

# MATCHING, UNANTICIPATED EXPERIENCES, DIVORCE, FLIRTING, REMATCHING, ETC.\*

Burkhard C. Schipper<sup>†</sup>      Tina Danting Zhang<sup>‡</sup>

February 11, 2025

## Abstract

We study dynamic decentralized two-sided matching in which players may make unanticipated experiences. As they become aware of these experiences, they may change their preferences over players on the other side of the market. Consequently, they may get “divorced” and rematch again with other agents, which may lead to further unanticipated experiences etc. A matching is stable if there is absence of pairwise common belief in blocking. Stable matchings can be destabilized by unanticipated experiences. Yet, we show that there exist self-confirming outcomes that are stable and do not lead to further unanticipated experiences. We introduce a natural decentralized matching process that at each period assigns probability  $1 - \varepsilon$  to the satisfaction of a mutual optimal blocking pair (if it exists) and picks any optimal blocking pair otherwise. The parameter  $\varepsilon$  is interpreted as a friction of the matching market. We show that for any decentralized matching process, frictions are necessary for convergence to stability even without unawareness. Our process converges to self-confirming stable outcomes. Further, we allow for bilateral communication/flirting that changes the awareness and say that a matching is flirt-proof stable if there is absence of communication leading to pairwise common belief in blocking. We show that our natural decentralized matching process converges to flirt-proof self-confirming outcomes.

**Keywords:** Decentralized matching, bi-lateral matching, unawareness, transformative experiences, endogenous preferences, disclosure.

**JEL-Classifications:** D83, C70.

---

\*We thank Kirill Rudov and Gaoji Hu for helpful discussions. Burkhard gratefully acknowledges financial support via ARO Contract W911NF2210282.

<sup>†</sup>Department of Economics, University of California, Davis. Email: bcschipper@ucdavis.edu

<sup>‡</sup>Department of Economics, University of California, Davis. Email: tdzhang@ucdavis.edu

# 1 Introduction

Many matching problems such as whom to marry, which school to choose, which profession to enter, where to take up a residency etc. involve transformative experiences (Paul, 2014) that change who we are, our beliefs, the things and issues we care about, and our preferences. These preference changes have implications for the stability of matchings. For instance, in the US marriage market, 43% of ever-married couples were divorced or widowed, and 23% of married couples are remarried couples (Livingstone, 2014). During marriage, spouses may become aware of intolerable attributes of the partner that they previously were oblivious to it. Or they may experience events in the family and workplace such as addiction, unemployment, domestic violence, mental health issues etc., some of which can only be fully grasped once experienced. For instance, marriage often involves parenting, itself a transformative experience, with profound changes in preferences as spouses become parents. When trying to carefully analyze the causal effect of children on relationship duration in Denmark, Svarer and Verner (2008) find a negative relationship.

Importantly, the transformative experiences, consequent preference changes, and implications for the matchings cannot be completely anticipated and comprehended before they are experienced. Again taking the marriage market as an example, Baker and Emery (1993) report that the median response to the question asking a non-representative sample of marriage license applicants to estimate the fraction of US couples who marry will divorce was 50%, while the median response assessing the likelihood that they personally would divorce was 0%. Maher (2003) finds similar numbers of 52% and 10% respectively, in non-representative samples of the general population, and 48% and 17%, respectively, for law students. Svarer and Verner (2008) find a significant negative impact on the dissolution of the relationship only for the first child but not for later children, pointing to the causal effect of unanticipated rather than anticipated events surrounding parenting on divorce.<sup>1</sup>

How to understand the dynamics and stability of matching markets with transformative experiences and unanticipated preferences changes? The elegant standard matching model of Gale and Shapley (1962) (see Roth and Sotomayor, 1990) gives little guidance as it is static. Models of decentralized matching processes (e.g., Knuth, 1976, Roth and Vande Vate, 1990, Ackermann et al., 2008, Rudov, 2024) focus on the path to stability without considering that experiences in matchings can lead to preference changes. With a few exceptions (e.g., Lazarova and Dimitrov, 2017, Chen and Hu, 2020), these approaches also lack incomplete information about preferences. Recent interesting models of matching under incomplete information allow mostly only for one-sided incomplete information, are static, and unrealistically assume that agents can anticipate all relevant future ex-

---

<sup>1</sup>There is evidence from other matching markets as well. For example in the labor market, on average among baby boomers, men held 12.8 jobs and women held 12.5 jobs from ages 18 to 56 (U.S. Department of Labor, 2023). Using data from employees in a financial institution, Holtom et al. (2017) state that employees report a substantial number of unanticipated shocks, both personal and organizational, and that only unanticipated shocks were significant predictors of staff turnover while none of the anticipated shocks were.

periences (e.g., Liu, 2020, Bikhchandani, 2017, Liu et al., 2014, Pomatto, 2022, Forges, 2004). Because of imperfect information, these models require also sophisticated solution concepts that make use of information revealed from absence of blocking and from counterfactual blocking, which add to the complexity of applying and analyzing matching under incomplete information. In order to capture transformative experience with unanticipated preference changes, we develop dynamic matching games under unawareness but perfect information in Section 3. Unawareness refers to the lack of conception rather than the lack of information and thus provides us with a robust notion of being “unanticipated”. We make use of unawareness structures introduced by Heifetz, Meier, and Schipper (2006, 2008, 2013) to capture asymmetric unawareness among players but simplify them to the case of perfect information with respect to what players are aware of. These structures are complemented with finite state machines that model the change of awareness in experienced matchings and the resulting preference changes.<sup>2</sup> Rather than allowing any arbitrary preference changes, we only consider preference changes due to players becoming aware of events during matchings.

We adapt the notion of stability to asymmetric unawareness by requiring absence of pairwise common (point)-belief in blocking. This reflects the idea that individual willingness to block, even when it coincidentally occurs with the blocking partner, by itself may not necessarily result in a new match but it takes the *agreement* of a pair to block. Such a notion of stability is consistent with Wilson’s (1978) notion of the coarse core that he introduced for exchange economies with asymmetric information and that has been extended to coalitional transferable utility games with unawareness by Bryan, Ryall, and Schipper (2022). Because experiences in a matching, even in a stable matching, can lead to changes in awareness and thus preferences changes, stability of the matching is not enough as a solution concept in the dynamic matching game with unawareness. We also need that awareness and thus preferences do not change in the stable matching. In other words, the state determining awareness and thus preferences must be absorbing (w.r.t. the finite state machine modeling the awareness dynamics based on experiences in matchings). A self-confirming outcome is a pair of a matching and state such that the matching is stable w.r.t. awareness and thus preferences at the state, and the state is absorbing given the matching (see Section 4).<sup>3</sup>

To model the process of (re-)matching, we first revisit in Section 2 the convergence of decentralized matching processes to stability for fixed preferences without unawareness. Knuth (1976) showed that a deterministic process of satisfying blocking pairs may lead to cycles. However, his example is unnatural in the sense that when there are multiple blocking pairs at a stage, he satisfies a blocking pair that is not mutually optimal. A blocking pair is optimal for a player if it is the best blocking pair for the player. It is

---

<sup>2</sup>Our class of dynamic matching games with unawareness have the flavor of stochastic games, another invention by Shapley (1953), in non-cooperative game theory except that ours are cooperative games of matching, state transitions are deterministic, and we allow for asymmetric unawareness.

<sup>3</sup>Our terminology is inspired by self-confirming equilibrium in non-cooperative game theory, which are outcomes in which players maximize expected utility w.r.t. beliefs consistent with their observations and these observations are generated by the play in equilibrium (see for instance Fudenberg and Levine (1993) and Battigalli and Guaitoli (1997) for games without unawareness and Schipper (2021) for games with unawareness).

mutually optimal if it is the best blocking pair for both players. In markets without frictions, we would expect that mutually optimal blocking pairs are satisfied whenever they exist. Knuth (1976) example leaves open the conjecture that a process of satisfying mutually optimal blocking pairs, when they exist, leads to the stability. We show with a new example that this is not the case. This implies that for the process of randomly picking blocking pairs to converge to stability as in Roth and Vande Vate (1990), it is necessary that strict positive probability is assigned to satisfying blocking pairs that are *not* mutually optimal blocking pairs. This yields an important corollary: Frictions are necessary for decentralized matching processes to converge to stability. This is contrary to the intuition in economics that markets function best in the absence of frictions.

Armed with the insights from our study of decentralized matching processes with full awareness, we introduce in Section 4 a decentralized matching process that w.r.t. any unstable matching satisfies with probability  $1 - \varepsilon$  a randomly picked mutually optimal blocking pair, if it exists, and picks an optimal blocking pair otherwise. W.r.t. states (and hence awareness), the dynamics follows the finite state machine mentioned above. We show that this decentralized process converges to a self-confirming outcome. Moreover, we demonstrate by example that in such a self-confirming outcome, players may remain unaware.

In real life, awareness may not just be raised via experiences in matches but also through communication. In Section 5, we extend the model to communication by allowing players to raise awareness of some other players with the intend to create pairwise common belief in blocking (i.e., “flirting”). Formally, this is modeled with another finite state machine. Communication can have two kinds of effects in our model: First, because it potentially raises awareness of some players, it can change their preferences and thus create blocking pairs. Second, it may raise awareness in such a way to augment blocking pairs with pairwise common belief in blocking. In both cases, such communication may invite further communication, leading to further changes of awareness, beliefs, and preferences etc. A flirt-proof stable outcome consists of a matching and state such that communication does not change the state given the matching and the matching is stable given the state.<sup>4</sup> When flirt-proof stable outcome is absorbing (w.r.t. the finite state state machine modeling changes in awareness given the experience in matchings), then we call it a flirt-proof self-confirming outcome. We show convergence of our decentralized matching process with communication to flirt-proof self-confirming outcome and that it refines the set of self-confirming outcomes.

It is often argued that divorce improves well-being as it avoids that estranged spouses need to suffer through a marriage. Longitudinal studies of divorce do not find clear-cut systematic improvement of well-being or mental health (Spanier and Furstenberg, 1982, Lucas, 2005, Symoens et al., 2013). While divorce allows one to escape a match, it is far from clear that the subsequent dynamic rematching process leads to a better outcome. One might hypothesize that the person initiating a divorce, let’s call him/her the divorcee, should be better off w.r.t. the her changed preferences while the person

---

<sup>4</sup>Flirt-proof stability is reminiscent of extensions of Wilson’s (1978) core concepts for exchange economies under asymmetric information that allow for various forms of information revelation in the core; see Forges and Serrano (2013) for a survey.

that is divorced, let's call him/her the divorced, might become worse off as he/she loses her/his partner that was matched in prior stable matching. While earlier studies find no evidence for improved well-being of the divorcee (Spanier and Furstenberg, 1982), more recent studies (Symoens et al., 2013) find a significant positive effect of being the one initiating the divorce. We show in Section 6 that the welfare of divorcees can go in any direction even for the same matching game and same initial conditions. That is, divorcees may end up better, equivalent, or worse off w.r.t. their preferences after an awareness change. The same we show for the divorced. So in terms of welfare effects of divorce, anything goes.

In Section 7 we study welfare in our setting of unanticipated experiences, changing awareness, and endogenous preferences. Since we consider unawareness but perfect information, we work with ordinal preferences, which makes Pareto efficiency a natural welfare criterion just like in the case of matching under complete information. However, Pareto efficiency is not straightforward in settings with endogenous preferences such as ours. Even if there are no Pareto improvements for a matching *before* the matching is experienced, there might be Pareto improvements to the matching *after* it is experienced (and vice versa). We define a notion of persistent Pareto efficiency. We argue that self-confirming outcomes may not be persistent Pareto efficient but every persistent Pareto efficient outcome is a self-confirming outcome.

In the final Section 8, we discuss the effect of infidelity. We also discuss how players can be confused about other players in self-confirming outcomes leading to (stable) awareness of unawareness (Schipper, 2024). Finally, we discuss the related literature and point to potential extensions of our work.

## 2 Decentralized Matching (without Unawareness) Revisited

In this section, we revisit standard decentralized matching without unawareness in order to show that arbitrary small frictions are necessary for stable outcomes to emerge in decentralized random matching. This will motivate our decentralized random matching process that we use in later sections on matching with unawareness.

Consider a standard two-sided marriage matching market with non-transferable utility (and without unawareness)  $\langle M, W, (\succ_m)_{m \in M}, (\succ_w)_{w \in W} \rangle$  with a nonempty finite set of men  $M$ , a nonempty finite set of women  $W$ , and for each man  $m \in M$ ,  $\succ_m$  is a strict preference relation over  $W \cup m$  while for each woman  $w \in W$ ,  $\succ_w$  is a strict preference relation over  $M \cup w$ .

A *matching* is a one-to-one function  $\mu : M \cup W \rightarrow M \cup W$  such that  $\mu(m) = w$  if and only if  $\mu(w) = m$ . If for  $i \in M \cup W$ ,  $\mu(i) = i$ , then  $i$  is unmatched. Given a matching  $\mu$ , a man  $m$  and a woman  $w$  form a *blocking pair* if  $\mu(m) \neq w$ ,  $m \succ_w \mu(w)$ , and  $w \succ_m \mu(m)$ . That is,  $m$  and  $w$  form a blocking pair if they are not matched to each other given  $\mu$  and they prefer each other over their current match, respectively. A matching  $\mu$  is individually rational if  $\mu(i) \succeq_i i$  for all  $i$ . A matching  $\mu$  is *stable* if it is

individually rational and there is no blocking pair.

Consider an initial matching  $\mu$  and a decentralized process of successively satisfying blocking pairs. A blocking pair is *satisfied* if they leave their current match and are match to each other, leading to a new matching. When does a process satisfying blocking pairs lead to a stable matching?

## 2.1 Knuth's Cycle

Knuth (1976) showed with the help of an example that the decentralized process of satisfying blocking pairs may not lead to a stable matching. We briefly present his cyclic example and discusses its weaknesses.

**Example 1 (Knuth, 1976)** Consider  $M = \{m_1, m_2, m_3\}$  and  $W = \{w_1, w_2, w_3\}$  with strict preferences given by the following rank order lists (from most preferred to least preferred, respectively):

$$\begin{aligned} \succ_{m_1}: & w_2, w_1, w_3 \\ \succ_{m_2}: & w_1, w_3, w_2 \\ \succ_{m_3}: & w_1, w_2, w_3 \\ \succ_{w_1}: & m_1, m_3, m_2 \\ \succ_{w_2}: & m_3, m_1, m_2 \\ \succ_{w_3}: & m_1, m_3, m_2 \end{aligned}$$

Knuth's (1976) example involves a cycle of eight matchings.<sup>5</sup> Let  $\mu_1 = \begin{pmatrix} m_1 & m_2 & m_3 \\ w_1 & w_2 & w_3 \end{pmatrix}$  be the first matching. That is, in matching  $\mu_1$ , man  $m_1$  is matched to woman  $w_1$ , man  $m_2$  is matched to  $w_2$ , etc. This matching is not stable because there are blocking pairs  $(m_1, w_2)$  and  $(m_3, w_2)$ . Satisfying  $(m_1, w_2)$  leaves their original matched partners,  $w_1$  and  $m_2$ , unmatched and leads a new matching  $\mu_2 = \begin{pmatrix} m_1 & m_2 & m_3 & w_1 \\ w_2 & m_2 & w_3 & w_1 \end{pmatrix}$ . Subsequent matchings of the process are shown in the bipartite graphs in Figure 1, where each matching is a bipartite graph, a match is represented with a solid line, and all blocking pairs are indicated with either dashed or dotted lines. The solid triangles represent the blocking pairs that are satisfied in order to reach the next matching.

As shown in Figure 1, when we satisfy blocking pairs in the sequence indicated by solid triangles, the process loops back to the initial matching after eight rounds. Thus, the example shows that deterministic decentralized matching processes can lead to cycles that prevent the emergence of a stable matching. Deterministic refers to the fact that when there are multiple blocking pairs, we do not choose one randomly to satisfy but pick a particular with probability one.  $\square$

---

<sup>5</sup>We use the exposition of Knuth's example from Roth and Vande Vate (1990), as it uses the prevalent notations while the original exposition does not.

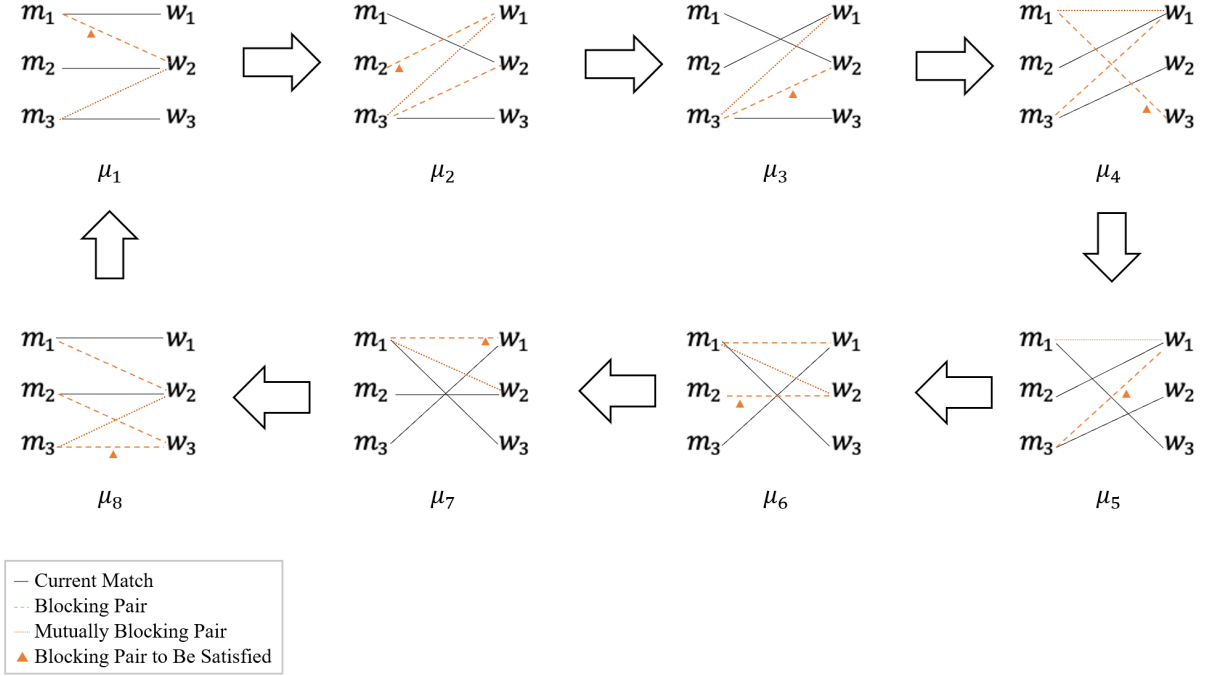


Figure 1: Knuth's Cycle

At a second glance, the Knuth's example is not fully convincing because it involves satisfying a sequence of rather unnatural blocking pairs. For example, in order to create Knuth's cycle, at matching  $\mu_1$  the blocking pair  $(m_1, w_2)$  needs to be satisfied instead of blocking pair  $(m_3, w_2)$ . Both blocking pairs involve woman  $w_2$ . Given that potentially both men,  $m_3$  and  $m_2$  compete for being matched to her, it is much more natural to satisfy blocking pair  $(m_3, w_2)$  because  $w_2$  prefers  $m_3$  over  $m_1$ . If  $w_2$  is asked to rematch, she would surely pick  $m_3$  instead of  $m_1$ .

