Commitment and matching in the marriage market

Berliant, Marcus and Thakur, Sounak

Washington University in St. Louis

11 September 2019

Commitment and Matching in the Marriage Market

Marcus Berliant [*] Sounak Thakur [†]

September 11, 2019

Abstract

The set of stable marriage matches is different depending on whether allocation within marriage is determined by binding agreements in the marriage market (BAMM) or by bargaining in marriage (BIM). With transferable utility, any stable matching is utilitarian efficient under BAMM, but not under BIM. Is it possible to implement the efficient matching under BIM? We show that if one side of the market is sufficiently sensitive relative to the other, if the more sensitive side can be ranked by sensitivity, and if their preferences are hierarchical, the top trading cycles algorithm results in an efficient matching.

Keywords: Two-sided matching, marriage, bargaining
JEL Classification Numbers: C78, D1,J12
Acknowledgment: We are grateful to Robert Pollak and SangMok Lee for helpful comments. All remaining errors are our own.

[*]Washington University in St. Louis. Email: berliant@wustl.edu
[†]Washington University in St. Louis. Email: sounak@wustl.edu

1 Introduction

The dominant paradigm in the marriage-matching literature considers marriage market equilibrium under **B**inding **A**greements in the **M**arriage **M**arket(**BAMM**). In the typical model of the marriage market (for instance, see Chiappori *et al.* (2018), Chiappori *et al.* (2017), Gayle & Shephard (2019)), the division of the marital surplus is negotiated at the time of marriage. It is assumed that the contract reached in the marriage market is binding upon the couple, i.e., it cannot be breached or re-negotiated under any state of the world that may occur in future. In other words, there is full commitment within marriage.

An empirically testable implication of BAMM is that unanticipated changes in laws governing exit from marriage, i.e., divorce, have no impact on behavior within marriage. However, this does not hold in the data. For example, Voena (2015) finds that change in divorce and property division laws in the United States reduced female labor force participation and increased rates of asset accumulation in marriages that had formed before the change in laws. Similarly, empirical evidence suggests that policy-induced changes in spousal incomes change household expenditure patterns (for example, see Lundberg *et al.* (1997)) — a finding that is at odds with couples having reached binding agreements in the marriage market.

While the empirical evidence is not consistent with BAMM, it can be rationalized using the **B**argaining **I**n **M**arriage(**BIM**) hypothesis. According to BIM, married couples play a cooperative game in each period. Given the threat points (whether internal as in Lundberg & Pollak (1993) or exit threats as in Voena (2015)) of this game, married couples attain efficient outcomes in each period of marriage, which can change if threat points are affected by exogenous changes in policy, examples of which include legal changes and government-administered welfare programs that affect relative spousal incomes.

The BAMM and BIM assumptions also entail potentially different marriage market equilibria, and have potentially different welfare implications. For instance, with transferable utility, any stable matching under BAMM yields the highest total utility (to all players) amongst all possible matchings. In other words, a stable matching under BAMM and transferable utility is *utilitarian efficient*. However, this does not necessarily hold under BIM. As Pollak (2019) demonstrates, the set of stable matchings under BIM do not necessarily coincide with the set of stable matchings under BAMM. In particular, he illustrates that the BAMM and BIM equilibria can be distinct.

If the notion of marriage market equilibrium in a BIM setting is stability, the appropriate algorithm to find the equilibrium/equilibria is the Gale-Shapley algorithm. This is the route taken by Pollak (2019). However, the few empirical papers that have tried to predict real world matches using the Gale-Shapley algorithm (see Hitsch *et al.* (2010), Banerjee *et al.* (2013), Lee (2009)) have failed to replicate patterns of assortative matching on several important dimensions. These results cast doubt on whether Gale-Shapley is the appropriate algorithm to use in order to find marriage market equilibria. Further, they leave open the possibility that the marriage market equilibrium under BIM is identical to the marriage market equilibrium under BAMM — a possibility, which, in our opinion, should be a subject of theoretical and empirical research.

In this paper, we explore the theoretical aspect of the problem. To be precise, we pose the following question: Using a matching algorithm different from the Gale-Shapley algorithm, can we implement the stable matching under BAMM (with transferable utility) even under BIM? An obvious candidate for implementing the BAMM assignment in the BIM world is the *top trading cycles* algorithm, which, in contrast to the Gale-Shapley algorithm, produces a Pareto efficient matching. We show that if agents on one side of the market are sufficiently sensitive to matches relative to the other side, if the more sensitive side can be ranked by sensitivity, and if preferences over members of the opposite sex are hierarchical, the top trading cycles algorithm results in a utilitarian efficient matching. As is obvious, utilitarian efficiency is achieved at the cost of stability — a tension that has been well-recognized in the literature (see Lee & Yariv (2018) for a recent example).

The exercise in the current paper is, in spirit, similar to the familiar second welfare theorem in general equilibrium theory (see Mas-Colell *et al.* (1995)), which provides conditions under which a Pareto optimal allocation can be supported as a competitive equilibrium (with taxes and transfers). In our setting, the counterpart to a Pareto optimal allocation is a stable matching under BAMM, which happens to be belong to the core of the assignment game; while the counterpart to decentralization using prices (as in the second welfare theorem) is the "decentralization" using the top trading cycles algorithm.

