Farsighted Objections and Maximality in One-to-one Matching Problems

Mert Kimya

September 2020
Farsighted Objections and Maximality in One-to-one Matching Problems

Mert Kimya †

University of Sydney, School of Economics, NSW 2006, Australia

September 23, 2020

Abstract

We characterize the set of stable matchings when individuals are farsighted and when they choose their objections optimally along a farsighted objection path. We use a solution concept called maximal farsighted set (MFS), which is an adaptation of the concepts developed in Dutta and Vohra (2017) and Dutta and Vartiainen (2020) to one-to-one matching problems. MFS always exists, but it need not be unique. There is a unique largest MFS that contains all other, which is equal to the largest consistent set of Chwe (1994). This implies that the largest consistent set embodies the idea of maximality in one-to-one matching problems.

1 Introduction

In their seminal paper, Gale and Shapley (1962) define the one-to-one matching model. Let \( N \) be a finite set of individuals partitioned into a set of men \( M \) and a set of women \( W \). Each individual \( i \in N \) has a strict preference relation \( P(i) \) over the individuals on the other side of the market and the prospect of being alone. We represent the preference of each agent by an ordered list. For instance, for \( m \in M \), \( P(m) = w_1, w_2, m, w_3, \ldots \) indicates that \( m \) prefers \( w_1 \) over \( w_2 \), \( w_2 \) over being single, and he prefers remaining single to matching with anyone else. Let \( P \) denote a preference profile specifying the preference relation of each \( m \in M \) and \( w \in W \). A one-to-one matching problem is a triple \((M, W, P)\).

A matching is a function \( \mu : N \to N \) such that \( i \) for all \( m \in M \), \( \mu(m) \in W \cup \{m\} \), \( ii \) for all \( w \in W \), \( \mu(w) \in M \cup \{w\} \), and \( iii \) for all \( i \in N \), \( \mu(\mu(i)) = i \).

The core of a one-to-one matching problem identifies all those matchings that cannot be improved upon by a coalition. The notion of the core assumes that individuals care about the immediate consequence of their deviation. But, one might anticipate that an initial objection might be followed by further objections, in which case the initial deviator should evaluate the ultimate consequence of its deviation. In this paper we are interested in such farsighted players when they choose their objections optimally along a farsighted objection path.

* I would like to thank Murali Agastya and Rajiv Vohra for helpful comments and suggestions.
† E-mail Address: mert.kimya@sydney.edu.au
An Illustrative Example

Consider a one-to-one matching problem with three men, \{m_1, m_2, m_3\}, and four women, \{w_1, w_2, w_3, w_4\}. The preferences of individuals are as follows:

\[
\begin{array}{cccccc}
  m_1 & m_2 & m_3 & w_1 & w_2 & w_3 \\
  w_3 & w_1 & w_1 & m_1 & m_2 & m_1 \\
  w_2 & w_3 & w_3 & m_3 & m_3 & m_2 \\
  w_1 & w_2 & w_2 & m_2 & m_2 & m_1 \\
  m_1 & w_4 & m_3 & w_1 & w_2 & w_3 \\
  w_4 & m_2 & w_4 & w_4 & w_4 & w_4 \\
\end{array}
\]

Consider the matching \(\mu = (m_1 \, m_2 \, m_3 \, w_4 \, w_1 \, w_2 \, w_3 \, w_4)\). There is an objection to this matching by the pair \((m_1, w_2)\). If we assume that \((m_1, w_2)\) cannot rearrange other matchings and cannot force singles to marry then after the deviation the resulting matching is \(\mu' = (m_1 \, m_2 \, m_3 \, w_1 \, w_4 \, w_2 \, w_3 \, w_4)\).

According to the notion of the core, existence of \(\mu'\) implies that \(\mu\) is unstable. But, this notion does not consider whether \(\mu'\) itself is immune to further deviations. Perhaps players can expect further deviations that could ultimately make either \(m_1\) or \(w_2\) worse off, preventing the initial deviation. This idea is formalized through the notion of a farsighted objection (see Chwe (1994) and Harsanyi (1974)). We say that a matching \(\mu''\) is a farsighted objection to a matching \(\mu'\) if there exists a sequence of deviations from \(\mu'\) to \(\mu''\) such that every coalition active in this sequence prefers \(\mu''\) to the matching it replaces.

Consider the farsighted objection path depicted in Figure 1. Observe that every coalition moving along this objection path prefers the final outcome \(\mu''\) to the matching it replaces, while the final matching makes \(w_2\) worse off compared to the initial matching \(\mu\). This seems to imply that the deviation by \((m_1, w_2)\) from \(\mu\) to \(\mu'\) can be prevented through the farsighted objection to \(\mu''\).

\[
\mu' = \begin{pmatrix} m_1 & w_2 \\ m_2 & m_2 \\ m_3 & w_3 \\ w_1 & w_1 \\ w_4 & w_4 \end{pmatrix} \xrightarrow{(m_3, w_1), (m_2, w_4)} \begin{pmatrix} m_4 & w_2 \\ m_2 & w_4 \\ m_3 & w_1 \\ w_3 & w_3 \end{pmatrix} \xrightarrow{(m_1, w_3)} \begin{pmatrix} m_1 & w_3 \\ m_2 & w_4 \\ m_3 & w_1 \\ w_2 & w_2 \end{pmatrix} = \mu''
\]

Figure 1: \(\mu''\) is a farsighted objection to \(\mu'\)

There is a crucial oversight in this argument. How are we to know that it is indeed optimal for each coalition moving along the path to take the prescribed action? Perhaps a coalition is assumed to enforce a matching at some point, but they would be better off enforcing some other matching expecting a better outcome from there. This is known as the problem of maximality (see Ray and Vohra (2014) for simple demonstrations of this problem in abstract games).

The depicted farsighted objection path does not satisfy a reasonable definition of maximality which allows the players moving along the path to refuse the action assigned
to them and instead take some other action with some other player. To see this, observe that the depicted farsighted objection path prescribes that in the last stage \((m_1, w_3)\) will match. Consider the deviation in which \(w_3\) refuses to take the prescribed action and instead deviates together with \(m_2\). Let \(\mu^*\) be the resulting matching. One can show that there is no farsighted objection to \(\mu^*\) that makes either \(m_2\) or \(w_3\) worse off than \(\mu''\). Hence, there is nothing that prevents \((m_2, w_3)\) from deviating from the initial farsighted objection path. This implies that the farsighted objection path depicted in Figure 1 cannot be used to prevent the initial deviation from \(\mu\) to \(\mu'\) by \((m_1, w_2)\).

A similar argument can be used to rule out any other farsighted objection to \(\mu'\) that makes either \(m_1\) or \(w_2\) worse off compared to \(\mu\). Hence, there is nothing to keep \((m_1, w_2)\) from deviating at \(\mu\), implying that \(\mu\) is not only myopically unstable it should also be farsightedly unstable (see Example 4 in Section 3.2 for the formal analysis of this example).

Our goal in this paper is to study the set of all farsightedly stable matchings once we assume that players take optimal actions along a farsighted objection path. It is clear that any matching such as \(\mu\) in this example cannot be stable, but it is not immediate from the example what solution concept is going to be used to tackle this question.

**Maximal Farsighted Set**

MFS is an adaptation of the history dependent strong rational expectations farsighted stable set of Dutta and Vartiainen (2020) to one-to-one matching problems, which itself is the history-dependent version of the concept developed in Dutta and Vohra (2017).

We assume that the agents have an expectation that either deems a matching stable or that defines a farsighted objection path to the matching. We further impose the property that it is indeed optimal for any coalition moving along an assigned farsighted objection path to take the particular move assigned by the expectation. That is, no coalition has an incentive to deviate from the assigned paths. MFS corresponds to the stable matchings of such an expectation.

**Overview of the Results**

In a one-to-one matching problem, typically each side of the market is thought to compete with each other to obtain desirable partners. There is a special form of a farsighted objection in which this is not the case. In particular, one side of the market might collaborate and initiate a farsighted objection to swap their partners for mutual improvement. We call such farsighted objections collision.

The other kind of farsighted objections that are useful are improving paths, in which each coalition moving along the path not only improves at the end of the path, but also immediately improves at the matching the coalition imposes.

We use collusion and improving paths in conjunction to define a concept, that we call obtainability that provides an almost complete characterization of farsighted objection paths. Roughly, we say that a matching \(\mu\) is obtainable from \(\mu'\) if \(\mu\) and \(\mu''\) can be partitioned into \(\mu_1, \mu_2\) and \(\mu'_1, \mu'_2\) such that \(\mu_1\) is obtainable from \(\mu'_1\) through collusion and \(\mu_2\) is obtainable from \(\mu'_2\) through an improving path. We say that a set of matchings \(V\) is an obtainable set if for any matching \(\mu\) outside of \(V\), there exists a matching inside of \(V\) that is obtainable from \(\mu\).
A set of matchings \( V \) is a credible set if for any matching \( \mu \in V \) and any deviation from \( \mu \), there exists a matching in \( V \) that is obtainable from the deviation that makes at least one of the deviators worse off. Proposition 1 shows that a set of matchings \( V \) is an MFS if and only if \( V \) is an obtainable credible set of matchings.

Proposition 1 implies that any credible set containing a stable matching is an MFS. In particular, any set of matchings composed of stable matchings is an MFS. Hence, an MFS always exists, but it need not be unique. Although we have these relations between MFS and stable matchings these do not imply that an MFS needs to contain a stable matching.

One problem with MFS is multiplicity. Even in simple examples with a unique stable matching, one can find a considerable number of MFS. Proposition 2 alleviates this problem by characterizing the unique largest MFS that contains all other. Consequently, it also provides a characterization of all matchings that can be supported as the stationary outcomes of an MFS.

The largest consistent set (see Chwe (1994)) has been an influential solution concept with nice properties such as existence and uniqueness under mild conditions. However, it has long been criticized for failing to incorporate the idea of maximality (see Xue (1998), Ray and Vohra (2014)). This refers to the observation that instead of considering the best course of action, individuals form extreme expectations based on pessimism leading to unreasonable predictions in some situations.

Proposition 3 shows that the largest consistent set is immune to this criticism in the space of one-to-one matching problems. That is, it actually embodies the idea of maximality here. This is shown through establishing the equivalence between the largest consistent set and the largest MFS. It is important to note that this does not imply an equivalence between consistent sets and MFS. Indeed, neither an MFS needs to be a consistent set nor a consistent set needs to be an MFS.

Relation to the Literature

The starting point of the farsighted stability literature can be traced back to Harsanyi (1974), who criticized von Neumann and Morgenstern’s stable set (1944) for being myopic. Chwe (1994) formalized Harsanyi’s criticism and developed the solution concepts of the farsighted stable set and the largest consistent set. Although these two concepts have been highly influential they have also been criticized for not embodying the idea of maximality (see Xue (1998), Ray and Vohra (2014)). This refers to the observation that instead of considering the optimal course of action, in these concepts players form extreme expectations based on optimism or pessimism leading to unreasonable predictions in some situations (see Xue (1998)).

There is a growing literature that tries to incorporate this idea of maximality into farsighted solution concepts. For examples, see Dutta and Vohra (2017), Dutta and Vartiainen (2020), Herings, Mauleon and Vannetelbosch (2004), Karos and Kasper (2018), Kimya (2020), and Konishi and Ray (2003). Within this literature we choose to use an adaptation of the solution concept developed by Dutta and Vohra (2017) and later extended by Dutta and Vartiainen (2020). These papers assume that agents have an expectation that either deems an outcome (in our context an outcome is a matching) stable or that defines a farsighted objection path to the outcome. They further impose the property that it is indeed optimal for any coalition moving along an assigned farsighted objection path to take the particular move assigned by the expectation, hence
incorporating the notion of maximality.\footnote{See Jordan (2006) for an earlier use of expectations.}

The literature that applies the insights of the farsighted stability literature into matching problems is more recent. One notable example and the paper closest to ours is Mauleon, Vannetelbosch and Vergote (2011), who completely characterize the farsighted stable set in one-to-one matching problems.\footnote{See Ehlers (2007) for the analysis of von Neumann and Morgenstern’s stable set in one-to-one matching problems.} Interestingly, they show that a set of matchings $V$ is a farsighted stable set if and only if $V$ is composed of a single stable matching. With our Proposition 1, this result implies that any set $V$ obtained as a union of farsighted stable sets is an MFS. Hence, all farsighted stable sets are included in the largest MFS.

