} = \{(m_{2i-1}, w_{2i}), (m_{2i}, w_{2i-1})\}$. It is easy to verify that the man-optimal stable matching is $\alpha_{11} \cup \alpha_{21} \cup \alpha_{31} \cup \alpha_{41}$, the woman-optimal stable matching is $\alpha_{12} \cup \alpha_{22} \cup \alpha_{32} \cup \alpha_{42}$, and the set of stable matchings of I is $\{\alpha_{1j_1} \cup \alpha_{2j_2} \cup \alpha_{3j_3} \cup \alpha_{4j_4},$ where $j_1, j_2, j_3, j_4 \in \{1, 2\}\}$. In other words, each agent has exactly two stable partners, and every stable matching of I is an extreme stable matching. Furthermore, the agents $m_3, m_4, m_5, m_6, w_1, w_2, w_7, w_8$ all have two unstable partners. We will now prove that $\tau^* = \alpha_{11} \cup \alpha_{22} \cup \alpha_{32} \cup \alpha_{41}$ is not reachable by ROM.

Let M_1 and W_1 denote the set of men and women who are matched to their best stable partners in τ^*. Thus, $M_1 = \{m_1, m_2, m_7, m_8\}$ while $W_1 = \{w_3, w_4, w_5, w_6\}$. If there is a permutation π of I's agents so that $\text{ROM}(\pi, I)$ outputs τ^*, the last agent in π must belong to $M_1 \cup W_1$.

Suppose that the last agent in π is m_1. Consider the instance prior to ROM processing m_1, $I_{-\{m_1\}}$. Since there is one more woman than man in the instance, at least one woman is unmatched in all the stable matchings of $I_{-\{m_1\}}$. In this case, it is w_3. (The reader can verify this by computing, say, the man-optimal matching of $I_{-\{m_1\}}$.) According to Fact 2(b), this means that when ROM finally processes m_1, the woman w_3 will be matched to her man-optimal stable partner in I, which is m_3. Thus, the output of $\text{ROM}(\pi, I)$ is not τ^*.

If the last agent in π is m_2, m_7, or m_8, the women w_4, w_5, w_6, respectively, will have to be matched to their man-optimal stable partner in I so τ^* is again not the output of $\text{ROM}(\pi, I)$. A similar argument can be used to show why none of the agents in W_1 can be the last agent in π. It follows that π does not exist. ∎

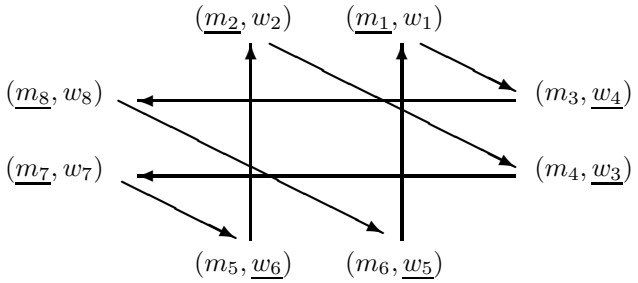

FIG. 1. *The labeled jealousy graph of the stable matching* $(m_1, w_1), (m_2, w_2), (m_3, w_4), (m_4, w_3), (m_5, w_6), (m_6, w_5), (m_7, w_7), (m_8, w_8)$ *in the SMI instance described in the proof of Lemma* 5. *The agents that are matched to their best stable partner are labeled* lucky *and underlined, while the agents that are matched to their worst stable partner are labeled* unlucky.

Consider the stable matching $\tau* = (m_1, w_1), (m_2, w_2), (m_3, w_4), (m_4, w_3), (m_5, w_6), (m_6, w_5), (m_7, w_7), (m_8, w_8)$ in the SMI instance described in the proof of Lemma 5. Label the agents matched to their best stable partner as *lucky* and the agents matched to their worst stable partner as *unlucky*. The graph in Figure 1 shows the labeled jealousy graph of τ^* using this labeling. Since the said stable matching is not ROM-reachable (and therefore not strongly ROM-reachable too), the labeled jealousy graph has directed cycles. On the other hand, suppose we modify the preference lists of the agents so that every one has at most one unstable partner. For example, assume m_3 and w_1, m_4 and w_7, m_5 and w_2, m_6 and w_8 are no longer in each other's preference lists. Then the edges from (m_1, w_1) to (m_3, w_4), from (m_4, w_3) to (m_7, w_7), from (m_5, w_6) to (m_2, w_2), and from (m_8, w_8) to (m_6, w_5) disappear from the labeled jealousy graph. That is, the labeled jealousy graph will no longer contain a directed cycle. This is so because according to Theorem 5, τ^*, still an extreme stable matching of the instance, is now strongly ROM-reachable.