The remainder of this paper is structured as follows: Section 2 describes the economic environment. Section 3 presents alternative desirable properties of a marriage market equilibrium. Section 4 discusses implementation of the utilitarian efficient assignment under BIM. Section 5 concludes with a brief discussion. All proofs are placed in the appendix.

2 The Economic Environment

There are a finite and equal number of men and women in the market. Formally, let $\mathcal{M}$ and $\mathcal{W}$ denote the set of men and women respectively and let $|\mathcal{M}| = |\mathcal{W}| = N$, where $|X|$ denotes the cardinality of the set X. Men and women play the following two-stage game: In the first stage, men and women match with one another. We assume that the matching is simultaneous, not sequential. In the second stage, matched couples play a cooperative game. In particular, each married couple decides on public and private consumption within marriage.

We assume that individual preferences over private and public consumption goods within marriage are such that utility is transferable within any couple. Formally, let $\{\succ_m\}_{m\in\mathcal{M}}, \{\succ_w\}_{w\in\mathcal{W}}$ denote individual preference orderings over bundles of private and public consumption goods. We assume that these orderings are such that for any

$m \in \mathcal{M}$, $w \in \mathcal{W}$, there exist cardinalizations, denoted by U_m and U_w, that represent $\succ_m$ and $\succ_w$ such that the utility possibility set is given by:

$$\mathbf{U} = \{(U_m, U_w) \in \mathbb{R}^2 : U_m + U_w \leq s_{m,w}\}. \tag{1}$$

where $s_{m,w}$ denotes the utility surplus produced if man m were to marry woman w. We assume that $s_{m,w} > 0 \; \forall m \in \mathcal{M}$, $w \in \mathcal{W}$. Thus, utility is transferable within each household.

While we shall not specify the household game that gives rise to the Pareto frontiers described here, we point out two important facts. First, it is well-known that (generalized) quasi-linear preference orderings satisfy the transferable utility property (see Bergstrom (1989), Chiappori (2017)). Second, transferability of utility does not require the Pareto frontier to be a hyperplane for every cardinalization of preferences. However, transferability of utility requires that there exist a cardinalization of preferences such that the Pareto frontier is a hyperplane as described above (see Bergstrom & Varian (1985), Chiappori (2017)). In our context, U_m and U_w are such well-chosen cardinalizations.

The primitives of the economic environment depend on whether we assume BIM or BAMM. Under BAMM, the primitives of the two-stage problem are given by the objects $< \mathcal{M}, \mathcal{W}, S >$, where S is a NXN utility-surplus matrix, whose m, w -th element, denoted by $s_{m,w}$, is the utility surplus if the couple (m, w) were to be formed. Further, we normalize the utility surplus from non-marriage to zero for each individual. By contrast, the primitives of the problem under BIM are given by the following objects: $< \mathcal{M}, \mathcal{W}, U^{\mathcal{M}}_{BIM}, U^{\mathcal{W}}_{BIM} >$, where $U^{\mathcal{M}}_{BIM}$ ($U^{\mathcal{W}}_{BIM}$) is an NXN matrix whose m, w-th element, denoted by $u^{\mathcal{M}}_{m,w}$ ($u^{\mathcal{W}}_{m,w}$), gives the payoff in marriage that will accrue to man m (woman w) if he (she) were to marry woman w (man m). These payoffs are the outcome of bargaining that would happen in marriage, were couple (m, w) to be formed. Moreover, the outcome of the bargaining game is correctly foreseen by all participants in the marriage market. Further, we assume that the anticipated outcome of the bargaining game induces a strict preference ordering over the set of men. Formally, for any w, $u^{\mathcal{W}}_{m,w} \neq u^{\mathcal{W}}_{m',w}$ whenever $m \neq m'$. Finally, in order to ensure comparability with BAMM, we set $U^{\mathcal{M}}_{BIM} + U^{\mathcal{W}}_{BIM} = S$.

The solution under BAMM consists of the following two objects: an assignment/matching of women to men[1] and a utility imputation vector for all possible couples that determine how the marital surplus will be split. Formally, the solution under BAMM consists of A_{BAMM} and I_{BAMM} where A_{BAMM} is a one-to-one onto mapping such that $A_{BAMM} : \mathcal{W} \mapsto \mathcal{M}$ and and a NXN matrix I_{BAMM}, whose $(m, w) - th$ element, denoted by $I_{BAMM}(m, w)$ is an ordered pair in$\{(U_m, U_w) \in \mathbb{R}^2 : m \in \mathcal{M}, w \in \mathcal{W}$ and $U_m + U_w \leq s_{m,w}\}$, $m \in \mathcal{M}$, $w \in \mathcal{W}$. By contrast, under BIM the solution to the matching game consists of only one object, namely, the assignment $A_{BIM} : \mathcal{W} \mapsto \mathcal{M}$ where A_{BIM} is one-to-one and onto. For any couple that may form, the utility to the man and the woman are as dictated by the primitives of the problem.

Notice that under BAMM, the splits of the marital surplus are decided in the marriage market. These contracts are inviolable, ie, they cannot be reneged in marriage. By contrast, under BIM, each individual, in the marriage market, correctly foresees his/her payoff in each possible match. As mentioned before, the potential payoffs result from bargaining in marriage, should the corresponding man-woman pair match. Most importantly, no contracts regarding the split of marital surplus can be made in the marriage market.