An interesting implication of Mauleon, Vannetelbosch and Vertoge (2011)’s result is that the farsighted stable set does not suffer from the problem of maximality in this domain.\footnote{For other examples in which the farsighted stable set does not suffer from the maximality critique see Hirai, Watanabe and Muto (2019) and Ray and Vohra (2015, 2019).} This is because, any singleton farsighted stable set is immune to the critique of maximality (see Ray and Vohra (2019)). Our paper complements the results of Mauleon, Vannetelbosch and Vertoge (2011) in two ways.

First, although Mauleon, Vannetelbosch and Vertoge (2011) show that the farsighted stable set is immune to the maximality critique, we still lack a characterization of the matchings that can be supported by an expectation that respects the idea of maximality. Indeed, in this paper we see that the farsighted stable set, and hence the stable matchings, are only a small subset of the set of matchings that can be supported.

Second, we still do not know whether the largest consistent set suffers from the problem of maximality in one-to-one matching problems. Although the largest consistent set has been fairly influential, it has long been criticized for not embodying the idea of maximality. This paper shows that this issue is not a problem for the largest consistent in the space of one-to-one matching problems.

A precursor to Mauleon, Vannetelbosch and Vertoge (2011)’s paper is Diamantoudi and Xue (2003) who study the farsighted stability properties of hedonic games, of which one-to-one matching problems is a special case. Most relatedly, they show that the core is a farsighted objection to any other outcome implying that each stable matching is a farsighted stable set in one-to-one matching problems. Recently, farsighted models have been extended to incorporate heterogeneous expectations. Bloch and van den Nouweland (2020) and Herings, Mauleon and Vannetelbosch (2017) provide applications of these models to one-to-one matching problems.\footnote{For other work on farsightedness in matching problems see Klijn and Masso (2003), and Mauleon, Molis, Vannetelbosch and Vertoge (2014).}

**Outline**

The paper starts with the formal description of the model and the solution concept in Section 2. Section 3 contains the main results including the characterization of MFS and its comparison to other farsighted solution concepts. Section 4 contains the Appendix, which includes the proofs.
2 The Model

2.1 One-to-one Matching Problems

Let $N$ be a finite set of individuals partitioned into a set of men $M = \{m_1, m_2, ..., m_r\}$ and a set women $W = \{w_1, w_2, ..., w_s\}$. Each agent $i \in N$ has a strict preference relation over the individuals on the other side of the market and the prospect of being alone. We represent the preference of each $m \in M$ by an ordered list, $P(m)$, on the set $W \cup \{m\}$. For instance, $P(m) = w_1, w_2, m, w_3, ..., w_s$ indicates that $m$ prefers $w_1$ to $w_2$, $w_2$ to $w_3$, and he prefers remaining single to matching with anyone else. Similarly, the preference of each $w \in W$ is represented by an ordered list, $P(w)$, on the set $M \cup \{w\}$. Let $P$ denote a preference profile specifying the preference relation of each $m \in M$ and $w \in W$. That is, $P = \{P(m_1), P(m_2), ..., P(m_r), P(w_1), P(w_2), ..., P(w_s)\}$, where $P(i)$ denotes individual $i$’s preference ordering. A one-to-one matching problem is a triple $(M, W, P)$.

A matching is a function $\mu : N \to N$ such that i) for all $m \in M$, $\mu(m) \in W \cup \{m\}$, ii) for all $w \in W$, $\mu(w) \in M \cup \{w\}$, and iii) for all $i \in N$, $\mu(\mu(i)) = i$. The set of all matchings is denoted by $\mathcal{M}$. We say that individual $i \in N$ prefers a matching $\mu$ to a matching $\mu'$, $\mu \succeq_i \mu'$, if $\mu(i)P(i)\mu'(i)$. An individual $i \in N$ is indifferent between a matching $\mu$ and a matching $\mu'$, $\mu \sim_i \mu'$, if $\mu(i) = \mu'(i)$. Finally, $i \in N$ weakly prefers a matching $\mu$ to a matching $\mu'$, $\mu \succeq_i \mu'$, if either $\mu \succeq_i \mu'$ or $\mu \sim_i \mu'$. Note that the relation $\succeq_i$ over $\mathcal{M}$ is complete and transitive.

2.2 The Solution Concept

2.2.1 Enforceability, Individual Rationality and the Core

A coalition $S \subseteq N$ is able to enforce a matching $\mu'$ over a matching $\mu$ if i) every player in $S$ is either single or matched with another player in $S$ at $\mu'$, and ii) every other player is matched to the same partner in $\mu'$ as they were at $\mu$ unless their partner in $\mu$ is in $S$, in which case they are single in $\mu'$. Formally, a coalition $S \subseteq N$ is able to enforce a matching $\mu'$ over a matching $\mu$ if i) for any $i \in S$, $\mu'(i) \in S$, and ii) for any $i \notin S$, $\mu'(i) = i$ if $\mu(i) \in S$ and $\mu'(i) = \mu(i)$ if $\mu(i) \notin S$. We say that a matching $\mu'$ is enforceable from a matching $\mu$ if there exists a coalition that is able enforce $\mu'$ over $\mu$.

This definition of enforceability is the same as the one used in Mauleon, Vannetelbosch and Vertoge (2011), and it respects what Ray and Vohra (2015) call coalitional sovereignty, which is the idea that the deviating coalition should not have the power to rearrange the matchings unrelated to the coalition.

A matching $\mu$ is individually rational if there does not exist an individual $i$ that can enforce $\mu'$ over $\mu$ and $\mu' \succ_i \mu$.

A matching $\mu$ is stable if it is individually rational and there does not exist $(m, w) \in M \times W$ that can enforce $\mu'$ over $\mu$ with the property that $\mu' \succ_m \mu$ and $\mu' \succ_w \mu$.

A matching $\mu$ is in the core if there does not exist a coalition $S \subseteq N$ that can enforce $\mu'$ over $\mu$ with the property that $\mu' \succ_S \mu$.\footnote{For any two matchings $\mu, \mu' \in \mathcal{M}$, we write $\mu' \succ_S \mu$ if $\mu' \succ_i \mu$ for all $i \in S$.}

2.2.2 Maximal Farsighted Set

The notions of core and stability assume that individuals only care about the immediate consequence of their deviation. But, one might anticipate that an initial objection might
be followed by further objections in which case the initial deviator should evaluate the ultimate consequence of its deviation. In this paper we are interested in such farsighted players.

We start with the definition of a farsighted objection, which was introduced by Harsanyi (1974) and formalized by Chwe (1994).

**Definition 1. Farsighted Objection**

A matching $\mu'$ is a farsighted objection to a matching $\mu$ if there exists a sequence of matchings $\mu_0, \mu_1, \ldots, \mu_k$, where $\mu_0 = \mu$ and $\mu_k = \mu'$, and a sequence of coalitions $S_0, S_1, \ldots, S_{k-1}$ such that $S_i$ is able to enforce $\mu_{i+1}$ over $\mu_i$ and $\mu_k \succeq S_i \mu_i$ for all $i = 0, 1, \ldots, k - 1$.

A farsighted objection formalizes the idea that when a coalition deviates from a matching it considers the ultimate consequence of its deviation. But this definition ignores one important issue. How are we to know that it is indeed optimal for each coalition moving along the path to take the prescribed action? In particular, perhaps a coalition moving along the path is assumed to enforce a matching at some point, but they would be better off enforcing some other matching expecting a better outcome from there. This is known as the problem of maximality (see Ray and Vohra (2014)).

Dutta and Vohra (2017) come up with an elegant solution to this problem. They assume that the agents have an expectation that either deems an outcome (in our context an outcome is a matching) stable or defines a farsighted objection path to the outcome. They further impose the property that it is indeed optimal for any coalition moving along an assigned farsighted objection path to take the particular move assigned by the expectation. In particular, if any coalition deviates to another outcome then the assigned path at that outcome would make the coalition worse off. Dutta and Vartiainen (2020) extend Dutta and Vohra (2017)’s solution to incorporate history-dependent expectations. We will use a simple variation of this history-dependent approach.

A history $h$ is a finite sequence of matchings along with the coalitions that generate transitions between matchings. If there is no change in matching then the empty coalition is recorded. For example, $h = (\mu_0, S_1, \mu_1, S_2, \mu_2, \emptyset, \mu_2)$ denotes the history that starts at the matching $\mu_0$ from which the coalition $S_1$ enforces $\mu_1$, from which $S_2$ enforces $\mu_2$ and no more transitions take place at $\mu_2$. Let $H$ denote the set of all histories.

Let $x(h)$ denote the last matching in history $h$. An expectation $E$ is a mapping that assigns a matching and a coalition pair to each history. For each $h$, $E(h) = \{S(h), y(h)\}$, where $y(h)$ is the matching that follows $x(h)$ and $S(h)$ is the coalition implementing the transition. Of course, $S(h)$ is assumed to be able to enforce $y(h)$ from $x(h)$. If $x(h) = y(h)$ then we set $S(h) = \emptyset$.

Let $E^0(h) = h$, $E^1(h) = E(h)$, $E^2(h) = E(h, E^1(h))$, and generally $E^{k+1}(h) = E(h, E^1(h), E^2(h), \ldots, E^k(h))$ for all $k = 0, 1, 2, \ldots$. Similarly, $S^k(h)$ and $y^k(h)$ denote the first and second components of $E^k(h)$. That is, $E^k(h) = (S^k(h), y^k(h))$. An expectation $E$ is absorbing if there exists a $k$ such that $S^k(h) = \emptyset$. We restrict attention to absorbing expectations.

For an absorbing expectation $E$ and a history $h$, let $k^*$ be the first $k$ such that $S^k(h) = \emptyset$. The continuation path generated by $E$ at history $h$ is $\bar{E}(h) = (x(h), E^1(h), \ldots, E^{k^*}(h))$. Let $E^T(h)$ denote the matching the continuation path at $h$ terminates in.

**Definition 2. Stable Expectations**

An expectation $E$ is stable if
1. for any \( h \in H \) such that \( S(h) \neq \emptyset \), \( E(h) \) is a farsighted objection path, and

2. for any \( h \in H \) such that \( S(h) = \emptyset \) there does not exist a coalition \( S \) that can enforce a matching \( \mu' \) from \( x(h) \) such that \( E^T(h, S, \mu') \succ_S x(h) \).

The first condition states that if a continuation path is assigned to a history then it should be a farsighted objection path. The second condition requires that if a matching is deemed stable by an expectation then there should not exist a coalition that can deviate anticipating that the resulting terminal matching will be preferred to the status quo.

On top of these conditions, we also require a maximality condition that ensures that each coalition moving along the expectation is indeed taking the optimal action given the expectation. Suppose \( \mu = \begin{pmatrix} m_1 & m_2 & m_3 \\ w_1 & w_2 & w_3 \end{pmatrix} \) and \( \mu' = \begin{pmatrix} m_1 & m_2 & m_3 & w_1 & w_2 \\ w_3 & m_2 & w_3 & w_1 & w_2 \end{pmatrix} \). The coalition \( \{m_1, m_2\} \) is able to enforce the matching \( \mu' \) over \( \mu \) by separating from their partners. Suppose that the expectation specifies this move from \( \mu \) to \( \mu' \). In such a situation it is reasonable to assume that \( m_1 \) might refuse to do his part of the move and instead decide to do something else. For example, he might decide to match with \( w_3 \) instead and \( w_3 \) might be willing to cooperate with this action. But, what would be the resulting matching if \( m_1 \) deviates by partnering with \( w_3 \)?

Note that the initial move by \( \{m_1, m_2\} \) is actually two independent moves. In one move \( m_1 \) separates from \( w_1 \) and in the other \( m_2 \) separates from \( w_2 \). If we assume that \( m_1 \) decides to deviate from his part, then it is natural to assume that \( m_1 \) has no say over the move by \( m_2 \). That is, \( m_1 \) cannot force \( m_2 \) to continue being a mate for \( w_3 \) by refusing the action he is supposed to take. This argument implies that the deviation by \( \{m_1, w_3\} \) from the prescribed move leads to the matching \( \begin{pmatrix} m_1 & m_2 & m_3 & w_1 & w_2 \\ w_3 & m_2 & w_3 & w_1 & w_2 \end{pmatrix} \).

