Int J Game Theory (2017) 46:29–49
DOI 10.1007/s00182-015-0519-1

ORIGINAL PAPER

Paths to stability in two-sided matching under uncertainty

Emiliya Lazarova[1] · **Dinko Dimitrov**[2]

Accepted: 18 November 2015 / Published online: 26 November 2015

Abstract We consider one-to-one matching problems under two modalities of uncertainty in which types are assigned to agents either with or without replacement. Individuals have preferences over the possible types of the agents from the opposite market side and initially know the 'name' but not the 'type' of their potential partners. In this context, learning occurs via matching and using Bayes' rule. We introduce the notion of a stable and consistent outcome, and show how the interaction between blocking and learning behavior shapes the existence of paths to stability in each of these two uncertainty environments. Existence of stable and consistent outcomes then follows as a side result.

Keywords Consistent outcomes · Paths to stability · Uncertainty · Two-sided matchings

We are thankful to Francis Bloch and Arunava Sen for fruitful discussions, to an anonymous referee for her/his helpful remarks, and to Riccardo Di Maria for his patience. We also thank conference participants at the Public Economic Theory Conference 2013 in Lisbon, the SAET Meeting 2013 in Paris, the Summer Meeting of the American Economic Association 2013 in Los Angeles, the Econometric Society European Meeting 2013 in Gothenburg, the 25th International Conference on Game Theory 2014 in Stony Brook, the International Workshop on Game Theory and Economic Applications 2014 in Sao Paulo, as well as seminar participants at the Indian Statistical Institute, Delhi, University of Birmingham, and University of East Anglia for constructive comments.

✉ Dinko Dimitrov
dinko.dimitrov@mx.uni-saarland.de

Emiliya Lazarova
E.Lazarova@uea.ac.uk

[1] School of Economics, University of East Anglia, Norwich, UK

[2] Chair of Economic Theory, Saarland University, Saarbrücken, Germany

JEL Classification C62 · C78 · D71 · D83

A second marriage is the triumph of hope over experience.
Samuel Johnson

1 Introduction

Since the seminal contribution of Gale and Shapley (1962), the analysis of equilibrium outcomes in two-sided markets has focused on markets with centralized mechanisms in place. The question whether such outcomes can be reached in a decentralized manner by successive myopic blockings was first studied in Knuth (1976) and generally answered into the negative. However, Roth and Vande Vate (1990) show that there is a process leading from any unstable matching to a stable one, provided that blocking pairs are chosen appropriately.[1] This result was generalized to the roommate problem (Chung 2000; Diamantoudi et al. 2004; Iñarra et al. 2008), to matching markets with couples (Klaus and Klijn 2007), and to the many-to-many matching problem (Kojima and Ünver 2008), while an analysis of the strategic considerations of random stable mechanisms can be found in Roth and Vande Vate (1991) for the marriage market and in Pais (2008) for the college admissions problem. More recently, Klaus et al. (2011) analyze the blocking dynamics in roommate markets when agents make mistakes in their myopic blocking decisions, while Chen et al. (2012) provide a convergence to stability result for job matchings with competitive salaries. In all these works, however, it is assumed that players have complete information about the type of the other agents on the market. In the present paper we re-visit the question whether an equilibrium outcome in the standard one-to-one, two-sided market can be reached in a decentralized manner when, realistically, the assumption of perfect information is removed. In our setup, market participants have preferences over the types of agents with whom they can be matched, but not over their identities. We keep information requirements to the minimum, that is, initially, players only know their own type, which is allowed to be independent of individual preferences. Thus, two agents of the same type may have different preferences. Agents gather information about the type of their partners in the process of matching and thus, each player's information set expands by matching with a new partner.

We define an outcome for such a two-sided matching problem under uncertainty to consist of a matching and a system of beliefs collecting each agent's beliefs about the type of the agents from the opposite side of the market. We focus on outcomes where the system of beliefs is consistent with the process of learning via matching and where there are no further profitable deviations for any pair of players. Our definition of a blocking opportunity in this context requires the existence of types for the pair members characterized by a positive probability that the corresponding type of each agent in the pair is ranked higher by the other agent relative to the type of his or her current partner.

[1] As shown by Ma (1996), the random stable mechanism suggested by Roth and Vande Vate (1990) does not always reach all stable matchings.

In the domain of possible blocking notions, the one adopted here is most permissive.[2] In this sense, the set of stable outcomes obtained in our analysis is a subset of the sets of stable outcomes obtained employing stricter notions of blocking. Moreover, our analysis indicates that this blocking notion provides a sufficient condition for a matching which is part of a stable outcome under uncertainty to be also stable in the corresponding problem under complete information. We present and discuss the stability notion based on this type of blocking behavior and the consistency of beliefs in detail in Sect. 2.

Using these main ingredients of our setup, we address the question whether it is possible to reach a stable and consistent outcome from any initial self-consistent outcome (as defined in Sect. 2), and answer it in the positive (Theorem 1). The construction of a path in this case is shaped by the interaction between blocking and learning behavior and builds on Roth and Vande Vate's (1990) algorithm for reaching a stable matching in environments with complete information. Since a self-consistent outcome always exists, the non-emptiness of the set of stable and consistent outcomes for any two-sided matching problem under uncertainty follows as a side result.

We then turn to the study of the links of a matching problem under uncertainty and the corresponding problem under complete information, where agents' preferences over individuals in the latter follow their preferences over types in the former. We can readily show that the matching part of any stable and consistent outcome for the problem under uncertainty is a stable matching for the problem under complete information (Theorem 2). In order to connect, however, a stable matching for the latter problem to a stable and consistent outcome of the former, we need to take into account the way in which types are attributed to agents. If types are assigned as random independent draws from the set of types without replacement, then there exists a belief updating process that transforms a stable matching for the problem under complete information into a stable and consistent outcome for the corresponding problem under uncertainty (Theorem 3). If, on the other hand, types are assigned to agents as random independent draws from the set of types with replacement, then two important features connecting the problem under uncertainty and its corresponding problem under complete information play a crucial role: (1) strict preferences over types do not imply strict preferences over potential partners any more, and (2) knowing the type of one partner is not informative about the probability with which other potential partners are ranked higher than the current one. We handle these issues by restricting our analysis to matching problems where agents of the same type have preferences which are dichotomously aligned, that is, we require for any two agents of the same type that the sets of their individually rational types coincide. Then, our final result (Theorem 4) relates any stable matching for the problem under complete information to a homomorphic matching and a consistent system of beliefs for the problem under uncertainty, provided that agents' preferences over types are dichotomously aligned. Here we define two matchings to be homomorphic if the number of same-type agents who are matched to a given type of agents on the other side of the market is equal in both matchings.

[2] Our approach is similar in spirit to the maximax criteria discussed in management theories on decision making under uncertainty.

Our work contributes to the study of matching markets under uncertainty and, to the best of our knowledge, it is the first attempt to analyze paths to stability in such context. The setup we present differs for instance from the one in Roth (1989) who considers a non-cooperative model, where agents know their own preferences for partners but do not know their potential partners' preferences. In contrast, the agents in our model are aware of their own preferences over types, but agents of the same type are allowed to have different preferences over the types of the agents on the opposite market side. This distinguishes our work from that of Liu et al. (2014) and Bikhchandani (2014) who study stable outcomes in many-to-one and in one-to-one matching problems with transferable and non-transferable utility, respectively. In addition, unlike Liu et al. (2014) and Chakraborty et al. (2010), we assume in our model that agents do not observe the entire matching[3] and thus, they learn and update their beliefs solely by being matched to different partners along a sequence of matchings. We believe that such minimal informational background allows us to more starkly contrast our framework with the complete information world.

The rest of the paper is organized as follows. In Sect. 2 we introduce the basic ingredients of our setup. In Sect. 3 we present two general results that hold independently of the way in which types are assigned to agents. The results for which the assignment function is a constraining factor are discussed in Sect. 4. We add some thoughts on how our framework can be used in future research as a concluding remark.

2 Notation and definitions

Our setup consists of the following basic ingredients.

Types and preferences

We consider two finite sets M and W of agents, called "men" and "women" , respectively. Agents can be of different types. We denote the finite set of all possible types by Θ. The function $\theta : M \cup W \to \Theta$ assigns a type to each agent such that men and women are of *different* types[4], i.e., $\theta(m) \neq \theta(w)$ holds for $m \in M$ and $w \in W$. Agents' *strict* preferences are defined over the set of all possible types.[5] A profile of such preferences is denoted by $\succeq = (\succeq_i)_{i \in M \cup W}$. When the assignment of types is known, agents can use their preferences over types to derive preferences over individuals on the other side of the market. Notice that, in general, strict preferences over types do not imply strict preferences over agents as some agents of the same sex can be of the same type.

[3] In our analysis, whether or not agents observe the entire matching, is immaterial. This would only become relevant if we further extended the agents' information set to include others' preferences.

[4] The assumption that types are gender specific allows us to keep the notation less cumbersome while preserving the generality of our setup. This implies, for instance, that the generic type "green eyes" is divided into female-green-eyes and male-green-eyes types.