If we are in a BIM environment, it is convenient to develop some further notation to denote the utility to an individual from an assignment. For any i, $i \in \mathcal{M} \cup \mathcal{W}$, we intend to have a function that provides the utility received by i under any given assignment A. This is accomplished by defining $\widetilde{U}_i : \mathcal{A} \mapsto \mathbb{R}_+$, where
$\mathcal{A} = \{A | A : \mathcal{W} \mapsto \mathcal{M}\}$ and $\widetilde{U}_i$ satisfies the following property:

[1] Since we have normalized the utility surplus from non-marriage to zero and assumed that each marriage produces a positive surplus, all individuals would marry under any reasonable solution concept in our setting. Also, we exclude polygamy by assumption.

1. For any $A \in \mathcal{A}$, $m \in \mathcal{M}$ and the ordered pair $(w, m) \in A$, $\widetilde{U_m}(A) = u^{\mathcal{M}}_{m,w;BIM}$

2. For any $A \in \mathcal{A}$, $w \in \mathcal{W}$ and the ordered pair $(w, m) \in A$, $\widetilde{U_w}(A) = u^{\mathcal{W}}_{m,w;BIM}$.

Notice that $\widetilde{U}_i(A)$ is the utility of individual i under assignment A. As is standard, in this basic framework, there are no externalities between matched couples.

It is worth emphasizing that under our set-up, both the matching and the split of the surplus accruing to each spouse in the second stage are determined in the first stage. Nonetheless, the second stage of the game is not superfluous. In other words, we cannot reduce the game we have described to a one-shot game, like the prisoners' dilemma, for example. The difference between a one-shot game, like the prisoners' dilemma and the current set-up is as follows: In a prisoners' dilemma, the prisoners are matched. By contrast, under the current set-up, the payoff matrix in the second stage is sensitive to the matching that occurs in the first stage.

Finally, we note that while the Gale-Shapley algorithm is usually used in a non-transferable utility framework, it can easily be adapted for use in a transferable utility setting in a Bargaining in Marriage (BIM) set-up. In doing so, we follow Pollak (2019), who points out that the anticipated outcome of bargaining provides the utility that agents foresee arising from different marriages. These numbers can be used to derive a ranking of potential partners, which are the primitives required to run the Gale-Shapley algorithm.

3 Marriage Market Equilibrium: Alternative Criteria and Welfare Implications

With a view to exploring the nature of the marriage market equilibrium under Binding Agreements in the Marriage Market(BAMM) and Bargaining in Marriage(BIM), we first introduce a few possible characteristics of an equilibrium assignment:

1. **Stability:** In a BIM setting, an assignment A_{BIM} is said to be stable if there

does not exist a pair (w,m), $w \in \mathcal{W}$, $m \in \mathcal{M}$ such that $A_{BIM}(w) \neq m$, $u^{\mathcal{W}}_{m,w} > u^{\mathcal{W}}_{A_{BIM}(w),w}$ and $u^{\mathcal{M}}_{m,w} > u^{\mathcal{M}}_{m,A^{-1}_{BIM}(m)}$. Analogously, in a BAMM setting, an assignment A_{BAMM} and associated imputations of utility $I_{BAMM}(m,w) = (u^*_m(m,w), u^*_w(m,w))$, $m \in \mathcal{M}$, $w \in \mathcal{W}$, is said to be stable if there does not exist a pair (w',m'), $w' \in \mathcal{W}$, $m' \in \mathcal{M}$ such that $A_{BAMM}(w') \neq m'$, $u^*_{w'}(m',w') > u^*_{w'}(A_{BAMM}(w'),w')$ and $u^*_{m'}(m',w') > u^*_{m'}(m', A^{-1}_{BAMM}(m'))$.

2. **Woman-Pareto Optimality:** In a BIM setting, an assignment A is said to be *woman Pareto optimal* if there does not exist another assignment $A' \in \mathcal{A}$ such that $\widetilde{U}_w(A') \geq \widetilde{U}_w(A)$ $\forall w \in \mathcal{W}$ and there is at least one $w' \in \mathcal{W}$ such that $\widetilde{U}_{w'}(A') > \widetilde{U}_{w'}(A)$. Similarly, one can define a *man Pareto optimal* assignment[2].

3. **Utilitarian Optimality**: In a BIM setting, an assignment A_{BIM} is said to satisfy *utilitarian optimality* if it is an assignment that a utilitarian social planner would choose, ie, $A_{BIM} \in \underset{A' \in \mathcal{A}}{\arg\max} \left[\sum_{m \in \mathcal{M}} \widetilde{U}_m(A') + \sum_{w \in \mathcal{W}} \widetilde{U}_w(A')\right]$. Analogously, in a BAMM setting, an assignment A_{BAMM} is said to satisfy *utilitarian optimality* if $A_{BAMM} \in \underset{A' \in \mathcal{A}}{\arg\max} \left[\sum_{\{(w,m):A'(w)=m\}} s_{m,w}\right]$.

It is well-known that with transferable utility and BAMM, stability is equivalent to *utilitarian optimality* (see Koopmans & Beckmann (1957), Shapley & Shubik (1971)). On the other hand, a stable assignment under BIM, which may be found by using the Gale-Shapley algorithm, is not utilitarian optimal in general (see Pollak (2019)). As mentioned before, we will implement the utilitarian efficient assignment in a BIM setting using the top trading cycles algorithm. To ensure comparability, we will restrict our attention to the *woman-proposing* Gale-Shapley algorithm and the *" woman-choosing" top trading cycles* algorithm. The two alternative matching algorithms are described below.