In general, assume that coalition \( S \) is supposed to move from \( \mu' \) to \( \mu'' \). Then a coalition \( T \) with \( T \cap S \neq \emptyset \) can deviate to a matching \( \mu \) if \( i \) every player in \( T \) is either single or matched with a player in \( T \) at \( \mu \), and \( ii \) all other individuals are matched to their partners in \( \mu'' \) unless their partner in \( \mu'' \) is in \( T \), in which case they are single. Note that, although coalition \( T \) is deviating from the prescribed move from \( \mu' \) to \( \mu'' \), effectively it is the coalition \( S \cup T \) that is enforcing the move from \( \mu' \) to \( \mu \).

The definition below formalizes the stable expectations that satisfy this notion of maximality and defines our solution concept.

**Definition 3. Maximal Farsighted Set**

An expectation \( E \) is maximal if \( E \) is stable and for any \( h \in H \) such that \( S(h) \neq \emptyset \), there does not exist a coalition \( T \) with \( T \cap S(h) \neq \emptyset \) and a matching \( \mu \) such that the following conditions hold

- \( i \) If \( i \in T \) then \( \mu(i) \in T \), and \( ii \) for any \( i \notin T \), \( \mu(i) = y(h)(i) \) if \( y(h)(i) \notin T \) and \( \mu(i) = i \) if \( y(h)(i) \in T \).

- \( E^T(h, S(h) \cup T, \mu) \succ_T E^T(h) \)

The set of stationary points of a maximal expectation is an MFS.

The first condition defines the set of matchings \( T \) can impose by deviating from the prescribed move, and the second condition requires the profitability of the deviation. A set of matchings is an MFS if there exists a maximal expectation that can support it as
the set of stationary matchings of the expectation. It is important to note that an MFS is not necessarily a farsighted stable set and it should not be confused with a farsighted stable set that satisfies the maximality property, as in Ray and Vohra (2019).

Dutta and Vohra (2017) and Dutta and Vartiainen (2020) consider two different solution concepts, rational expectations farsighted stable set and strong rational expectations farsighted stable set, based on two different notions of maximality. The notion of maximality that MFS uses is in the same spirit as the one corresponding to the strong rational expectations farsighted stable set, but it is different. Strong rational expectations farsighted stable set assumes that if $S$ is supposed to move then any coalition $T$ with $S \cap T \neq \emptyset$ might deviate, but when $T$ deviates they also annul the actions taken by $S \setminus T$. In the specific environment of matching problems, since each coalition might be taking an action that is composed of several independent actions we find it reasonable to assume that when $T$ deviates they cannot annul any independent action taken by $S \setminus T$. This is the only difference between the notion that we use and the notion Dutta and Vartiainen (2020) uses.\(^6\) On the other hand, the notion of maximality rational expectations farsighted stable set uses is weaker than the notion we use. Please see Section 3.4.3 for the formal definition of rational expectations farsighted stable set and its characterization in one-to-one matching problems.

3 The Results

3.1 Collusion and Improving Paths

In a matching problem, typically each side of the market is thought to compete with each other to obtain desirable partners. There is a special form of a farsighted objection in which this is not the case. In particular, one side of the market might collaborate and initiate a farsighted objection to swap their partners for mutual improvement. We call such farsighted objections collusion.

**Definition 4. Collusion**

A matching $\mu'$ can be obtained from a matching $\mu$ through collusion if there exists a coalition $S \subseteq N$ such that the following conditions hold:

- Either $S \subseteq M$ or $S \subseteq W$
- $\mu' \succ_S \mu$
- $\mu'(i) \neq i$ for all $i \in S$
- $\mu'(i) = \mu(i)$ if $i \notin S$ and $\mu(i) \notin S$

A matching $\mu'$ can be obtained from a matching $\mu$ through collusion if there exists a coalition composed of only men or only women, who can swap their partners among each other in such a way that every individual within the coalition improves. The following example demonstrates the concept of collusion.

\(^6\)Dutta and Vartiainen (2020) could not have proposed to use our notion of maximality, because they were working on an abstract environment to define their solution as general as possible. In contrast, in our environment the meaning of enforceability is quite clear and this allows us to modify the maximality condition in a suitable way. This flexibility provided by the language of expectations is one of the strengths of the framework developed by Dutta and Vohra (2017).
**Example 1.** Consider a matching problem with $M = \{m_1, m_2, m_3\}$ and $W = \{w_1, w_2, w_3\}$. The preferences of individuals are as follows:

<table>
<thead>
<tr>
<th></th>
<th>$w_1$</th>
<th>$w_2$</th>
<th>$w_3$</th>
<th>$m_1$</th>
<th>$m_2$</th>
<th>$m_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w_1$</td>
<td>$w_1$</td>
<td>$w_2$</td>
<td>$w_3$</td>
<td>$m_1$</td>
<td>$m_2$</td>
<td>$m_3$</td>
</tr>
<tr>
<td>$w_2$</td>
<td>$w_3$</td>
<td>$w_2$</td>
<td>$w_3$</td>
<td>$m_2$</td>
<td>$m_3$</td>
<td>$m_3$</td>
</tr>
<tr>
<td>$w_3$</td>
<td>$w_1$</td>
<td>$w_1$</td>
<td>$w_1$</td>
<td>$m_3$</td>
<td>$m_2$</td>
<td>$m_1$</td>
</tr>
</tbody>
</table>

Observe that the coalition $\{w_1, w_2, w_3\}$ can obtain the matching $\mu' = (m_1 \ m_2 \ m_3) w_1 \ w_3 \ w_2$ from $\mu = (m_1 \ m_2 \ m_3) w_3 \ w_2 \ w_1$ through collusion. To see this one simply needs to observe that i) each woman in the coalition $\{w_1, w_2, w_3\}$ is better off at $\mu'$, and ii) each woman in the coalition $\{w_1, w_2, w_3\}$ is matched with a man who was initially matched with a woman in the coalition $\{w_1, w_2, w_3\}$. In other words, as we move from $\mu$ to $\mu'$, the individuals in coalition $\{w_1, w_2, w_3\}$ are swapping their partners for mutual-improvement.

It is easy to see that if $\mu'$ can be obtained from a matching $\mu$ through collusion and if $\mu'$ is individually rational then $\mu'$ is a farsighted objection to $\mu$.

**Lemma 1.** Suppose $\mu'$ is individually rational and it can be obtained from $\mu$ through collusion. Then $\mu'$ is a farsighted objection to $\mu$.

Suppose that $\mu'$ is a farsighted objection to $\mu$. Then it turns out that either i) $\mu'$ is obtainable from $\mu$ through collusion, or ii) there exists a pair $(i, j)$, where possibly $i = j$, that is matched in $\mu'$ and they prefer each other to their partners in $\mu$.

Diamantoudi and Xue (2003), and Mauleon, Vannetelbosch and Vertoge (2011) have shown that every stable matching is a farsighted objection to any other matching. Coupled with the observation above this implies that any stable matching can be obtained from any other stable matching through collusion. Hence, the existence of a collusion-proof stable matching implies the uniqueness of a stable matching. The lemma below formally states these results.

**Lemma 2.**

- Suppose that $\mu'$ is a farsighted objection to $\mu$. Then either $\mu'$ is obtainable from $\mu$ through collusion or there exists $i, j \in N$, where possibly $i = j$, such that $\mu'(i) = j$ and $\mu' \succ_{\{i,j\}} \mu$.
- Let $\mu$ and $\mu'$ be two distinct stable matchings. Then $\mu'$ is obtainable from $\mu$ through collusion.
- There exists a unique stable matching if there exists a collusion proof stable matching.

The other kind of farsighted objections that are going to be useful are improving paths in which each coalition not only improves at the end of the path, but also immediately improves at the matching the coalition imposes.
Definition 5. Improving Paths

A farsighted objection path \((\mu_0, S_1, m_1, \ldots, S_k, \mu_k)\) is an improving path if \(\mu_i \succ_S \mu_{i-1}\) for all \(i = 1, 2, \ldots, k\).

The fact that improving paths are farsighted objection paths are embedded in the definition. Hence, both collusion and improving paths form farsighted objections in the space of individually rational matchings. On the other hand, it is easy to construct farsighted objection paths that are neither improving paths nor collusive paths. Nevertheless, we can use collusion and improving paths in conjunction to define a concept, that we call obtainability, to provide an almost complete characterization of farsighted objection paths.

Given a one-to-one matching problem \((M, W, P)\) and any subset \(N'\) of \(N\), the problem restricted to \(N'\) is the one-to-one matching problem \((M \cap N', W \cap N', P_{N'})\), where \(P_{N'}\) is the preference profile restricted to \(N'\). We say that a matching \(\mu'\) in the game restricted to \(N'\) agrees with the matching \(\mu\) in the original game if \(\mu'(i) = \mu(i)\) for all \(i \in N'\).

Given a matching \(\mu\), we say that \(\mu_1\) and \(\mu_2\) form a partition of \(\mu\) if there exists a partition of \(\mu\), \(\mu'_1\) and \(\mu'_2\), such that \(\mu'_1\) can be obtained from \(\mu_1\) through collusion and \(\mu'_2\) can be obtained from \(\mu_2\) through an improving path.

Example 2. Consider a matching problem with \(M = \{m_1, m_2, m_3, m_4\}\) and \(W = \{w_1, w_2, w_3, w_4\}\). The preferences of individuals are as follows:
Consider the matching $\mu = \left( \begin{array}{cccc} m_1 & m_2 & m_3 & m_4 \\ w_1 & w_2 & w_3 & w_4 \end{array} \right)$ and $\mu' = \left( \begin{array}{cccc} m_1 & m_2 & m_3 & m_4 \\ w_2 & w_1 & w_3 & w_4 \end{array} \right)$. To see that $\mu'$ is obtainable from $\mu$, observe that $\mu_1 = \left( \begin{array}{cc} m_1 & m_2 \\ w_4 & w_2 \end{array} \right)$ and $\mu_2 = \left( \begin{array}{cc} m_3 & m_4 \\ w_1 & w_3 \end{array} \right)$ forms a partition of $\mu$. Furthermore, $\mu'_1 = \left( \begin{array}{cc} m_1 & m_2 \\ w_2 & w_4 \end{array} \right)$ is obtainable from $\mu_1$ through an improving path and $\mu'_2 = \left( \begin{array}{cc} m_3 & m_4 \\ w_3 & w_1 \end{array} \right)$ is obtainable from $\mu_2$ through collusion. Finally, when we merge $\mu'_1$ and $\mu'_2$, we obtain $\mu'$. See Figure 2, where this is demonstrated.

Since obtainability partitions a matching and then uses the definitions of improving paths and collusion, it is immediate that if $\mu'$ is individually rational and obtainable from $\mu$, then $\mu'$ is a farsighted objection to $\mu$. It turns out that the converse is also true. That is, if an individually rational matching $\mu'$ is a farsighted objection to $\mu$ then it is obtainable from $\mu$.

**Lemma 3.** Let $\mu'$ be an individually rational matching. Then $\mu'$ is obtainable from $\mu$ if and only if $\mu'$ is a farsighted objection to $\mu$.

We have already seen why if an individually rational $\mu'$ is obtainable from $\mu$ then $\mu'$ is farsighted objection to $\mu$. The other direction can be shown through repeated application of Lemma 2. Please see the Appendix for the proof.

### 3.2 Maximal Farsighted Set

In this section we provide a complete characterization of MFS in one-to-one matching problems and then we study the implications of this characterization.

**Definition 7.** Let $V$ be a set of matchings. A matching $\mu \in \mathcal{M}$ is credible given $V$ if for any $S \subseteq N$ and $\mu'$ such that $\mu'$ is enforceable over $\mu$ by $S$, there exists $\mu'' \in V$, where either $\mu'' = \mu'$ or $\mu''$ is obtainable from $\mu'$, such that $\mu'' \not\succ_S \mu$.

A set of matchings $V$ is credible if every $\mu \in V$ is credible given $V$.