In order to formalize our observation, we follow Bennett (1994) in saying that pair  $(i, j)$  is the *optimal blocking pair* for  $i$  at matching  $\mu$  if  $j$  is  $i$ 's most preferred individual among the set of individuals who forms a blocking pair with agent  $i$  at the matching  $\mu$ . Furthermore, we call  $(i, j)$  a *mutually optimal blocking pair* at  $\mu$  if it is an optimal blocking pair for both  $i$  and  $j$  at  $\mu$ . The mutually optimal blocking pairs are represented in Figure 1 by the dotted lines.

Knuth's cycle avoids choosing the optimal blocking pairs whenever they exist, which is in every round. We think this is very unnatural because in a frictionless marriage market optimal blocking pairs should be able "to meet" and consequently form a match. If in Knuth's example a mutually optimal blocking pair is chosen at any round, the process is guaranteed to get out from the cycle. For example, if we satisfy the mutually optimal blocking pair  $(m_3, w_2)$  at  $\mu_1$  instead of  $(m_1, w_2)$ , we reach the matching  $\mu'_2 = \begin{pmatrix} m_1 & m_2 & m_3 & w_3 \\ w_1 & m_2 & w_2 & w_3 \end{pmatrix}$ , which has only one blocking pair,  $(m_2, w_3)$ . Satisfying

$(m_2, w_3)$  leads to a stable matching  $\mu'_3 = \begin{pmatrix} m_1 & m_2 & m_3 \\ w_1 & w_3 & w_2 \end{pmatrix}$ . Thus, while Knuth's example demonstrates that deterministic decentralized matching may lead to a cycle, it leaves open the possibility that deterministic decentralized matching in which more naturally at each step a mutual optimal blocking pair is satisfied, whenever it exist, leads to a stable outcome. Unfortunately, we show in the next section that this is not the case.

## 2.2 A Cycle with Mutually Optimal Blocking Pairs

In this section, we improve upon Knuth's cycle. We present a new example with a cycle that involves satisfying mutually optimal blocking pairs. At every stage, there exists a *unique* mutually optimal blocking pair. When satisfying the unique mutually optimal blocking pair at every stage, the process ends up in a cycle. Compared to Knuth's cycle, we need now four participants on each side of the market.

**Example 2** Consider a marriage market with four men  $M = \{m_1, m_2, m_3, m_4\}$  and four women  $W = \{w_1, w_2, w_3, w_4\}$ , with the strict preference profiles, respectively, given by the rank order lists:

$$\begin{aligned} \succ_{m_1}: & w_2, w_4, w_3, w_1 \\ \succ_{m_2}: & w_4, w_2, w_1, w_3 \\ \succ_{m_3}: & w_1, w_3, w_2, w_4 \\ \succ_{m_4}: & w_3, w_1, w_4, w_2 \\ \succ_{w_1}: & m_1, m_2, m_4, m_3 \\ \succ_{w_2}: & m_4, m_3, m_2, m_1 \\ \succ_{w_3}: & m_2, m_1, m_3, m_4 \\ \succ_{w_4}: & m_3, m_4, m_1, m_2 \end{aligned}$$

Let the first matching be given by  $\mu_1 = \begin{pmatrix} m_1 & m_2 & m_3 & m_4 \\ w_2 & w_1 & w_3 & w_4 \end{pmatrix}$ . It has exactly one blocking pair  $(m_2, w_2)$ , which is also a mutually optimal blocking pair. Satisfying this blocking yields the second matching  $\mu_2 = \begin{pmatrix} m_1 & m_2 & m_3 & m_4 & w_1 \\ m_1 & w_2 & w_3 & w_4 & w_1 \end{pmatrix}$ . Subsequent steps are shown in the sequence of bipartite graphs (each representing a matching) in Figure 2. After nine rounds of satisfying the unique mutually optimal blocking pairs, respectively, we reach the matching  $\mu_{10} = \mu_2$ , completing the cycle. This example improves Knuth's example by demonstrating that even when satisfying unique mutually optimal blocking at each step of the process, we can get into a cycle.  $\square$

More generally, we conclude:

**Proposition 1** *There does not exist a decentralized process of satisfying blocking pairs that always chooses mutually optimal blocking pairs when they exist and always reaches a stable matching.*



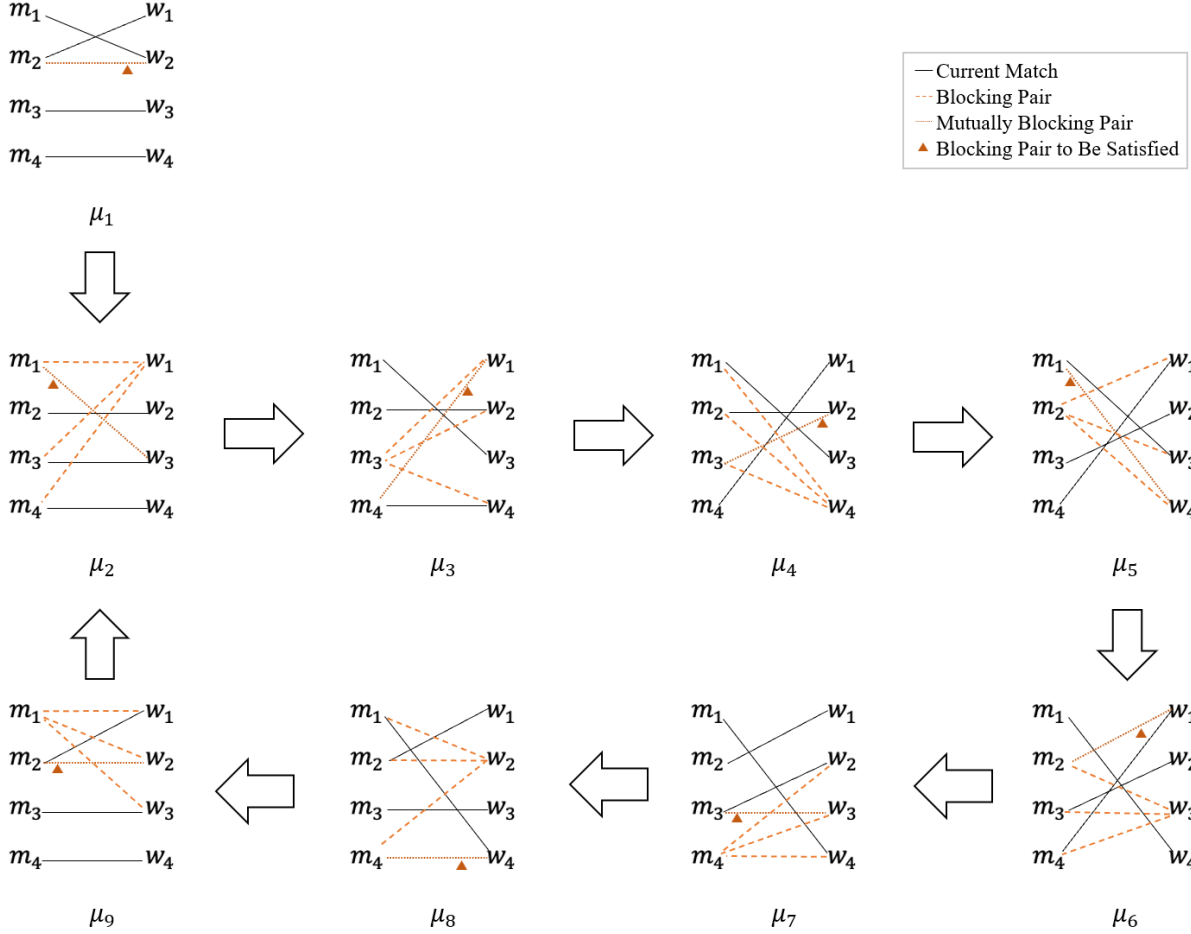


Figure 2: Cycle with Unique Mutually Optimal Blocking Pairs Only

Kirill Rudov kindly informed us he has independently shown the same observation Rudov (2024, Proposition 4 in the online appendix). His counterexample makes use of a five-by-five market while we find a counterexample for four-by-four markets.

### 2.3 Random Paths to Stability

Prior section showed that we need to go beyond deterministic processes in order to guarantee stable outcomes in decentralized matching markets. Roth and Vande Vate (1990) showed that we can reach stable outcomes when a process that randomly picks among blocking pairs. In particular, they showed that starting from an arbitrary matching  $\mu$ , there exist a finite sequence of matchings  $\mu_1, \dots, \mu_k$  such that  $\mu = \mu_1$  and  $\mu_k$  is stable, and for  $\mu_i, i = 1, \dots, k - 1$ , there is a blocking pair  $(m_i, w_i)$  that if satisfied yields matching  $\mu_{i+1}$ . As a corollary, they show that the process of satisfying randomly chosen blocking pairs will converge to a stable matching with probability one.

Our example in Section 2.2 shows that the process of randomly chosen blocking pairs

of Roth and Vande Vate (1990) does not always choose a mutually optimal blocking pairs even when they exists. Thus, we conclude that the result by Roth and Vande Vate (1990) cannot be strengthened to prioritizing *mutually optimal* blocking pairs (if they exist). More formally:

**Proposition 2** *The process of satisfying randomly chosen blocking pairs must put positive probability on sub-optimal blocking pairs (where one of the involved agent prefers another blocking pair that he/she is in) in order to converge to a stable matching.*

This observation has a profound economic corollary: *Frictions that prevent mutually optimal blocking pairs to be satisfied are necessary for any decentralized matching processes to reach stability.* This observation runs counter to standard economic wisdom that frictions hamper the functioning of decentralized markets. In contrast, for decentralized matching markets, small frictions are necessary for reaching stable matchings. In Appendix A, we present another application of our cyclic example to entry in matching markets.

## 2.4 Arbitrarily Small Frictions are Enough

While we just have shown that frictions are necessary to reach stable outcomes in decentralized random matching, we will now show that we can make these frictions arbitrarily small.

We define the unperturbed process as follows: Starting from an arbitrary matching, if this matching is stable, no change occurs; if mutually optimal blocking pair(s) exist(s), satisfy one of them with equal probability; otherwise, satisfy one of the optimal blocking pair(s) with equal probability, which always exists when the matching is unstable. Apply the same rules to the next matching.

The perturbed process we define as follows: Starting from an arbitrary matching, if this matching is stable, no change occurs; if mutually optimal blocking pair(s) exist(s), with probability  $1 - \varepsilon$  select randomly one mutually optimal blocking pair and satisfy it, and with probability  $\varepsilon$  satisfy a randomly selected optimal blocking pair that is not mutually optimal; otherwise, satisfy a randomly selected optimal blocking pair. Apply the same rules to the next matching.

For the unperturbed process, both stable matchings and cycles formed by satisfying unique mutually optimal blocking pairs are absorbing sets. However, we argue that for an arbitrarily small  $\varepsilon > 0$ , which captures the frictions that prevent mutually optimal blocking pairs to match, the perturbed process converges to a stable matching in finite time with probability 1. Given a marriage market  $\langle M, W, (\succ_i)_{i \in M \cup W} \rangle$ , cycle(s) formed by satisfying unique mutually optimal blocking pairs at each step may or may not exist. If there is no such cycle, our claim is trivially true. If a cycle by satisfying the unique mutually optimal blocking pair at each step exists, conditional on entering such a cycle, the probability that the process stays in this cycle for exactly  $k$  period is  $(1 - \varepsilon)^{k-1} \varepsilon$ . Hence the probability that the process stays in this cycle forever is  $\lim_{k \rightarrow \infty} (1 - \varepsilon)^{k-1} \varepsilon = 0$ , which

means that the process leaves the cycle in finite time with probability 1. After leaving the cycle, the process might come back to this cycle or enter another cycle. However, the same argument applies, and the process leaves the cycle again in finite time with probability 1. As soon as the process catches one of Ackermann et al. (2008)'s optimal blocking pairs, it is trapped in a basin of attraction of a stable matching. Note that by the  $\varepsilon$ -events, any optimal blocking pair has a strict positive probability. Thus, we can make use of the result by Ackermann et al. (2008) according to which for any unstable matching, there exist a finite sequence of satisfying optimal blocking pairs that leads to a stable matching. The process may potentially spend long periods in the cycle(s), but only the stable matchings are absorbing

The discussion so far motivates for the later sections our choice the process of randomly chosen blocking pairs that prioritizes mutual optimal blocking pairs when they exist but at each step puts arbitrarily small probability on satisfying just optimal blocking pairs.

### 3 Matching under Unawareness

We continue to consider a two-sided marriage matching market with non-transferable utility with an equal number of men and women. However, different from the previous section we now allow for asymmetric awareness among players. The preference rankings of players can now depend on their awareness. We model asymmetric awareness using a simplified version of unawareness structures by Heifetz et al. (2006, 2013). Consider a finite lattice of disjoint finite state spaces  $(\mathcal{S}, \supseteq)$ . Denote by  $\check{S} = \bigvee_{S \in \mathcal{S}} S$  the join of the lattice. For any spaces  $S, S' \in \mathcal{S}$  with  $S' \supseteq S$ , there is a surjective projection  $r_S^{S'} : S' \rightarrow S$ . Projections commute, i.e., for any  $S, S', S'' \in \mathcal{S}$  with  $S'' \supseteq S' \supseteq S$  we have  $r_S^{S''} = r_S^{S'} \circ r_{S'}^{S''}$ . Moreover, for any  $S \in \mathcal{S}$ ,  $r_S^S$  is the identity on  $S$ . Let  $\Omega := \bigcup_{S \in \mathcal{S}} S$ . We sometimes use  $\omega_S$  to denote the projection of  $\omega$  to space  $S$ . We also write  $S_\omega$  for the space that contains  $\omega$ .

For any  $S \in \mathcal{S}$  and  $D \subseteq S$ , denote by  $D^\uparrow := \bigcup_{S' \supseteq S} (r_S^{S'})^{-1}(D)$ . An *event*  $E \subseteq \Omega$  is defined by a base-space  $S \in \mathcal{S}$  and a base  $D \subseteq S$  such that  $E := D^\uparrow$ . For any  $S \in \mathcal{S}$ , denote by  $\Sigma(S)$  the set of events with base-space  $S$ .

Awareness affects preferences of players. To make it explicit, we let preferences of players and their (point-)beliefs depend on states. To this end, we introduce for each player a point-belief type mapping  $t_i : \Omega \rightarrow \Omega$  such that

- (i) For any  $S \in \mathcal{S}$ ,  $\omega \in S$  implies  $t_i(\omega) = r_{S'}^S(\omega)$  for some  $S' \preceq S$ .
- (ii) For any  $S, S', S'' \in \mathcal{S}$  with  $S'' \supseteq S' \supseteq S$ ,  $\omega \in S''$ ,  $t_i(\omega) \in S'$  implies  $t_i(\omega_S) = r_S^{S'}(t_i(\omega))$ .
- (iii) For any  $S, S', S'' \in \mathcal{S}$  with  $S'' \supseteq S' \supseteq S$ ,  $\omega \in S''$  and  $t_i(\omega_{S'}) \in S$  implies  $S_{t_i(\omega)} \supseteq S$ .

Property (i) means that a player cannot be aware of more than what is described by a state. Properties (ii) and (iii) are consistency conditions on how awareness is related

across states. These properties specialize the properties of type mappings in Heifetz, Meier, and Schipper (2013) to our case of point-belief type mappings. That is, the structure so far is a special case of unawareness structures in Heifetz, Meier, and Schipper (2013). Here we only allow for point-beliefs (rather than non-degenerate beliefs) and unawareness. We refer to  $S_{t_i(\omega)}$  as player  $i$ 's awareness level at state  $\omega$ . Player  $i$  with point-belief  $t_i(\omega)$  can reason over all states in  $S_{t_i(\omega)}$  and in spaces  $S \leq S_{t_i(\omega)}$ . In particular, such a player is aware of all events with a base-space less expressive than  $S_{t_i(\omega)}$ .

For simplicity, we restrict unawareness structures by ruling out redundancies.

**Assumption 1 (No Redundancies)** For any  $\omega, \omega' \in \Omega$  with  $\omega \neq \omega'$  there exist  $i \in M \cup W$  such that  $S_{t_i(\omega)} \neq S_{t_i(\omega')}$ .

That is, different states imply that some player's awareness must differ.

We also impose a richness condition according to which any combination of awareness among players is feasible.

**Assumption 2 (Richness)** For any profile of spaces  $(S_i)_{i \in M \cup W} \in \mathcal{S}^{|M \cup W|}$ , there exist  $\omega \in \tilde{S}$  such that for any  $i \in M \cup W$ ,  $S_{t_i(\omega)} = S_i$ .