The *woman-proposing* Gale-Shapley algorithm proceeds as follows: In the first round, each woman proposes to her favorite man. Each man tentatively accepts (ie, "dates") the woman that he prefers most amongst the women who have proposed to

[2] We do not need to define Pareto optimality for a BAMM setting.

him. He rejects all other proposals. In any subsequent round, each woman who is not currently "dating" a man proposes to her most preferred man from amongst the set of men have not rejected her at any previous round. If a man prefers his current partner to all the proposals he receives in the current round, he rejects all proposals and continues "dating" his existing partner. On the other hand, if a woman who has proposed to a man is more attractive to him than his current partner, he ends his "engagement" with his current partner and starts "dating" the most preferred woman who proposed to him in the current round. He rejects all other proposals. The algorithm stops when there are no more rejections by men.

The *" woman-choosing" top trading cycles* algorithm proceeds as follows. In the first step, each man points to his favorite woman and each woman points to her favorite man. If $(m_1, w_1, m_2, w_2, ..., m_k, w_k)$ form a cycle, each woman pairs with the man she points to. Matched men and women are removed and the algorithm proceeds until everyone is matched.

As originally shown by Gale & Shapley (1962), the Gale-Shapley algorithm produces a stable match. On the other hand, the *top trading cycles* algorithm produces a *woman-Pareto optimal* assignment, ie, an assignment of men to women such that by changing the assignment, no woman can be made better off without making at least one other woman worse off. The Gale-Shapley assignment is not necessarily *woman-Pareto optimal* while the *top trading cycles* assignment is not necessarily stable (see Abdulkadiroğlu & Sönmez (2003) for an illustration).

A matching mechanism is said to be strategy proof if it is a dominant strategy for all agents to reveal their true preferences under that mechanism. Since the woman-proposing Gale-Shapley mechanism is woman-optimal, it is a dominant strategy for each woman to state her true preferences (see Roth & Sotomayor (1990), Theorem 4.7, page 90). However, with strict preferences, whenever more than one stable assignment exists, there will always be an incentive for some man to misrepresent his preferences under the woman-proposing Gale-Shapley algorithm (see Roth &

Sotomayor (1990), Corollary 4.12, page 96). A similar result applies to the top trading cycles algorithm. Abdulkadiroğlu & Sönmez (2003)[3] prove that the top trading cycles mechanism is strategy proof for women while the example in our Appendix A shows that it is not strategy proof for men.

4 Implementing the BAMM Assignment in a BIM Framework

Proposition 1 below provides a sufficient condition under which the **BAMM** assignment may coincide with the assignment under **BIM** with *top-trading cycles.* In order to state Proposition 1, we must first introduce some notation and establish a lemma.

For $\mathcal{J} = \mathcal{M}, \mathcal{W}$, define $U_{\mathcal{J}} = \{u \in \mathbb{R} : \exists\ A \in \mathcal{A} \text{ s.t. } u = \sum_{i \in \mathcal{J}} \widetilde{U}_i(A)\}$

$U^*_{\mathcal{J}} := \max\limits_{A \in \mathcal{A}} \sum_{i \in \mathcal{J}} \widetilde{U}_i(A) \quad$ and $\quad A^*_{\mathcal{J}} = \arg\max\limits_{A \in \mathcal{A}} \sum_{i \in \mathcal{J}} \widetilde{U}_i(A),$.

In words, $A^*_{\mathcal{J}}$ is the set of assignments, each element of which maximizes the sum of utilities of all individuals belonging to set $\mathcal{J}$.

$$U^{*,(-1)}_{\mathcal{J}} := \max_{A \in \mathcal{A} \setminus A^*_{\mathcal{J}}} \sum_{i \in \mathcal{J}} \widetilde{U}_i(A), \qquad \mathcal{J} = \mathcal{M}, \mathcal{W}$$

In words, if the unique values of the sum of utilities (over all possible assignments) of individuals in set $\mathcal{J}$ (to members of the opposite sex) were to be ranked in descending order, $U^{*,(-1)}_{\mathcal{J}}$ would be the second element.

Further, define **CONDITION A** as follows:

CONDITION A: $U^*_{\mathcal{W}} - U^{*,(-1)}_{\mathcal{W}} > U^*_{\mathcal{M}}$, and $A^*_{\mathcal{W}}$ is a singleton.

With a view toward understanding the condition intuitively, let us define the first-best assignment for any gender as the assignment that maximizes the sum of utilities of all members of that gender across all possible assignments. Then, **CONDITION A** translates into the requirement that the first-best assignment for men entail a lower total utility (to men) than the difference in utility (to women) between the

[3] See their Proposition 4, pg. 738

first-best and the second-best assignment for women. Since we have normalized the utility of non-marriage to zero, **CONDITION A** implies that men, as a group, are less sensitive to marriage than women. Also, note that **CONDITION A** is a cardinal property, ie, whether it holds depends on the choice or cardinalization of the utility function representing preferences. Given underlying preference orderings over private and public consumption goods in marriage, it has to hold for a well-chosen cardinalization of utility such that the Pareto frontier is a hyperplane as in (1).

As an illustration of **CONDITION A**, consider the following example.