A credible set of matchings have the property that any deviation from the set can be countered by a matching in the set that is obtainable from the deviation. The definition is close to the definition of a consistent set (see Chwe (1994)). Note that a set $V$ is credible if $\mu \in V$ only if $\mu$ is credible given $V$. The main difference between a credible set and a consistent set is that a consistent set is defined with an ‘if and only if’, instead of an ‘only if’, which implies that credibility is a weaker notion than consistency. A detailed comparison to consistent sets is available in Section 3.3.

Our first main result provides a complete characterization of MFS through obtainability and credible sets.
Proposition 1. A set of matchings $V$ is an MFS if and only if $V$ is an obtainable credible set of matchings.

The formal proof of Proposition 1, which uses the notion of coherent set of paths developed in Dutta and Vartiainen (2020), can be found in the Appendix. Here, we provide a sketch of the proof.

First, we will show that any MFS needs to be an obtainable credible set of matchings. If a matching is not individually rational, then it cannot be a stationary point of an MFS. This is because, the singleton with an unacceptable partner can always deviate to a matching in which he is single and no farsighted objection from this matching can assign an unacceptable partner to this individual. Since MFS is always a subset of individually rational matchings and since each MFS needs to contain a farsighted objection to any matching outside the MFS, by Lemma 3 any MFS is an obtainable set of matchings. Finally, an MFS needs to be a credible set, as otherwise a deviation from a stationary matching cannot be prevented.

For the other direction, suppose that $V$ is an obtainable credible set of matchings. We will construct a maximal expectation $E$ that supports $V$ as its stationary outcomes, implying that $V$ is an MFS.

First, as in Ray and Vohra (2019) we define a canonical farsighted objection path. Observe that whenever an individually rational matching $\mu'$ is obtainable from $\mu$ then it can be obtained from $\mu$ through a canonical farsighted objection path in which first the coalition of every non-single individual who is better off in $\mu'$ separates from her partner. All matched pairs in the matching we obtain are matched to their partners in $\mu'$. Now, the coalition composed of all singletons with a partner in $\mu'$ can match to form $\mu'$. Any farsighted objection path in this form will be called a canonical objection path.

For any initial history $h = \mu$, if $\mu \in V$ we set $E(h) = (\emptyset, \mu)$, and if $\mu \notin V$ then we assign a canonical objection path to a matching in $V$ that is obtainable from $\mu$.

We need to ensure that there is no deviation from the assigned paths. As an example, take any history $h$ such that $S(h) = \emptyset$ and $x(h) = \mu \in V$. If a coalition $S$ deviates to a matching $\mu'$ then by the definition of a credible set there exists another matching $\mu'' \in V$ that is obtainable from $\mu'$ that makes one of the deviators worse off. We assign the canonical objection path from $\mu'$ to $\mu''$ to this deviation.

These paths would deter any deviation from a stationary outcome in $V$. But, we should also make sure that these paths themselves are immune to deviation. To get an idea on how this is achieved consider the assigned canonical objection path from $\mu'$ to $\mu''$ above. Suppose that coalition $S$ is supposed to move at some point, but coalition $T$ with $T \cap S \neq \emptyset$ deviates to the matching $\mu_1$. Note that if $T$ doesn’t deviate, it expects to obtain the outcome $\mu''$. Furthermore, if $T$ deviates at $\mu''$ then the deviation can be prevented by some matching $\mu'''$ that makes somebody in $T$ worse off. One can show that since $\mu'''$ is obtainable from the matching $T$ imposes at $\mu''$, it is also obtainable from $\mu_1$. Hence, a canonical objection path from $\mu_1$ to $\mu'''$ would prevent this deviation.

We can continue this argument to also make sure that there are no deviations from these ‘punishment paths’, and the paths that are used to punish deviations from punishment paths, and so on. The formal proof uses the notion of coherent set of paths developed by Dutta and Vartiainen (2020) to achieve this.

Now, we discuss the implications of Proposition 1. Note that any credible set has to be a subset of individually rational matchings. This is because, if a matching is not individually rational then the singleton with an unacceptable partner can always deviate to a matching in which he is single and no matching that is obtainable from this matching
can assign an unacceptable partner to this individual. Hence, every MFS is composed of individually rational matchings.

Any set containing a stable matching is an obtainable set, because a stable matching is a farsighted objection to any other matching (see Mauleon, Vannetelbosch and Vertogen (2011), and Diamantoudi and Xue (2003)). This implies that any credible set containing a stable matching is an MFS.

Furthermore, any set composed of stable matchings is a credible set as any deviation from a stable matching can be prevented by a farsighted objection back to the stable matching itself. Since sets including stable matchings are also obtainable, we have that any set composed of stable matchings is an MFS. This implies that an MFS always exists, but it need not be unique.

Though we have these relationships between stable matchings and MFS, these do not imply that an MFS needs to contain a stable matching. See Example 3 below for an MFS that does not contain a stable matching. The corollary below summarizes these implications of Proposition 1.

**Corollary 1.**

- If a set of matchings \( V \) is an MFS, then \( V \) is a subset of individually rational matchings.

- Any credible set of matchings containing a stable matching is an MFS.

- Any set of matchings composed of stable matchings is an MFS. Hence, an MFS always exists, but need not be unique.

- An MFS might have an empty intersection with the set of stable matchings.

Before seeing an example that demonstrates these results, we state the following lemma that makes it easier to check whether a set is credible. Given a set of matchings \( V \), a matching \( \mu \), a coalition \( S \) and a matching \( \mu' \) enforceable by \( S \) over \( \mu \), we say that the deviation of \( S \) from \( \mu \) to \( \mu' \) is credible if for every \( \mu'' \in V \), with either \( \mu'' = \mu' \) or \( \mu'' \) is obtainable from \( \mu' \), we have that \( \mu'' \succeq_S \mu \).

**Lemma 4.**

Let \( V \) be an obtainable set of individually rational matchings. If \( \mu \in V \) is not credible given \( V \) then there exists a coalition \( S \) that has a credible deviation to \( \mu' \), where

- \( S \) does not include any player for whom \( \mu \) is the most preferred matching in \( V \),

- there exists a pair \( (m, w) \in M \times W \) such that \( \{m, w\} \subseteq S \), \( \mu'(m) = w \), and \( \mu' \succ_S \mu \), and

- for every \( i \in S \), \( \mu'(i) \neq i \).

The example below shows an MFS that does not contain any stable matching.

**Example 3.** Consider a matching problem with \( M = \{m_1, m_2, m_3, m_4\} \) and \( W = \{w_1, w_2, w_3, w_4\} \). The preferences of individuals are as follows:
Let $\mu^* = \left( \begin{array}{cccc} m_1 & m_2 & m_3 & m_4 \\ w_2 & w_1 & w_4 & w_3 \end{array} \right)$, $\mu_1 = \left( \begin{array}{cccc} m_1 & m_2 & m_3 & m_4 \\ w_3 & w_2 & w_4 & w_1 \end{array} \right)$, $\mu_2 = \left( \begin{array}{cccc} m_1 & m_2 & m_3 & m_4 \\ w_1 & w_2 & w_4 & w_3 \end{array} \right)$, and $\mu_3 = \left( \begin{array}{cccc} m_1 & m_2 & m_3 & m_4 \\ w_2 & w_4 & w_1 & w_3 \end{array} \right)$.

$\mu^*$ is the unique stable matching in this problem. We will show that the set $V = \{ \mu_1, \mu_2, \mu_3 \}$, which does not contain a stable matching, is an MFS. Observe that $\mu_1$ is obtainable from the stable matching $\mu^*$ through collusion by $\{m_1, m_2, m_4\}$. Similarly, one can show that for any $\mu \notin V$, there exists a $\mu' \in V$ that is obtainable from $\mu$.\footnote{Note that $\mu_1$ is the most preferred matching for $m_1, m_2$ and $m_4$ and since the only pair that can object to $\mu_1$ is $(m_3, w_1)$, by Lemma 4 if there is a credible deviation from $\mu_1$ then the deviation by $(m_3, w_1)$ to $\left( \begin{array}{cccc} m_1 & m_2 & m_3 & m_4 \\ w_3 & w_2 & w_1 & m_4 \end{array} \right)$ is credible. But, $\mu_2$ is obtainable from this matching through an improving path, and $\mu_2 \not\sim_{m_3} \mu_1$. Hence, $\mu_1$ is credible given $V$.}

Consider the matching $\mu_2$. The only pair that can object to $\mu_2$ is $(m_1, w_2)$, and $\mu_2$ is the most preferred matching in $V$ for $w_1, w_3$ and $w_4$. By Lemma 4, this implies that if there is a credible deviation from $\mu_2$ then the deviation by $(m_1, w_2)$ to $\left( \begin{array}{cccc} m_1 & m_2 & m_3 & m_4 \\ w_3 & w_2 & w_1 & m_4 \end{array} \right)$ is credible. But, $\mu_1$ is obtainable from this matching through an improving path, and $\mu_1 \not\sim_{w_2} \mu_2$. Hence, $\mu_2$ is credible given $V$.

Finally, consider the matching $\mu_3$. By Lemma 4 if there is a credible deviation to $\mu_3$ then there is a credible deviation that includes the pair $(m_2, w_1)$ and that does not include the player $m_3$.\footnote{This is because $(m_2, w_1)$ is the only pair that has an objection to $\mu_3$ and $\mu_3$ is the most preferred matching for $m_3$ in $V$.} Note that any such deviation leaves $m_3$ single. But, since $\mu_1$ is the most preferred matching for $m_1, m_2$ and $m_4$, it is obtainable from any matching in which $m_3$ is single, and $\mu_1 \not\sim_{w_1} \mu_3$. Hence, $\mu_3$ is also credible given $V$.

This shows that $V$ is a credible set. Since $V$ is also obtainable, it is an MFS that does not contain any stable matching.

The example above also demonstrates the multiplicity problem of MFS. We have seen that $V = \{ \mu_1, \mu_2, \mu_3 \}$ is an MFS. Since $\mu^*$ is a stable matching, $V_1 = \{ \mu^* \}$ is also an MFS. Yet another MFS is $V_2 = \{ \mu^*, \mu_1, \mu_2, \mu_3 \}$.\footnote{Since $V$ is credible and $V_1 \supseteq V$, any $\mu \in V$ is also credible given $V_2$. Since $\mu^*$ is stable, it is also credible given $V_2$. Hence, $V_2$ is a credible set. By Corollary 1, any credible set containing a stable matching is an MFS.}
It turns out that the problem of multiplicity is alleviated by the fact that there exists a unique largest MFS that contains all other. The following proposition characterizes this largest MFS, consequently it also provides a characterization of all matchings that can be supported through an MFS.

**Proposition 2.** Let $\mathcal{M}_0$ denote the set of all individually rational matchings. Suppose $\mathcal{M}_k$ is defined for all $k = 0, 1, \ldots, n - 1$. Define $\mathcal{M}_n$ as
\[
\mathcal{M}_n = \{ \mu \in \mathcal{M}_{n-1} | \mu \text{ is credible given } \mathcal{M}_{n-1} \}
\]

Let $k^*$ be the first integer $k$ for which $\mathcal{M}_{k^*} = \mathcal{M}_{k^*+1}$, and let $\mathcal{M}^* = \mathcal{M}_{k^*}$.

- Take any $\mu \in \mathcal{M}$ There exists an MFS that includes $\mu$ if and only if $\mu \in \mathcal{M}^*$.
- $\mathcal{M}^*$ is an MFS. Furthermore, it is the unique MFS that contains all other.

Here we provide a sketch of the proof, for details please see the formal proof in the Appendix. It is immediate that $\mathcal{M}^*$ is a credible set. Furthermore, $\mathcal{M}^*$ is an obtainable set as it always includes all stable matchings. By Proposition 1, $\mathcal{M}^*$ is an MFS.

Now we need to show that no matching $\mu \notin \mathcal{M}^*$ can be a part of an MFS. If $\mu \notin \mathcal{M}_1$ then there exists a coalition that can enforce $\mu'$ from $\mu$ and every individually rational matching that is obtainable from $\mu'$ would make this coalition better off. Since any MFS needs to be a subset of individually rational matchings, this observation implies that $\mu$ cannot be a part of any MFS. Now that we established any MFS needs to be a subset of $\mathcal{M}_1$, a similar argument can be applied to show that no $\mu \notin \mathcal{M}_2$ can be a part of an MFS. Continuing this way, we can establish that no matching $\mu \notin \mathcal{M}^*$ can be a part of an MFS.