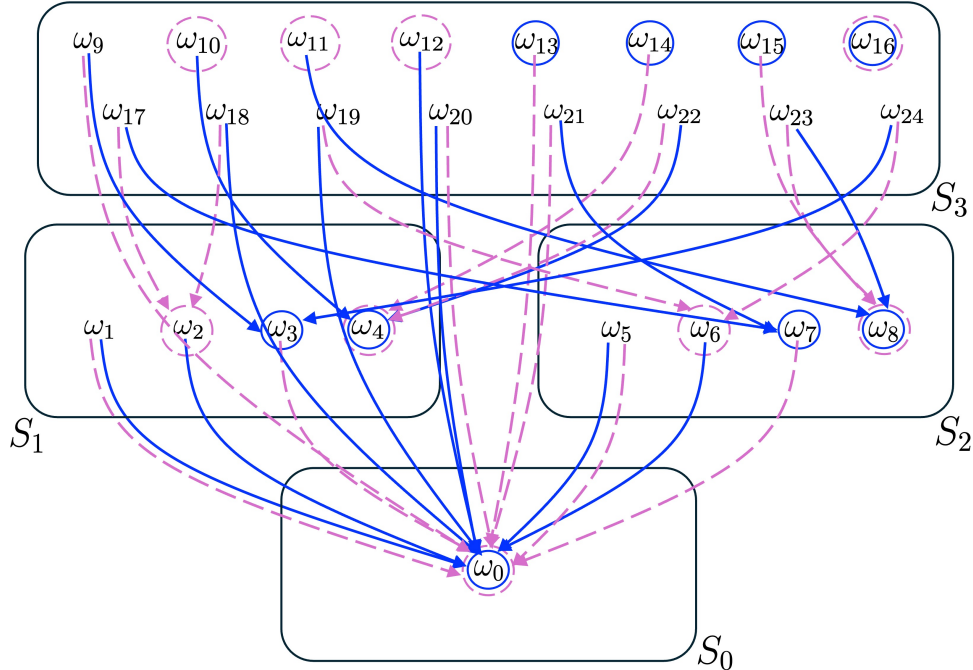


Figure 3: Illustration of an Unawareness Structure with Two Players and Two Characteristics

We illustrate an unawareness structure with point-beliefs for two players satisfying Assumptions 1 and 2 in Figure 3. There are two characteristics of which players may or may not be unaware. Thus, we have four spaces. The left space  $S_1$  models the situation

in which only the first characteristic is expressible, while the right space  $S_2$  models the situations when only the second characteristic is expressible. On the upmost space  $S_3$  both characteristics are expressible. In the lowest space  $S_0$  none of the characteristics is expressible. States models combinations of awareness of both agents in a consistent way as described by the type mapping and the conditions imposed on the type mapping. The type mapping of one player (i.e., the man) is blue solid while the one of the other player is pink dash (i.e., the woman). For instance, at state  $\omega_{10}$  the woman is aware of both characteristics, because her type mapping maps  $\omega_{10}$  to itself, while the man is only aware of the first characteristics, because his type mapping maps  $\omega_{10}$  to  $\omega_4$  in  $S_1$ . At  $\omega_4$  both players are aware of the first characteristic. Thus, at  $\omega_{10}$  the man point-believes that the woman is aware of the first characteristics, which is consistent with his awareness at  $\omega_{10}$  since he can only envision the first characteristic but not the second at  $\omega_{10}$ . More interesting is for instance state  $\omega_{24}$ . At that state, the man is aware of the first characteristic only while the woman is only aware of the second characteristics only because their type mappings map to  $\omega_3$  and  $\omega_6$ , respectively. At  $\omega_3$ , the man correctly believes that the woman is unaware of characteristic 1 while at  $\omega_6$  the woman correctly believes that the man is unaware of characteristic 2. This is indicated by the type mappings that map the respective states to  $\omega_0$ . We omit the projections in order not to clutter the figure further. Yet, it should be clear that for instance  $\omega_{13}, \omega_{14}$  and  $\omega_{15}$  all project to  $\omega_7$  etc. Assumptions 1 and 2 are easy to verify for instance in spaces  $S_1$  and  $S_2$  (albeit less obvious in  $S_3$ ). Note that any unawareness structure for correct point-beliefs of two players and two characteristics satisfying Assumptions 1 and 2 must look like Figure 3. With more players, which we need in non-trivial matching games, or more characteristics, they become more complex. However, in our examples we will often just focus on the states that are relevant for the feature we aim to illustrate with the example.

Let  $\mathcal{P}_i$  be the set of all strict preferences over  $W \cup \{i\}$  if  $i \in M$  and  $M \cup \{i\}$  if  $i \in W$ . For every player  $i$  there is a preference mapping defined by  $\succ_i: \Omega \rightarrow \mathcal{P}_i$ . That is, every player at each state has preferences over players of the other side and outcomes in which (s)he stays alone. Since preferences shall only be affected by awareness, we require that each player's preference mapping is constant within each space/awareness level. That is, for any  $S \in \mathcal{S}$  and  $\omega, \omega' \in S$ ,  $\succ_i(\omega) = \succ_i(\omega')$ . We assume for simplicity that if for  $\omega \in S$  and  $\omega' \in S'$  we have  $j \succ_i(\omega) j'$  and  $j \succ_i(\omega') j'$ , then  $j \succ_i(\omega'') j'$  for  $\omega'' \in S \vee S'$ . That is, if  $i$  prefers  $j$  to  $j'$  with awareness level  $S$  and also with awareness level  $S'$ , then  $i$  also prefers  $j$  to  $j'$  with the joint awareness level  $S \vee S'$ .

Recall that a *matching* is a one-to-one function  $\mu: M \cup W \rightarrow M \cup W$  such that  $\mu(m) = w$  if and only if  $\mu(w) = m$ . If for  $i \in M \cup W$ ,  $\mu(i) = i$ , then  $i$  is unmatched. Let  $\mathbf{M}$  denote the set of all matchings.

Matchings may allow players to discover new characteristics that subsequently change their preferences. That is, matchings may lead to changes awareness and thus beliefs and preferences. This is modeled via a finite state machine  $\langle \Omega, \mathbf{M}, \tau \rangle$  with transition function  $\tau: \Omega \times \mathbf{M} \rightarrow \Omega$  defined as follows:

- (i) For any  $\omega \in \check{S}$  and  $\mu \in \mathbf{M}$ , we require  $\tau(\omega, \mu) \in \check{S}$  such that  $S_{t_i(\tau(\omega, \mu))} \supseteq S_{t_i(\omega)}$  for

$i \in M \cup W$ . That is, every player's awareness can never decrease.

- (ii) We extend  $\tau$  to all states in  $\Omega$  by for any  $S \in \mathcal{S}$ ,  $\omega \in \check{S}$ , and  $\mu \in \mathbf{M}$ ,  $\tau \left( r_S^{\check{S}}(\omega), \mu \right) = r_S^{\check{S}}(\tau(\omega, \mu))$ .

Note that we allow  $i$  not only to become aware from her/his own match but also from a change of the matching that does not involve  $i$ . Note further that states implicitly encode four features: First, they encode the point-belief of each player via the type-mapping. Point beliefs are correct up to differences in awareness. Second, states encode awareness of each player, also via the type-mapping. The awareness level is given by the space in which the value of the player's type is located. Third, states encode preferences of each player via the preference map. Finally, they encode the transition across states (and thus the change of point-beliefs, awareness, and preferences of each player) conditional on matchings via the finite state machines. Assumption 1 has implications beyond awareness. For transitions, the assumption implies that for each profile of awareness and matching, each player has a unique experience and thus a transition to unique potentially different profile of awareness. I.e., the assumption rules out the case where the same profile of awareness and matching can give rise to different experiences for one player.

**Definition 1** *A finite dynamic two-sided matching game with unawareness is defined by*

$$\langle (\mathcal{S}, \supseteq), (r_S^{S'})_{S' \supseteq S}, M, W, (t_i)_{i \in MUW}, (\succ_i)_{i \in MUW}, \mathbf{M}, \tau \rangle.$$

The model introduced so far can be interpreted for instance as modeling unawareness of preference-relevant characteristics of players and their dynamics. E.g., consider a set of preference-relevant characteristics. For each subset of characteristics, there is a state space modeling everything relevant to the players but only pertaining to this subset of characteristics. The lattice order  $\supseteq$  on state spaces is induced by set inclusion on the set of characteristics. A state describes now for each player of which characteristics in the subset he/she is aware of. For each space, preferences are constant in states because they are driven by the characteristics associated with the state space. All what differs from state to state is the awareness of characteristics by players. At one state in the state space the player may be aware of all characteristics associated with the space, while at another state of the state space the player may be unaware of some and thus "live" in an even less expressive state space. The preference of the player is now given by her preference in the less expressive state space.

## 4 Self-Confirming Stable Outcomes

In this section, we will define step-by-step our solution concept. Our aim is a solution concept that features a stable matching given beliefs and stable beliefs given the matching.

For any  $m \in M$  and  $w, w' \in W$ , define  $[w \succ_m w'] := \{\omega \in \Omega : w \succ_m(\omega) w'\}$ . Analogously, define  $[m \succ_w m']$  for any  $w \in W$  and  $m, m' \in M$ . Given matching  $\mu \in \mathbf{M}$ ,

the set of states in which  $(m, w)$  forms a *blocking pair* is  $[m \succ_w \mu(w)] \cap [w \succ_m \mu(m)]$ . Similarly, for any  $i \in M \cup W$  and matching  $\mu \in \mathbf{M}$ , define  $[i \succ_i \mu(i)] := \{\omega \in \Omega : i \succ_i(\omega) \mu(i)\}$ . This is the set of states in which player  $i$  prefers to stay alone rather than stay with her/his current partner in the matching  $\mu$ .

For any event  $E \subseteq \Sigma$  and  $i \in M \cup W$ , let  $K_i(E) = \{\omega \in \Omega : t_i(\omega) \in E\}$ , if there exists  $\omega \in \Omega$  such that  $t_i(\omega) \in E$ . Otherwise, let  $K_i(E) = \emptyset^{S(E)}$  for the  $K_i(E)$  is the set of states in which player  $i$  believes  $E$ . By the properties of the  $t_i$ , if  $E$  is an event in  $\Sigma$ , then  $K_i(E)$  is an event. For any pair  $(m, w)$ , let  $K_{m,w}(E) = K_m(E) \cap K_w(E)$ . That is,  $K_{m,w}(E)$  is the event that  $E$  is mutual belief among  $m$  and  $w$ . Finally, let  $CK_{m,w}(E) = \bigcap_{n \geq 1} K_{m,w}^n(E)$ . This is the event that  $E$  is common belief among  $m$  and  $w$ . These are the usual (pair-wise) mutual belief and common belief operators specialized to our setting.

In a standard matching model with complete information and full awareness, a matching is stable if there is no blocking pair. In our setting, there might be a blocking pair without it being common belief among the pair that they form a blocking pair. Consequently, they may not agree to block. We say that in the matching  $\mu \in \mathbf{M}$ , at state  $\omega$  it is common belief among  $m$  and  $w$  that they form a blocking pair if  $\omega \in CK_{m,w}([m \succ_w \mu(w)] \cap [w \succ_m \mu(m)])$ .<sup>6</sup>

**Definition 2 (Stability)** *We say that  $\mu$  is stable at  $\omega$  if*

- (i) *there does not exist  $(m, w) \in M \times W$  such that  $\omega \in CK_{m,w}([m \succ_w \mu(w)] \cap [w \succ_m \mu(m)])$ , and*
- (ii) *there does not exist  $i \in M \cup W$  such that  $\omega \in [i \succ_i \mu(i)]$ .*

That is, we say that  $\mu$  is stable at  $\omega$  if there does not exist a pair  $(m, w)$  such that at  $\omega$  it is common belief among  $m$  and  $w$  that  $(m, w)$  forms a blocking pair. And there should also not exist a player who prefers to stay alone over her current match.

In a standard matching model with complete information and full awareness, stable matchings are in the core of the matching game. In our setting with asymmetric unawareness, stable matchings are in the coarse core. The coarse core has been introduced by Wilson (1978) for exchange economies with asymmetric information. It has been extended to general TU games with incomplete information and unawareness by Bryan et al. (2022). Similar ideas can be used to extend it to NTU games with incomplete information and unawareness like our matching games.

Since our stability notion generalizes absence of blocking pairs to absence of pairwise common belief in blocking, we illustrate this novel feature with a simple example.

**Example 3** There are two men and women each,  $M = \{m_1, m_2\}$  and  $W = \{w_1, w_2\}$ .

---

<sup>6</sup>In terms of terminology, we often just refer to “pair-wise common belief” if the pair of players is clear from the context.

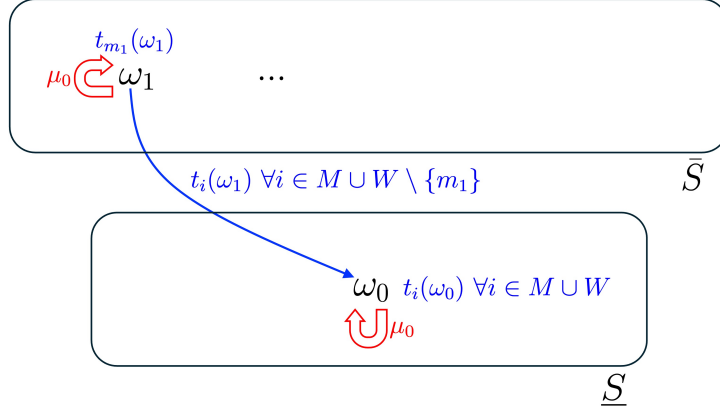


Figure 4: Unawareness Structure of Example 3

The preference mappings are given by the following rank order lists:

$$\begin{array}{ll}
 \succ_{m_1}: & w_1, w_2 & \succ_{m_1}: & w_2, w_1 \\
 \succ_{m_2}: & w_1, w_2 & \succ_{m_2}: & w_1, w_2 \\
 \succ_{w_1}: & m_1, m_2 & \succ_{w_1}: & m_1, m_2 \\
 \succ_{w_2}: & m_1, m_2 & \succ_{w_2}: & m_1, m_2 \\
 & \omega_0 & & \omega_1
 \end{array}$$

All players have preferences constant in the states except for man  $m_1$ . The unawareness structure is depicted in Figure 4. There are two spaces, the richer space  $\bar{S}$  and the poorer space  $\underline{S}$ . We just focus on the two states relevant to our example, which is  $\omega_1$  in  $\bar{S}$  and  $\omega_0$  in  $\underline{S}$ . (We omit the projections as  $\omega_1$  projects to  $\omega_0$ .) The type mappings for the point beliefs are depicted in blue.<sup>7</sup> Importantly, at state  $\omega_1$ , men  $m_1$ 's point belief is at  $\omega_1$  while all others have point-belief  $\omega_0$ . Thus, all except man  $m_1$  are unaware.

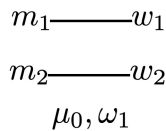


Figure 5: Stable matching at  $\omega_1$

Consider now the matching  $\mu_0$  given in Figure 5. In this matching, man  $m_1$  is matched to  $w_1$  while at  $\omega_1$  he strictly prefers  $w_2$  over  $w_1$ . In fact, they form a blocking pair in the standard sense. Nevertheless, this matching is stable in our sense at  $\omega_1$  because at  $\omega_1$  it is *not* a common belief among  $m_1$  and  $w_2$  that they form a blocking pair. This is because woman  $w_2$  is unaware and believes that man  $m_1$  strictly prefers  $w_1$  over herself. This example illustrates the difference between absence of blocking and pair-wise common belief in absence of blocking and thus the difference between the standard notion of stability and our notion of stability.  $\square$

<sup>7</sup>From now on we suppress circles to avoid clutter.



Stability itself is not sufficient for a satisfactory solution because  $\mu$  being stable at  $\omega$  does not rule out that some players discover something in the matching  $\mu$  at  $\omega$  that changes their preferences and consequently destabilizes the previously stable matching. In other words, some players may discover something in a stable matching that makes them change their preferences so that they want to get divorced. To obtain a satisfactory solution, we also need that beliefs are stable given the matching.

**Definition 3** We say that a state  $\omega \in \Omega$  is absorbing given  $\mu$  if  $\tau(\omega, \mu) = \omega$ .

Putting these two ideas together yields our solution concept:

**Definition 4 (Self-confirming outcome)** We say that an outcome  $(\omega, \mu)$  is self-confirming if

- (i)  $\mu$  is stable at  $\omega$ , and
- (ii)  $\omega$  is absorbing given  $\mu$ .

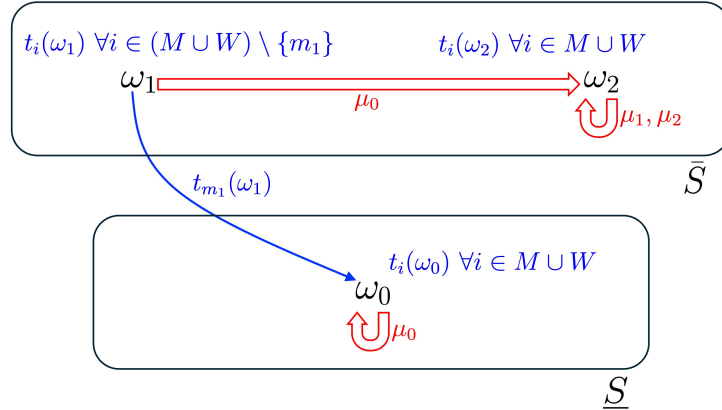


Figure 6: Unawareness Structure of Example 4

Since the state to be absorbing given the matching is a novel feature of the solution concept, we illustrate it with the following simple example.