Example 1

An economy consists of three men and three women with the following preferences[4] Suppose the utilities from different assignments are given by

Table 1

	Woman 1	Woman 2	Woman 3
Man 1	(1,5)	(0,10)	(0.5,1)
Man 2	(0.5,10)	(1,5)	(0,0.5)
Man 3	(0.5,2)	(1,2)	(0,0)

In Table 1 above, the entry in the cell with index (i, j) is the ordered pair (U_i^j, U_j^i) where U_i^j denotes the utility of man i if he were to marry woman j and U_j^i denotes the utility of woman j if she were to marry man i.

It is easy to check that **CONDITION A** holds in the example above (see Appendix B).

As stated earlier, **CONDITION A** requires that the loss in total utility to women by moving from the assignment that is first-best for women to the assignment that is second-best for women exceed the total utility (to men) of the first-best assignment for men. Thus, it seems intuitive that this requirement implies that

[4]The preference ordering in this example is a slight alteration of Example 1 in Abdulkadiroğlu & Sönmez (2003), pg. 736. Cardinal utility values consistent with the ordering have been added by us.

the assignment that is first-best for women also maximizes the total utility of all individuals (of both sexes). This is formally demonstrated in the lemma below.

Lemma 1

*If **CONDITION A** holds,* $A^* \in A^*_{\mathcal{W}} \implies A^* \in \underset{A \in \mathcal{A}}{\arg\max} \left[\sum_{m \in \mathcal{M}} \widetilde{U}_m(A) + \sum_{w \in \mathcal{W}} \widetilde{U}_w(A) \right]$

Proof: See Appendix C.

Proposition 1

If the top-trading cycles algorithm produces an assignment $A^* \in A^*_{\mathcal{W}}$ *and **CONDITION A** holds, then the assignment under top-trading cycles coincides with an equilibrium **BAMM** assignment.*

Proof: See Appendix D.

As an illustration of the proposition above, note that in Example 1 the *man-pointing, woman-choosing top trading cycles* algorithm converges to the following assignment.

$$M_1 \to W_2, \quad M_2 \to W_1, \quad M_3 \to W_3 \tag{2}$$

By part 1 of Proposition 1, (2) is also the equilibrium assignment under **BAMM**.

The *woman-proposing* variant of the Gale-Shapley algorithm converges to the unique stable assignment, which is the following:

$$M_1 \to W_1, \quad M_2 \to W_2, \quad M_3 \to W_3 \tag{3}$$

Notice that assignments (2) and (3) are distinct. Thus, A^* is not stable in general.

Note that Proposition 1 requires that the top trading cycles assignment coincide with an element in $A^*_{\mathcal{W}}$. While that may be the case, it is not guaranteed to happen. In particular, the top trading cycles algorithm can converge to an assignment that is Pareto optimal but does not belong to $A^*_{\mathcal{W}}$. As an illustration of this consider the following simple example.

Example 2

Suppose there are two individuals of each gender in the economy and their utilities from alternative assignments are as shown in Table 2 below:

Table 2

	Woman 1	Woman 2
Man 1	(0,10)	(1,5)
Man 2	(0,0)	(1,0)

If we run the top trading cycles, in the first step both women point at man 1. Both men point at woman 2. Man 1 and woman 2 form the only cycle, so they match. In the next step man 2 matches with woman 1. Thus the top trading cycles algorithm produces the following assignment.

$$W_1 \to M_2, \quad W_2 \to M_1 \tag{4}$$

Notice that assignment (4) yields a total utility of 6, which is lower than the total utility of 11 yielded by assignment (5) below

$$W_1 \to M_1, \quad W_2 \to M_2 \tag{5}$$

Hence, assignment (5) is the unique BAMM assignment in this economy, and the top trading cycles algorithm reaches a different assignment.

Since our interest lies in implementing the BAMM assignment, that leaves open the issue as to whether there are conditions on preferences under which the top trading cycles algorithm or some variant thereof can implement the BAMM assignment.

Example 2 above illustrates why the top trading cycles algorithm may fail to converge to the **BAMM** assignment. If the preferences of men are such that they prefer women who lose lower amounts of utility when they are matched with a lower ranked partner as opposed to women who lose higher amounts of utility when forced to make the corresponding change in respect to their partner, the assignment resulting from the top trading cycles is different from the **BAMM** assignment. In order to ensure that the top trading cycles implements the BAMM assignment, we need to make more assumptions on preferences. To that end, we first introduce a definition.

Definition:

Woman i is more sensitive than woman i' if and only if the following holds:
$\widetilde{\mathcal{U}}_i^{(j+1)} - \widetilde{\mathcal{U}}_i^{(j)} > N.\big(\widetilde{\mathcal{U}}_{i'}^{(N)} - \widetilde{\mathcal{U}}_{i'}^{(1)}\big) \quad \forall j \in \{1, 2, .., N-1\}$
where $\widetilde{\mathcal{U}}_i := \{u \in \mathbb{R}_+ | \exists A \in \mathcal{A} \text{ s.t.} u = \widetilde{U}_i(A)\}$ and $\widetilde{\mathcal{U}}_i^{(j)}$ denotes the j-th order statistic of $\widetilde{\mathcal{U}}_i$.

In words, if a more "sensitive" woman were to be matched with a partner one rank below rather than with a partner of the rank (according to her ordering) under consideration, she would lose more utility than all "less" sensitive woman could gain by switching from her worst partner to her best partner.