The following example demonstrates Proposition 2.

**Example 4.** Consider a matching problem with $M = \{m_1, m_2, m_3\}$ and $W = \{w_1, w_2, w_3, w_4\}$. The preferences of individuals are as follows:

\[
\begin{array}{ccccccc}
  & m_1 & m_2 & m_3 & w_1 & w_2 & w_3 & w_4 \\
 w_3 & w_1 & w_1 & m_1 & m_1 & m_2 & m_1 \\
w_2 & w_3 & w_3 & m_3 & m_3 & m_3 & m_2 \\
w_1 & w_2 & w_2 & m_2 & m_2 & m_1 & m_3 \\
m_1 & m_1 & m_3 & w_3 & w_2 & w_3 & w_4 \\
w_4 & m_2 & w_4 \\
\end{array}
\]

Let $\mu_1 = \begin{pmatrix} m_1 & m_2 & m_3 & w_1 \\ w_2 & w_3 & w_1 & w_4 \end{pmatrix}$, $\mu_2 = \begin{pmatrix} m_1 & m_2 & m_3 & w_4 \\ w_1 & w_3 & w_2 & w_4 \end{pmatrix}$, $\mu_3 = \begin{pmatrix} m_1 & m_2 & m_3 & w_4 \\ w_3 & w_1 & w_2 & w_4 \end{pmatrix}$.

$\mu_1$ is the unique stable matching in this problem and the set $V = \{\mu_1, \mu_2, \mu_3\}$ is the largest MFS. To see why, first take any matching $\mu$ in which $\mu(m_3) = w_1$ and $\mu(m_2) \neq w_3$. Consider the deviation from $\mu$ by $\{m_2, w_3\}$ to $\mu'$ in which they are matched. There is no matching that is obtainable from $\mu'$ that makes either $m_2$ or $w_3$ worse off. Hence, the deviation is credible and $\mu \notin \mathcal{M}_1$.

Take any matching $\mu$ in which $\mu(m_1) = w_2$ and $\mu(m_3) \neq w_1$ and consider the deviation by $\{m_3, w_1\}$ to $\mu'$ in which they are matched. There is no matching that is obtainable from $\mu'$ that makes either $m_3$ or $w_1$ worse off. Hence, the deviation is credible and $\mu \notin \mathcal{M}_1$. 

16
We have shown that in any matching in \( M_1 \), if \( m_3 \) is matched with \( w_1 \) then \( m_2 \) is matched with \( w_3 \), and if \( m_1 \) is matched with \( w_2 \) then \( m_3 \) is matched with \( w_1 \). Take any matching \( \mu \) in which \( \mu(m_3) = w_3 \) and \( \mu(m_1) \neq w_2 \) and consider the deviation by \( \{m_1, w_2\} \) to \( \mu' \) in which they are matched. The only matchings that are obtainable from \( \mu' \) and that could make either \( m_1 \) or \( w_2 \) worse off are the ones in which \( m_3 \) is matched with \( w_1 \) and \( m_1 \) is matched with \( w_3 \). But we have already shown that in any matching in \( M_1 \) in which \( m_3 \) is matched with \( w_1 \), we also have that \( m_2 \) is matched with \( w_3 \). Hence, there is no matching in \( M_1 \) that is obtainable from \( \mu' \) and that makes either \( m_1 \) or \( w_2 \) worse off, implying that the deviation is credible given \( M_1 \) and \( \mu \notin M_2 \).

Similar arguments can be used to eliminate all matchings outside \( V \) in the first two stages of elimination. That is, \( M_2 \subseteq V \). Now, we will show that any \( \mu \in V \) is credible given \( V \), which implies that \( \mu \) is credible given \( M_2 \) and \( V \) is indeed the largest MFS.

Since \( \mu_1 \) is the stable matching it is credible given any set, including \( V \). By Lemma 4 if there is a credible deviation from \( \mu_2 \) then the deviation by \( \{m_1, w_2\} \) is credible. But, \( \mu_3 \) is obtainable through an improving path from the matching \( \{m_1, w_2\} \) imposes, and \( \mu_3 \neq \{m_2, w_2\} \). Hence, \( \mu_2 \) is credible given \( V \). Consider \( \mu_3 \). By Lemma 4 if there is a credible deviation from \( \mu_3 \) then the deviation by \( \{m_3, w_1\} \) or by \( \{m_3, w_3\} \) should be credible. In either case \( \mu_2 \) is obtainable from the matching imposed after the deviation and \( \mu_2 \neq \mu_1 \mu_3 \). Hence, \( \mu_3 \) is also credible given \( V \). This establishes that the largest MFS is indeed \( V = \{\mu_1, \mu_2, \mu_3\} \).

Although there is a unique largest MFS, it is still of interest to find conditions under which there truly exists a unique MFS. It can be shown that a unique MFS only exists if there is a collusion-proof stable matching. However, it turns out that the existence of a collusion-proof stable matching does not guarantee the uniqueness of MFS. A simple condition that guarantees uniqueness is Banerjee, Konishi and Sonmez (2001)’s top coalition property.

In the context of one-to-one matching problems, the top coalition property states that for any set of players \( S \subseteq N \), either i) there exists \((m, w) \in M \times W \) with \( m, w \in S \) such that \( m \) is the most preferred partner for \( w \) in \( S \) and \( w \) is the most preferred partner of \( m \) in \( S \), or ii) there exists \( i \in N \) that prefers to remain single over matching with anyone else in \( S \). This pair or singleton is called the top coalition in \( S \). The top coalition matching is \( \mu^* = \{S_1, S_2, ..., S_n\} \), where \( S_1 \) is the top coalition in \( N \), \( S_2 \) is the top coalition in \( N \setminus S_1 \), and so on.

The top coalition property implies the uniqueness of a stable matching (see Banerjee, Konishi and Sonmez (2001)) and the farsighted stable set (see Mauleon, Vannetelbosch and Vertoge (2011)). It also implies that the largest consistent set is composed of the top coalition matching (see Diamantoudi and Xue (2003)). Unsurprisingly, these results also extend to the MFS. The lemma below establishes this.

**Lemma 5.** There exists a unique MFS only if there exists a collusion-proof stable matching. Furthermore, if the matching problem satisfies the top coalition property then there exists a unique MFS composed of the unique collusion-proof stable matching, which is the top coalition of the matching problem.

The characterization of MFS with obtainable credible sets is reminiscent of the definition of the consistent set of Chwe (1994). Furthermore, the largest MFS reminds one of the largest consistent set. In general, neither an MFS needs to be a consistent set nor a consistent set needs to be an MFS, but the largest MFS turns out to be equal to the largest consistent set. We discuss this relation in the next section.
3.3 The Largest Consistent Set

The largest consistent set (Chwe (1994)) has been an influential solution concept with nice properties such as uniqueness and existence under mild conditions. However, it has been criticized for not incorporating the idea of maximality (see Xue (1998) and Ray and Vohra (2014)). This refers to the observation that instead of considering the best course of action, individuals form extreme expectations based on pessimism leading to unreasonable predictions in some situations.

The analysis in this section shows that the largest consistent set embodies the idea of maximality in the space of one-to-one matching problems. We know this, because Proposition 3 below establishes that the largest consistent set is equal to the largest MFS in one-to-one matching problems. We start with the definition of the largest consistent set.

**Definition 8.** A set of matchings $V \subseteq M$ is a consistent set if $\mu \in V$ if and only if for all $\mu'$ and coalition $S$ such that $S$ can enforce $\mu'$ over $\mu$, there exists $\mu'' \in V$, where $\mu'' = \mu'$ or $\mu''$ is a farsighted objection to $\mu'$, such that $\mu'' \not\succ S \mu$.

The definition of a consistent set is reminiscent of the definition of a credible set, but the two are different. The essential difference is that a credible set is defined with an ‘only if’ instead of an ‘if and only if’, which implies that credibility is a weaker notion than consistency. On the other hand, MFS is characterized together with credibility and obtainability, whereas a consistent set need not be obtainable. Hence, neither an MFS needs to be a consistent set nor a consistent set needs to be an MFS. The following examples demonstrate this. We start with an example of an MFS that is not a consistent set.

**Example 5.** Consider a matching problem with $M = \{m_1, m_2, m_3\}$ and $W = \{w_1, w_2, w_3\}$. The preferences of individuals are as follows:

<table>
<thead>
<tr>
<th></th>
<th>$m_1$</th>
<th>$m_2$</th>
<th>$m_3$</th>
<th>$w_1$</th>
<th>$w_2$</th>
<th>$w_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w_1$</td>
<td>$w_2$</td>
<td>$w_3$</td>
<td>$m_2$</td>
<td>$m_3$</td>
<td>$m_1$</td>
<td></td>
</tr>
<tr>
<td>$w_2$</td>
<td>$w_3$</td>
<td>$w_1$</td>
<td>$m_3$</td>
<td>$m_1$</td>
<td>$m_2$</td>
<td></td>
</tr>
<tr>
<td>$w_3$</td>
<td>$w_1$</td>
<td>$w_2$</td>
<td>$m_1$</td>
<td>$m_2$</td>
<td>$m_3$</td>
<td></td>
</tr>
<tr>
<td>$m_1$</td>
<td>$m_2$</td>
<td>$m_3$</td>
<td>$w_1$</td>
<td>$w_2$</td>
<td>$w_3$</td>
<td></td>
</tr>
</tbody>
</table>

Let $\mu_1 = \begin{pmatrix} m_1 & m_2 & m_3 \\ w_1 & w_2 & w_3 \end{pmatrix}$, $\mu_2 = \begin{pmatrix} m_1 & m_2 & m_3 \\ w_3 & w_1 & w_2 \end{pmatrix}$, $\mu_3 = \begin{pmatrix} m_1 & m_2 & m_3 \\ w_2 & w_3 & w_1 \end{pmatrix}$.

Both $\mu_1$, $\mu_2$ and $\mu_3$ are stable matchings. By Corollary 1, the set $V = \{\mu_1, \mu_2\}$ is an MFS, hence $V$ is also a credible set. However, $V$ is not a consistent set. This is because, for any coalition and for any matching $\mu'$ this coalition can enforce from $\mu_3$, there exists a matching $\mu \in V$ that is a farsighted objection to $\mu'$ and that makes at least somebody in the deviating coalition worse off.

The next example shows a consistent set that is not an MFS.

**Example 6.** Consider the matching problem in Example 3. For convenience we rewrite the preferences below.
Let \( \mu_1 = \begin{pmatrix} m_1 & m_2 & m_3 & m_4 \\ w_3 & w_2 & w_1 & w_4 \end{pmatrix} \), \( \mu_2 = \begin{pmatrix} m_1 & m_2 & m_3 & m_4 \\ w_1 & w_2 & w_4 & w_3 \end{pmatrix} \).

We will show that the set \( V' = \{ \mu_1, \mu_2 \} \) is a consistent set that is not an MFS. First we will show that \( V' \) is a consistent set. Note that the arguments used in Example 3 to show that \( \mu_1 \) and \( \mu_2 \) are credible do not use any matching outside \( V' \), so those arguments still apply implying that \( V' \) is a credible set. This means that we only need to show the ‘if’ part of the definition of a consistent set.\(^{10}\)

First observe that in any matching in which someone is single, two singletons can together deviate by matching. No matching in set \( V' \) can make the deviators worse off since in all matchings in \( V' \) the deviators are matched with someone, who they prefer to being single. Observe that in any matching in which \( m_2 \) is matched with someone other than \( w_2 \), \( m_2 \) can deviate alone. No matching in set \( V' \) can make the deviator worse off since in all matchings in \( V' \), \( m_2 \) is matched with his most preferred individual. In any matching in which \( w_4 \) is matched with \( m_1 \) or \( m_2 \), \( w_4 \) can deviate alone. No matching in set \( V' \) can make the deviator worse off since in all matchings in \( V' \), \( w_4 \) is matched with \( m_3 \) who is preferred to \( m_1 \) and \( m_2 \). In any matching in which \( m_4 \) is matched with \( w_4 \), \( m_4 \) can deviate alone. No matching in set \( V' \) can make the deviator worse off since in all matchings in \( V' \), \( m_4 \) is matched with someone who he prefers to \( w_4 \). This exhausts all matchings outside the set \( V' \). Hence, from any matching outside \( V' \) there exists a deviation that cannot be prevented by a threat from \( V' \), implying that \( V' \) is indeed a consistent set.