[5] Strict preferences are a common assumption in the matching literature. Recently, some authors have departed from this assumption and have studied preference profiles with indifferences, e.g., Erdil and Haluk (2008) and Abdulkadiroğlu et al. (2009).

Initially, individuals know their own type (and thus, the 'type' of the possibility of remaining single) and only the 'name' of all individuals from the opposite market side but not their types. The reader can think of an analogy with a phone-directory where the listing of registered users provides an index of names but no description of qualities. We assume, instead, that each agent has a prior about the types of the players on the other side of the market. For the purposes of our analysis it is not necessary that agents on the same market side or those of the same type hold a common prior. Thus, priors can be individual-specific and they may not reflect the true distribution of types in the population of agents. We denote the prior agent i has about agent j being of type $t \in \Theta$ by $\pi_i(j,t)$ with $\pi_i(j,t) > 0$ holding for all $t \in \Theta$ and all $i, j \in M \cup W$ who are from opposite market sides. A *one-to-one matching problem under uncertainty* then consists of two finite sets of agents, a finite set of types, assignment function, individual priors, as well as a strict preference profile over types.

Beliefs updating and consistent outcomes

An *outcome of the matching problem under uncertainty* is a pair (μ, α) consisting of a *matching function* μ and a *system of beliefs* α. The matching function $\mu : M \cup W \to M \cup W$ is such that $\mu(i) \in W \cup \{i\}$, $\mu(j) \in M \cup \{j\}$, and $\mu^2(k) = k$ hold for $i \in M$, $j \in W$, and $k \in M \cup W$. The interpretation of $\mu(k) = k$ for some $k \in M \cup W$ is that the corresponding agent is single under μ. The system of beliefs α contains all agents' beliefs about the type of each agent on the opposite side of the market. In particular, we use the notation $\alpha_i(j,t)$ to denote the belief agent i holds about j being of type $t \in \Theta$. Clearly, $\sum_{t\in\Theta} \alpha_i(j,t) = 1$ and, since agents know their own types, $\alpha_i(i, \theta(i)) = 1$ and $\alpha_i(i, \theta') = 0$ holds for each $i \in M \cup W$ and all $\theta' \neq \theta(i)$.

Let us next define the notion of a *self-consistent* outcome $(\mu, \alpha_{|\mu})$, where $\alpha_{|\mu}$ stands for a system of beliefs which is consistent *only* with respect to the individual priors and the knowledge gained from the types of the corresponding matching partners under μ. More precisely, suppose that agent i's partner under μ is of type t', i.e., $\theta(\mu(i)) = t'$. Take an agent j from i's opposite market side. Then i's belief $(\alpha_{|\mu})_i(j,t)$ about j being of type t is the conditional probability $Prob_i(\theta(j) = t \mid \theta(\mu(i)) = t')$ agent i assigns to the event that j is of type t, provided that his/her partner under μ is of type t'. Thus,

$$\begin{aligned}(\alpha_{|\mu})_i(j,t) &= \frac{Prob_i\left(\theta(j) = t \cap \theta(\mu(i)) = t'\right)}{Prob_i\left(\theta(\mu(i)) = t'\right)} \\ &= \frac{Prob_i\left(\theta(j) = t \cap \theta(\mu(i)) = t'\right)}{\pi_i\left(\mu(i), t'\right)},\end{aligned}$$

where $Prob_i(\theta(j) = t \cap \theta(\mu(i)) = t')$ is the joint probability of j being of type t and $\mu(i)$ being of type t'. Notice that $j = \mu(i)$ implies

$$\begin{aligned}&Prob_i\left(\theta(j) = t \cap \theta(\mu(i)) = t'\right) \\ &= \begin{cases} Prob_i\left(\theta(\mu(i)) = t'\right) = \pi_i\left(\mu(i), t'\right) & \text{if } t = t' \\ Prob_i\left(\theta(\mu(i)) = t \cap \theta(\mu(i)) = t'\right) = 0 & \text{if } t \neq t'; \end{cases}\end{aligned}$$

and therefore

$$(\alpha_{|\mu})_i(\mu(i), t) = \begin{cases} 1 & \text{if } t = t' \\ 0 & \text{if } t \neq t'. \end{cases}$$

In other words, when an agent's system of beliefs is consistent with respect to a matching μ, then the agent knows the type of his/her partner under μ.

Moreover, the assumption on how types are assigned to agents—whether types are assigned to agents as random independent draws from the set of types *with* or *without* replacement—has implications for the updating of the system of beliefs $(\alpha_{|\mu})_i(j, t)$ for all $j \neq \mu(i)$. In particular, the updating of the belief that an agent $j \neq \mu(i)$ is of the same type as i's partner, agent $\mu(i)$ where we let $\theta(\mu(i)) = t'$, implies

$$\begin{aligned} &Prob_i\left(\theta(j) = t' \cap \theta(\mu(i)) = t'\right) \\ &= \begin{cases} 0 & \text{if without replacement} \\ \pi_i\left(j, t'\right) \times \pi_i\left(\mu(i), t'\right) & \text{if with replacement;} \end{cases} \end{aligned}$$

and therefore

$$(\alpha_{|\mu})_i(j, t') = \begin{cases} 0 & \text{if without replacement} \\ \pi_i\left(j, t'\right) & \text{if with replacement.} \end{cases}$$

Clearly, when types are assigned to agents as random independent draws from the set of types *without* replacement, learning about the types of agents on the opposite market side occurs via direct matching with such an agent (partners know each other's type) and via Bayesian updating of one's beliefs given the type of their parters. In the case when types are assigned to agents as random independent draws from the set of types *with* replacement, in contrast, agents gain information about the type of someone from the opposite market side only if they are matched to each other.

Thus, we call the system of beliefs α *consistent with respect to the matching* μ (denoted by $\alpha_{|\mu}$) if each agent $i \in M \cup W$,

(1) uses Bayes' rule to update his/her beliefs about the type of each agent on the other side of the market as explained above, and
(2) there is no belief updating if the agent is single under μ (i.e., $\mu(i) = i$ implies $(\alpha_{|\mu})_i(j, t) = \pi_i(j, t)$ for any j from i's opposite market side and any $t \in \Theta$).

Next, we generalize the notion of consistent updating and define the consistency of an outcome with respect to a given sequence of matchings in order to incorporate the fact that the beliefs an agent holds evolve with the search for an optimal partner. For this, let us start with the meaning of the notion 'satisfying a blocking pair' (cf. Roth and Vande Vate 1990). If the pair (m, w) is blocking an outcome with matching function μ, we say that a new matching ν is obtained from μ by satisfying the blocking pair if m and w are married under ν, their partners under μ (if any) are unmatched at ν, and all other agents are matched to the same mates under ν as they were under μ. We will consider then an outcome (μ, α) to be *consistent with respect to a self-consistent initial outcome* $(\mu_0, \alpha_{|\mu_0})$ if there is a sequence of outcomes $(\mu_1, \alpha_{|\mu_1}), \ldots, (\mu_k, \alpha_{|\mu_1,\ldots,\mu_k})$ with $(\mu_1, \alpha_{|\mu_1}) = (\mu_0, \alpha_{|\mu_0})$ and $(\mu_k, \alpha_{|\mu_1,\ldots,\mu_k}) = (\mu, \alpha)$ such that for $\ell = 1, \ldots, k-1$:

(1) there is a blocking pair (m_ℓ, w_ℓ) for $(\mu_\ell, \alpha_{|\mu_1,\ldots,\mu_\ell})$ such that $\mu_{\ell+1}$ is obtained from μ_ℓ by satisfying (m_ℓ, w_ℓ);
(2) there is a consistent Bayesian updating of beliefs $\alpha_{|\mu_1,\ldots,\mu_{\ell+1}}$ such that for $\ell = 1, \ldots, k-1$:
  (2.1) the agents in each blocking pair along the sequence update their beliefs with respect to the types of the agents on the opposite side of the market; that is, for $i \in \{m_\ell, w_\ell\}$ with $\theta(\mu_{\ell+1}(i)) = t'$, for every agent j from i's opposite market side, and for every type $t \in \Theta$, we have that $(\alpha_{|\mu_1,\ldots,\mu_{\ell+1}})_i(j,t)$ is the conditional probability agent i assigns to the event that j is of type t given that his/her match under $\mu_{\ell+1}$ is of type t'. Notice that the probability that i assigns to his/her partner under $\mu_{\ell+1}$ is of type t' equals agent i's belief (as being updated along the path from $(\mu_1, \alpha_{|\mu_1})$ to $(\mu_\ell, \alpha_{|\mu_1,\ldots,\mu_\ell})$) about the type of his/her partner in $\mu_{\ell+1}$ being t'. Thus,

$$\begin{aligned}(\alpha_{|\mu_1,\ldots,\mu_{\ell+1}})_i\,(j,t) &= Prob_i(\theta(j) = t \mid \theta\,(\mu_{\ell+1}(i)) = t') \\ &= \frac{Prob_i\left(\theta(j) = t \cap \theta\,(\mu_{\ell+1}(i)) = t'\right)}{Prob_i\,(\theta\,(\mu_{\ell+1}(i)) = t')} \\ &= \frac{Prob_i\left(\theta(j) = t \cap \theta\,(\mu_{\ell+1}(i)) = t'\right)}{(\alpha_{|\mu_1,\ldots,\mu_\ell})_i\,(\mu_{\ell+1}(i), t')};\end{aligned}$$

  (2.2) at each step of the sequence agents who are not part of a blocking pair do not update their believes; that is, for any $i \in (M \cup W) \setminus \{m_\ell, w_\ell\}$ and any j from i 's opposite market side, $(\alpha_{|\mu_1,\ldots,\mu_{\ell+1}})_i(j,t) = (\alpha_{|\mu_1,\ldots,\mu_\ell})_i(j,t)$ holds for each $t \in \Theta$.