**Example 4** There are two men and women each,  $M = \{m_1, m_2\}$  and  $W = \{w_1, w_2\}$ . The preference mappings are given by the following rank order lists:

$$\begin{array}{ll}
 \succ_{m_1}: & w_1, w_2 & \succ_{m_1}: & w_2, w_1 \\
 \succ_{m_2}: & w_1, w_2 & \succ_{m_2}: & w_1, w_2 \\
 \succ_{w_1}: & m_1, m_2 & \succ_{w_1}: & m_1, m_2 \\
 \succ_{w_2}: & m_1, m_2 & \succ_{w_2}: & m_1, m_2 \\
 & \omega_0 & & \omega_1, \omega_2
 \end{array}$$

All players have preferences constant in the states except for man  $m_1$ . The unawareness structure is depicted in Figure 6. There are two spaces, the richer space  $\bar{S}$  and the poorer

space  $\underline{S}$ . We focus on two states in the richer spaces,  $\omega_1$  and  $\omega_2$ , and one state in the poor space,  $\omega_0$ . (We omit the projections as they are trivial. Both states  $\omega_1$  and  $\omega_2$  project to  $\omega_0$ .) The type mappings for the point beliefs are depicted in blue. Importantly, at state  $\omega_1$ , men  $m_1$ 's point belief is at  $\omega_0$ . That is, he is unaware of something that obviously effects his preferences because his preference in states in  $\bar{S}$  differs from his preference at  $\underline{S}$ .

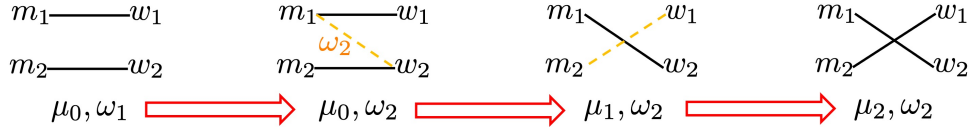


Figure 7: Process of Example 4

The relevant state transitions are depicted by red arrows in Figure 6. We print the matching(s) that facilitate the state transitions beside the red arrows. To make the process transparent, we depict it the separate Figure 7. Let  $\omega_1$  be the initial state. At that state all players except  $m_1$  believe in  $\omega_1$ . Man  $m_1$  believes in state  $\omega_0$ . Let matching  $\mu_0$  be the initial matching. It is easy to verify that it is stable given  $\omega_1$  and also stable given  $\omega_0$  as there is no pair who has common belief in blocking. Yet, as argued above, stability is not sufficient for a satisfactory solution concept. Given state  $\omega_1$  and matching  $\mu_0$ , man  $m_1$  becomes aware as the process transitions to state  $\omega_2$  at which  $m_1$ 's point-belief is correctly  $\omega_2$ . (Note that initially at  $\omega_0$  man  $m_1$  did not anticipate this transition.) Consequently,  $m_1$ 's preference changes. At  $\omega_2$ ,  $m_1$  and  $w_2$  form now a blocking pair, this is common belief (not just among  $m_1$  and  $w_2$  but among all players in this case), and thus  $\mu_0$  is not stable anymore. When satisfying the blocking pair, both  $m_1$  and  $w_2$  have to get divorced from their current partners and we reach matching  $\mu_1$  (at state  $\omega_2$ ). This matching is not stable since both  $m_2$  and  $w_1$  prefer to be matched to each other rather than staying single. Thus they form a blocking pair and this is common belief. Satisfying this blocking pair, we reach matching  $\mu_2$  (at  $\omega_2$ ). This matching is stable given  $\omega_2$  as their no blocking pairs. Moreover, state  $\omega_2$  is absorbing given  $\mu_2$ , i.e., awareness or preferences do not change given  $\mu_2$ . This example shows that stability of matching at  $\omega_1$  is not enough for a solution concept. Our solution concept also requires the state to be absorbing like at outcome  $(\mu_2, \omega_2)$ .  $\square$

**Proposition 3** *Every finite dynamic two-sided matching game with unawareness has a self-confirming outcome.*

PROOF. (i): Call the two-sided matching game with unawareness at a *given* state the stage game. Any stage game of the two-sided matching game with unawareness involves the same finite number of finite state spaces. Any finite state machine on a finite space must have absorbing sets. What is left to show is that there is an absorbing set that is a singleton: Consider  $\omega \in \check{S}$  such that  $S_{t_i(\omega)} = \check{S}$  for all  $i \in M \cup W$ . Such a state exists in  $\check{S}$  by Assumption 2. We have  $\tau_i(\omega, \mu) = \omega$  for all  $\mu$ . To see this, suppose to the contrary that  $\tau_i(\omega, \mu) = \omega'$  for some  $\omega' \neq \omega$  and  $\mu \in M$ . By Assumption 1, there exists  $i \in M \cup W$

such that  $S_{t_i(\omega')} \neq S_{t_i(\omega)}$ . Since  $S_{t_i(\omega)} = \check{S}$ , we must have  $S_{t_i(\omega')} \triangleleft \check{S}$ . But this contradicts that assumption that  $S_{t_i(\tau(\omega, \mu))} \supseteq S_{t_i(\omega)}$ . We conclude that  $\omega$  is absorbing.

(ii): Every state  $\omega \in \Omega$  pins down a strict preference profile of all agents. Consider the absorbing state from part (i). By Gale and Shapley (1962), a stable matching in the standard sense exist, i.e. there is no blocking pair. Consequently, there can also be no common belief in blocking. Hence, it is stable.

By (i) we argued that there exists an absorbing state  $\omega$ . By (ii), there is a stable match at  $\omega$ . Hence, there exists a self-confirming outcome.  $\square$

The proof is straightforward: Awareness can only go up. Once all players are aware of everything, there must exist an absorbing state. At that state there must exist a stable match in the standard sense of Gale and Shapley (1962). Absence of blocking pairs means also no common belief in blocking. Hence, it is also stable in our sense.

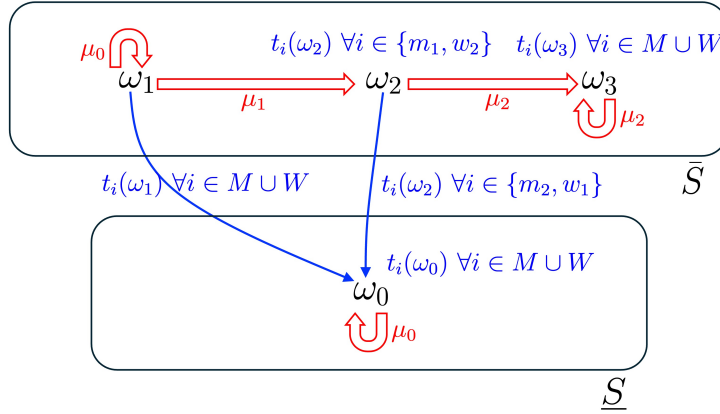


Figure 8: Trapped in Unawareness by Marriage

While for showing existence it is enough to argue with the upmost space, there can be absorbing states and self-confirming outcomes that involve unawareness. In the following simple example we illustrate how players are trapped by marriage in their unawareness.

**Example 5** There are two men and women each,  $M = \{m_1, m_2\}$  and  $W = \{w_1, w_2\}$ . The preference mappings are given by the following rank order lists:

$$\begin{array}{ll}
 \succ_{m_1}: & w_1, w_2 \\
 \succ_{m_2}: & w_1, w_2 \\
 \succ_{w_1}: & m_1, m_2 \\
 \succ_{w_2}: & m_1, m_2 \\
 & \omega_0
 \end{array}
 \qquad
 \begin{array}{ll}
 \succ_{m_1}: & w_2, w_1 \\
 \succ_{m_2}: & w_1, w_2 \\
 \succ_{w_1}: & m_2, m_1 \\
 \succ_{w_2}: & m_1, m_2 \\
 & \omega_1, \omega_2, \omega_3
 \end{array}$$

The unawareness structure is given by Figure 8. At  $\omega_1$ , all players' point-belief is  $\omega_0$  in  $S$ . That is, all players are unaware. Let  $\mu_0$  as in the example earlier, i.e., the matching given in Figure 5. This matching is stable given  $\omega_1$  because at that state all players are unaware and their point-belief is  $\omega_0$ . Thus, there is no pair who has common belief in blocking.

Since the  $\omega_1$  is absorbing given  $\mu_0$ , the outcome  $(\mu_0, \omega_1)$  is a self-confirming outcome despite all players being unaware. We also observe in Figure 8 that there would be other matchings like  $\mu_1$  that would allow male  $m_1$  and  $w_2$  to become aware (and subsequently matching  $\mu_2$  that allow the remaining players to become aware). However, since  $(\mu_0, \omega_1)$  is absorbing, these states are not reached. Players are trapped in unawareness in the marriages. We will return to this example later in Subsection 8.1 when discussing infidelity.  $\square$

More interesting than showing existence of self-confirming outcomes is the process of finding those self-confirming outcomes with transformative experiences and discoveries in matchings, leading to changes of preferences, further attempts of trying to find stable matches, further preference-perturbing discoveries etc. We are interested in a natural decentralized matching process that leads to self-confirming outcomes. From Section 2 we know that even without unawareness, it is not straightforward for decentralized matching processes to reach a stable matching. With asymmetric unawareness, changes of awareness may also lead to changes of preferences which complicates the process is even further.

For any  $\omega \in \Omega$ ,  $\mu \in \mathbf{M}$  and  $w \in W$ , let

$$M_w(\omega, \mu) := \{m \in M : \omega \in CK_{m,w}([m \succ_w \mu(w)] \cap [w \succ_m \mu(m)])\} \cup \{w : w \succ_w(t_w(\omega)) \mu(w)\}.$$

The first term,  $\{m \in M : \omega \in CK_{m,w}([m \succ_w \mu(w)] \cap [w \succ_m \mu(m)])\}$ , is the set of men  $m$  for which it is common belief among  $w$  and  $m$  at  $\omega$  and matching  $\mu$  that  $m$  and  $w$  form a blocking pair. The second term,  $\{w : w \succ_w(t_w(\omega)) \mu(w)\}$  is nonempty only if woman  $w$  prefers to stay alone at  $\omega$  rather than with her partner in the matching  $\mu$ .

We say that at  $\omega$  and  $\mu$ ,  $w$  and  $i$  are a  $(w, i)$ -commonly believed  $w$ -best blocking pair if  $i \in M_w(\omega, \mu)$  and  $i \succ_w(t_w(\omega)) j$  for any  $j \in M_w(\omega, \mu)$  with  $j \neq i$ . Analogously, define  $W_m(\omega, \mu)$  and commonly believed  $m$ -best blocking pairs. We say that at  $\omega$  and  $\mu$ ,  $m$  and  $w$  are a  $(i, j)$ -commonly believed *mutual best blocking* pair if it is both a  $(i, j)$ -commonly believed  $i$ -best blocking pair and  $(i, j)$ -commonly believed  $j$ -best blocking pair. Of course, our earlier assumption ensures that either  $i$  and  $j$  are members of different sides of the market or  $i = j$ . Denote by  $B(\omega, \mu) \subseteq M \times W$  the commonly believed best blocking pairs at  $\omega$  and  $\mu$ . Denote by  $MB(\omega, \mu) \subseteq M \times W$  the commonly believed mutual best blocking pairs at  $\omega$  and  $\mu$ .

**Lemma 1** *For any  $\omega$  and  $\mu$ , if  $\mu$  is not stable at  $\omega$ , then  $B(\omega, \mu) \neq \emptyset$ .*

PROOF. For any  $\omega \in \Omega$ , if  $\mu$  is not stable, then there exists  $i \in W$  such that  $M_i(\omega, \mu) \neq \emptyset$  or  $j \in M$  such that  $W_j(\omega, \mu) \neq \emptyset$ . Note also that since  $M_i(\omega, \mu)$  and  $W_j(\omega, \mu)$  are finite, the commonly believed  $i$ -best blocking pair or the commonly believed  $j$ -best blocking pair exist. Therefore,  $B(\omega, \mu) \neq \emptyset$ .  $\square$

In contrast to  $B(\omega, \mu)$ , the set  $MB(\omega, \mu)$  may be empty even if  $\mu$  is not stable. Therefore, define

$$\widehat{MB}(\omega, \mu) := \begin{cases} MB(\omega, \mu) & \text{if } MB(\omega, \mu) \neq \emptyset \\ B(\omega, \mu) & \text{otherwise.} \end{cases}$$

Fix small  $\varepsilon \in (0, 1)$  and define a matching process by transition probabilities  $P^\varepsilon$  such that for any  $\omega$  and  $\mu$ ,

$$P^\varepsilon(\mu', \omega' \mid \mu, \omega) := \begin{cases} \frac{1-\varepsilon}{|\widehat{MB}(\omega, \mu)|} & \text{if } \mu' \text{ differs from } \mu \text{ by satisfaction of exactly one} \\ & \text{pair in } \widehat{MB}(\omega, \mu) \text{ and } \tau(\omega, \mu') = \omega'; \\ \frac{\varepsilon}{|B(\omega, \mu)|} & \text{if } \mu' \text{ differs from } \mu \text{ by satisfaction of exactly one} \\ & \text{pair in } B(\omega, \mu) \text{ and } \tau(\omega, \mu') = \omega'; \\ 1 & \text{if } \mu' = \mu \text{ and } \mu \text{ is stable given } \omega \text{ and } \tau(\omega, \mu') = \omega'; \\ 0 & \text{else.} \end{cases}$$

We define the matching process by transition probabilities on the product space  $\Omega \times \mathbf{M}$ . This warrants some explanations. Naturally, the process would proceed as follows: Given a current state  $\omega$  and awareness/preferences at  $\omega$ , the current match  $\mu$  may feature some blocking pairs w.r.t. preferences at  $\omega$ . Satisfying some blocking pair (i.e., preferably some mutual optimal blocking pair) would lead to another match  $\mu'$ . At this match  $\mu'$  and state  $\omega$ , there may be discoveries leading to another state  $\tau(\omega, \mu') = \omega'$ . At this state and corresponding awareness/preferences, there might be blocking pairs. Satisfying a blocking pair may lead to yet another match  $\mu'' \dots$  While in this explanation, we let states transit *after* the transition of the matches, the above formal description of the process involves a simultaneous transition of states and matches. It can be thought of just describing the process at every “even” period. This is well-defined because states move deterministically according to a finite state machine given the prior state and the new match. The process on  $\Omega \times \mathbf{M}$  is not deterministic though because blocking pairs are selected randomly with  $(1 - \varepsilon)$ -priority given to mutual best blocking pairs if they exist.

**Proposition 4** *For any (initial) outcome  $(\mu, \omega)$  and  $\varepsilon \in (0, 1)$ , the matching process  $P^\varepsilon$  converges almost surely to a self-confirming outcome.*

PROOF. First, we argue that each absorbing set of  $P^\varepsilon$  cannot involve more than one state. To see this, consider an absorbing set of  $P^\varepsilon$  that involves two states. Denote them by  $\omega^*$  and  $\omega^{**}$ . By the definition of absorbing set, each outcome (consisting both of a state and a matching) is reachable from any other outcomes in the absorbing set via a finite number of transitions. Thus,  $\omega^*$  must be reachable from  $\omega^{**}$  and vice versa. Since for each agent  $i$ , awareness can only increase along the process, we must have  $S_{t_i(\omega^*)} = S_{t_i(\omega^{**})}$  for all  $i$ . The assumption of no redundancies, Assumption 1, implies now that  $\omega^* = \omega^{**}$ .

Second, we argue that each absorbing set of  $P^\varepsilon$  cannot involve more than one matching and that matching must be stable. Suppose to the contrary that there exists an absorbing set that involves an unstable matching denoted by  $\mu_1$ .

**Claim:** Consider w.l.o.g.  $\mu_1$  in the absorbing set for  $\omega^*$ . We claim that there exist a sequence of matchings  $\mu_1, \dots, \mu_k$ , with  $k \leq 2|W| \cdot |M|$ , such that  $\mu_k$  is stable, and for each  $i = 1, \dots, k - 1$ , there exists a pairwise commonly believed best blocking pair  $(m_i, w_i)$  such that  $\mu_{i+1}$  is obtained from  $\mu_i$  by satisfying  $(m_i, w_i)$ .

To prove the claim note first that the preference profile is fixed by the absorbing state  $\omega^*$ . For any women  $w_i$ , since the preference of women  $w_i$  is an ordered list over  $M \cup w_i$  of length  $|M| + 1$ , we consider the order from the worst to the best, with the worst ranking 0 and the best has rank  $|M|$ . Define the payoff of agent  $w_i$  in matching  $\mu$  by  $p_{w_i}(\mu) = k \in \{0, \dots, |M|\}$  if  $\mu(w_i)$  is at the  $k$ th place of the order. That is, if  $w_i$  is matched to her most preferred agent in  $M \cup w_i$ , then here payoff is  $|M|$ .

Let  $X(\mu)$  be the set of matched women at  $\mu$ . Define a “potential” function  $\Phi(\mu) = \sum_{w \in X(\mu)} (|M| - p_w(\mu))$ . Note that this potential function is minimized when all women are matched to their “best” agent on their lists, respectively. Function  $\Phi$  is bounded above by  $|W| \cdot |M|$  as  $|X(\mu)| \leq |W|$  and  $p_w(\mu) \geq 0$  for  $w \in X(\mu)$ .