We will now use the SMI instance in the proof of Lemma 5 as a gadget to prove the next theorem.

THEOREM 6. *Let I be an SMI instance where agents can have two or more unstable partners. Let μ be an extreme stable matching of I. Then determining if μ is ROM-reachable is NP-complete.*

Proof. Let Φ be a 3-SAT instance with n variables $x_1, x_2, \ldots, x_n$ and q clauses $C_1, C_2, \ldots, C_q$. For each variable x_i, create the subinstance

$$\begin{array}{llllll} a_{i1}\colon & b_{i1} & b_{i2} & \cdots & b_{i1}\colon & a_{i2} \quad a_{i1} \\ a_{i2}\colon & b_{i2} & b_{i1} & \cdots & b_{i2}\colon & a_{i1} \quad a_{i2} \end{array}$$

where the dots indicate that a_{i1} and a_{i2} may have more agents in their preference lists. For each clause C_j, create the subinstance

$$\begin{array}{llllll} m_{j1}\colon & w_{j1} & w_{j2} & & & w_{j1}\colon & m_{j6} \quad m_{j2} \quad m_{j1} \quad m_{j3} \\ m_{j2}\colon & w_{j2} & w_{j1} & & & w_{j2}\colon & m_{j5} \quad m_{j1} \quad m_{j2} \quad m_{j4} \\ m_{j3}\colon & w_{j1} & w_{j3} & w_{j4} & w_{j8} & w_{j3}\colon & m_{j4} \quad \mathbf{z_{j1}} \quad \mathbf{z_{j3}} \quad m_{j3} \\ m_{j4}\colon & w_{j2} & w_{j4} & w_{j3} & w_{j7} & w_{j4}\colon & m_{j3} \quad \mathbf{z_{j2}} \quad m_{j4} \\ m_{j5}\colon & w_{j7} & w_{j5} & w_{j6} & w_{j2} & w_{j5}\colon & m_{j6} \quad m_{j5} \\ m_{j6}\colon & w_{j8} & w_{j6} & w_{j5} & w_{j1} & w_{j6}\colon & m_{j5} \quad m_{j6} \\ m_{j7}\colon & w_{j7} & w_{j8} & & & w_{j7}\colon & m_{j4} \quad m_{j8} \quad m_{j7} \quad m_{j5} \\ m_{j8}\colon & w_{j8} & w_{j7} & & & w_{j8}\colon & m_{j3} \quad m_{j7} \quad m_{j8} \quad m_{j6} \end{array}$$

where z_{jk}, $k = 1, 2, 3$ is based on the kth literal in C_j. If this literal is x_i, set z_{jk} to a_{i1}; otherwise, if the literal is $\overline{x_i}$, set z_{jk} to a_{i2}. Add w_{j3} or w_{j4} to to the preference list of z_{jk} depending on whose preference list z_{jk} appears in. For $j = 1, \ldots, m$, let

$$\tau_j^* = \{(m_{j1}, w_{j1}), (m_{j2}, w_{j2}), (m_{j3}, w_{j4}), (m_{j4}, w_{j3}), (m_{j5}, w_{j6}), \\ (m_{j6}, w_{j5}), (m_{j7}, w_{j7}), (m_{j8}, w_{j8})\}$$

and for $i = 1, \ldots, n$ let $\beta_{i1} = \{(a_{i1}, b_{i1}), (a_{i2}, b_{i2})\}$. It is easy to verify that

$$\mu^{**} = \tau_1^* \cup \tau_2^* \cup \cdots \cup \tau_m^* \cup \beta_{11} \cup \beta_{21} \cup \cdots \cup \beta_{n1}$$

is an extreme stable matching of I_Φ.

Claim. The 3-SAT instance Φ has a satisfying assignment if and only if there is a permutation π of I_Φ's agents such that the output of ROM(π, I_Φ) is μ^{**}.

We omit the proof of the claim as it is very similar to that of Theorem 1. Given 3-SAT instance Φ with n variables and q clauses, we have created an instance I_Φ that has $4n + 16q$ agents so that Φ is satisfiable if and only if μ^{**}, an extreme stable matching of I_Φ, is reachable by ROM. The theorem follows. ∎

6. Conclusion. We investigated the stable matchings that Ma's ROM [15] can reach starting from the empty matching. Since ROM induces a probability distribution on an instance's set of stable matchings, we were equivalently interested in the stable matchings that are in the *support* of ROM. In the first half of the paper, we showed that it is NP-complete to determine if a particular stable matching lies in the support of ROM, but it is computationally easy to determine if *some* nontrivial stable matching is in the support of ROM in the case when all agents have at least two stable partners.