Condition (1) above defines a 'legitimate' path of search for an optimal partner. We take an outcome to be consistent with respect to an initial self-consistent outcome if it can be derived from it by satisfying blocking pairs. Condition (2), on the other hand, describes a sound 'learning process', i.e., the updating of beliefs along the path of blocked matchings. The agents participating in a blocking pair know the type of their partners and use Bayesian updating to re-calculate the probability with which any other agent on the opposite side of the market is of any given type; and last, agents who do not participate in a blocking pair do not update their beliefs as they do not gain any additional information.[6]

Using the above definitions, we call an outcome (μ, α) *consistent* if there exists an initial self-consistent outcome $(\mu_0, \alpha_{|\mu_0})$ with respect to which it is consistent.

Stable outcomes

Since agents' preferences in a matching problem are defined over types, the definitions of individual rationality and unilateral blocking are straightforward. We will say that

[6] Notice that agents in our setup know their preferences over types, while through the matching process they learn the type of their corresponding partner from the opposite market side. For an alternative way of modelling uncertainty where agents' preferences are defined over partnership plans, and are assumed to be not completely known to them, we refer the reader to the recent work of Kadam and Kotowski (2014). The focus in these authors' work is mainly on the existence of dynamically stable matchings.

an outcome (μ, α) is *individually rational* if for each $i \in M \cup W, \theta(\mu(i)) \succeq_i \theta(i)$. On the other hand, if $\theta(i) \succ_i \theta(\mu(i))$, we say that agent i *unilaterally blocks* the outcome (μ, α). Notice that, although not explicitly mentioned, the notion of an individually rational (consistent) outcome (μ, α) implicitly makes use of the system of beliefs α as in the matching μ each agent knows his/her own type and the type of his/her partner. Clearly then, the individual rationality of an outcome (μ, α) implies the individual rationality of any other (consistent) outcome (μ', α') with $\mu' = \mu$. Hence, in what follows, when talking about the individual rationality of a matching μ in a problem under uncertainty we will mean the individual rationality of the outcome (μ, α) for any system of beliefs α which is consistent with respect to μ.

Finally, a *pair of agents* (m, w) with $m \in M$ and $w \in W$ is blocking the outcome (μ, α) if there are types $t_1, t_2 \in \Theta$ such that the following two conditions hold:

(1) $t_1 \succ_m \theta(\mu(m))$ and $t_2 \succ_w \theta(\mu(w))$;
(2) $\alpha_m(w, t_1) > 0$ and $\alpha_w(m, t_2) > 0$.

Our definition of a blocking pair needs further discussion. We require that each member of a blocking pair assigns some positive probability to the fact that the other member of the pair is of a type ranked higher than the type of his or her current match. Certainly, the validity of this blocking rule hinges upon a behavioral model of extreme optimism and no costs of switching as even the tiniest perceived positive probability that an agent can be better off in the new matching is enough to induce blocking. Numerous other behavioral models can be studied including those of the "extreme pessimists" who would only leave a partner if they know with certainty that their new partner is higher ranked than the current one; or of a more 'balanced' approach where, for instance, agents block a matching if their potential partners are more likely to be of a type that is higher ranked than the corresponding current one, rather than ranked lower. Since under our assumption the blocking possibilities are the most permissive, however, an outcome which cannot be blocked in our sense cannot be blocked under any other more demanding blocking notion.

Notice finally that Condition (2) does not imply that our blocking notion is independent of (the updating process of) agents' beliefs. Admittedly, as we assume that $\pi_i(j, t) > 0$ for all $t \in \Theta$ and all $i, j \in M \cup W$ who are from opposite market sides, one might think that agents block matchings irrespective of the learning process. However, this impression is misleading, firstly, because agents learn each others' types when matched and thus they will not form a blocking pair with an agent whose type they know with certainty to be less desirable than the type of their current partner. Furthermore, in the case when types are assigned to agents without replacement, an agent may deduce that another agent with whom she has never been matched is of a less desirable type than her current match by updating her beliefs using Bayes' rule. To see the argument, take a consistent outcome $(\mu, \alpha_{|\mu_0,\ldots,\mu})$ and notice that $(\alpha_{|\mu_0,\ldots,\mu})_i(j, t) = 0$ will be true for some i and j from the opposite market sides and some type t whenever agent i has been matched to an agent of type t along the path $\mu_0, \ldots, \mu$. Thus, due to the updating of i's beliefs on the possible types of j, the probability that i and j form a blocking pair will be zero even if type t is i's most preferred type and the type of his/her current match is second-best and i does not know j's type with certainty.

In what follows we will focus on outcomes which are both consistent and *stable*, that is, outcomes where the corresponding beliefs updating process has taken place and for which there are no blocking pairs.

3 Paths to stability

We start by asking the question whether there exists a stable and consistent outcome with respect to any initial self-consistent outcome and answer it to the affirmative (independently of the way in which types are assigned to agents) by means of a constructive proof. The existence of stable and consistent outcomes in our setting becomes then a direct corollary of our first result.

The construction of a path in this case is shaped by the interaction between blocking and learning behavior and uses, in part, Roth and Vande Vate's (1990) algorithm for reaching a stable matching in environments with complete information. More precisely, Roth and Vande Vate's algorithm is applied at each step, where there is a blocking pair consisting of agents who know each other's type. Correspondingly, if there is a pair whose blocking behavior is based on the hope, albeit a tiny one, that the other agent is of a higher ranked type, we let the corresponding pair marry such that the pair members can convince each other. The interplay between these two types of blocking can be explained as follows. Suppose that, at a given step along the path, there are only blocking pairs whose members know each other's type and thus, letting one of these pairs marry, does not change agents' beliefs. It may still happen that at the next step there are new blocking pairs whose members do not know each other's type. The reason for this is that in a matching where an agent is married to her most preferred partner, she would not form a blocking pair even though she does not know the type of all men; but if her partner divorces her, she may engage in a learning experiment if she hopes that a marriage with an unknown man would make her better off compared to being alone or marrying a man whose type she knows.

Theorem 1 *Let* $(\mu_0, \alpha_{|\mu_0})$ *be a self-consistent outcome of a given matching problem under uncertainty. Then the matching problem has a stable outcome which is consistent with respect to* $(\mu_0, \alpha_{|\mu_0})$.

Proof Take $(\mu_0, \alpha_{|\mu_0})$ as above. If there is an agent $i \in M \cup W$ for whom this outcome is not individually rational, consider the outcome $(\mu_1, \alpha_{|\mu_0})$ that differs from $(\mu_0, \alpha_{|\mu_0})$ only by the fact that i and $\mu(i)$ are now 'divorced'; notice that in such a case no agent learns the type of any other agent on the opposite side of the market and thus, there is no update of agents' beliefs. Continuing in this way, and as the sets of agents are finite, we can finally reach an individually rational outcome $(\mu_k, \alpha_{|\mu_0})$ which is consistent with the initial self-consistent outcome $(\mu_0, \alpha_{|\mu_0})$.

Thus, without loss of generality, we proceed by assuming that $(\mu_0, \alpha_{|\mu_0})$ is an individually rational self-consistent outcome. Let us collect in the set $B(0)$ all agents who form blocking pairs for $(\mu_0, \alpha_{|\mu_0})$ such that the corresponding pair members know each other's type, and let $L(0)$ be the analogous set in which the members of a blocking pair do not know each other's type, i.e., there is a possibility of learning. If there is no blocking pair at all for $(\mu_0, \alpha_{|\mu_0})$, we are done. Given the individual

rationality and self-consistency of $(\mu_0, \alpha_{|\mu_0})$, we have $B(0) = \emptyset$.[7] So, if there is a blocking pair for $(\mu_0, \alpha_{|\mu_0})$, then it must contain agents only from $L(0)$.