To construct a sequence of matchings  $\mu_1, \dots, \mu_k$ , we divide it into two phases. The first phase features the sequence  $\mu_1, \dots, \mu'$  and the second  $\mu', \dots, \mu_k$ . In the first phase, we move from  $\mu_i$  to  $\mu_{i+1}$  by satisfying pairwise commonly believed best blocking pairs that involve matched women only. With each satisfied pairwise commonly believed best blocking pair,  $\Phi$  decreases by at least 1 because: (1) Starting from  $\mu_i$ , if this woman gets rematched to an unmatched man, the set  $X(\mu_{i+1}) = X(\mu_i)$ , and her payoff increases. (2) If she gets rematched to a matched man, then  $X(\mu_{i+1}) \subsetneq X(\mu_i)$  (i.e., her new matching partner leaves his current partner), and her payoff increases. (3) If she becomes unmatched, her term in the potential function is dropped, and that term must have been positive before. Hence, at most after  $|W| \cdot |M|$  steps, no matched woman can improve her payoff, which implies that no matched woman has a commonly believed blocking pair. Furthermore, if there is no pairwise commonly believed blocking pair, there is no pairwise commonly believed best blocking pair. The first phase terminates with a matching  $\mu'$  in which no matched woman has a pairwise commonly believed (best) blocking pair. Observe that the process  $P^\varepsilon$  allows for above sequence of pairwise commonly believed best blocking pairs to be satisfied because at any step the process puts strict positive probability on any pairwise commonly believed best blocking pair at that step. Moreover, we assumed that  $(\mu_1, \omega^*)$  is in the absorbing set of  $P^\varepsilon$ . Since we reached  $(\mu', \omega^*)$  with  $P^\varepsilon$ , it implies that  $(\mu', \omega^*)$  is in the absorbing set of  $P^\varepsilon$  as well.

In the second phase, suppose we start from the matching  $\mu'$ . If there is no pairwise commonly believed blocking pair in  $\mu'$  (also among unmatched women), then let  $\mu' = \mu_k$  and the second phase terminates. Otherwise, the second phase continues as follows. Since no matched woman has a pairwise commonly believed blocking pair in  $\mu'$ ,  $\{(m, w) : \mu'(w) \neq w, \omega^* \in CK_{m,w}([m \succ_w \mu'(w)] \cap [w \succ_m \mu'(m)])\} = \emptyset$ . That is, the set of pairwise commonly believed blocking pairs that involves a matched woman is empty.

Satisfy a pairwise commonly believed best blocking pair of an unmatched woman,  $(m^*, w^*)$ , and call the resulting matching  $\mu''$ . We argue that in  $\mu''$ , no matched woman can have a pairwise commonly believed blocking pair, i.e.  $\{(m, w) : \mu''(w) \neq w, \omega^* \in CK_{m,w}([m \succ_w \mu''(w)] \cap [w \succ_m \mu''(m)])\} = \emptyset$ . Since the set of matched woman only change by adding  $w^*$  and removing  $\mu'(m^*)$  if  $m^*$  is matched under  $\mu'$ , the set of matched women is now  $\{w : \mu''(w) \neq w\} = \{w^*\} \cup \{w : \mu'(w) \neq w\} \setminus \{\mu'(m^*) : \mu'(m^*) \neq m^*\}$ . Therefore, the set of pairwise commonly believed blocking pairs that involves a matched

woman is now

$$\begin{aligned} & \{(m, w) : \mu''(w) \neq w, \omega^* \in CK_{m,w}([m \succ_w \mu''(w)] \cap [w \succ_m \mu''(m)])\} = \\ & \{(m, w) : w = w^*, \omega^* \in CK_{m,w}([m \succ_w \mu''(w)] \cap [w \succ_m \mu''(m)])\} \cup \\ & \{(m, w) : \mu'(w) \neq w, \omega^* \in CK_{m,w}([m \succ_w \mu''(w)] \cap [w \succ_m \mu''(m)])\} \setminus \\ & \{(m, w) : w = \mu'(m^*) \neq m^*, \omega^* \in CK_{m,w}([m \succ_w \mu''(w)] \cap [w \succ_m \mu''(m)])\} \end{aligned}$$

We show that the r.h.s. is empty. Consider first the set

$$\{(m, w) : w = w^*, \omega^* \in CK_{m,w}([m \succ_w \mu''(w)] \cap [w \succ_m \mu''(m)])\}.$$

For  $w^*$ , since  $(m^*, w^*)$  is her pairwise commonly believed best blocking pair given  $\mu'$ , she cannot have any pairwise commonly believed blocking pair after satisfying  $(m^*, w^*)$ , i.e.,  $\{(m, w) : w = w^*, \omega^* \in CK_{m,w}([m \succ_w \mu''(w)] \cap [w \succ_m \mu''(m)])\} = \emptyset$ .

Next consider the set

$$\begin{aligned} & \{(m, w) : \mu'(w) \neq w, \omega^* \in CK_{m,w}([m \succ_w \mu''(w)] \cap [w \succ_m \mu''(m)])\} \setminus \\ & \{(m, w) : w = \mu'(m^*) \neq m^*, \omega^* \in CK_{m,w}([m \succ_w \mu''(w)] \cap [w \succ_m \mu''(m)])\}. \end{aligned}$$

This set concerns pairs that involve matched women under  $\mu'$  except the women that was possibly matched under  $\mu'$  to the man that got rematched under  $\mu''$ .

Recall that for all  $m \neq m^*$  we have  $\mu''(m) = \mu'(m)$ . Moreover,  $\mu''(m) \succ_{m^*}(t_{m^*}(\omega^*)) \mu'(m)$ . Thus, every man believes that he gets a weakly better match under  $\mu''$ . Thus, we have that for all  $m \in M$ ,  $\{w : w \succ_m(t_{m^*}(\omega^*)) \mu''(m)\} \subseteq \{w : w \succ_m(t_w(\omega^*)) \mu'(m)\}$  (with strict “ $\subset$ ” for  $m^*$  and “ $=$ ” for all  $m \neq m^*$ ). Similarly, for all  $w$  such that  $w \neq w^*$  and  $w \neq \mu'(m^*)$  (if  $\mu'(m^*) \neq m^*$ ), we have  $\mu''(w) = \mu'(w)$ . Thus, every woman who is matched in  $\mu''$  believes that she gets a weakly better match under  $\mu''$  than under  $\mu'$ . Therefore, for all  $w$  with  $\mu'(w) \neq w$  and  $w \neq \mu'(m^*)$ ,  $\{m : m \succ_w(t_w(\omega^*)) \mu''(w)\} \subseteq \{m : m \succ_w(t_w(\omega^*)) \mu'(w)\}$ .

Therefore, if there exists a pair  $(mw)$  with  $w \neq \mu'(w)$  and  $w \neq \mu'(m^*)$  such that  $\omega^* \in CK_{m,w}([m \succ_w \mu''(w)] \cap [w \succ_m \mu''(m)])$ , then by the arguments above we must have  $\omega^* \in CK_{m,w}([m \succ_w \mu'(w)] \cap [w \succ_m \mu'(m)])$ . However, this contradicts our earlier conclusion from the first phase that  $\{(m, w) : \mu'(w) \neq w, \omega^* \in CK_{m,w}([m \succ_w \mu'(w)] \cap [w \succ_m \mu'(m)])\} = \emptyset$ . We conclude that  $\{(m, w) : \mu''(w) \neq w, \omega^* \in CK_{m,w}([m \succ_w \mu''(w)] \cap [w \succ_m \mu''(m)])\} = \emptyset$ .

The analogous arguments apply inductively to next steps of the second phase. Since only unmatched woman are able to block in the second phase, men are never left and can only improve, which can only happen at most  $|W| \cdot |M|$  times. When the second phase terminates, there is no pairwise commonly believed blocking pair, and the matching is stable given  $\omega^*$ . Observe that the process  $P^\varepsilon$  allows for second-phase sequence of pairwise commonly believed best blocking pairs to be satisfied because at any step the process puts strict positive probability on any pairwise commonly believed best blocking pair at that step. Moreover, we have already shown that  $(\mu', \omega^*)$  is in the absorbing set of  $P^\varepsilon$ . Since we reached  $(\mu_k, \omega^*)$  with  $P^\varepsilon$ , it implies that  $(\mu_k, \omega^*)$  is in the absorbing set of  $P^\varepsilon$  as well. This completes the proof of the claim.

The claim implies that each absorbing set of  $P^\varepsilon$  must be a singleton  $(\mu, \omega)$  consisting of an absorbing state  $\omega$  given  $\mu$  and a stable  $\mu$  given  $\omega$ . To see this, note that each absorbing set cannot have more than one matching. Otherwise, each such matchings must be reachable by the process from each other with a finite number of transitions. This would imply that all these matchings of the absorbing set are unstable. However, the claim shows that there is a sequence of matchings allowed by the process leading to a stable matching, a contradiction.

**Claim:** An outcome  $(\mu, \omega)$  is self-confirming if and only if it is an absorbing outcome of  $P^\varepsilon$ .

To prove the claim, consider first “ $\Leftarrow$ ” direction: Let  $(\mu, \omega)$  be an absorbing outcome of  $P^\varepsilon$ . Then  $P^\varepsilon(\mu, \omega \mid \mu, \omega) = 1$ . By the construction of  $P^\varepsilon$ ,  $\mu$  is stable given  $\omega$  and  $\omega$  is absorbing given  $\mu$ . Thus,  $(\mu, \omega)$  is self-confirming.

“ $\Rightarrow$ ”: Let  $(\mu, \omega)$  be self-confirming. Then  $\mu$  is stable at  $\omega$ , and  $\omega$  is absorbing given  $\mu$ . Hence,  $P^\varepsilon(\mu, \omega \mid \mu, \omega) = 1$ , i.e.  $(\mu, \omega)$  be an absorbing outcome of  $P^\varepsilon$ . This completes the proof of the claim.

To complete the proof of Proposition 4, it is enough to note that  $P^\varepsilon$  must converge almost surely to an absorbing outcome.  $\square$

The proof shows first that any absorbing set must exactly involve a single state because awareness must be constant within each absorbing set. Second, it shows that any absorbing set can at most involve one matching and this matching must be stable given the state. This part of the proof slightly extends an argument by Ackermann et al. (2008) to a sequence of satisfying pairwise commonly believed best blocking pairs.

## 5 Allowing for Flirting

The prior solution concept features very conservative blocking behavior. Only if there is common belief among a pair that they want to block, there is a chance that the pair is selected for blocking. There are situations in which for instance man  $m$  believes that although there is currently absence of common belief in blocking with women  $w$ , if he were to talk to her and raise her awareness, it would result in common belief in blocking. That is, communication involving raising awareness is natural in this setting. Such a communication could be a feature of flirting behavior. Flirting can destabilize outcomes in two ways: First, raising awareness may change preferences and thus allow for blocking pairs. This is illustrated with the help of the following example.

**Example 6** This is a variation of prior Example 4. There are two men and women each,  $M = \{m_1, m_2\}$  and  $W = \{w_1, w_2\}$ . The preference mappings are as before given



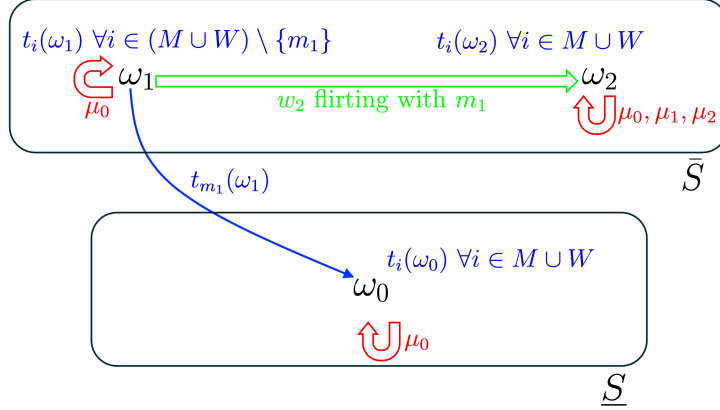


Figure 9: Unawareness Structure of Example 6

by the following rank order lists:

$$\begin{array}{ll}
 \succ_{m_1}: w_1, w_2 & \succ_{m_1}: w_2, w_1 \\
 \succ_{m_2}: w_1, w_2 & \succ_{m_2}: w_1, w_2 \\
 \succ_{w_1}: m_1, m_2 & \succ_{w_1}: m_1, m_2 \\
 \succ_{w_2}: m_1, m_2 & \succ_{w_2}: m_1, m_2 \\
 \omega_0 & \omega_1, \omega_2
 \end{array}$$

The unawareness structure is depicted in Figure 9. The type mappings are as in Example 4. What differs are the state transitions. Recall that  $\mu_0$  is stable given  $\omega_1$  in Example 4. Now we have also that  $\omega_1$  is absorbing given  $\mu_0$  as shown by the red arrow from  $\omega_1$  to itself. Thus,  $(\mu_0, \omega_1)$  is a self-confirming outcome. However, it can be destabilized by flirting. Observe that women  $w_2$  has an incentive to raise man  $m_1$  awareness and thus changing his preference in her favor. This is indicated in Figure 9 by the green transition arrow. This leads to  $\omega_2$  upon which  $m_1$  and  $w_2$  block, yielding the matching  $\mu_1$ . This is followed by blocking from  $m_2$  and  $w_1$ , yielding matching  $\mu_2$ , which is stable given  $\omega_2$ . This process is depicted in Figure 10. Since  $\omega_2$  is absorbing, we have reached a new self-confirming outcome. However,  $(\mu_2, \omega_2)$  can not be destabilized by further flirting. It is a “flirt-proof” self-confirming outcome while  $(\mu_0, \omega_1)$  is just a self-confirming outcome (that, as we have shown, is not flirt-proof).  $\square$

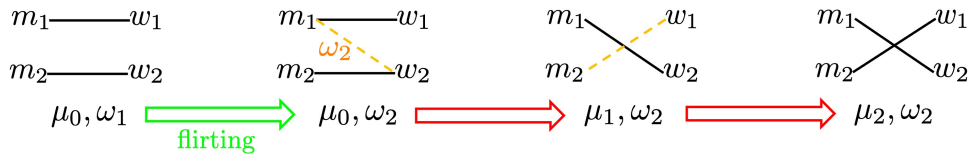


Figure 10: Process of Example 6

The example shows how flirting can lead to change of preferences through raising awareness and subsequently a blocking pair and common belief in blocking by this pair of players. A more subtle effect of flirting pertains just to the last feature. There are

situation in which there is already a blocking pair but no common belief in blocking by this pair (e.g., Example 3). In such a case, flirting can help creating this pairwise common belief in blocking without any change of preferences. This is illustrated in the next example.

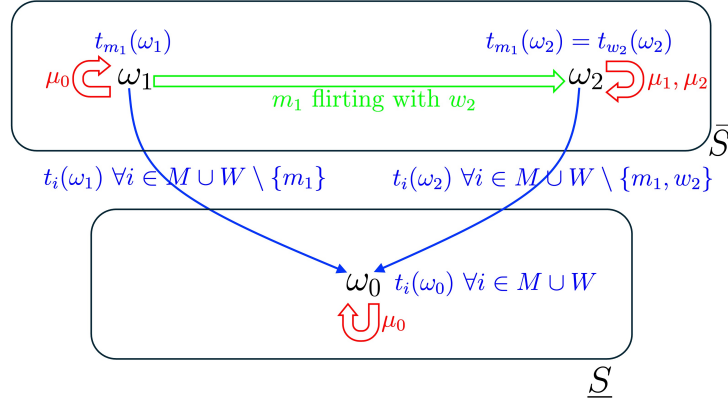


Figure 11: Unawareness Structure of Example 7

**Example 7** This example can be understood as an extension of Example 3. There are two men and women each,  $M = \{m_1, m_2\}$  and  $W = \{w_1, w_2\}$ . The preference mappings are as before given by rank order lists of the prior Example 6. The unawareness structure is depicted in Figure 11. Note that different from the prior example, at  $\omega_1$  all players are unaware except man  $m_1$ . This is like in Example 3. Recall that  $\mu_0$  is stable given  $\omega_1$  in Example 3 because there is absence of pair-wise common belief between  $m_1$  and  $w_2$  of blocking. However, at  $\omega_1$ , man  $m_1$  could communicate with  $w_2$  and raise her awareness, as indicated by the green state transition. At  $\omega_2$  both are now aware and now  $m_1$  and  $w_2$  do not just form a blocking pair but there is also common belief among them of blocking. The resulting match  $\mu_1$  leads to the final match  $\mu_2$  at  $\omega_2$ , which is a self-confirming outcome. This outcome  $(\mu_2, \omega_2)$  can not be destabilized by further flirting. It is a “flirt-proof” self-confirming outcome while  $(\mu_0, \omega_1)$  is just a self-confirming outcome (that, as we have shown, is not flirt-proof). The process can be depicted as in Figure 10 except that now  $m_1$  flirts with  $w_2$  rather than the other way around. Thus, in this example flirting does not change preferences of the player who is flirted with but creates pair-wise common belief of blocking (by making a player aware that *others* have different preferences).  $\square$

We are interested in outcomes that are stable w.r.t. flirting, or as we alluded to already in the examples, are “flirt-proof stable”. To define the refined stability notion, we need to model communication that raises awareness of potential blocking partners. This changes (point-)beliefs from one space to a richer space and thus consists of a transition to another state. We model this with an another finite state machine. In contrast to the transition function  $\tau$  defined earlier, the transition due to communication will be part of the solution concept rather than the primitives of the dynamic matching

game with unawareness. To this end, we require some notation. For any man  $m \in M$ , define

$$H_m(\omega, \mu, w) := \{\omega' \in S_{t_m(\omega)} : \omega' \in CK_{m,w}([m \succ_w \mu(w)] \cap [w \succ_m \mu(m)]), S_{t_w(\omega')} \supseteq S_{t_w(\omega)}\}$$

as the set of hypothetical states considered by  $m$  at state  $\omega$  in matching  $\mu$  such that if  $m$  would suitably raise  $w$ 's awareness, then there would be common belief in blocking among  $w$  and  $m$ . Likewise, for any  $w \in W$ , define  $H_w(\omega, \mu, m)$ .

Next, for any man  $m \in M$ , define

$$S_m(\omega, \mu, w) := \bigvee_{\omega' \in H_m(\omega, \mu, w)} S_{t_w(\omega')}$$

as the awareness that is raised by  $m$  to  $w$  at state  $\omega$  in matching  $\mu$ . Likewise, for any  $w \in W$ , define  $S_w(\omega, \mu, m)$ . Note that we assume that if there are alternative ways to raise awareness in order to obtain common belief in blocking, then the the highest awareness is communicated.