We assume that women can be ranked by order of sensitivity. We state this formally in **Assumption 1**[5] below.

Assumption 1: Sensitivity

The following statements hold:

1. $\min \widetilde{\mathcal{U}}_i = C, \quad C \in \mathbb{R}_+ \quad \forall i \in \mathcal{W}$

2. i *is more sensitive than* $i+1$ $\forall i \in \{1, ..., N-1\},\ i \in \mathcal{W}$

[5] These are not the weakest possible assumptions that implement the utilitarian efficient assignment, but weaker assumptions are more complicated.

Assumption 2: Hierarchy

1. There exist a finite number of groups labeled $1, 2, .., K$, where $K \leq N$, ranked hierarchically; group 1 being the highest and group K being the lowest. Formally, let the family of sets $P_1, P_2, ..., P_K$ be a partition of the set of all agents in the game, i.e., $\mathcal{M} \cup \mathcal{W}$.

2. There are an equal number of men and women in each set P_k, $k = 1, 2, ..., K$.

3. Given $w_1, w_2 \in \mathcal{W}$, if w_1 is more sensitive than w_2, w_1 is in the same level as w_2 or at a higher level than w_2.

4. For men and women in a group k, $k < K$, the following holds: For each woman(man), there is a distinct man(woman) in her(his) level whom she(he) strictly prefers to all other men(women) in her(his) level or below her(his) level. For men and women in a group $k > 1$, the following holds: Each woman(man) strictly prefers any man(woman) above her(his) level to any man(woman) in her(his) level.

One can think of at least two real-world scenarios in which it is plausible that Assumption 2 holds. The first is a school assignment context where a school may have a priority for students who live in the attendance area of the school, or has siblings attending the same school (see Abdulkadiroğlu & Sönmez (2003)). The second example, and the one that is more closely related to the current context, is the Indian marriage market, where there are quite a few castes ranked hierarchically (see Anderson (2003)). Interestingly, Anderson (2003) uses a quality-of-groom (as perceived by the bride) function which is such that a bride prefers grooms of a higher caste to those of a lower caste. Such a quality-of-groom function is consistent with Assumption 2.

We are now in a position to state the central proposition in this paper.

Proposition 2

If Condition A, Assumption 1 and Assumption 2 hold, the top trading cycles algorithm produces the BAMM assignment.

Proof: See Appendix E.

As an illustration of Proposition 2, consider the following example[6]:

Example 3

There are three women and three men with preference orderings given below.

$m_1 : w_1 \succ w_3 \succ w_2$

$m_2 : w_2 \succ w_1 \succ w_3$

$m_3 : w_2 \succ w_1 \succ w_3$

$w_1 : m_2 \succ m_1 \succ m_3$

$w_2 : m_1 \succ m_2 \succ m_3$

$w_3 : m_1 \succ m_2 \succ m_3$

There are two levels in society, ie $K = 2$. Level 1 consists of $\{m_1, m_2, w_1, w_2\}$ and level two consists of $\{m_3, w_3\}$. Notice that this preference ordering satisfies Assumption 2 if we further assume w_1 is more sensitive than w_2, who is more sensitive than w_3. To see this, observe that the most preferred woman for m_1 and m_2 are both from level 1. The same holds for w_1 and w_2. Further, w_3 is m_3's worst choice. Similarly, m_3 is w_3's worst choice.

The top trading cycles algorithm on this particular preference ordering proceeds as follows: At Step 1, there is exactly one cycle, which is the following: (w_1, m_2, w_2, m_1). Notice that this cycle is nested within level 1. Further, all members from level 2, ie m_3 and w_3 point to some member in level 1, but neither m_3 nor w_3 is part of any cycle. At Step 1, w_1 is matched with m_2 and w_2 is matched with m_1.

[6] The example is adapted from Abdulkadiroğlu & Sönmez (2003), Example 1, Pg. 736

At Step 2 of the algorithm the only cycle is (m_3, w_3). Thus, m_3 and w_3 are paired at Step 2 of the algorithm, and the algorithm terminates.

While the matching produced by the top trading cycles algorithm is woman Pareto optimal, it is not stable. For example, m_1 prefers w_3 over his current match and w_3 prefers m_1 over her current match. Two aspects of the matching produced by the top trading cycles algorithm are worth emphasizing. First, all matches are nested within levels. This is consistent with caste endogamy observed in the Indian marriage market. Second, given the assignment produced by the top trading cycles, the profitable bilateral deviation is between the two levels, not within a given level. This is a result that holds generally. The proposition below states this formally.

Proposition 3

If Assumption 2 holds, all bilaterally profitable deviations from the matching produced by the top trading cycles are across levels.

Proof: See Appendix F.

The fact that bilaterally profitable deviations are across levels has the following interpretation in the Indian marriage context: If the marriage matching process in society produces a utilitarian efficient matching, individuals may have an incentive to deviate from the efficient matching. To prevent those one would need strict social norms, for example, brutal punishments to couples who bilaterally deviate. However, these punishments would not be necessary if the matching produced were stable. Thus, the existence of costly-to-implement social sanctions against inter-caste marriages is consistent with our framework, but cannot be rationalized if the marriage matching in the Indian market were to be thought of as being produced by the Gale-Shapley algorithm.