In this case we can construct a sequence of consistent outcomes $(\mu_0, \alpha_{|\mu_0})$, $(\mu_1, \alpha_{|\mu_0,\mu_1})$, $\ldots$, $(\mu_k, \alpha_{|\mu_0,\mu_1,\ldots,\mu_k})$ along which individuals can learn the type of the agents on the opposite side of the market by forming blocking pairs only with such agents with whom they have not been matched before. Here k is the smallest integer for which $L(k) = \emptyset$, i.e., there is no possibility for learning. Consider the consistent outcome $(\mu_k, \alpha_{|\mu_0,\mu_1,\ldots,\mu_k})$ and note that if $B(k) = \emptyset$, then we are done.

If $B(k) \neq \emptyset$, then pick up at random a woman $w_k \in B(k)$ and one of w_k's most preferred partners in $B(k)$, say m_k, and construct the consistent outcome $(\mu_{k+1}, \alpha_{|\mu_0,\mu_1,\ldots,\mu_{k+1}})$ by satisfying the blocking pair (m_k, w_k) and updating the system of beliefs $\alpha_{|\mu_0,\mu_1,\ldots,\mu_{k+1}} = \alpha_{|\mu_0,\mu_1,\ldots,\mu_k}$. Set $A(k+1) = \{m_k, w_k\}$ to be the set of satisfied blocking pairs where agents knew each other's type prior to this matching.

If $L(k+1) = \emptyset$ and $B(k+1) = \emptyset$, then we are done. If $L(k+1) \neq \emptyset$, however, then construct μ_{k+2} by satisfying a blocking pair in $L(k+1)$ and update the beliefs in a consistent manner. Set $A(k+2) = \emptyset$. Notice that $L(q) = \emptyset$ in some finite steps q due to the finiteness of the sets M and W, i.e., men and women will eventually learn the types of all agents on the opposite side of the market. And if $L(k+1) = \emptyset$, but $B(k+1) \neq \emptyset$, then notice that $w_k \notin B(k+1)$ because m_k is one of w_k's most preferred partners in $B(k)$ and she cannot form any new blocking pairs in μ_{k+1} that she could not form in μ_k. Then pick a blocking pair at random from the set $B(k+1)$, say (w_{k+1}, m_{k+1}) and form the matching μ_{k+2} by satisfying this blocking pair. Let $\alpha_{|\mu_0,\mu_1,\ldots,\mu_{k+2}} = \alpha_{|\mu_0,\mu_1,\ldots,\mu_{k+1}} = \alpha_{|\mu_0,\mu_1,\ldots,\mu_k}$. Set $A(k+2) = A(k+1) \cup \{m_{k+1}, w_{k+1}\}$ and note that $A(k+1) \subseteq A(k+2)$.

Thus, if there is no subsequent step r with $L(r) \neq \emptyset$ (i.e., there are no possibilities for learning any more), we can adopt Roth and Vande Vate's (1990) algorithm to construct an increasing sequence of sets that contain no blocking pairs until a stable matching is found. This is possible because, the lack of possibility for learning implies that all agents involved in blocking have complete information about their potential blocking partners, i.e., they either know all agents whose type is higher ranked than the type of their current partner or if there is such agent in the set $i \in B(r)$ whose type they do not know but with whom they cannot form a blocking pair, then i must know all agents whose type is higher ranked than the type of i's current partner and therefore i cannot be their potential blocking partner. Since only blocking pairs with no learning are satisfied along the path following μ_k and reaching a stable matching, we construct a stable and consistent outcome that consists of the stable matching just obtained and the system of beliefs $\alpha_{|\mu_0,\mu_1,\ldots,\mu_k}$. □

Given that a self-consistent outcome of any two-sided matching problem under uncertainty always exists, the following corollary to Theorem 1 immediately follows.

[7] Notice that for a self-consistent outcome $(\mu, \alpha_{|\mu})$, and agents $m \in M$ and $w \in W$ with $\mu(m) \neq w$, we have that $(\alpha_{|\mu})_m(w, \theta(w)) = (\alpha_{|\mu})_w(m, \theta(m)) = 1$ (i.e., m and w know each other's type) holds only if $|M|, |W| \leq 2$, $\mu(i) \neq i$ for some $i \in M \cup W$, and types are assigned to agents as random independent draws from the set of types without replacement. In what follows, we exclude this trivial case.

Corollary 1 *The set of stable and consistent outcomes for any matching problem under uncertainty is non-empty.*

4 Links with the complete information world

In this section we discuss the relation between the set of stable and consistent outcomes for a two-sided matching problem under uncertainty and the set of stable matchings for its corresponding two-sided matching problem under complete information. Recall that a one-to-one matching problem under complete information is a tuple $(M, W, \succeq')$, where M and W are the sets of men and women as defined above and $\succeq'$ denotes a preference profile that collects the preferences men and women hold over their potential partners in a matching. Given a matching problem under uncertainty as defined above, we say that the matching problem under complete information $(M, W, \succeq')$ *corresponds* to it if the sets of agents coincide and the preference profile is such that for each agent it induces the same ranking of potential partners. That is, for $m \in M$ and $w_i, w_j \in W$, $w_i \succeq'_m w_j$ if and only if $\theta(w_i) \succeq_m \theta(w_j)$; $w_i \succeq'_m m$ if and only if $\theta(w_i) \succeq_m \theta(m)$, and similarly, for $w \in W$ and $m_i, m_j \in M$, $m_i \succeq'_w m_j$ if and only if $\theta(m_i) \succeq_w \theta(m_j)$ and $m_i \succeq'_w w$ if and only if $\theta(m_i) \succeq_w \theta(w)$.

We also recall two commonly used notions with regards to matching under complete information. A matching μ is *individually rational* if $\mu(i) \succeq'_i i$ for each $i \in M \cup W$. An individually rational matching μ is *stable* if there does not exist a pair (m, w) of agents such that $w \succ'_m \mu(m)$ and $m \succ'_w \mu(w)$.

Remark 1 It is easy to see that μ is individually rational for a matching problem under complete information if and only if, for any system of beliefs α with $\alpha_i(\mu(i), \theta(\mu(i))) = 1$ for $i \in M \cup W$, the outcome (μ, α) is individually rational for the corresponding matching problem under uncertainty.

Theorem 2 *If (μ, α) is a stable and consistent outcome for a given matching problem under uncertainty, then μ is a stable matching for the corresponding problem under complete information.*

Proof Let (μ, α) be as above and suppose that μ is not stable for the corresponding matching problem under complete information. By Remark 1, μ is individually rational. Hence, there should exist a pair (m, w) of agents who are not matched to each other under μ and prefer to be matched to each other than to their current partners: $w \succ'_m \mu(m)$ and $m \succ'_w \mu(w)$. This implies that $t_1 := \theta(w) \succ_m \theta(\mu(m))$ and $t_2 := \theta(m) \succ_w \theta(\mu(w))$. Suppose now that $\alpha_m(w, t_1) = 0$. By $\pi_m(w, t_1) > 0$, the consistency of agents' beliefs, and $t_1 \succ_m \theta(\mu(m))$, we have that $\alpha_m(w, t_1) = 0$ can happen only if types are assigned without replacement such that $\alpha_m(w', t_1) = 1$ for $w' \neq w$ in contradiction to $t_1 := \theta(w)$. We conclude then that $\alpha_m(w, t_1) > 0$ should hold. By an analogous argument, $\alpha_w(m, \theta(m)) > 0$ also holds. Therefore, we have established that (m, w) is a blocking pair for the outcome (μ, α) under uncertainty, too. Thus, we have a contradiction. □

Notice that the above result may not hold in a behavioral model that implies the existence of less blocking possibilities than those discussed in Sect. 2. In such models,

the set of stable outcomes would be larger than the one studied here, thus, there may be an outcome which is stable and consistent under uncertainty without its matching part being stable under complete information. In this sense, Theorem 2 provides a sufficient condition for stability and consistency under uncertainty to imply stability under complete information.

Let us now change the starting point of our analysis and consider the following situation. Suppose that the matching part of an initial self-consistent outcome for a matching problem under uncertainty is stable for its corresponding problem under complete information. Then, in view of Theorem 1, we can reach a stable and consistent outcome for the problem under uncertainty. What are then the conditions allowing us to conclude that the matching part of the latter outcome is in some sense 'similar' to the stable matching under complete information? In order to tackle this issue we need to take a closer look at how types are assigned to agents.

In what follows we will explicitly distinguish between problems where types are assigned to agents as random independent draws from the set of types without replacement and with replacement. As already mentioned, the first crucial difference is that in the former case learning occurs via matching (partners know each other's type) and belief updating, while in the latter case agents gain information about the type of someone from the opposite market side only if they are matched to each other. The second important difference between the two modalities of uncertainty concerns how preferences over types (in the problem under uncertainty) are translated into preferences over individuals (in the corresponding problem under complete information): when types are assigned without replacement, strict preferences over types imply strict preferences over individuals; however, when types are assigned with replacement, agents' preferences over potential partners can contain indifferences even if their preferences over types are strict since many agents can be assigned the same type.

4.1 Assignment without replacement

Without further ado we can describe the process of beliefs' updating that transforms a stable matching for the problem under complete information into a stable and consistent outcome for the corresponding problem under uncertainty when types are assigned as random independent draws without replacement.