Now, define the *communication function*  $f : \Omega \times \mathbf{M} \rightarrow \Omega$  by for any  $\omega \in \Omega$ ,  $\mu \in \mathbf{M}$ , we have that  $f(\omega, \mu)$  with for any  $w \in W$ ,  $S_{t_w(f(\omega, \mu))} = S_{t_w(\omega)} \bigvee_{m \in M} S_m(\omega, \mu, w)$  and  $m \in M$ ,  $S_{t_m(f(\omega, \mu))} = S_{t_m(\omega)} \bigvee_{w \in W} S_w(\omega, \mu, m)$ . By Assumption 2, such a state  $f(\omega, \mu)$  exists.

After communication that potentially raises awareness of players and consequently change preferences, players may now want to communicate further. For any  $(\omega, \mu)$  define recursively,  $f^1(\omega, \mu) = f(\omega, \mu)$ , and for  $n > 1$ ,  $f^n(\omega, \mu) = f(f^{n-1}(\omega, \mu), \mu)$  for any  $\omega \in \Omega$  and  $\mu \in \mathbf{M}$ . That is,  $f^n(\omega, \mu)$  captures  $n$  rounds of communication starting from state  $\omega$  and matching  $\mu$ . Since the model is finite and awareness can never decrease via communication, we observe:

**Lemma 2** *For any  $\omega \in \Omega$  and  $\mu \in \mathbf{M}$ , there exist a unique absorbing state for  $f$  denoted by  $f^\infty(\omega, \mu)$ .*

In the absorbing state of the communication function, further communication does not change awareness of any player.

**Definition 5 (Flirt-proof stability)** *We say that matching  $\mu$  is flirt-proof stable at  $\omega$  if  $\omega = f^\infty(\omega, \mu)$  and it is stable at  $\omega$ .*

**Definition 6 (Flirt-proof self-confirming outcome)** *We say that an outcome  $(\omega, \mu)$  is flirt-proof self-confirming if*

- (i)  $\mu$  is flirt-proof stable at  $\omega$ , and
- (ii)  $\omega$  is absorbing given  $\mu$ .

**Proposition 5** *Every finite dynamic two-sided matching game has a flirt-proof self-confirming outcome.*

The proof is analogous to the proof of existence of a self-confirming outcome (Proposition 3) and thus omitted.

For a fixed small  $\varepsilon \in (0, 1)$ , define a matching and flirting process by transition probabilities  $Q^\varepsilon$  such that for any  $\omega$  and  $\mu$ ,

$$Q^\varepsilon(\mu', \omega' \mid \mu, \omega) := \begin{cases} \frac{1-\varepsilon}{|\widehat{MB}(f^\infty(\omega, \mu), \mu)|} & \text{if } \mu' \text{ differs from } \mu \text{ by satisfaction of exactly one} \\ & \text{pair in } \widehat{MB}(f^\infty(\omega, \mu), \mu) \text{ and } \tau(f^\infty(\omega, \mu), \mu') = \omega'; \\ \frac{\varepsilon}{|B(f^\infty(\omega, \mu), \mu)|} & \text{if } \mu' \text{ differs from } \mu \text{ by satisfaction of exactly one} \\ & \text{pair in } B(f^\infty(\omega, \mu), \mu) \text{ and } \tau(f^\infty(\omega, \mu), \mu') = \omega'; \\ 1 & \text{if } \mu' = \mu \text{ and } \mu \text{ is flirt-proof stable given } \omega \text{ and} \\ & \tau(f^\infty(\omega, \mu), \mu') = \omega'; \\ 0 & \text{else.} \end{cases}$$

Naturally, the process would proceed as follows: Given a current state  $\omega$  and awareness/preferences at  $\omega$ , the current matching  $\mu$ , some players may have incentives to raise awareness of others, and the state evolves through communication to  $f(\omega, \mu)$  and further until communication quiets down and reaches state  $f^\infty(\omega, \mu)$ . Now there may be some blocking pairs w.r.t. preferences at  $f^\infty(\omega, \mu)$ . Satisfying some blocking pair, with  $(1 - \varepsilon)$ -priority given to a mutually optimal blocking pair, would lead to another matching  $\mu'$ . At this matching  $\mu'$  and state  $f^\infty(\omega, \mu)$ , there may be discoveries leading to another state  $\tau(f^\infty(\omega, \mu), \mu') = \omega'$ . At this state and corresponding awareness/preferences, some players may have an urge to communicate and raise other players' awareness. This goes on till communication quiets down again and there might be now common belief of blocking pairs. Satisfying such a blocking pair, with  $(1 - \varepsilon)$ -priority given to a mutual optimal best blocking pair, may lead to yet another match  $\mu'' \dots$  We let states transit *after* the transition of the matching, modeling transition through transformative experiences in the matching. We also let states transit *before* the transition of the matching, modeling transition through communication. Above formal description of the process involves a simultaneous transition of states and matches. It can be thought of just describing the process at every “even” period. In a sense, the implicit “odd” period consists of long shatter of communication between players and the transition of the matching and the implicit “even” period consists of the transformative experiences.

**Lemma 3** *An outcome  $(\mu, \omega)$  is flirt-proof self-confirming if and only if it is an absorbing outcome of  $Q^\varepsilon$ .*

PROOF.  $\Leftarrow$ : Let  $(\mu, \omega)$  be an absorbing outcome of  $Q^\varepsilon$ . Then  $Q^\varepsilon(\mu, \omega \mid \mu, \omega) = 1$ . By the construction of  $Q^\varepsilon$ ,  $(\mu, \omega)$  is flirt-proof self-confirming.

$\Rightarrow$ : Let  $(\mu, \omega)$  be flirt-proof self-confirming. Then  $\mu$  is flirt-proof stable at  $\omega$ , and  $\omega$  is absorbing given  $\mu$ . Hence,  $Q^\varepsilon(\mu, \omega \mid \mu, \omega) = 1$ , i.e.  $(\mu, \omega)$  be an absorbing outcome of  $Q^\varepsilon$ .  $\square$

**Proposition 6** *For any (initial) outcome  $(\mu, \omega)$  and  $\varepsilon \in (0, 1)$ , the matching and flirting process  $Q^\varepsilon$  convergences to a flirt-proof self-confirming outcome almost surely.*

PROOF. First, we argue that each absorbing set of  $Q^\varepsilon$  cannot involve more than one state (denoted with  $\omega^*$ ). Suppose by contradiction that an absorbing set of  $Q^\varepsilon$  that involves two different states. Denote them by  $\omega^*$  and  $\omega^{**}$  with  $\omega^* \neq \omega^{**}$ . By the definition of absorbing set, each outcome (consisting both of a state and a matching) is reachable from any other outcomes in the absorbing set via a finite number of transitions. Thus,  $\omega^*$  must be reachable from  $\omega^{**}$  and vice versa. That is,  $\tau(\dots\tau(f^\infty(\omega^*, \cdot), \cdot)\dots, \cdot) = \omega^{**}$  and  $\tau(\dots\tau(f^\infty(\omega^{**}, \cdot), \cdot)\dots, \cdot) = \omega^*$ . Since for each agent  $i$ , awareness can only increase along the process, we must have  $S_{t_i(\omega^*)} = S_{t_i(\omega^{**})}$  for all  $i$ . The assumption of no redundancies, Assumption 1, implies now that  $\omega^* = \omega^{**}$ , a contradiction.

The rest of the proof is analogous to the proof of Proposition 4. □

Recall that  $(\mu, \omega)$  is self-confirming if (i)  $\mu$  is stable at  $\omega$  and (ii)  $\omega$  is absorbing given  $\mu$ , i.e.  $\tau(\omega) = \omega$ . Meanwhile  $(\mu, \omega)$  is flirt-proof self-confirming if (i) matching  $\mu$  is stable at  $\omega$ , (iii)  $\omega = f^\infty(\omega, \mu)$ , and (iii)  $\tau(\omega) = \omega$  absorbing given  $\mu$ . Since (i)-(iii) implies (i)-(ii), flirt-proof self-confirming implies self-confirming. Thus, we conclude:

**Proposition 7** *If an outcome  $(\mu, \omega)$  is flirt-proof self-confirming, then it is also a self-confirming outcome.*

Examples 6 and 7 show that the converse does not hold.

## 6 Does Divorce Improve Welfare?

Will divorcees, i.e. players that initiate the divorce upon experiencing a preference change, necessarily become better off? More precisely, suppose a player is enlightened during a matching, changes his/her preferences, and consequently divorces his/her current match. Will such a player become necessarily better off in the resulting rematching process where the player's welfare is evaluated using his/her new preferences? Similarly, will the divorced, i.e., the player who is divorced by the divorcee, necessarily become worse off? We show by example that this is not the case. In terms of welfare of the divorcee and the divorced anything goes even within the same matching game and the same initial condition.

**Example 8** Suppose there are five men  $M = \{m_1, m_2, m_3, m_4, m_5\}$  and five women  $W = \{w_1, w_2, w_3, w_4, w_5\}$ . Consider the following profile of strict preferences given by

the rank order lists:

- $\succ_{m_1}: w_1, w_2, w_4, w_3, w_5$
- $\succ_{m_2}: w_2, w_3, w_4, w_1, w_5$
- $\succ_{m_3}: w_1, w_3, w_2, w_4, w_5$
- $\succ_{m_4}: w_4, w_5, w_3, w_2, w_1$
- $\succ_{m_5}: w_5, w_1, w_3, w_2, w_4$
- $\succ_{w_1}: m_5, m_1, m_3, m_4, m_2$
- $\succ_{w_2}: m_3, m_1, m_2, m_5, m_4$
- $\succ_{w_3}: m_2, m_3, m_1, m_4, m_5$
- $\succ_{w_4}: m_1, m_2, m_4, m_3, m_5$
- $\succ_{w_5}: m_4, m_5, m_1, m_2, m_3$

and initial matching  $\mu_1 = \begin{pmatrix} m_1 & m_2 & m_3 & m_4 & m_5 \\ w_1 & w_2 & w_3 & w_4 & w_5 \end{pmatrix}$ , which is stable given the above preference profile.

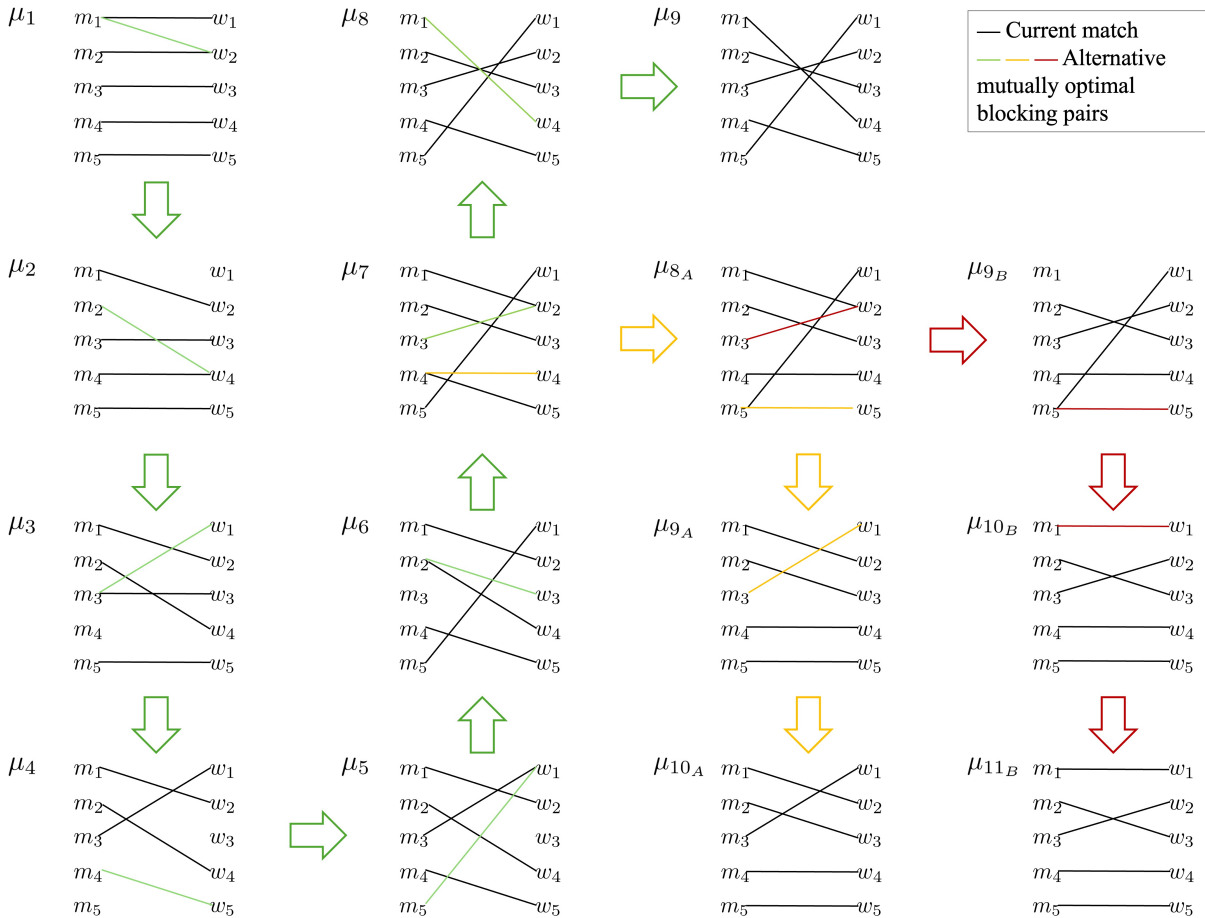


Figure 12: Processes in Example 8

Suppose  $m_1$  is enlightened and changes his preference such that his new rank order list is  $w_2, w_1, w_4, w_3, w_5$ . For simplicity, suppose that this is the only change in the players'

preferences along the entire process and that all players have correct belief about every other agent's preference. With  $m_1$ 's new preference, he blocks with  $w_2$ . We show that if  $m_1$  rematches to  $w_2$ , the process of satisfying mutually optimal blocking pairs may lead to any of the three stable matchings, depending on which mutually optimal blocking pair is satisfied at stages 7 or 8 of the process (see Figure 12). Furthermore, in these stable matchings,  $m_1$  can be matched to  $w_2$  (in  $\mu_{10A}$ ),  $w_1$  (in  $\mu_{11B}$ ), or  $w_4$  (in  $\mu_9$ ), which means that he may be better off, same, or worse off.

Now observe woman  $w_1$ . She is the partner that gets initially divorced by the divorcee  $m_1$ . After the rematching process, she may be matched to  $m_5$  (in  $\mu_9$ , her most preferred counterpart, making her strictly better off compared to the initial matching. She may also be matched to  $m_3$  (in  $\mu_{10A}$ ), making her worse off compared to her initial match. Finally, she may also be matched to  $m_1$  again (in  $\mu_{11B}$ ), her initial match, resulting in no change of welfare for her. Thus, the example demonstrates that for both the divorcee and divorced may be better, worse, or equal as well off as in the initial matching, even within the same matching game and the from the same initial condition.  $\square$

## 7 Pareto Efficiency

The previous section begs the question about how to evaluate welfare in our setting. Under complete information, Pareto efficiency is used as welfare criterion for NTU matching markets. Focusing on unawareness but perfect information allows us to consider ordinal preferences just as in the case of complete information. However, the extension of Pareto efficiency to our setting is not straightforward because unanticipated experiences raise awareness and potentially change preferences. We need a notion of Pareto efficiency for endogenous preferences. Focusing for simplicity on matching without flirting, consider the following notion of Pareto improvement: Given an outcome  $(\mu, \omega)$ , the matching  $\mu' \neq \mu$  is a Pareto improvement if for all  $i \in M \cup W$ ,  $\mu'(i) \neq \mu(i)$  implies  $\mu'(i) \succ_i(t_i(\omega)) \mu(i)$ . In order to make sense of the definition recall that throughout the paper we assumed for simplicity that players have strict preferences. Thus, when a matching  $\mu'$  makes some players better off without making anybody worse off as compared to matching  $\mu$ , matching  $\mu'$  must make all players better off (compared to  $\mu$ ) for which the partner under  $\mu'$  differs from their partner under matching  $\mu$ . While the definition of Pareto improvement is a straightforward extension of Pareto improvement under complete information, it seems too weak in our context of endogenously changing preferences. In some sense, it is an *ex ante* notion of Pareto improvement *before* matching  $\mu$  has been experienced by the players. We may want to require that there is also no Pareto improvement *after*  $\mu$  has been experienced at  $\omega$ , awareness may have changed and with it preferences. That is, given an outcome  $(\mu, \omega)$ , say that matching  $\mu' \neq \mu$  is a  $\mu$ -experienced Pareto improvement if for all  $i \in M \cup W$ ,  $\mu'(i) \neq \mu(i)$  implies  $\mu'(i) \succ_i(t_i(\tau(\omega, \mu))) \mu(i)$ . It should be clear that  $\mu'$  might be a Pareto improvement given  $(\mu, \omega)$  but not a  $\mu$ -experienced Pareto improvement because  $\mu'$  might be worse for some players than  $\mu$  after having experienced  $\mu$  and “developed a taste for it.” Conversely,  $\mu'$  might be a  $\mu$ -experienced Pareto improvement but not a Pareto improvement because after experiencing  $\mu$  players discover

much to hate about it and prefer now  $\mu'$ . Thus, either notion does not imply the other. Requiring both notions may still not be enough: There could be a Pareto improvement after the matching  $\mu$  has been experienced at  $\omega$  and experienced again at  $\tau(\omega, \mu)$ , etc.

To define a notion of Pareto efficiency as persistent absence of Pareto improvements, let for all  $\omega \in \Omega$  and  $\mu \in \mathbf{M}$ ,  $\tau^0(\omega, \mu) := \omega$ ,  $\tau^1(\omega, \mu) := \tau(\omega, \mu)$ , and recursively for any  $n = 1, 2, \dots$ ,  $\tau^n(\omega, \mu) := \tau(\tau^{n-1}(\omega, \mu), \mu)$ .