We now provide a partial converse to Proposition 1. To that end, define **CONDITION B** as follows:

CONDITION B: $U^*_{\mathcal{W}} - \sum_{w\in\mathcal{W}} \widetilde{U}_w(A^*_{\mathcal{M}}) > U^*_{\mathcal{M}}$, $A^*_{\mathcal{W}}$ is a singleton, and $A^*_{\mathcal{M}} \cap A^*_{\mathcal{W}} = \emptyset$.

In words, **CONDITION B** requires that the total loss in utility to women by moving from the assignment that is first-best for women to the assignment that is first-best for men exceed the total utility to all men from the assignment that is first-best for men. If we assume $A^*_{\mathcal{M}} \cap A^*_{\mathcal{W}} = \emptyset$, CONDITION A $\implies$ CONDITION B To see why this is true, assume that $A^*_{\mathcal{M}} \cap A^*_{\mathcal{W}} = \emptyset$ and CONDITION A holds. So, $U^{*,(-1)}_{\mathcal{W}} \geq \sum_{w\in\mathcal{W}} \widetilde{U}_w(A^*_{\mathcal{M}})$. Hence,
$U^*_{\mathcal{W}} - U^{*,(-1)}_{\mathcal{W}} > U^*_{\mathcal{M}} \implies U^*_{\mathcal{W}} - \sum_{w\in\mathcal{W}} \widetilde{U}_w(A^*_{\mathcal{M}}) > U^*_{\mathcal{M}}$
Hence, $A^*_{\mathcal{M}} \cap A^*_{\mathcal{W}} = \emptyset$ and **CONDITION A** $\implies$ **CONDITION B**. Before we introduce the next proposition, we need to develop some notation. Denote by A^*_{TTC} the assignment that the *top trading cycles algorithm* produces.

Proposition 4

*If $\mathcal{A}^*_{BAMM} = A^*_{TTC} = A^*_{\mathcal{W}}$, $A^*_{\mathcal{M}} \cap \mathcal{A}^*_{BAMM} = \emptyset$ and $\mathcal{A}^*_{BAMM}$ is a singleton, then **CONDITION B** holds.*

Proof: See Appendix G

Next, we illustrate through an example that Assumption 2 is necessary for implementing the utilitarian efficient matching through the top trading cycles algorithm. Consider the example below:

Example 4

Table 3

	Woman 1	Woman 2	Woman 3
Man 1	(5,500)	(0.1,25)	(1,6)
Man 2	(1,1000)	(5,50)	(0.1,5)
Man 3	(1,0.1)	(5,0.1)	(0.1,0.1)

In Table 3 above, the entry in the cell with index (i, j) is the ordered pair (U_i^j, U_j^i) where U_i^j denotes the utility of man i if he were to marry woman j and U_j^i denotes the utility of woman j if she were to marry man i. Note that in the example above, the preference ordering satisfies CONDITION A and Assumption 1, but fails to satisfy Assumption 2 (see Appendix H for details). Further, as we show in Appendix H, the top trading cycles algorithm produces the following assignment:

$$W_1 \to M_1, \quad W_2 \to M_2, \quad W_3 \to M_3$$

The assignment above results in a total utility of 560.2 to all agents, which is lower than 1026.3 produced by the following assignment:

$$W_1 \to M_2, \quad W_2 \to M_1, \quad W_3 \to M_3$$

Hence, in the example above, the top trading cycles algorithm does not result in a utilitarian efficient assignment/matching.

Finally, note that in the "unusual" case where a stable assignment is also Pareto optimal, the equilibrium under **BIM** with the Gale-Shapley algorithm could coincide with the equilibrium with the top-trading cycles algorithm. This equilibrium could be distinct from the equilibrium under **BAMM**. As an illustration, consider the following example[7].

Example 5

Suppose the economy consists of two men and two women whose preferences can be represented by the cardinal utility shown in Table 4 below.

The equilibrium **BAMM** assignment is the following:

$$M_1 \to W_1, \quad M_2 \to W_2 \tag{6}$$

[7]The example is a slight modification of the example in Pollak (2019), pg 23.

Table 4

	Woman 1	Woman 2
Man 1	(11,1)	(2,2)
Man 2	(2,2)	(0,0)

Irrespective of whether one uses the *top trading cycles* algorithm or the Gale-Shapley algorithm, the **BIM** assignment is the following:

$$M_1 \to W_2, \quad M_2 \to W_1 \tag{7}$$

Note that assignments (6) and (7) are distinct.

5 Conclusion

The set of stable marriage matches, and their welfare implications, are different depending on whether allocation within marriage is determined by binding agreements in the marriage market (BAMM) or by bargaining in marriage (BIM) with no commitment. With transferable utility, any stable matching is utilitarian efficient under BAMM. This, however, does not hold under BIM, which appears to be a more (empirically) plausible assumption than BAMM. In this paper we showed that it is possible to implement the utilitarian efficient matching even in a BIM setting. If agents on one side of the market are sufficiently sensitive to matches relative to the other side, the more sensitive side can be ranked by sensitivity, and preferences over members of the opposite sex are hierarchical, the top trading cycles algorithm results in a utilitarian efficient matching.