Theorem 3 *Let a matching problem under uncertainty with types assigned without replacement be given and μ be stable for the corresponding matching problem under complete information. Then there exists a stable outcome (μ, α) for the problem under uncertainty which is consistent with respect to $(\mu, \alpha_{|\mu})$.*

Proof Let μ be as above and consider the self-consistent outcome $(\mu, \alpha_{|\mu})$. If there are no blocking pairs for it, then we have shown what we need. Notice further that, in view of Remark 1, it is impossible for an agent to unilaterally block $(\mu, \alpha_{|\mu})$.

Suppose now that there is a pair (m, w) that blocks $(\mu, \alpha_{|\mu})$. Then, by satisfying this pair, we can construct the consistent outcome $(\mu_1, \alpha_{|\mu,\mu_1})$. This cannot be a stable outcome. In order to see that, notice first that, since μ is a stable matching for the problem under complete information, either m or w weakly prefers his or her

partner under μ over w and m, respectively. Let, w.l.o.g., $\mu(m) \succeq'_m w = \mu_1(m)$ hold. Moreover, $\mu(m) \sim'_m w$ is ruled out since it would imply that $\theta(\mu(m)) \sim_m \theta(w)$ and given the antisymmetry of agents' preferences over types and that types are assigned without replacement, $\theta(\mu(m)) = \theta(w)$ would contradict $\mu(m) \neq w$. Hence, $\mu(m) \succ'_m w = \mu_1(m)$ holds and thus, $\theta(\mu(m)) \succ_m \theta(\mu_1(m))$. Notice that m and $\mu(m)$ know each other's type as they were partners in μ, thus $(\alpha_{|\mu,\mu_1})_m(\mu(m), \theta(\mu(m))) = 1$. It follows, therefore, that m wants to form a blocking pair with $\mu(m)$ in μ_1.

It is straightforward to show that $\mu(m)$ also wants to form a blocking pair with m in μ_1. Given the individual rationality of μ, $m \succeq'_{\mu(m)} \mu(m) = \mu_1(\mu(m))$, thus, in view of Remark 1, $\theta(m) \succeq_{\mu(m)} \theta(\mu(m))$. Firstly, notice that $\theta(m) \sim_{\mu(m)} \theta(\mu(m)$ is only possible if $\mu(m) = m$ due to the antisymmetry of agents' preferences over types and the fact that types are assigned without replacement. In this case, m is single under μ and the analysis in the paragraph above implies that μ_1 is blocked unilaterally by m as it is not individually rational. If, on the other hand, $\theta(m) \neq \theta(\mu(m))$, then we have $\theta(m) \succ_{\mu(m)} \theta(\mu(m)) = \theta(\mu_1(\mu(m)))$. Recalling that m and w know each other's type, we have shown that the pair $(m, \mu(m))$ blocks $(\mu_1, \alpha_{|\mu,\mu_1})$. Thus, we can construct the consistent outcome $(\mu_2, \alpha_{|\mu,\mu_1,\mu_2})$ by satisfying either m or $(m, \mu(m))$.

If w is also single in μ, (i.e., $\mu_2(w) = w = \mu(w)$) then we have $\mu_2 = \mu$. Alternatively, if $\mu(w) \neq w$, then we can show that $(\mu(w), w)$ blocks μ_2 using the same logical steps with which we showed that $(m, \mu(m))$ blocked μ_1 as w's partner in μ, $\mu(w)$, is also single in matching μ_2. We can then construct the consistent outcome $(\mu_3, \alpha_{|\mu,\mu_1,\mu_2,\mu_3})$ with $\mu_3 = \mu$ and $\alpha_{|\mu,\mu_1,\mu_2,\mu_3} = \alpha_{|\mu,\mu_1}$ by satisfying the blocking pair $(\mu(w), w)$.

Notice that the pair (m, w) cannot block the consistent outcome $(\mu, \alpha_{|\mu,\mu_1})$ because in the process of beliefs' updating m has learned the type of w and knows that he prefers to be with his partner under matching μ than with w. If there is no blocking pair for $(\mu, \alpha_{|\mu,\mu_1})$, then this is a stable and consistent outcome and we have shown what we need. If there is a blocking pair for $(\mu, \alpha_{|\mu,\mu_1})$, then this was also blocking the self-consistent outcome $(\mu, \alpha_{|\mu})$ as only the beliefs of m and w have changed. Then, following the same procedure as above, we can construct a path by satisfying the blocking pairs that will lead to a consistent outcome that comprises of μ and a system of beliefs in which at most two agents use Bayes' rule to update their beliefs in a consistent manner. This process will continue along the path until all agents who form blocking pairs for $(\mu, \alpha_{|\mu})$ have learned the type of their partners in the blocking pair. Due to the finiteness of the sets M and W, this path will terminate in a finite number of steps with a stable and consistent outcome that contains μ. □

4.2 Assignment with replacement

As already mentioned, the assignment of types with replacement may induce indifferences in agents' preferences over individuals although their preferences over types are strict. The presence of indifferences makes two distinct matchings qualitatively indistinguishable in terms of the blocking opportunities of same-type agents. To make this point clear, let's take a simple matching problem in which there is one woman of an 'orange' type and two men who are of the same 'green' type. Then, it is clear that the

two distinct matchings in which the woman is married to either man are equivalent in terms of the type of the matched pairs (i.e., in both matchings, the orange woman is married to a green man and a green man is single), though they are not equivalent in terms of the identity of the matched individuals. Formally, we summarize this equivalence in the notion of *homomorphic matchings*. We call two matchings μ and μ' *homomorphic* if the number of agents of a given type $\bar{t} \in \Theta$ who are matched under μ to an agent from type $t \in \Theta$ is equal in μ and μ'. Clearly, 'being homomorphic' is a transitive binary relation on the set of all possible matching functions defined over $M \cup W$.

We turn now to the question whether it is possible to start from a self-consistent outcome containing a stable matching for the problem under complete information and reach a stable outcome for the problem under uncertainty such that the two matching parts are homomorphic. As our first example illustrates, some restrictions of agents' preferences over types are needed for the process of consistent belief's updating to deliver such an outcome.

Example 1 The set of men is $\{m_1, m_2, m_3\}$ with each of them being of distinct type, i.e., $\theta(m_1) = t_1$, $\theta(m_2) = t_2$, $\theta(m_3) = t_3$. The set of women is $\{w_1, w_2, w_3\}$ with $\theta(w_1) = \theta(w_2) = s_1$, $\theta(w_3) = s_2$. Consider the following preference profile where only the individually rational types are indicated.

$$m_1 : s_2 \succ s_1 \succ t_1$$
$$m_2 : s_2 \succ s_1 \succ t_2$$
$$m_3 : s_2 \succ s_1 \succ t_3$$

$$w_1 : t_1 \succ s_1$$
$$w_2 : t_1 \succ t_2 \succ s_1$$
$$w_3 : t_3 \succ s_2$$

The corresponding problem under complete information is given below.

$$m_1 : w_3 \succ w_1 \sim w_2$$
$$m_2 : w_3 \succ w_1 \sim w_2$$
$$m_3 : w_3 \succ w_1 \sim w_2$$

$$w_1 : m_1$$
$$w_2 : m_1 \succ m_2$$
$$w_3 : m_3$$

Observe that if an outcome is stable for the problem under uncertainty, then, due to Theorem 2, its matching part should be stable for the problem under complete information, too. Hence, this matching part is either μ defined by $\mu(m_1) = w_1$, $\mu(m_2) = w_2$, $\mu(m_3) = w_3$, or μ' defined by $\mu'(m_1) = w_2$, $\mu'(m_2) = m_2$, $\mu'(w_1) = w_1$, $\mu'(m_3) = w_3$ as only these two matchings are stable for the complete information problem.

Consider first the initial self-consistent outcome $(\mu, \alpha_{|\mu})$ and observe that the pair (m_1, w_2) is blocking it as its both members are not matched under μ, respectively, to

partners of their most preferred type. Moreover, note that (m_1, w_2) is the only blocking pair for this outcome. Thus, starting from the initial outcome $(\mu, \alpha_{|\mu})$ we reach the outcome (μ', α'), where the difference between $\alpha_{|\mu}$ and α' is that, under α', man m_1 knows that w_1 and w_2 are of the same type (s_1).

Next, consider outcome (μ', α'). Clearly, the only blocking pair for this outcome is (w_1, m_2) as w_1 and m_2 do not know each other's type, and, therefore, each one of them holds a strictly positive belief that the other one is of his/her most preferred type: $\alpha'_{w_1}(m_2, t_1) > 0$ and $\alpha'_{m_2}(w_1, s_2) > 0$. By satisfying this blocking pair, we reach outcome (μ'', α'') where agents w_1 and m_2 learn each other's true type, and this makes the difference between α'' and α'.

Outcome (μ'', α''), however, is not individually rational as woman w_1 prefers to be by herself than matched to a man of type t_2. Moreover w_1 is the only blocking agent that blocks this outcome. By satisfying this blocking pair, we reach again matching μ' in the outcome (μ', α''), and we note that no agent updates his/her beliefs in this process.