**Definition 7** *We say that outcome  $(\mu, \omega)$  is persistent Pareto efficient if for any  $n = 0, 1, \dots$ , there does not exist a matching  $\mu' \neq \mu$  such that for all players  $i \in M \cup W$ ,*

$$\mu'(i) \neq \mu(i) \text{ implies } \mu'(i) \succ_i(t_i(\tau^n(\omega, \mu))) \mu(i).$$

That is, an outcome is persistent Pareto efficient if starting from the outcome for every repetition of the matching there does not exist a Pareto improvement. This notion can be extended to allow for flirting but for simplicity we focus on the case without communication.

We are interested in analogues to the welfare theorems. It is easy to see that the analogue to the first welfare theorem fails. If  $(\mu, \omega)$  is a self-confirming outcome, then it may not be persistent Pareto efficient. It may not even be Pareto efficient in the sense of absence of a Pareto improvement. Stability of  $\mu$  means that there is absence of pairwise common belief in blocking. Yet, as Example 3 demonstrates, there can be cases of self-confirming outcomes where each player of a pair likes to block but this is not pairwise common belief among them, causing the outcome to be Pareto inefficient in the sense that there is a Pareto improvement at  $(\mu, \omega)$ . Yet, we do have the analogue to the second welfare theorem:

**Proposition 8** *If  $(\mu, \omega)$  is persistent Pareto efficient, then  $(\mu, \lim_{n \rightarrow \infty} \tau^n(\omega, \mu))$  is a self-confirming outcome.*

**PROOF.** Any two-sided matching game with unawareness involves a finite number of finite state spaces. Any finite state machine on a finite space must have absorbing sets. Since awareness can only go up, we must have that  $\lim_{n \rightarrow \infty} \tau^n(\omega, \mu)$  is an absorbing state of  $\tau$  given  $\mu$ . Since  $(\mu, \omega)$  is persistent Pareto efficient, for any  $n$  there is no Pareto improvement  $\mu'$ . Thus, there is also no Pareto improvement at  $\lim_{n \rightarrow \infty} \tau^n(\omega, \mu)$ . Hence, there is no blocking pair and thus no pair that has pairwise common belief in blocking. We conclude that  $\mu$  is stable at  $\lim_{n \rightarrow \infty} \tau^n(\omega, \mu)$ . Since we already observed that  $\lim_{n \rightarrow \infty} \tau^n(\omega, \mu)$  is an absorbing state given  $\mu$ , we have that  $(\mu, \lim_{n \rightarrow \infty} \tau^n(\omega, \mu))$  is a self-confirming outcome.  $\square$

Our notion of persistent Pareto efficiency of  $(\mu, \omega)$  takes serious that the players experiences  $\mu$ . Alternatively (or additionally) one may require that also Pareto improvements are experienced. Given an outcome  $(\mu, \omega)$ , we say that matching  $\mu' \neq \mu$  is a  $\mu'$ -experienced Pareto improvement if for all players  $i \in M \cup W$ ,  $\mu'(i) \neq \mu(i)$  implies  $\mu'(i) \succ_i(t_i(\tau(\omega, \mu'))) \mu(i)$ . Note that given outcome  $(\mu, \omega)$ , matching  $\mu' \neq \mu$  being a



$\mu$ -experienced Pareto improvement does not imply that  $\mu'$  is a  $\mu'$ -experienced Pareto improvement and vice versa. For instance, a matching  $\mu'$  may not be a  $\mu$ -experienced Pareto improvement but when exposed to  $\mu'$ , players may discover that they like it more than  $\mu$  due to unanticipated experiences, awareness changes, and induced preference changes. We believe that a notion of Pareto efficiency that requires absence of  $\mu'$ -experienced Pareto improvements is too demanding and conceptually problematic especially in larger matching markets. It would require an experiment of experiencing *all* matchings. But since no two matchings can be experienced at the same time, it raises the question in which order they would be experienced.

## 8 Discussion

### 8.1 Infidelity

Neither in self-confirming outcomes nor flirt-proof self-confirming outcomes, players may become fully aware. The reason is that stability given the state prevents them from making transformative experiences and absorbing states does not allow for further changes of preferences in their stable matching. This has been illustrated in prior Example 5. One extra source of experiences is experimentation. In the marriage market it may be dubbed “infidelity”. Consider again Example 5. Suppose that in the self-confirming outcome  $(\mu_0, \omega_1)$  both man  $m_1$  and  $w_2$  temporary match despite  $\mu_0$  being stable given  $\omega_1$ . (Recall that in matching  $\mu_0$  given by the first matching in Figure 13, man  $m_1$  is matched to woman  $w_1$  and woman  $w_2$  is matched to man  $m_2$ .) Then man  $m_1$  and  $w_2$  would become aware that they are each others’ first choice. The state would transit to  $\omega_2$ . Consequently,  $m_1$  and  $w_2$  form a blocking pair and since this is also common belief among  $m_1$  and  $w_2$ , the original matching  $\mu_0$  would be destabilized. Divorcing the players and satisfying the blocking pair yields matching  $\mu_1$  at state  $\omega_2$ ; see Figure 13. This may be dubbed the “direct” effect of infidelity. Yet, there is also an indirect effect. Matching  $\mu_1$  is not stable given  $\omega_2$  because both man  $m_2$  and woman  $w_1$  are unmatched and now form a blocking pair and this is common belief. Consequently, satisfying this blocking pair yields matching  $\mu_2$ ; see Figure 13 for the resulting sequence of matchings. In matching  $\mu_2$ , both man  $m_2$  and woman  $w_2$  would become also aware that they are each others’ first choice. This is the “indirect” effect of infidelity of  $m_1$  and  $w_2$  on  $m_2$  and  $w_2$ . Contrary to the typical view of infidelity, it is a positive external effect in this case. (Obviously, this is not generally the case.) We reach a self-confirming outcome  $(\mu_2, \omega_2)$ , in which all parties are fully aware. In this example, infidelity allows all to escape the unawareness trap by marriage.

### 8.2 Stable Confusion About Others

Consider again Example 7. In outcome  $(\mu_2, \omega_2)$ , man  $m_1$  is matched to woman  $w_2$  and man  $m_2$  is matched to woman  $w_1$ . And while it is common belief among  $m_1$  and  $w_2$  that they prefer each other over the other counterpart, it is not common belief among

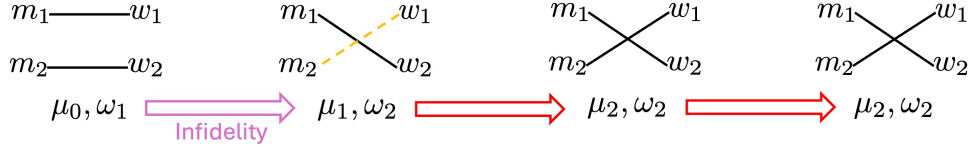


Figure 13: Infidelity in Example 5

all players. In particular, since at state  $\omega_2$  the point-belief of both man  $m_1$  and woman  $w_2$  is  $\omega_0$  (see Figure 11), they do not understand why  $m_1$  and  $w_2$  formed a blocking pair and why the matching changed from  $\mu_0$  to  $\mu_2$ . In other words, they are confused about  $m_1$  and  $w_2$ . Note also that players  $m_1$  and  $w_2$  can not gain anything from enlightening players  $m_2$  and  $w_2$ . So there is no force, either by communication or blocking actions, to change the situation. The outcome is flirt-proof self-confirming. It emphasizes the fact that a matching can be stable given beliefs and beliefs can be stable given the matching despite some players being confused. The confusion itself is stable because there is nothing they can do about within the game. We believe that this is often quite realistic in matching markets. It is also not surprising from a theoretical point of view because solutions concepts to coalitional games like stability or the core are not concepts akin to rationalizability in non-cooperative game theory.

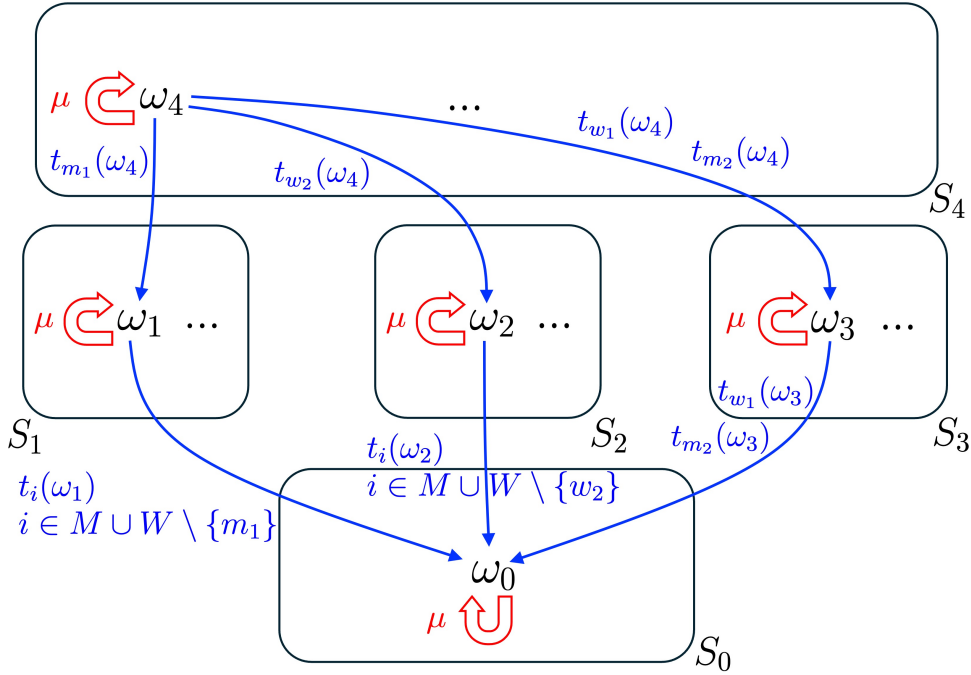


Figure 14: Unawareness Structure of Example 9

Confusion about other players that are *not* own partners like in Example 7 should be distinguished from confusion about your *own* partner. Such a outcome is also feasible in (flirt-proof) self-confirming outcomes but not reachable via our matching process unless we start in it already. We demonstrate in the following example:

**Example 9** There are two men  $M = \{m_1, m_2\}$  and two women  $W = \{w_1, w_2\}$ . The unawareness structure is given by Figure 14. There are five spaces. Spaces  $S_1, S_2$ , and  $S_3$  are all pairwise incomparable. Space  $S_4$  is the join while space  $S_0$  is the meet of the lattice. We just single out the states of interest in each space. At  $\omega_4$  in  $S_4$ , man  $m_1$ 's point-belief is  $\omega_1$  in  $S_1$  and woman  $w_2$ 's point-belief is  $\omega_2$ . The point-belief of  $m_2$  and  $w_1$  at  $\omega_4$  is  $\omega_3$  in the  $S_3$ . This is also common belief among them. Moreover, they (commonly) believe that  $m_1$  and  $w_2$  have point-belief  $\omega_0$  in  $S_0$ . Finally,  $m_1$  and  $w_2$  believe that all other players' point-belief is  $\omega_0$  in  $S_0$ .

The players' preferences at various states are given by the following rank-order lists:<sup>8</sup>

|                         |                         |                         |                         |
|-------------------------|-------------------------|-------------------------|-------------------------|
| $\succ_{m_1}: w_1, w_2$ | $\succ_{m_1}: w_2, w_1$ | $\succ_{m_1}: w_1, w_2$ | $\succ_{m_1}: w_1, w_2$ |
| $\succ_{m_2}: w_2, w_1$ | $\succ_{m_2}: w_2, w_1$ | $\succ_{m_2}: w_2, w_1$ | $\succ_{m_2}: w_1, w_2$ |
| $\succ_{w_1}: m_1, m_2$ | $\succ_{w_1}: m_1, m_2$ | $\succ_{w_1}: m_1, m_2$ | $\succ_{w_1}: m_2, m_1$ |
| $\succ_{w_2}: m_2, m_1$ | $\succ_{w_2}: m_2, m_1$ | $\succ_{w_2}: m_1, m_2$ | $\succ_{w_2}: m_2, m_1$ |
| $\omega_0$              | $\omega_1$              | $\omega_2$              | $\omega_3$              |

Consider the matching  $\mu$  given in Figure 16. We claim that  $(\mu, \omega_4)$  is a flirt-proof self-

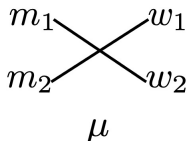


Figure 15: Matching in the Flirt-proof Self-confirming Outcome of Example 9

confirming outcome. At  $\omega_4$ , man  $m_1$  point-believes  $\omega_1$ . Given preferences at  $\omega_1$ , he is matched to his most preferred partner in  $\mu$ . At  $\omega_4$ , woman  $w_2$  point-believes  $\omega_2$ , at which her most preferred partner is  $m_1$ , to which she is matched to in  $\mu$ . Finally, at  $\omega_4$ , both man  $m_2$  and woman  $w_1$  point-believe  $\omega_3$ . At this state, both are matched to each other, the most most preferred partners. Thus, at  $\omega_4$ , nobody forms a blocking pair and hence there can also not be pair-wise common belief in blocking. Moreover, nobody has an incentive to raise somebody else awareness. Finally,  $\omega_4$  is absorbing. Hence  $(\mu, \omega_4)$  is a flirt-proof self-confirming outcome as claimed. However, this is not common belief, even not among partners. Consider man  $m_1$  at  $\omega_4$ . He point-believes  $\omega_1$  and thus believes that his partner in matching  $\mu$ , woman  $w_2$  point-believes  $\omega_0$ . Yet, at  $\omega_0$ , woman  $w_2$  forms a blocking pair with man  $m_2$  and this is also common belief. That is, at  $\omega_1$ , man  $m_1$  does not understand why his partner, woman  $w_2$  would not want to leave him. A similar situation occurs with woman  $w_2$ . She also does not understand why  $m_1$  does not leave her. That is, at  $\omega_4$  both partners are confused about each other. They are staying together because neither has an incentive to leave the other and the partner, even though (s)he is expected to leave by her/his partner, does not leave. To emphasize, different from the prior example, here  $m_1$  and  $w_2$  are confused about their *own* partner.<sup>9</sup> While we

<sup>8</sup>We omit their rank-order lists for state  $\omega_3$  as these preferences will not play any role in the discussion.

<sup>9</sup>W.r.t.  $m_2$  and  $w_1$ , at  $\omega_4$ , it is commonly believed among  $m_2$  and  $w_1$  that both are match to each other's most preferred partner. Yet, they are confused about  $m_1$  and  $w_2$ .

think that this might be a realistic feature of some matching situations, we believe it is less compelling than the prior confusion. Typically, within a match, partners know each other to the extent that they should rationalize each other's actions.  $\square$

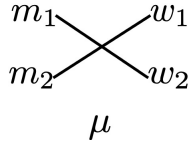


Figure 16: Matching in the Flirt-proof Self-confirming Outcome of Example 9

Self-confirming outcomes with confusion about the *own* partner can be ruled out with *non-trivial* matching dynamics, which is in contrast to self-confirming outcomes with confusion about others that are not partners. While  $(\mu, \omega_4)$  is a flirt-proof self-confirming outcome and thus our matching process is absorbing, it can never emerge when the process starts at an outcome *different* from  $(\mu, \omega_4)$ . The reason is that for such matching to be reached from another matching, we must satisfy a pair for which it is pair-wise common belief in blocking. So pair  $m_1$  and  $w_2$  could not have been matched in a penultimate stage before reaching  $(\mu, \omega_4)$  in our matching process. In conclusion, this example shows that such self-confirming outcomes with confusion about the *own* partner are less plausible because it cannot be reached unless the process trivially start already in this outcome.

While self-confirming outcomes do not rationalize the confusion, we may wonder about the possible state of mind of players who are confused. There are two possible ways to rationalize such confusion. The first is that others make mistakes: They should block but mistakenly do not do so. The second explanation keeps the assumption that others are rational but explains the confusion with awareness of unawareness: When a player expects others to block but such blocking does not happen, the player may suspect that (e)he her/himself is unaware of something that some others are aware. That is, the player may become aware that (s)he is unaware of something. While our model can be extended to model awareness of unawareness explicitly using the tools presented in Schipper (2024), it would not change any of our conclusions unless players can do something about discovering what they might miss. This would require enriching the problem with additional structure such as infidelity as discussed above or individual actions of engaging in gossiping, asking for counseling, etc.

### 8.3 Related Literature

Our work is related to four stands of literature. The first strand is the literature on matching with incomplete information. There are several papers on matching with incomplete information under the non-transferable utility (NTU) framework. Under two-sided uncertainty, Lazarova and Dimitrov (2017) study stability where a pair of agents blocks if each of the agents believe that he/she can do better with positive probability. This is a very demanding notion of stability as a stable matching must be robust to beliefs that

put a tiny probability on agents do better. Our notion of stability, in contrast, is absence of pairwise common (point-)belief in blocking. In order to block a pair must *agree* to block, i.e., it must be common belief among them that both prefer each other over their current partner. Under one-sided uncertainty where the workers' types distribution and the firms' types are common knowledge, Bikhchandani (2017) investigates two notions of stability: Ex ante stability where agents block if they are better off with all admissible types and Bayesian stability where agents block if they have higher expected utilities. Ex ante stability implies that participation in a block makes it common certainty that each is better off in a block. The notion of ex ante stability is adapted from a notion of stability introduced for transferable utility (TU) matching games with one-sided incomplete information by Liu et al. (2014). Pomatto (2022) provides an epistemic non-cooperative counterpart to Liu et al. (2014). Chen and Hu (2020) construct a learning process leading to an extension of the notion of stability by Liu et al. (2014). Alston (2020) shows that belief restrictions imposed on top of the stability notion by Liu et al. (2014) may lead to non-existence (see also Bikhchandani, 2017). Liu (2020) offers a stability notion for TU matching games with one-sided incomplete information that is explicit about the on-path and off-path beliefs. The stability notion closest to ours in the literature on TU matching with incomplete information is Forges (2004), who studies the extension of the incentive compatible coarse core of Vohra (1999) to assignment games. There is also the literature on centralized matching with incomplete information from a mechanism design perspective; see for instance Roth (1989), Majumdar (2003), Ehlers and Massó (2007), Yenmez (2013), and Fernandez, Rudov, and Yariv (2022).