To see that \( V' \) is not an MFS, let \( \mu \) be a matching in which \( \mu(m_3) = w_1 \) and \( \mu(m_1) = w_2 \). Neither \( \mu_1 \) nor \( \mu_2 \) is obtainable from \( \mu \), implying that \( V' \) is not an obtainable set. By Proposition 1, \( V' \) is not an MFS.

Although neither an MFS needs to be a consistent set nor a consistent set needs to be an MFS, the largest MFS turns out to be equal to the largest consistent set. Perhaps more importantly, this implies that the largest consistent set does not suffer from the problem of maximality in one-to-one matching problems.

**Proposition 3.** A set of matchings \( V \) is the largest consistent set if and only if \( V \) is the largest MFS.

To prove this result, we first show that the set \( \mathcal{M}^* \) defined in Proposition 2 is a consistent set, which implies that the largest MFS is a subset of the largest consistent set. Then, we show that if the largest consistent set is a subset of \( \mathcal{M}_k \) (defined in Proposition 2) then it is necessarily a subset of \( \mathcal{M}_{k+1} \). Coupled with the observation that the largest consistent set is a subset of \( \mathcal{M}_0 \), this implies that the largest consistent set is a subset of \( \mathcal{M}^* \). Hence, the largest consistent set is a subset of the largest MFS.

\(^{10}\)In the space of individually rational matchings the only difference between the definition of a credible set and a consistent set is that the former uses ‘only if’, whereas the latter uses ‘if and only if’. Since \( V' \) is a subset of individually rational matchings and since it is also credible, we need only check the if part of the definition to show that \( V' \) is a consistent set.
3.4 Relation to Other Farsighted Solutions

3.4.1 Farsighted Core

A matching $\mu \in \mathcal{M}$ is in the farsighted core if and only if there exists no farsighted objection to $\mu$. This is an exclusive solution concept that is often empty even when there is a unique stable matching. But when it exists, it is guaranteed to be included in any MFS since any MFS needs to be an obtainable set.

By Lemma 2, if a matching $\mu'$ is a farsighted objection to a matching $\mu$ then either i) $\mu'$ is obtainable from $\mu$ through collusion, or ii) there exists a pair matched in $\mu'$ who prefer each other to their partners in $\mu$. This implies that the existence of a collusion-proof stable matching is necessary and sufficient for the existence of the farsighted core. Note that the existence of a collusion-proof stable matching implies the uniqueness of a stable matching.

Lemma 6. The farsighted core exists if and only if a collusion-proof stable matching exists. When it exists, the farsighted core is composed of the unique collusion-proof stable matching, and it is included in every MFS.

3.4.2 Farsighted Stable Set

The farsighted stable set (see Chwe (1994)) is the stable set (see von Neumann and Morgenstern (1944)) defined with farsighted objections. Specifically, a set of matchings $V$ is a farsighted stable set if i) there does not exist $\mu, \mu' \in V$ such that $\mu$ is a farsighted objection to $\mu'$, and ii) for any $\mu' \notin V$, there exists a $\mu \in V$ that is a farsighted objection to $\mu'$.

Part i) of the definition is known as the internal stability condition, which states that no matching within the farsighted stable set has a farsighted objection within the set. Part ii) is known as the external stability condition which states that every matching outside the farsighted stable set must have a farsighted objection within the set.

Mauleon, Vannetelbosch and Vertoge (2011) completely characterized the farsighted stable set in one-to-one matching problems. Interestingly, they show that a set of matchings is a farsighted stable set if and only if the set is composed of a single stable matching. This also implies that the farsighted stable set does not suffer from the problem of maximality in the context of one-to-one matching problems.\footnote{Any singleton farsighted stable set satisfies stronger notions of maximality than the one we impose (see Ray and Vohra (2019)).} Combined with our results, Mauleon, Vannetelbosch and Vertoge (2011)’s characterization of the farsighted stable set implies the following lemma.

Lemma 7. Let $V$ be a union of farsighted stable sets, then $V$ is an MFS. Hence, the largest MFS includes all farsighted stable sets.

3.4.3 Rational Expectations Farsighted Stable Set

Rational expectations farsighted stable set (see Dutta and Vohra (2017) and Dutta and Vartiainen (2020)) uses a maximality notion that is weaker than the one associated with MFS. It simply assumes that if coalition $S$ is supposed to move at some point, then only coalition $S$ itself can deviate from this move.
Definition 9. Rational Expectations Farsighted Stable Set

An expectation \( E \) is rational if \( E \) is stable and for any \( h \in H \) such that \( S(h) \neq \emptyset \), there does not exist a matching \( \mu \) that \( S(h) \) can enforce from \( x(h) \) such that \( E^T(h, S(h), \mu) \succ_{\{S(h)\}} E^T(h) \).

The stationary points of a rational expectation is a rational expectations farsighted stable set.

Since this notion of maximality is weaker than the notion MFS uses, it is immediate that each MFS is necessarily a rational expectations farsighted stable set. By Proposition 1, this implies that if a set is obtainable and credible then it is a rational expectations farsighted stable set. It turns out that the other direction also holds.

Lemma 8. A set of matchings is a rational expectations farsighted stable set if and only if it is an MFS. Hence, \( V \) is a rational expectations farsighted stable set if and only if \( V \) is an obtainable credible set of matchings.

We already discussed why if \( V \) is obtainable and credible then it is an REFS. For the other direction first note that if a matching is not individually rational, then it cannot be a stationary point of a rational expectations farsighted stable set. This is because, the singleton with an unacceptable partner can always deviate to a matching in which he is single and no farsighted objection from this matching can assign an unacceptable partner to this individual. Since rational expectations farsighted stable set is always a subset of individually rational matchings and since it also needs to contain a farsighted objection to any matching outside it, by Lemma 3 any rational expectations farsighted stable set is an obtainable set of matchings. Finally, it also needs to be a credible set, as otherwise a deviation from a stationary matching cannot be prevented.

4 Appendix

The following lemma is a restatement of Lemma 1 in Mauleon, Vannetelbosch and Vertoge (2011), it will be useful throughout the proofs.

Lemma 9. Consider any two matchings \( \mu, \mu' \in \mathcal{M} \), where \( \mu' \) is individually rational. Then \( \mu' \) is a farsighted objection to \( \mu \) if and only if there does not exist a pair \( \{i, \mu(i)\} \) such that \( \mu \succ_{\{i, \mu(i)\}} \mu' \).

Proof of Lemma 1. Suppose \( \mu' \) is individually rational and it can be obtained from \( \mu \) through collusion. Without loss of generality assume that \( S \subseteq \mathcal{M} \) can obtain \( \mu' \) from \( \mu \) through collusion. By the definition of collusion, for any pair \( \{i, \mu(i)\} \), either \( \mu(i) = \mu'(i) \) or \( \{i, \mu(i)\} \cap S \neq \emptyset \). In the former case we have \( \mu \sim_i \mu' \), and in the latter case we have \( \mu' \succ_{\{i, \mu(i)\} \cap S} \mu \). Hence, by Lemma 9 \( \mu' \) is a farsighted objection to \( \mu \).

Proof of Lemma 2. To prove the first statement suppose that \( \mu' \) is a farsighted objection to \( \mu \), but it is not obtainable from \( \mu \) through collusion. Towards a contradiction assume that there does not exist \( i, j \in N \), where possibly \( i = j \), such that \( \mu'(i) = j \) and \( \mu' \succ_{\{i,j\}} \mu \).

Since \( \mu' \) is a farsighted objection to \( \mu \), there exists at least one individual \( i \in N \) that prefers his partner in \( \mu' \) to his partner in \( \mu \). Without loss of generality assume that \( m_1 \in \mathcal{M} \) is such an individual. By the contradiction hypothesis, \( m_1 \) is not single in \( \mu' \) and \( \mu'(m_1) \) does not prefer \( m_1 \) to \( \mu(\mu'(m_1)) \). But since \( \mu' \) is a farsighted objection to \( \mu \) this implies that \( \mu(\mu'(m_1)) \) is also not single in \( \mu' \) and prefers his partner in \( \mu' \) to
his partner in $\mu$. Continuing this way, we will obtain a set of men who all prefer their partners in $\mu'$ to their partners in $\mu$ and that can also swap their partners to reach $\mu'$. This is a contradiction to the initial assumption that $\mu'$ is not obtainable from $\mu$ through collusion. This proves the first statement in Lemma 2.

Suppose $\mu$ and $\mu'$ are two distinct stable matchings. Diamantoudi and Xue (2003), and Mauleon, Vannetelbosch and Vertoge (2011) have shown that a stable matching is a farsighted objection to any other matching, so $\mu'$ is a farsighted objection to $\mu$. Since there is no pair that can object to $\mu$, by the first statement in Lemma 2 we have that $\mu'$ is obtainable from $\mu$ through collusion.

Since any two stable matchings can be obtained from each other through collusion, a sufficient condition for the existence of a unique stable matching is the existence of a collusion-proof stable matching.

**Proof of Lemma 3.** We have already seen that if an individually rational matching $\mu'$ is obtainable from $\mu$ then $\mu'$ is a farsighted objection to $\mu$. Here, we will show that if an individually rational matching $\mu'$ is a farsighted objection to $\mu$ then it is obtainable from $\mu$. So, suppose that $\mu'$ is individually rational and it is a farsighted objection to $\mu$.

If $\mu'$ is obtainable from $\mu$ through collusion or an improving path, then we are done. Assume that this is not the case. By Lemma 2 there exists $i, j \in N$, with possibly $i = j$, such that $\mu'(i) = j$ and $\mu' \succ_{i,j} \mu$. Pick one such pair and call them $(i_1, j_1)$. Consider the path $(\mu, \{i_1, j_1\}, \mu_1)$, where $\mu_1$ is such that $\mu_1(i_1) = j_1$. Observe that $\mu'$ is still a farsighted objection to $\mu_1$. By Lemma 2 either $\mu'$ is obtainable from $\mu$ through collusion or there exists $i, j \in N$, with possibly $i = j$, such that $\mu'(i) = j$ and $\mu' \succ_{i,j} \mu_1$.\footnote{It cannot be the case that $\mu_1 = \mu'$, because that would imply that $\mu'$ is obtainable from $\mu$ through an improving path.} If $\mu'$ is still not obtainable through collusion, then pick one such pair and call them $(i_2, j_2)$. Now, we can consider that path $(\mu, \{i_1, j_1\}, \mu_1, \{i_2, j_2\}, \mu_2)$. Continuing this way, we will obtain a path $(\mu, \{i_1, j_1\}, \mu_1, \{i_2, j_2\}, \mu_2, \ldots, \{i_k, j_k\}, \mu^*)$ such that $\mu'$ is obtainable from $\mu^*$ through collusion and $(i_t, j_t)$ are matched to each other both in $\mu^*$ and $\mu'$ for all $t = 1, \ldots, k$.

Let $N_1 = \{i \in N | \mu^*(i) = \mu'(i)\}$ and let $N_2 = N \setminus N_1$. Note that $N_2$ only includes those individuals who can obtain $\mu'$ from $\mu^*$ through collusion and their partners, and all of these individuals are matched with the same person at $\mu$ and $\mu^*$. Hence, every $i \in N_t$ is matched with some $j \in N_t$ at both $\mu$ and $\mu'$, for $t = 1, 2$.

Let $\mu_i$ be the matching that agrees with $\mu$ in the game restricted to $N_i$, for $i = 1, 2$. Then $\mu_1$ and $\mu_2$ form a partition of $\mu$. Similarly, let $\mu'_i$ be the matching that agrees with $\mu'$ in the game restricted to $N_i$, for $i = 1, 2$. Then $\mu'_1$ and $\mu'_2$ form a partition of $\mu'$.

$\mu'_1$ agrees with $\mu^*$ and since $\mu^*$ is obtainable from $\mu$ through an improving path, $\mu'_1$ is obtainable from $\mu_1$ through an improving path.