Finally we point out that in outcome (μ', α'') both women w_1 and w_2 know the true types of the men m_1 and m_2, and vise versa, and that w_3 and m_3 (who only know each other's type) are matched to a partner of their top type. The latter is also the reason why there are no blocking pairs for (μ', α'') and thus, this outcome is stable for the matching problem under uncertainty. Notice however, that μ' is not homomorphic to μ: there is one agent of type s_1 assigned under μ to the single agent of type t_2, while there is no such agent under μ'.

We reach a similar conclusion when we start our analysis with the initial self-consistent outcome $(\mu', \alpha_{|\mu'})$. With respect to this outcome, note that the only pair blocking it is (m_1, w_1). We can then construct the outcome (μ'', α'') with $\mu''(m_1) = w_1$, $\mu''(m_2) = m_2$, $\mu''(w_2) = w_2$, $\mu''(m_3) = w_3$ where the difference between $\alpha_{|\mu'}$ and α'' is that, under α'', man m_1 knows that w_1 and w_2 are of the same type (s_1). Finally, the outcome (μ'', α'') is blocked only by (m_2, w_2) and thus, we can reach the outcome (μ''', α''') where $\mu''' = \mu$ and, additionally to α'', man m_2 knows the type of w_2. Clearly, (μ, α''') is a stable outcome as, in contrast to $(\mu, \alpha_{|\mu})$, the pair (m_1, w_2) is not blocking it due to the fact that the knowledge of man m_1 has expanded along the unique path from $(\mu', \alpha_{|\mu'})$ to (μ, α'''). As we recall, however, μ' and μ are not homomorphic matchings.

It is worth pointing out that in the above example the agents of the same type (w_1 and w_2) differ with respect to their sets of individually rational types. The condition on agents' preference we define next imposes a certain degree of correlation between the preferences of agents of the same type. For $i \in M \cup W$, let $IR(i) = \{t \in \Theta : t \succeq_i \theta(i)\}$ be the set of individually rational types for i. We say that agents' preferences in a matching problem under uncertainty are *dichotomously aligned* if for $i, j \in A$, $A \in \{M, W\}$, we have that $\theta(i) = \theta(j)$ implies $IR(i) = IR(j)$. Notice that the condition of dichotomously aligned preferences imposes no restriction on how two agents of the same type rank their individually rational types.[8] This condition turns

[8] In that sense, preference dichotomous alignment is a much weaker condition than pairwise preference alignment applied to our context. The latter condition was introduced in Pycia (2012) and shown to be necessary and sufficient for core stability in general coalition formation games.

out to be sufficient for an individually rational matching to generate the individual rationality of all matchings that are homomorphic to it.

Lemma 1 *Let a matching problem under uncertainty with types assigned with replacement be given and μ be individually rational for it. If agents' preferences are dichotomously aligned, then all matchings homomorphic to μ are individually rational.*

Proof Suppose that agents' preferences are dichotomously aligned and that, on the contrary, there are a matching μ', which is homomorphic to μ, and an agent $i \in M \cup W$ such that $\theta(i) \succ_i \theta(\mu'(i))$. If $\theta(\mu'(i)) = \theta(\mu(i))$, we have a direct contradiction to the individual rationality of μ. If $\theta(\mu'(i)) \neq \theta(\mu(i))$, then, by μ' and μ being homomorphic, there is an agent k with $\theta(k) = \theta(i)$ such that $\theta(\mu(k)) = \theta(\mu'(i))$. By the individual rationality of $\mu, \theta(\mu(k)) \succeq_k \theta(k)$. However, by agents' preferences being dichotomously aligned, $\theta(\mu'(i)) \succeq_i \theta(i)$. Thus, we have again a contradiction. □

As our next example shows, we cannot expect the individual rationality of a single class of homomorphic matchings to imply agents' preferences being dichotomously aligned, that is, the reverse statement to the one in Lemma 1 does not hold.

Example 2 The set of men is $\{m_1, m_2, m_3\}$ with each of them being of distinct type, i.e., $\theta(m_1) = t_1, \theta(m_2) = t_2$, and $\theta(m_3) = t_3$. The set of women is $\{w_1, w_2, w_3\}$ with $\theta(w_1) = \theta(w_2) = s_1$,and $\theta(w_3) = s_2$. Consider the following preference profile

$$m_1 : s_2 \succ s_1 \succ t_1$$
$$m_2 : s_1 \succ s_2 \succ t_2$$
$$m_3 : s_1 \succ s_2 \succ t_3$$

$$w_1 : t_1 \succ t_2 \succ s_1 \succ t_3$$
$$w_2 : t_1 \succ t_2 \succ t_3 \succ s_1$$
$$w_3 : t_3 \succ t_2 \succ t_1 \succ s_2$$

and take the (largest) class $\{\mu, \mu'\}$ of homomorphic matchings $\mu(m_1) = w_1$, $\mu(m_2) = w_2$, $\mu(m_3) = w_3$, and $\mu'(m_1) = w_2$, $\mu'(m_2) = w_1$, $\mu'(m_3) = w_3$. Notice that these matchings are individually rational; however, agents' preferences are not dichotomously aligned as $t_3 \in IR(w_2) \setminus IR(w_1)$.

Theorem 4 *Let a matching problem under uncertainty with types assigned with replacement and dichotomously aligned preferences be given and μ be stable for the corresponding matching problem under complete information. Then there exist a matching μ^* which is homomorphic to μ, and a system of beliefs α^* such that (μ^*, α^*) is consistent with respect to $(\mu, \alpha_{|\mu})$ and stable for the matching problem under uncertainty.*

Proof Let μ be as above and consider the self-consistent outcome $(\mu, \alpha_{|\mu})$. If there are no blocking pairs for $(\mu, \alpha_{|\mu})$, then we have shown what we need. Notice further that, in view of Remark 1, it is impossible for an agent to unilaterally block $(\mu, \alpha_{|\mu})$.

Suppose now that there is a pair (m, w) that blocks $(\mu, \alpha_{|\mu})$. We show first that it is possible to construct a path leading from $(\mu, \alpha_{|\mu})$ to a consistent outcome containing a matching which is homomorphic to μ.
Consider the consistent outcome $(\mu_1, \alpha_{|\mu,\mu_1})$, where μ_1 is obtained from μ by satisfying the pair (m, w). Since μ is stable for the problem under complete information, we have either $\mu(m) \succeq'_m w = \mu_1(m)$ or $\mu(w) \succeq'_w m = \mu_1(w)$. Suppose, w.l.o.g., that $\mu(m) \succeq'_m w = \mu_1(m)$ holds. The following four cases are possible.

Case 1 ($\mu(m) = m$ and $\mu(w) = w$). Notice that $\mu(m) = m \succeq'_m w = \mu_1(m)$ implies that $\theta(m) \succeq_m \theta(w)$. Moreover, since agents' preferences over types are antisymmetric and men and women are of different types, $\theta(m) \succeq_m \theta(w)$ implies $\theta(m) \succ_m \theta(w)$. We have further that $(\alpha_{|\mu,\mu_1})_m(m, \theta(m)) = 1$ holds since agent m knows his own type. Thus, m unilaterally blocks $(\mu_1, \alpha_{|\mu,\mu_1})$. We can then construct the consistent outcome $(\mu_2, \alpha_{|\mu,\mu_1,\mu_2})$ from $(\mu_1, \alpha_{|\mu,\mu_1})$ by satisfying m. Notice that μ_2 and μ are homomorphic as they coincide. Clearly, the only difference between $\alpha_{|\mu}$ and $\alpha_{|\mu,\mu_1,\mu_2}$ is that m and w know each other's types in $\alpha_{|\mu,\mu_1,\mu_2}$. Thus, if an agent has an incentive to form a blocking pair for $(\mu_2, \alpha_{|\mu,\mu_1,\mu_2})$ with some agent from the opposite market side, then this incentive was also present at the outcome $(\mu = \mu_2, \alpha_{|\mu})$. In other words, the set of pairs blocking $(\mu_2, \alpha_{|\mu,\mu_1,\mu_2})$ coincides with the set of pairs blocking $(\mu, \alpha_{|\mu})$ up to the pair (m, w) which does not block $(\mu_2, \alpha_{|\mu,\mu_1,\mu_2})$ as m has learned the fact that w is not individually rational for him.