The second strand is the literature on decentralized matching. Roth and Vande Vate (1990) showed that from any matching, there exists a sequence of matching by satisfying blocking pairs that leads to a stable matching. Ackermann et al. (2008) showed that from any matching, there exists a sequence of matching by satisfying optimal blocking pairs that leads to a stable matching. As corollaries, the process of satisfying random (optimal) blocking pairs leads to a stable matching with probability one.<sup>10</sup> These results are partly driving our decentralized matching process to converge to a self-confirming outcome: As we perturb the process that prioritizes mutually optimal blocking pairs with  $\varepsilon$  probability to select an optimal but not necessarily mutually optimal blocking pair even when a mutually blocking pair exists, we always have positive probability over all optimal blocking pairs. This is very different from Klaus, Klijn, and Walzl (2010) who perturb their process with  $\varepsilon$  probability to match a non-blocking pair together. Another difference between our random path of self-confirming outcomes and random path to stability in the literature is that our notion of blocking differs slightly since we require pairwise common (point-)belief in blocking. Another insightful paper on random path to stability is Rudov (2024), who shows that under some conditions, any unstable matching can reach any stable matching through the process of satisfying random (optimal) blocking

---

<sup>10</sup>Applications of Roth and Vande Vate (1990) on random paths to stability include decentralized market processes for stable job matching with competitive salaries (Chen, Fujishige, and Yang, 2011) and path to stability under incomplete information (Lazarova and Dimitrov, 2017, for under NTU matching games and Chen and Hu, 2020, for TU matching games). It has also been extended to more general contexts such as matching with couples (Klaus and Klijn, 2007), many-to-many matching (Kojima and Ünver, 2008), and many-to-many matching with contracts (Millán and Pepa Risma, 2018).

pairs. As mentioned in already in Section 2.2, he also observed using a five-by-five market that a process that satisfies only mutual optimal blocking pairs whenever they exist may fail to converge to a stable matching. From non-cooperative or search perspectives, Lauermann and Nöldeke (2014), Wu (2015), and Ferdowsian, Niederle, and Yariv (2025) show that decentralized interactions only lead to stable outcomes when there is a unique stable matching or when preferences are highly correlated. In this context note that our observation in Section 2 that without restrictions on the marriage market structure, prioritizing mutually optimal blocking pairs can lead to cycles, is consistent with these findings. Adachi (2003) shows that equilibrium outcomes converge to stable matchings as search frictions vanish. Search frictions in Adachi (2003) are represented by a time discounting while we interpret frictions as any factor that favors a non-mutual optimal blocking pair over a mutually optimal blocking pair. In contrast to Adachi (2003), we show that frictions can be arbitrary small but should not vanish for convergence to a stable matching. Finally, Doval (2022) investigates a notion of dynamic stability when matching opportunities arrive over time and matching is irreversible.

The third strand is the literature on the core. When modeling a matching problem under complete information, the core of the marriage market is the set of stable matchings (Roth and Sotomayor, 1990, Sasaki and Toda, 1992). For exchange economies with asymmetric information, Wilson (1978) introduced the notions of the coarse core and the fine core, which involves no information sharing or maximal information sharing respectively. Since both the core allocations and counterfactual blocking may be informative, these core concepts have been further refined (e.g., Vohra, 1999, Forges, 2004, Dutta and Vohra, 2005, Ray and Vohra, 2015). For a survey of the core under incomplete information, see Forges, Minelli, and Vohra (2002) and Forges and Serrano (2013). The coarse core has been extended to coalitional games with unawareness by Bryan, Ryall, and Schipper (2022). Our notion of stability is inspired by the coarse core. Our notion of flirt-proof stability is inspired by refinements of the coarse core that feature intentional communication. In some sense, flirt-proof stability combines the idea of the coarse core and with the idea of non-cooperative disclosure games (e.g., Milgrom and Roberts, 1986) but keeps the spirit of cooperative game theory.

The final strand is the literature on unawareness. Since this is the first paper on matching under unawareness, we contribute to the recent growing literature on exploring the implications of unawareness in economics. Other applications of unawareness pertain to disclosure, moral hazard, contract theory, screening, efficient mechanism design, auctions, procurement, delegation, speculation, financial market microstructure, default in general equilibrium, electoral campaigning, business strategy, and conflict resolution; for a bibliography, see Schipper (2025).

## 8.4 Possible Extensions

While we focus in this paper on unawareness and *perfect information* for simplicity, ordinality of preferences, and our desire to isolate phenomena related to unawareness independent from imperfect information, it would be interesting to allow for imperfect

information as well. Players could be uncertain about the awareness of others (i.e., including the experiences they made in the past) and their preferences. Matchings and absence of blocking become informative. Moreover, both matchings and flirting can be tools for experimentation with the aim not just to learn about what one is imperfectly informed about but also discover what one is unaware. While it is straightforward to extend the unawareness structure to allow for imperfect information using unawareness type spaces introduced in Heifetz, Meier, and Schipper (2013), finding the appropriate solution concept is not straightforward. As it has been observed in the literature on the core under asymmetric information and matching with incomplete information, a satisfactory solution concept to matching under incomplete information should take into account the information revealed in matchings and counterfactual blocking (for surveys, see Forges and Serrano, 2013, Forges, Minelli, and Vohra, 2002). Even without unawareness, progress has been made almost exclusively only in matching games with *one-sided* incomplete information (e.g., Liu et al, 2014, Bikhchandani, 2017, Liu, 2020).

While our analysis is mainly cast in terms of the marriage market (again for simplicity and as a guide for intuitions), we believe that it can be extended to other matching problems including settings with transferable utility and many-to-many matching. Unanticipated experiences may happen in employee-employer relationships, school choice, with “experience” goods, etc. Similarly, within the current setting of the marriage market, alternative divorce rules may be explored. Currently, divorce occurs when a player unilaterally separates the match. Yet, human history also saw other divorce rules such as the requirement of bilateral agreement to separate the link. We conjecture that such alternative divorce rule leads to less divorces, less experiences and thus less awareness within the market.

This paper focuses on decentralized matching under unawareness. It would be interesting to consider centralized matching in matching mechanisms. One desideratum for such mechanisms would be to pool awareness of all participants and implement an efficient matching at the pooled awareness level. Recently, Pram and Schipper (2025) introduced dynamic direct elaboration mechanisms with VCG-type transfers in quasi-linear environments to implement efficiently under pooled awareness. We currently consider extensions to matching.

## A Application of the Cyclic Example to Two-Sided Market Entry

In this section, we present another application of our observation in Section 2.2 to market entry in matching markets. Bennett (1994) considers adding a *single* agent (say, a woman) to a marriage market with stable matching, and shows that the process (i.e., optimal divorce and remarriage procedure) of satisfying the optimal blocking pair for the unmatched woman terminates in finite number of steps and reaches a stable matching. We consider an extended problem by adding *both* a woman and a man to marriage market with a stable matching and allow satisfying of mutually optimal blocking pairs.

In contrast to Bennett’s problem of adding a *single* agent only, a modification of our example shows that this process can now lead to a cycle.

The initial matching market consists of three man  $\{m_1, m_3, m_4\}$ , and three woman  $\{w_2, w_3, w_4\}$ . Let the preference profile be consistent with the preferences in Example 2. That is, we consider the rank-order lists as in Example 2 but cross out  $m_2$  and  $w_1$ :

$$\begin{aligned}
 \succ_{m_1}: & w_2, w_4, w_3, \cancel{w_1} \\
 \cancel{\succ_{m_2}}: & \cancel{w_4}, \cancel{w_2}, \cancel{w_1}, \cancel{w_3} \\
 \succ_{m_3}: & \cancel{w_1}, w_3, w_2, w_4 \\
 \succ_{m_4}: & w_3, \cancel{w_1}, w_4, w_2 \\
 \cancel{\succ_{w_1}}: & \cancel{m_1}, \cancel{m_2}, \cancel{m_4}, \cancel{m_3} \\
 \succ_{w_2}: & m_4, m_3, \cancel{m_2}, m_1 \\
 \succ_{w_3}: & \cancel{m_2}, m_1, m_3, m_4 \\
 \succ_{w_4}: & m_3, m_4, m_1, \cancel{m_2}
 \end{aligned}$$

Let the initial matching be  $\mu_1 = \begin{pmatrix} m_1 & m_3 & m_4 \\ w_2 & w_3 & w_4 \end{pmatrix}$ , which is stable (without  $m_2$  and  $w_1$ ). Add  $m_2$  and  $w_1$  to the market. The unique mutually optimal blocking pair is  $(m_2, w_2)$ .

Satisfying  $(m_2, w_2)$  leads to the matching  $\mu_2 = \begin{pmatrix} m_1 & m_2 & m_3 & m_4 & w_1 \\ m_1 & w_2 & w_3 & w_4 & w_1 \end{pmatrix}$ , which is the same as in the cycle in Example 2. As illustrated before, there is always a unique mutually optimal blocking pair along this path, and it cycles back to  $\mu_2$ . This observation shows that Bennett’s optimal divorce and remarriage procedure cannot easily be extended to situations where both a man and a woman are added to the market at the same time.

## References

- [1] Ackermann, H., Goldberg, P.W., Mirrokni, V.S., Röglin, H., Vöcking, B. (2008). Uncoordinated two-sided matching markets, Proceedings of the 9th ACM Conference on Electronic Commerce, 256–263.
- [2] Adachi, H. (2003). A search model of two-sided matching under nontransferable utility, Journal of Economic Theory 113, 182–198.
- [3] Alston, M. (2020). On the non-existence of stable matches with incomplete information, Games and Economic Behavior 120, 336–344.
- [4] Bikhchandani, S. (2017). Stability with one-sided incomplete information, Journal of Economic Theory 168, 372–399.
- [5] Baker, L.A., Emery, R.E. (1993). When every relationship is above average. Perceptions and expectations of divorce at the time of marriage, Law and Human Behavior 17, 439–450.



- [6] Battigalli, P., Guaitoli, D. (1997). Conjectural equilibria and rationalizability in a game with incomplete information, in: Battigalli, P., Montesano, A., Panunzi, F. (eds.), *Decisions, games and markets*, Dordrecht: Kluwer Academic Publishers, 97–124.
- [7] Bennett, E. (1994). Entry, divorce and re-marriage in matching markets, in: Megiddo, N. (ed.), *Essays in game theory. In honor of Michael Maschler*, Springer-Verlag, 29–38.
- [8] Bryan, K., Ryall, M., Schipper, B.C. (2022). Value capture in the face of known and unknown unknowns, *Strategy Science* 7, 157-189.
- [9] Chen, B., Fujishige, S., Yang, Z. (2011). Decentralized market processes to stable job matchings with competitive salaries, Kyoto University, Research Institute for Mathematical Sciences.
- [10] Chen, Y.-C., Hu, G. (2020). Learning by matching, *Theoretical Economics* 15, 29–56.
- [11] Doval, L. (2022). Dynamically stable matching, *Theoretical Economics* 17, 687–724.
- [12] Dutta, B., Vohra, R. (2005). Incomplete information, credibility and the core, *Mathematical Social Sciences* 50, 148–165.
- [13] Ehlers, L., Massó, J. (2007). Incomplete information and singleton cores in matching markets. *Journal of Economic Theory*, 136(1), 587-600.
- [14] Ferdowsian, A., Niederle, M., Yariv, L. (2025). Strategic decentralized matching: The effects of information frictions, Princeton and Stanford Universities.
- [15] Fernandez, M.A., Rudov, K., Yariv, L. (2022). Centralized matching with incomplete information, *American Economic Review: Insights* 4, 18–33.
- [16] Forges, F. (2004). The ex ante incentive compatible core of the assignment games, *Mathematical Social Sciences* 47, 135–151.
- [17] Forges, F., Minelli, E., Vohra, R. (2002). Incentives and the core of an exchange economy, *Journal of Mathematical Economics* 38, 1–41.
- [18] Forges, F., Serrano, R. (2013). Cooperative games with incomplete information: Some open problems, *International Game Theory Review* 15, 1340009.
- [19] Fudenberg, D., Levine, D.K. (1993). Self-confirming equilibrium, *Econometrica* 61, 523–545.
- [20] Gale, D., Shapley, L.S. (1962). College admissions and the stability of marriage, *American Mathematical Monthly* 69, 9–15.
- [21] Heifetz, A., Meier, M., Schipper, B.C. (2013). Unawareness, beliefs, and speculative trade, *Games and Economic Behavior* 77, 100–121.

- [22] Heifetz, A., Meier, M., Schipper, B.C. (2008). A canonical model for interactive unawareness, *Games and Economic Behavior* 62, 305–324.
- [23] Heifetz, A., Meier, M., Schipper, B.C. (2006). Interactive unawareness, *Journal of Economic Theory* 130, 78–94.
- [24] Holtom, B., Goldberg, C.B., Allen, D.G., Clark, M.A. (2017). How today’s shocks predict tomorrow’s leaving, *Journal of Business Psychology* 32, 59–71.
- [25] Klaus, B., Klijn, F. (2007). Paths to stability for matching markets with couples, *Games and Economic Behavior* 58, 154–171.
- [26] Klaus, B., Klijn, F., Walzl, M. (2010). Stochastic stability for roommate markets. *Journal of Economic Theory*, 145, 2218-2240.
- [27] Knuth, D.E. (1976). *Marriages stables*, Les Presses de l’université de Montréal.
- [28] Kojima, F., Ünver, M.U. (2008). Random paths to pairwise stability in many-to-many matching problems: a study on market equilibration, *International Journal of Game Theory* 36, 473–488.
- [29] Laueremann, S., Nöldeke, G. (2014). Stable marriages and search frictions, *Journal of Economic Theory* 151, 163–195.
- [30] Lazarova, E., Dimitrov, D. (2017). Paths to stability in two-sided matching under uncertainty, *International Journal of Game Theory* 46, 29–49.
- [31] Livingston, G. (2014). Four-in-ten couples are saying ‘I do,’ again, Washington, D.C.: Pew Research Center.
- [32] Liu, Q. (2020). Stability and Bayesian consistency in two-sided markets. *American Economic Review*, 110, 2625-2666.
- [33] Liu Q.M., Mailath, G.J., Postlewaite, A., Samuelson, L. (2014). Stable matching with incomplete information, *Econometrica* 82, 541–587.
- [34] Lucas, R.E. (2005). Time does not heal all wounds. A longitudinal study of reaction and adaptation to divorce, *Psychological Science* 16, 945–950.
- [35] Maher, H. (2003). Why are there so few prenuptial agreements?, Harvard Law School Discussion Paper No. 436.
- [36] Majumdar, D. (2003). Ordinally Bayesian incentive compatible stable matching.
- [37] Milgrom, P and J. Roberts (1986). Relying on the information of interested parties, *Rand Journal of Economics* 107, 18–32.
- [38] Millán, B., Pepa Risma, E. (2018). Random path to stability in a decentralized market with contracts, *Social Choice and Welfare* 51, 79–103.

- [39] Paul, L.S. (2014). Transformative experience, Oxford University Press.
- [40] Pomatto, L. (2022). Stable matching under forward-induction reasoning, *Theoretical Economics* 17, 1619–1649.
- [41] Pram, K., Schipper, B.C. (2025). Efficient mechanisms under unawareness, University of California, Davis.
- [42] Ray, D., Vohra, R. (2015). The farsighted stable set, *Econometrica* 83, 977–1011.
- [43] Roth, A.E. (1989). Two-sided matching with incomplete information about others' preferences, *Games and Economic Behavior* 1, 191–209.
- [44] Roth, A.E., Sotomayor, M.A. (1990). Two-sided matching. A study in game-theoretical modeling and analysis, Cambridge University Press.
- [45] Roth, A.E., Vande Vate, J.H. (1990). Random paths to stability in two-sided matching, *Econometrica* 58, 1475–1480.
- [46] Rudov, K. (2024). Fragile stable matchings, arXiv:2403.12183.
- [47] Sasaki, H., Toda, M. (1992). Consistency and characterization of the core of two-sided matching problems. *Journal of Economic Theory*, 56, 218–227.
- [48] Schipper, B.C. (2025). The Unawareness Bibliography, <https://faculty.econ.ucdavis.edu/faculty/schipper/unaw.htm>.
- [49] Schipper, B.C. (2024). Interactive awareness of unawareness, University of California, Davis.
- [50] Schipper, B.C. (2021). Discovery and equilibrium in games with unawareness, *Journal of Economic Theory* 198, 105365.
- [51] Shapley, L.S. (1953). Stochastic games, *Proceedings of the National Academy of Sciences of the USA* 39, 1095–1100.
- [52] Spanier, G.B., Furstenberg, F.F. (1982). Remarriage after divorce: A longitudinal analysis of well-being, *Journal of Marriage and Family* 44, 709–720.
- [53] Svarer, M., Verner, M. (2008). Do children stabilize relationships in Denmark, *Journal of Population Economics* 21, 395–417.
- [54] Symoens, S., Bastiaens, K., Mortelmans, D., Bracke, P. (2013). Breaking up, breaking hearts? Characteristics of the divorce process and well-being after divorce, *Journal of Divorce & Remarriage* 54, 177–196.
- [55] U.S. Department of Labor (2023). Number of jobs, labor market experience, marital status, and health for those born 1957-1964, USDL-23-1854
- [56] Vohra, R. (1999). Incomplete information, incentive compatibility, and the core, *Journal of Economic Theory* 86, 123–147.

- [57] Wilson, R. (1978). Information, efficiency, and the core of an economy, *Econometrica* 46, 807–816.
- [58] Wu, Q. (2015). A finite decentralized marriage market with bilateral search, *Journal of Economic Theory* 160, 216–242.
- [59] Yenmez, M.B. (2013). Incentive-compatible matching mechanisms: consistency with various stability notions, *American Economic Journal: Microeconomics* 5, 120–141.