Similarly, $\mu_2$ agrees with $\mu^*$, and since $\mu'$ is obtainable from $\mu^*$ through collusion, $\mu'_2$ is obtainable from $\mu_2$ through collusion. Hence, $\mu'$ is obtainable from $\mu$.

The following lemmas will be helpful in proving the rest of the results.

**Lemma 10.** If $V$ is the stationary points of a stable expectation then $V$ is a subset of individually rational matchings.

**Proof.** Let $V$ be the stationary points of a stable expectation $E$. Towards a contradiction suppose that there exists a matching $\mu \in V$ that is not individually rational. Since $\mu \in V$, there exists a history $h$ with $x(h) = \mu$ and $E(h) = (\emptyset, \mu)$. Let $i$ be an individual matched
with an unacceptable partner in $\mu$. Consider the history $(h, i, \mu')$. Either $E^T(h, i, \mu') = \mu'$ or $\bar{E}(h, i, \mu')$ is a farsighted objection path. No farsighted objection path can assign an unacceptable partner to $i$. But then $E^T(h, i, \mu') \succ_i x(h) = \mu$, a contradiction to the second property of a stable expectation. Hence, $V$ is a subset of individually rational matchings.}

\begin{lemma}
Suppose $V$ is the stationary points of a stable expectation, then $V$ is an obtainable credible set.
\end{lemma}

\begin{proof}
Let $E$ be a stable expectation that supports $V$. For any $\mu \notin V$, $E$ assigns a farsighted objection path that terminates in some matching $\mu' \in V$. By Lemma 10, $\mu'$ is an individually rational matching, and since $\mu'$ is a farsighted objection to $\mu$, by Lemma 3, $\mu'$ is obtainable from $\mu$. Hence, $V$ is an obtainable set of matchings.

Towards a contradiction suppose that $V$ is not a credible set. Then there exists $\mu \in V$ and a coalition $S$ that can enforce some $\mu'$ such that for every $\mu'' \in V$, where either $\mu'' = \mu'$ or $\mu''$ is obtainable from $\mu'$, we have $\mu'' \succ_S \mu$.

Since $\mu$ is a stationary point, there exists an history $h$ with $x(h) = \mu$ and $S(h) = \emptyset$. Consider $E^T(h, S, \mu')$, which is necessarily in $V$ and hence it is individually rational. Either $E^T(h, S, \mu') = \mu' \in V$ or $E^T(h, S, \mu')$ is obtainable from $\mu'$. Either case, $E^T(h, S, \mu') \succ_S \mu$, which is a contradiction to the second property of a stable expectation. Hence, $V$ is a credible set.
\end{proof}

\begin{definition}
We say that a matching $\mu$ is a refinement of a matching $\mu'$ if $\mu(i) \neq i$ implies that $\mu(i) = \mu'(i)$, i.e. every individual that is not single in $\mu$ is matched with her partner in $\mu'$.
\end{definition}

\begin{lemma}
Suppose that an individually rational matching $\mu'$ is a farsighted objection to $\mu$. Then there exists a farsighted objection path from $\mu$ to $\mu'$ either in the form $(\mu_0, S_1, \mu_1, S_2, \mu_2)$ or in the form $(\mu_0, S_2, \mu_2)$, where

\begin{itemize}
  \item $\mu_0 = \mu$, $\mu_2 = \mu'$ and
  \item $\mu_1$ is a refinement of both $\mu$ and $\mu'$.
\end{itemize}

Paths in these forms are called canonical objection paths.
\end{lemma}

\begin{proof}
Suppose $\mu'$ is a farsighted objection to $\mu$. First, suppose that $\mu$ is a refinement of $\mu'$. Let $S_2$ be the set of individuals who are single at $\mu$, but matched in $\mu'$. By the individual rationality of $\mu'$, $\mu' \succ_S \mu$. Furthermore, $S_2$ can enforce $\mu'$ from $\mu$. Hence, $(\mu, S_2, \mu')$ is a path in the required form.

Now, suppose that $\mu$ is not a refinement of $\mu'$. Then there exists $i \in N$ such that $\mu(i) \neq i$ and $\mu(i) \neq \mu'(i)$. For every pair $(i, j) \in M \times W$ such that $\mu(i) = j$, but $\mu'(i) \neq j$, either $\mu' \succ_i \mu$ or $\mu' \succ_j \mu$. Choose one such individual for each such pair, and let $S_1$ be the set of these individuals. Let $\mu_1$ be the matching $S_1$ can impose over $\mu$ by separating from their partners. Let $(m_1, w_1), (m_2, w_2), \ldots, (m_t, w_t)$ be the set of pairs of men and women that are matched in $\mu'$, but not matched in $\mu$, and let $S_2 = \cup_{t=1}^{t} \{m_t, w_t\}$. The path $(\mu_0, S_1, \mu_1, S_2, \mu_2)$, where $\mu_0 = \mu$ and $\mu_2 = \mu'$ is a farsighted objection path in the required form.
\end{proof}

\begin{lemma}
Suppose an individually rational matching $\mu'$ is obtainable from $\mu$. If $\mu''$ is a refinement of $\mu$, then $\mu'$ is also obtainable from $\mu''$
\end{lemma}

23
Proof. Suppose an individually rational matching $\mu'$ is obtainable from $\mu$ and $\mu''$ is a refinement of $\mu$, but $\mu'$ is not obtainable from $\mu''$. Since $\mu'$ is individually rational, by Lemma 3, $\mu'$ is not a farsighted objection to $\mu''$. But then, by Lemma 9 there exists $m \in M$ such that $\mu''(m) = w$ and $\mu'' \succ_{\{m,w\}} \mu'$. Since $\mu''$ is a refinement of $\mu$, we also have that $\mu(m) = w$ and hence $\mu \succ_{\{m,w\}} \mu'$. By Lemma 9, $\mu'$ is not a farsighted objection to $\mu$. Since $\mu'$ is individually rational, by Lemma 3 this also implies that $\mu'$ is not obtainable from $\mu$, a contradiction.

The following definition of an $m$-coherent set of paths is adapted to this setting from Dutta and Vartiainen (2020). For any path $p$, let $\mu(p)$ denote the matching the path $p$ terminates in.

**Definition 11.** A collection $P$ of farsighted objection paths is $m$-coherent if

1. For each $\mu \in M$, there exists a $p \in P$ that starts with $\mu$.
2. If $(\mu_0, S_1, \mu_1, S_2, \ldots, \mu_k) \in P$ then $(\mu_i, S_i, \mu_{i+1}, S_{i+1}, \ldots, \mu_k) \in P$ for all $i = 0, 1, \ldots, k$.
3. For any path $(\mu) \in P$ there does not exist a coalition $S$ that can enforce a matching $\mu'$ such that $\mu(p) \succ_S \mu$ for all paths $p \in P$ that starts with $\mu'$.
4. For any $(\mu_0, S_1, \mu_1, \ldots, \mu_l) \in P$, coalition $T$ with $T \cap S_1 \neq \emptyset$ and a matching $\mu$ such that i) if $i \in T$ then $\mu(i) \in T$, and ii) for any $i \notin T$, $\mu(i) = \mu_1(i)$ if $\mu_1(i) \notin T$ and $\mu(i) = i$ if $\mu_1(i) \in T$, there exists a path $p$ that starts with $\mu$ such that $\mu(p) \not\succ_T \mu_1$.

The following lemma is due to Dutta and Vartiainen (2020). For a set of paths $P$, let $\mu(P) = \{\mu \in M| \mu(p) = \mu \text{ for some } p \in P\}$

**Lemma 14.** If $P$ is an $m$-coherent set of paths then $\mu(P)$ is an MFS.

Proof. This lemma is the direct counterpart of Lemma 2 and Theorem 2 in Dutta and Vartiainen (2020), where the only difference is the maximality condition (the final condition) of m-coherent set of paths, which reflects the maximality condition imposed by MFS. This makes no difference in their proof.

**Lemma 15.** Let $V$ be an obtainable credible set of matchings. Then the collection of paths $P$ defined with the rules below is an $m$-coherent set of paths.

- For any $\mu \in V$ let $P$ include the path $(\mu)$.
- For any matching $\mu$ let $P$ include all canonical objection paths to $V$.
- $P$ does not include any other paths.

Proof.

**Step 1:** Show that for each $\mu \in M$, there exists a $p \in P$ that starts with $\mu$.

For any $\mu \in V$, the path $(\mu) \in P$. For any $\mu \notin V$, since $V$ is an obtainable set there exists $\mu' \in V$ that is obtainable from $\mu$. Every canonical objection path from $\mu$ to $\mu'$ is included in $P$. 

24
Step 2: Show that if \((\mu_0, S_1, \ldots, \mu_k) \in P\) then \((\mu_i, S_i, \mu_{i+1}, S_{i+1}, \ldots, \mu_k) \in P\) for all \(i = 0, 1, \ldots, k\).

Suppose that \((\mu_0, S_1, \ldots, \mu_k) \in P\). Then \((\mu_0, S_1, \ldots, \mu_k)\) is a canonical objection path terminating at some matching in \(V\). Since the continuation of any canonical objection path itself is a canonical objection path, \((\mu_i, S_i, \mu_{i+1}, S_{i+1}, \ldots, \mu_k) \in P\) for all \(i = 0, 1, \ldots, k\).

Step 3: Show that for any path \((\mu) \in P\) there does not exist a coalition \(S\) that can enforce a matching \(\mu'\) such that \(\mu(p) >_S \mu\) for all paths \(p \in P\) that starts with \(\mu'\).

Suppose \((\mu) \in P\). By the definition of \(P\), this means that \(\mu \in V\). Take any coalition \(S\) and \(\mu'\) that this coalition can enforce from \(\mu\). By the credibility of \(V\), there exists a matching \(\mu'' \in V\) such that either \(\mu'' = \mu'\) or \(\mu'' \not\in S\). If \(\mu'' = \mu'\) then take \(p = (\mu'') \in P\). If \(\mu''\) is obtainable from \(\mu'\) then since \(P\) includes all canonical objection paths to \(P\), there exists \(p \in P\) that starts with \(\mu'\) such that \(\mu'' = \mu(p)\), take one such path \(p\). Either case, we have that \(p \in P\) and \(\mu'' = \mu(p) \not\in S\).

Step 4: Show that for any \((\mu_0, S_1, \mu_1, \ldots, \mu_0) \in P\), coalition \(T\) with \(T \cap S_1 \neq \emptyset\) and a matching \(\mu\) such that \(i\) if \(i \in T\) then \(\mu(i) \in T\), and \(ii\) for any \(i \not\in T\), \(\mu(i) = \mu_1(i)\) if \(\mu_1(i) \not\in T\) and \(\mu(i) = i\) if \(\mu_1(i) \in T\), there exists a path \(p\) that starts with \(\mu\) such that \(\mu(p) \not\in T\) \(\mu_i\).

Take any path \(p = (\mu_0, S_1, \ldots, \mu_0) \in P\). Since we only include canonical objection paths, we know that either \(p = (\mu_0, S_1, \mu_1, \mu_2, \mu_2)\) or \(p = (\mu_0, S_1, \mu_1)\).

First suppose that \(p = (\mu_0, S_1, \mu_1, S_2, \mu_2)\). Since this is a canonical objection path, \(\mu_1\) is such that for all \(i \in N\), if \(\mu_1(i) \neq i\) then \(\mu_0(i) = \mu_2(i)\). Take any coalition \(T\) with \(T \cap S_1 \neq \emptyset\) and a matching \(\mu\) such that \(i\) if \(i \in T\) then \(\mu(i) \in T\), and \(ii\) for any \(i \not\in T\), \(\mu(i) = \mu_1(i)\) if \(\mu_1(i) \not\in T\) and \(\mu(i) = i\) if \(\mu_1(i) \in T\).

Since \(\mu_1\) is such that any \(i \in N\) who is not single is matched with her partner in \(\mu_2\), the conditions on \(\mu\) imply that \(\mu\) is such that any \(i \in N \setminus T\) who is not single is matched with her partner in \(\mu_2\).