Case 2 ($\mu(m) = m$ and $\mu(w) \neq w$). We can proceed as in Case 1 and construct the consistent outcome $(\mu_2, \alpha_{|\mu,\mu_1,\mu_2})$ from $(\mu_1, \alpha_{|\mu,\mu_1})$ by satisfying m. Further, $\theta(\mu(w)) \succeq_w \theta(w)$ follows from the individual rationality of μ and Remark 1. As agents' preferences over types are antisymmetric and men and women are of different types, we have $\theta(\mu(w)) \succ_w \theta(w)$. By the same reasoning, $\theta(w) \succ_{\mu(w)} \theta(\mu(w))$. Since w and $\mu(w)$ know each other's type as they were matched under μ, $(\alpha_{|\mu,\mu_1})_w(\mu(w), \theta(\mu(w))) = (\alpha_{|\mu,\mu_1})_{\mu(w)}(w, \theta(w)) = 1$ and w and $\mu(w)$ are single under μ_2, the pair $(\mu(w), w)$ blocks $(\mu_2, \alpha_{|\mu,\mu_1,\mu_2})$. We can then construct the consistent outcome $(\mu_3, \alpha_{|\mu,\mu_1,\mu_2,\mu_3})$ from $(\mu_2, \alpha_{|\mu,\mu_1,\mu_2})$ by satisfying $(\mu(w), w)$. Clearly, μ_3 and μ are homomorphic as they coincide. By the same reasoning as in Case 1, the set of pairs blocking $(\mu_3, \alpha_{|\mu,\mu_1,\mu_2,\mu_3})$ coincides with the set of pairs blocking $(\mu = \mu_3, \alpha_{|\mu})$ up to the pair (m, w) which does not block $(\mu_3, \alpha_{|\mu,\mu_1,\mu_2,\mu_3})$.

Case 3 ($\mu(m) \neq m$ and $\mu(w) = w$). Given that $\mu(m) \succeq'_m w = \mu_1(m)$, it must be that $\mu(m) \succ'_m w$ or $\mu(m) \sim'_m w$.

Case 3.1 Suppose $\mu(m) \succ'_m w$ and notice that $\theta(\mu(m)) \succ_m \theta(w)$ then follows. In addition, re-call that m knows the type of his partner in μ (i.e., $(\alpha_{|\mu,\mu_1})_m (\mu(m), \theta(\mu(m))) = 1$). Therefore, in matching μ_1, agent m wants to form a blocking pair with his partner under matching μ. It is easy to see that $\mu(m)$ also wants to form a blocking pair with m under matching μ_1: $\theta(m) \succ_{\mu(m)} \theta(\mu(m)) = \theta(\mu_1(\mu(m)))$ follows from the individual rationality of μ, Remark 1, and again by the fact that agents' preferences over types are antisymmetric and men and women are of different types. Since $\mu(m)$ also knows with certainty m 's type, we have established that the pair $(m, \mu(m))$ blocks $(\mu_1, \alpha_{|\mu,\mu_1})$. We can then construct the consistent outcome $(\mu_2, \alpha_{|\mu,\mu_1,\mu_2})$ by satisfying $(m, \mu(m))$. Note that $\mu_2 = \mu$ and thus, the two

matchings are homomorphic. As in the previous two cases, the set of pairs blocking $(\mu_2, \alpha_{|\mu,\mu_1,\mu_2})$ coincides with the set of pairs blocking $(\mu = \mu_2, \alpha_{|\mu})$ up to the pair (m, w) which does not block $(\mu_2, \alpha_{|\mu,\mu_1,\mu_2})$.

Case 3.2 If $\mu(m) \sim'_m w$, then $\theta(\mu(m)) \sim_m \theta(w)$ holds. Together with agents' preferences over types being antisymmetric, and $\mu(w) = w$, this establishes that $\theta(\mu(m)) = \theta(w)$. Therefore, μ_1 is homomorphic to μ . Moreover, since agents' preferences are dichotomously aligned, the fact that μ_1 is homomorphic to μ implies by Lemma 1 that it is also individually rational. Since w and $\mu(m)$ are of the same type, the pair $(m, \mu(m))$ does not block $(\mu_1, \alpha_{|\mu,\mu_1})$. Notice that in this particular case we can say that if a pair of agents with types (t_1, t_2) blocks $(\mu_1, \alpha_{|\mu,\mu_1})$, then a pair of agents with the same types also blocks $(\mu, \alpha_{|\mu})$. As to see the reason, notice first that any blocking pair for $(\mu_1, \alpha_{|\mu,\mu_1})$ containing agents only from the set $(M \cup W) \setminus \{m, w, \mu(m)\}$ is also blocking $(\mu, \alpha_{|\mu})$ since there is no change in the beliefs of these agents. On the other hand, if a blocking pair for $(\mu_1, \alpha_{|\mu,\mu_1})$ contains m (note that $(m, \mu(m))$ does not block $(\mu_1, \alpha_{|\mu,\mu_1})$ since $\theta(\mu(m)) = \theta(w)$ and m know the type of both w and $\mu(m)$), then it also blocks $(\mu, \alpha_{|\mu})$ since $\theta(\mu(m)) = \theta(w)$ is not a top type for m (otherwise, (m, w) would not block $(\mu, \alpha_{|\mu})$) and types are assigned with replacement. Analogously, if the corresponding blocking pair contains w, then that pair also blocks $(\mu, \alpha_{|\mu})$ as w is single under μ. Finally, suppose that $(m', \mu(m))$ is a blocking pair for the outcome $(\mu_1, \alpha_{|\mu,\mu_1})$ and $\theta(m)$ is the top type for agent $\mu(m)$. Then $(m', \mu(m))$ is not blocking $(\mu, \alpha_{|\mu})$ as $\mu(m)$ is matched under μ to an agent of her most preferred type and thus, she has no incentive to divorce him. Notice however that the pair (m', w) with $\theta(w) = \theta(\mu(m))$ is blocking $(\mu, \alpha_{|\mu})$, the reason being that w is single under μ and the preferences of w and $\mu(m)$ are dichotomously aligned (implying that $\theta(w)$ is not a top type for w), the outcome $(\mu, \alpha_{|\mu})$ is self-consistent (implying that m' does not know the type of w as he is not matched to her under μ), and the beliefs of m' are the same in $\alpha_{|\mu}$ and $\alpha_{|\mu,\mu_1}$.

Case 4 ($\mu(m) \neq m$ and $\mu(w) \neq w$). As in Case 3, we have either $\mu(m) \succ'_m w$ or $\mu(m) \sim'_m w$.

Case 4.1 If $\mu(m) \succ'_m w = \mu_1(m)$, then we can proceed along the line of the discussion of Case 3.1 and construct the consistent outcome $(\mu_2, \alpha_{|\mu,\mu_1,\mu_2})$ from $(\mu_1, \alpha_{|\mu,\mu_1})$ by satisfying $(m, \mu(m))$. We can establish that w, too, wants to return to her partner under μ, $\mu(w)$. Notice that w and $\mu(w)$ are single under μ_2 and, moreover, that $\theta(\mu(w)) \succeq_w \theta(w)$ follows from the individual rationality of μ and Remark 1. As agents' preferences over types are antisymmetric and men and women are of different types, we have $\theta(\mu(w)) \succ_w \theta(w)$. By the same reasoning, $\theta(w) \succ_{\mu(w)} \theta(\mu(w))$. Since w and $\mu(w)$ are married under μ, they know each other's type, hence, $(\alpha_{|\mu,\mu_1,\mu_2})_w(\mu(w), \theta(\mu(w))) = (\alpha_{|\mu,\mu_1,\mu_2})_{\mu(w)}(w, \theta(w)) = 1$. Thus the pair $(\mu(w), w)$ is blocking $(\mu_2, \alpha_{|\mu,\mu_1,\mu_2})$. We can then construct the consistent outcome $(\mu_3, \alpha_{|\mu,\mu_1,\mu_2,\mu_3})$ from $(\mu_2, \alpha_{|\mu,\mu_1,\mu_2})$ by satisfying this pair. Again, μ_3 and μ are homomorphic as they coincide. Notice that since m and w have learned each other's types, they do not form a blocking pair under $(\mu_3, \alpha_{|\mu,\mu_1,\mu_2,\mu_3})$. Thus, as in the Case 1, 2, and 3.1, the set of pairs blocking $(\mu_3, \alpha_{|\mu,\mu_1,\mu_2,\mu_3})$ coincides with the set of pairs blocking $(\mu = \mu_3, \alpha_{|\mu})$ up to the pair (m, w).