In the second half of the paper, we introduced the notion of a *strongly ROM-reachable stable matchings* which are stable matchings that ROM can reach in a "direct" manner. We provided a nice characterization and presented an efficient recognition algorithm for these stable matchings. Interestingly, strongly ROM-reachable stable matchings are also relevant to the EBL mechanism we described in the introduction. Suppose μ is a strongly ROM-reachable stable matching and π is the permutation found by CheckDirectROM. It is not difficult to show that when the reverse of π, π^r, is the input to the EBL, the output is again μ. That is, in the context of strongly ROM-reachable stable matchings, ROM and EBL are "equivalent" to each other.

Question: *What are the stable matchings μ for which there is a permutation π so that ROM$(\pi, I) =$ EBL$(\pi^r, I) = \mu$? Do they have to be strongly ROM-reachable? What precisely are the stable matchings that are both ROM-reachable and EBL-reachable?*

Last, we defined the class of *extreme stable matchings* and showed that the computational complexity of determining if ROM can output an extreme stable matching is dependent on the number of unstable partners of the agents. One interesting avenue of research is to investigate the stable matchings that can be reached by ROM when agents are allowed to enter as well as leave the market.

Question: *Might some stable matchings which were not reachable by ROM in our current setting be reachable in the setting where agents are also allowed to leave?*

Acknowledgments. I would like to thank David Manlove, Péter Biró, Daniel Moeller, and Ágnes Cseh, whose insightful questions, comments, and suggestions helped shape the direction of this research.

REFERENCES

[1] A. Abdulkadiroğlu, P. Pathak, and A. Roth, *The New York City high school match*, Amer. Econom. Rev., 95 (2006), pp. 364–367.

[2] A. Abdulkadiroğlu, P. Pathak, A. Roth, and T. Sönmez, *Changing the Boston School-Choice Mechanism*, NBER working paper 11965, 2006.

[3] H. Ackermann, P. Goldberg, V. Mirrokni, H. Röglib, and B. Vöcking, *Uncoordinated two-sided matching markets*, SIAM J. Comput., 40 (2011), pp. 92–106.

[4] B. Aldershof, O. Carducci, and D. Lorenc, *Refined inequalities for stable marriage*, Constraints, 4 (1999), pp. 281–292.

[5] H. Aziz, F. Brandt, and M. Brill, *The computational complexity of random serial dictatorship*, Econom. Lett., 121 (2013), pp. 341–345.

[6] P. Berman, M. Karpinski, and A. Scott, *Approximation Hardness of Short Symmetric Instances of Max-3sat*, in Electronic Colloquium on Computational Complexity Report, 2003.

[7] P. Biro, K. Cechlárová, and T. Fleiner, *The dynamics of stable matchings and half-matchings*, Internat. J. Game Theory, 36 (2008), pp. 333–352.

[8] Y. Blum, A. Roth, and U. Rothblum, *Vacancy chains and equilibration in senior-level labor markets*, J. Econom. Theory, 76 (1997), pp. 362–411.

[9] K. Cechlárová, *Randomized Matching Mechanism Revisited*, Manuscript, Institute of Mathematics, P.J. Safárik University, Slovakia, 2002.

[10] D. Gale and L. Shapley, *College admissions and the stability of marriage*, Amer. Math. Monthly, 69 (1962), pp. 9–15.

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

[12] M. Hoffman, D. Moeller, and R. Paturi, *Jealousy graphs: Structure and complexity of decentralized stable matching*, in Proceedings of WINE '13: The 9th Workshop on Web and Internet Economics, 2013.

[13] B. Klaus and F. Klijn, *Procedurally fair and stable matching*, Econom. Theory, 27 (2006), pp. 431–447.

[14] D. Knuth, *Mariages Stables*, Les Presses de L'Université de Montréal, 1976.

[15] J. Ma, *On randomized matching mechanisms*, Econom. Theory, 8 (1996), pp. 377–381.

[16] D. Manlove, *Algorithmics of Matching Under Preferences*, World Scientific, River Edge, NJ, 2013.

[17] Web document available at http://econ.core.hu/kutatas/jatek_resz.html#is_dor.

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

[19] A. Roth, *The origins, history, and design of the resident match*, J. Amer. Med. Assoc., 289 (2003), pp. 909–912.

[20] A. Roth and M. Sotomayor, *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*, Econom. Soc. Monogr. 18, Cambridge University Press, Cambridge, 1990.

[21] A. Roth and J. V. Vande Vate, *Random paths to stability in two-sided matching*, Econometrica, 58 (1990), pp. 1475–1480.