Let \(\mu'\) be the matching \(T\) can enforce from \(\mu_2\) such that \(\mu'(i) = \mu(i)\) for all \(i \in T\). By the credibility of \(V\), there exists \(\mu'' \in V\) with either \(\mu'' = \mu'\) or \(\mu'' \not\in T\) \(\mu_2\). Observe that \(\mu\) is a refinement of \(\mu'\). By Lemma 13, \(\mu''\) is obtainable from \(\mu\), and \(P\) includes a canonical objection path \(p\) from \(\mu\) to \(\mu''\). Hence, there exists \(p\) that starts with \(\mu\) such that \(\mu(p) = \mu'' \not\in T\) \(\mu_2\).

Now, assume that \(p = (\mu_0, S_1, \mu_1)\). Take any coalition \(T\) with \(T \cap S_1 \neq \emptyset\) and a matching \(\mu\) such that \(i\) if \(i \in T\) then \(\mu(i) \in T\), and \(ii\) for any \(i \not\in T\), \(\mu(i) = \mu_1(i)\) if \(\mu_1(i) \not\in T\) and \(\mu(i) = i\) if \(\mu_1(i) \in T\). These conditions also imply that \(\mu\) is enforceable by \(T\) from \(\mu_1\).

By the credibility of \(V\), there exists \(\mu'' \in V\) with either \(\mu'' = \mu\) or \(\mu''\) is obtainable from \(\mu\) such that \(\mu'' \not\in T\) \(\mu_1\). If \(\mu'' = \mu\) then \((\mu'') \in P\) and if \(\mu''\) is obtainable from \(\mu\) then \(P\) includes a canonical objection path \(p\) from \(\mu\) to \(\mu''\). Either way, there exists \(p\) that starts with \(\mu\) such that \(\mu(p) = \mu'' \not\in T\) \(\mu_1\).

\[\square\]

Proof of Proposition 1. First, suppose that \(V\) is an MFS. Then \(V\) is the stationary points of a maximal expectation. Since every maximal expectation is a stable expectation,
V is also the stationary points of a stable expectation. By Lemma 11, V is an obtainable credible set of matchings.

For the other direction suppose that V is an obtainable credible set of matchings. By Lemma 15, there exists an m-coherent set of paths P such that μ(P) = V. Hence, by Lemma 14, V is an MFS.

**Proof of Corollary 1.** The first, second and the fourth statement of the corollary are proven in the text. Here, we will prove the third statement.

Let V be a set composed of stable matchings. Take any matching μ ∈ V, a coalition S and a matching μ′ enforceable from μ by S. By the stability of μ, μ is obtainable from μ′ and μ ∉ S. Hence, μ is credible given V, implying that V is a credible set. Furthermore, since any set containing a stable matching is an obtainable set, V is also obtainable. By Proposition 1, V is an MFS.

**Proof of Lemma 4.** Let V be an obtainable set composed of individually rational matchings. Suppose μ ∈ V is not credible given V. By the definition of a credible set there exists a coalition T and a matching μ′ that T can enforce over μ such that for any μ′′ ∈ V, where μ′′ = μ′ or μ′′ is obtainable from μ′, we have μ′′ ≻ T μ. Clearly, this implies that T does not contain any player for whom μ is the most preferred matching in V.

Suppose that there exists i ∈ T such that μ′(i) = i. Let T′ ⊆ T be all i ∈ T such that μ′(i) = i. Note that T′ ≠ T, as otherwise μ would be obtainable from μ′ and μ ∉ T μ. Let S = T \ T′ and let μ′ be the matching enforceable over μ by S, where μ′(i) = μ′′(i) for all i ∈ S.

Since S ⊆ T and since μ′ is a refinement of μ′, by Lemma 13 we have that for any matching μ′′ ∈ V such that either μ′′ = μ′ or μ′′ is obtainable from μ′, we have μ′′ ≻ S μ.

Finally, we will show that there exists {m, w} ∈ S that are matched in μ′ and such that μ′ ≻ {m, w} μ. Towards a contradiction assume that there exists no such pair. But then by Lemma 9, μ ∈ V is obtainable from μ′ and μ ∉ S μ. A contradiction.

**Proof of Proposition 2.** It is immediate that M* is a credible set, it is also obtainable as it always includes a stable matching. Hence, by Proposition 1, M* is an MFS.

We need to show that no matching μ ∉ M* can be a stationary point of an MFS. If μ ∉ M1 then there exists a coalition S that can enforce some matching μ′ such that in any individually rational matching μ′′ such that either μ′′ = μ′ or μ′′ is obtainable from μ′, we have that μ′′ ≻ S μ. Since any MFS needs to be a subset of individually rational matchings, this implies that μ cannot be a stationary point of any MFS.

Now, take μ ∉ M2. There exists a coalition S that can deviate to some matching μ′ such that any matching μ′′ ∈ M1 such that either μ′′ = μ′ or μ′′ is obtainable from μ′, we have that μ′′ ≻ S μ. Since we have shown that any MFS needs to be a subset of M1, this implies that μ cannot be a stationary point of any MFS. The result is established through induction.

**Proof of Lemma 5.** First we will show that there exists a unique MFS only if there exists a collusion proof stable matching. Towards a contradiction assume that V is the unique MFS, but there does not exist a collusion-proof stable matching. First note that since any set composed of stable matchings is an MFS, and since the set V is the unique MFS, it needs to be the case that V is a singleton stable matching, i.e. V = {μ*}, where μ* is a stable matching. By the contradiction hypothesis there exists μ that is reachable.
from $\mu^*$ through collusion. We will show that the set $V' = \{\mu^*, \mu\}$ is an MFS, which is a contradiction to uniqueness.

Since $V'$ contains a stable matching, it is an obtainable set. We need to show that $V'$ is also credible. Since $\mu^*$ is stable, $\mu^*$ is credible given $V'$. Towards a contradiction assume that $\mu$ is not credible given $V'$. By Lemma 4 there exists a coalition $S$ that can enforce a matching $\mu''$ such that $i)$ there exists $m \in M$, $w \in W$ with $\{m, w\} \subseteq S$, $\mu''(m) = w$ and $\mu'' \succ_{\{m, w\}} \mu$, and $ii)$ $\mu^* \succ_{\{m, w\}} \mu$.

Without loss of generality assume that the set $M' \subseteq M$ can obtain $\mu$ from $\mu^*$ through collusion. Observe that $\mu \succeq M \mu^*$. But, this contradicts condition $ii)$ above which states that $\mu^* \succ_{\{m, w\}} \mu$. Hence, $\mu$ is credible given $V$. Since $V$ is also obtainable, $V$ is an MFS. This proves that there exists a unique MFS only if there exists a collusion proof stable matching.

Now suppose that the matching problem satisfies the top-coalition property and $\mu^*$ is the top-coalition matching. Diamantoudi and Xue (2003) have shown that in any hedonic game satisfying the top-coalition property the largest consistent set is composed of $\mu^*$. By Proposition 3 this implies that the largest MFS is composed of $\mu^*$, hence MFS is unique.

**Proof of Proposition 3.** First we will show that $\mathcal{M}^* = \{\mu \in \mathcal{M} | \mu \text{is credible given } \mathcal{M}^*\}$. By the definition of $\mathcal{M}^*$, if $\mu \in \mathcal{M}^*$ then $\mu$ is credible given $\mathcal{M}^*$. Hence, $\mathcal{M}^* \subseteq \{\mu \in \mathcal{M} | \mu \text{is credible given } \mathcal{M}^*\}$. Assume that $\mu \in \mathcal{M}$ and $\mu$ is credible given $\mathcal{M}^*$. Then $\mu \in \mathcal{M}_0$. Observe that by the definition of credibility if $\mu$ is credible given $V$ then $\mu$ is credible given $V'$ for all $V' \supseteq V$. But then $\mu$ is credible given $\mathcal{M}_0$, hence $\mu \in \mathcal{M}_1$. $\mu$ is also credible given $\mathcal{M}_1$, so $\mu \in \mathcal{M}_2$. Continuing this way, we have that $\mu \in \mathcal{M}^*$. Hence, $\mathcal{M}^* \supseteq \{\mu \in \mathcal{M} | \mu \text{is credible given } \mathcal{M}^*\}$.

We established that $\mu \in \mathcal{M}^*$ if and only if $\mu$ is credible given $\mathcal{M}^*$. $\mathcal{M}^*$ is a subset of individually rational matchings and by Lemma 3 in the domain of individually rational matchings $\mu$ is a farsighted objection to $\mu'$ if and only if $\mu$ is obtainable from $\mu'$. Replacing ‘obtainability’ with ‘farsighted objections’ in the definition of credibility, we have that $\mu \in \mathcal{M}^*$ if and only if for any coalition $S$ and $\mu'$ such that $\mu'$ is enforceable over $\mu$ by $S$, there exists $\mu'' \in \mathcal{M}^*$, where either $\mu'' = \mu'$ or $\mu''$ is a farsighted objection to $\mu'$, such that $\mu'' \not\succeq_S \mu$. This is the definition of a consistent set. Hence, $\mathcal{M}^*$ is a consistent set. This implies that $\mathcal{M}^*$ is a subset of the largest consistent set.

Now, we will show that the largest consistent set is a subset of $\mathcal{M}^*$. Assume that $\mu$ is a member of the largest consistent set. Then $\mu \in \mathcal{M}_0$, i.e. $\mu$ is individually rational. Suppose that every member of the largest consistent set is included in $\mathcal{M}_k$, we will show that every member of the largest consistent set is also included in $\mathcal{M}_{k+1}$. Towards a contradiction assume that $\mu$ is a member of the largest consistent set, but $\mu \notin \mathcal{M}_{k+1}$. Then $\mu$ is not credible given $\mathcal{M}_k$. Hence, there exists $S$ and $\mu'$ such that $\mu'$ is enforceable over $\mu$ by $S$, and for every $\mu'' \in \mathcal{M}_k$, where either $\mu'' = \mu'$ or $\mu''$ is obtainable from $\mu'$, we have $\mu'' \succ_S \mu$. Since $\mathcal{M}_k$ contains the largest consistent set, there exists $S$ and $\mu'$ such that $\mu'$ is enforceable over $\mu$ by $S$, and for every $\mu''$ in the largest consistent set, where either $\mu'' = \mu'$ or $\mu''$ is obtainable from $\mu'$, we have $\mu'' \succ_S \mu$. This is a contradiction to $\mu$ being a member of the largest consistent set. Hence $\mu \in \mathcal{M}_{k+1}$, and consequently all members of the largest consistent set is in $\mathcal{M}^*$, implying that the largest consistent set is a subset of $\mathcal{M}^*$.

---

13 Since the deviation by $S$ to $\mu''$ is credible, both $m$ and $w$ should prefer any matching in $V$ that is obtainable from $\mu''$. $\mu^* \in V$ and $\mu''$ is stable, so $\mu^*$ is obtainable from $\mu''$, hence we need $\mu'' \succ_{\{m, w\}} \mu$.

14 If not then the individual with an unacceptable partner can deviate. No matching that assigns an unacceptable partner to this individual can be obtainable from this objection.
Proof of Lemma 6. If $\mu$ is in the farsighted core then immediately $\mu$ is a collusion-proof stable matching. For the other direction suppose that $\mu$ is a collusion-proof stable matching. By Lemma 2, there does not exist a matching that is a farsighted objection to $\mu$. Hence, $\mu$ is in the farsighted core.

Now, suppose that $V$ is the farsighted core. We already established that $V$ should be composed of a single collusion-proof stable matching. By Lemma 2, there does not exist any matching that is obtainable from a collusion-proof stable matching. Since by Proposition 1 any MFS is obtainable, $V$ should be a subset of any MFS.

Proof of Lemma 7. Mauleon, Vannetelbosch and Vertoge (2011) showed that $V$ is a farsighted stable set if and only if $V$ is composed of a single stable matching. Hence, the result follows from Corollary 1.

Proof of Lemma 8. We have already seen why every MFS is a rational expectations farsighted stable set. For the other direction, observe that every rational expectations farsighted stable set is the stationary points of a stable expectation. By Lemma 11, this implies that every rational expectations farsighted stable set is an obtainable credible set of matchings.

References


15Lemma 2 established that any stable matching is obtainable from any other stable matching through collusion. Hence, the existence of a collusion-proof stable matching implies the uniqueness of a stable matching.