Case 4.2 On the other hand, if $\mu(m) \sim'_m w$, then from $\theta(\mu(m)) \sim_m \theta(w)$ and the antisymmetry of agents' preferences over types, it follows that $\theta(\mu(m)) = \theta(w)$. Let us consider the pair $(\mu(w), \mu(m))$ and show that it blocks the outcome $(\mu_1, \alpha_{|\mu,\mu_1})$. By the individual rationality of μ, $w \succeq'_{\mu(w)} \mu(w)$, hence, $\theta(w) \succeq_{\mu(w)} \theta(\mu(w))$. Since agents' preferences over types are antisymmetric, and men and women are of different types, $\theta(w) \succ_{\mu(w)} \theta(\mu(w))$ follows. This implies that $\theta(\mu(m)) \succ_{\mu(w)} \theta(\mu(w))$ as we already established that $\theta(\mu(m)) = \theta(w)$. Moreover, $(\alpha_{|\mu,\mu_1})_{\mu(w)}(\mu(m), \theta(\mu(m))) > 0$ holds since types are assigned with replacement. Therefore, $\mu(w)$ wants to form a blocking pair with $\mu(m)$. In order to show that $\mu(m)$ wants to form a blocking pair with $\mu(w)$, notice that by the individual rationality of μ, $\mu(w) \succeq'_w w$ and thus, $\theta(\mu(w)) \succeq_w \theta(w)$. Since $\theta(\mu(m)) = \theta(w)$ and agents' preferences are dichotomously aligned, $\theta(\mu(w)) \succeq_{\mu(m)} \theta(\mu(m))$ and as men and women are of different types, $\theta(\mu(w)) \succ_{\mu(m)} \theta(\mu(m))$ follows. We have finally $(\alpha_{|\mu,\mu_1})_{\mu(w)}(\mu(m), \theta(\mu(m))) > 0$ since types are assigned with replacement. Recalling the fact that both $\mu(w)$ and $\mu(m)$ are single under μ_1, we have shown that the pair $(\mu(w), \mu(m))$ blocks $(\mu_1, \alpha_{|\mu,\mu_1})$ indeed. We can then construct the consistent outcome $(\mu_2, \alpha_{|\mu,\mu_1,\mu_2})$ by satisfying $(\mu(w), \mu(m))$. Notice that, in the matchings μ and μ_2, m and $\mu(w)$ are married to a woman of the same type as $\theta(\mu(m)) = \theta(w)$. Thus μ_2 is homomorphic to μ. In addition, as μ is individually rational and agents' preferences are dichotomously aligned, the fact that μ_2 is homomorphic to μ implies by Lemma 1 that μ_2 is individually rational, too. Let us finally show that, as in Case 3.2, if a pair of agents with types (t_1, t_2) blocks $(\mu_2, \alpha_{|\mu,\mu_1,\mu_2})$, then a pair of agents with the same types also blocks $(\mu, \alpha_{|\mu})$. As indicated above, it is enough for this to focus on blocking pairs for $(\mu_2, \alpha_{|\mu,\mu_1,\mu_2})$ which contain a member from the set $\{m, w, \mu(m), \mu(w)\}$. Notice first that neither $(m, \mu(m))$ nor $(\mu(w), w)$ is a blocking pair for $(\mu_2, \alpha_{|\mu,\mu_1,\mu_2})$ due to $\theta(\mu(m)) = \theta(w)$ with m and $\mu(w)$ knowing the types of $\mu(m)$ and w. Suppose next that a blocking pair for $(\mu_2, \alpha_{|\mu,\mu_1,\mu_2})$ contains either m or w (and a corresponding second member from the set $(M \cup W) \setminus \{m, w, \mu(m), \mu(w)\}$). Since neither $\theta(\mu(m))$ is a top type for m nor $\theta(\mu(w))$ is a top type for w (otherwise, (m, w) would not be a blocking pair for $(\mu, \alpha_{|\mu})$), we have that the corresponding pair also blocks $(\mu, \alpha_{|\mu})$. If a blocking pair for $(\mu_2, \alpha_{|\mu,\mu_1,\mu_2})$ contains $\mu(w)$ (and thus, implying that $\theta(\mu(m))$ is not a top type for $\mu(w)$), then it also blocks $(\mu, \alpha_{|\mu})$ by $\theta(w) = \theta(\mu(m))$. Finally, if $\mu(m)$ belongs to a blocking pair for $(\mu_2, \alpha_{|\mu,\mu_1,\mu_2})$ and the corresponding man is m', then that pair also blocks $(\mu, \alpha_{|\mu})$, provided that $\theta(m)$ is not the top type for $\mu(m)$. However, if $\theta(m)$ is the top type for $\mu(m)$, then the pair (m', w) with $\theta(w) = \theta(\mu(m))$ is blocking $(\mu, \alpha_{|\mu})$ since $\theta(\mu(w))$ is not the top type for w and $(\mu, \alpha_{|\mu})$ is self-consistent.

Thus, in all possible cases we have reached an individually rational outcome containing a matching homomorphic to μ. If there is no blocking pair for the correspondingly constructed outcome, then this outcome is stable and we have established what we need. Moreover, we have shown that, in any of the above cases, if a pair blocks the corresponding outcome with matching part homomorphic to μ, then either the same pair or a pair of the same types of agents was also blocking the self-consistent and individually rational outcome $(\mu, \alpha_{|\mu})$. Then, using the same case separation and logical steps as above, we can construct a path by satisfying the blocking pairs that will lead to a consistent outcome that comprises of a matching homomorphic to μ and

a system of beliefs in which at most four agents (two men and two women) update their beliefs in a consistent manner. The process will continue along the path until all types of agents who form blocking pairs in $(\mu, \alpha_{|\mu})$ have learned the type of their partners in the blocking pair. Due to the finiteness of the sets M and W, this path will terminate in a finite number of steps with a stable and consistent outcome that contains a homomorphic matching to μ. □

5 Conclusion

In this paper, we embed the standard one-to-one matching problem in an environment of uncertainty. We show that it is possible to reach stability from any self-consistent outcome with only minimal information requirements. The study of the links between stability under uncertainty and stability under complete information, however, requires a special attention on how types are assigned to agents. For all but one of our results agents' types and preference are allowed to be completely independent.[9]

Thus, one can view agents' types as that part of their identity that is relevant to the way they are seen and classified by everyone else. Agents' preferences, on the other hand, are the part of their identity that dictates how they judge everyone else. We suggest that our approach to de-couple these two sides of an agent's identity, besides being more realistic, is well suited to inform further investigation into the sources of instability in other hedonic coalition formation problems such as the roommate problem, assignment problem, and in general hedonic games.

The focus of our analysis has been on the existence and construction of paths to a stable outcome under uncertainty. For the purposes of this work we adopted a specific assumption on the decision criteria agents use when deciding how to move along the path. This assumption allowed us to establish strong links between the set of stable matchings under uncertainty and the benchmark set of stable matchings of the corresponding problem under complete information. There are, of course, other possible decision rules, and, more realistically, different agents could adopt different decision rules. We claim, however, that our results on the links between the worlds of uncertainty and certainty would no longer hold in general, should some of the agents adopt different decision criteria. The use and analysis of other behavioral models within our framework, nevertheless, could provide valuable insights, particularly, when investigating the role of memory, the speed of learning, and the appropriate institutions that could facilitate the search along a path to stability in a decentralized manner.

References

Abdulkadiroğlu A, Pathak P, Roth AE (2009) Strategy-proofness versus efficiency in matching with indifferences: redesigning the NYC High school match. Am Econ Rev 99(5):1954–1978

Bikhchandani S (2014) Two-sided matching with incomplete information, working paper

[9] Even in Theorem 4, when we need a formal interdependence between the sets of types and preferences, this relation is as weak as possible. Indeed, this form of interdependence is all that is required to guarantee that the individual rationality of a matching implies that all other indistinguishable matchings, i.e., matchings in which the same types of agents are matched, are also individually rational.

Chakraborty A, Citanna A, Ostrovsky M (2010) Two-sided matching with interdependent values. J Econ Theory 145(1):85–105
Chen B, Fujishige S, Yang Z (2012) Decentralized market processes to stable job matchings with competitive salaries, working paper
Chung K-S (2000) On the existence of stable roommate matchings. Games Econ Behav 33:206–230
Diamantoudi E, Miyagawa E, Xue L (2004) Random paths to stability in the roommate problem. Games Econ Behav 48(1):18–28
Erdil A, Haluk E (2008) What's the matter with tie-breaking? Improving efficiency in school choice. Am Econ Rev 98(3):669–689
Gale D, Shapley LS (1962) College admissions and the stability of marriage. Am Math Mon 69(1):9–15
Iñarra E, Larrea C, Molis E (2008) Random paths to p-stability in the roommate problem. Int J Game Theory 36:461–471
Kadam SV, Kotowski M (2014) Multi-period matching, working paper
Klaus B, Klijn F (2007) Paths to stability in matching markets with couples. Games Econ Behav 58:154–171
Klaus B, Klijn F, Walzl M (2011) Stochastic stability for roommate markets. J Econ Theory 145:2218–2240
Knuth D (1976) Marriages stables. Les Presses de l'Universit é Montréal
Kojima F, Ünver U (2008) Random paths to pairwise stability in many-to-many matching problems: a study on market equilibration. Int J Game Theory 36:473–488
Liu Q, Mailath G, Postlewaite A, Samuelson L (2014) Matching with incomplete information. Econometrica 82(2):541–588
Ma J (1996) On randomized matching mechanisms. Econ Theory 8:377–381
Pais J (2008) Random matching in the college admissions problem. Econ Theory 35:99–116
Pycia M (2012) Stability and preference alignment in matching and coalition formation. Econometrica 80(1):323–362
Roth A (1989) Two-sided matching with incomplete information about others' preferences. Games Econ Behav 1(2):191–209
Roth A, Vande Vate J (1990) Random paths to stability in two-sided matching. Econometrica 58(6):1475–1480
Roth A, Vande Vate J (1991) Incentives in two-sided matching with random stable mechanisms. Econ Theory 1:31–44