Given that the assignments produced by using the alternative algorithms of Gale-Shapley and top trading cycles under BIM could be different, it is of great interest to examine the empirical evidence on which algorithm better represents the real world marriage market. After all, these algorithms are not meant to serve as literal descriptions of the matching process, but rather as constructive proofs of the existence of a matching with desirable properties — stability in the case of the Gale-Shapley algorithm and Pareto efficiency in the case of the top trading cycles algorithm. Thus, the choice of one matching algorithm over the other should not be based on the consideration as to whether one provides a better literal description of marriage matching but on whether one algorithm is better able to rationalize the data as compared to the other. However, this question has hardly been addressed in the literature.

There are only a few empirical papers dealing with marriage-matching in a *non-transferable* utility setting that use the Gale-Shapley algorithm. For instance, Hitsch *et al.* (2010) estimate mate preferences from matches observed on a dating site. Then they use the Gale-Shapley algorithm to predict matches on the dating site and do fairly well. They also attempt to use the estimated preferences to predict matches in the real world, again using the Gale-Shapley algorithm. In the real world, the Gale-Shapley algorithm underpredicts assortative matching on several dimensions. Lee (2009) performs a similar exercise using data from an online matchmaking platform in South Korea. In her exercise, she estimates preferences with matchmaker data, and the Gale-Shapley algorithm does a fair job of predicting matches amongst users of online services. Gale-Shapley predictions, however, are somewhat off in terms of predicting matches in the real world. Banerjee *et al.* (2013) use data on matches in the marriage market from India to estimate preferences for partner attributes, most notably for caste of the partner. They use their estimated preferences to simulate matches using the Gale-Shapley algorithm to clear the marriage market. While moments from their simulated data match data fit real world matches on several dimensions, their simulations overpredict (by a substantial margin) caste homogamy relative to that in the data.

In summary, the Gale-Shapley algorithm does not do a stellar job in predicting matches in the real world. Further, while there is sufficient evidence to suggest that BIM, rather than BAMM, is an appropriate framework to model ongoing marriages, there are, to the best of our knowledge, no empirical studies that investigate whether the marriage market equilibrium in the real world is substantially different from the BAMM equilibrium. Moreover, social norms governing marriage and courtship vary widely across the world, and there may exist social norms that violate stability. For example, in Kyrgyzstan, men routinely kidnap women, often without their consent, for marriage (see Kleinbach *et al.* (2005), Handrahan (2004) and Nedoluzhko & Agadjanian (2015)). In India, caste endogamy and clan exogamy are widely prevalent. It is possible that social norms relating to endogamy and exogamy serve to facilitate efficient matches even though these matches may not be stable. Further, the fact that they are often enforced by brutal social punishments to couples that deviate could be on account of the fact that the utilitarian efficient matching is not robust to bilateral deviation. In all these cases, the *top trading cycles* algorithm, which results in an assignment that is Pareto optimal, could be a better predictor of matches in the real world than the *Gale-Shapley* algorithm, which results in an assignment that is stable. In future research, we intend to estimate models using the *top trading cycles* algorithm instead of the Gale-Shapley algorithm, for example, with the data in the studies cited above.

References

Abdulkadiroğlu, Atila, & Sönmez, Tayfun. 2003. School choice: A mechanism design approach. *American Economic Review*, **93**(3), 729–747.

Anderson, Siwan. 2003. Why dowry payments declined with modernization in Europe but are rising in India. *Journal of Political Economy*, **111**(2), 269–310.

Banerjee, Abhijit, Duflo, Esther, Ghatak, Maitreesh, & Lafortune, Jeanne. 2013. Marry for what? Caste and mate selection in modern India. *American Economic Journal: Microeconomics*, **5**(2), 33–72.

Bergstrom, Theodore C. 1989. A fresh look at the rotten kid theorem–and other household mysteries. *Journal of Political Economy*, **97**(5), 1138–1159.

Bergstrom, Theodore C, & Varian, Hal R. 1985. When do market games have transferable utility? *Journal of Economic Theory*, **35**(2), 222–233.

Chiappori, Pierre-André. 2017. *Matching with transfers: The economics of love and marriage*. Princeton University Press.

Chiappori, Pierre-André, Salanié, Bernard, & Weiss, Yoram. 2017. Partner choice, investment in children, and the marital college premium. *American Economic Review*, **107**(8), 2109–67.

Chiappori, Pierre-André, Dias, Monica Costa, & Meghir, Costas. 2018. The Marriage Market, Labor Supply, and Education Choice. *Journal of Political Economy*, **126**(S1), S26–S72.

Gale, David, & Shapley, Lloyd S. 1962. College admissions and the stability of marriage. *The American Mathematical Monthly*, **69**(1), 9–15.

Gayle, George-Levi, & Shephard, Andrew. 2019. Optimal taxation, marriage, home production, and family labor supply. *Econometrica*, **87**(1), 291–326.

Handrahan, Lori. 2004. Hunting for women: bride-kidnapping in Kyrgyzstan. *International Feminist Journal of Politics*, **6**(2), 207–233.

Hitsch, Gunter J, Hortaçsu, Ali, & Ariely, Dan. 2010. Matching and sorting in online dating. *American Economic Review*, **100**(1), 130–63.

Kleinbach, Russell, Ablezova, Mehrigiul, & Aitieva, Medina. 2005. Kidnapping for marriage (ala kachuu) in a Kyrgyz village. *Central Asian Survey*, **24**(2), 191–202.

Koopmans, Tjalling C, & Beckmann, Martin. 1957. Assignment problems and the location of economic activities. *Econometrica*, **25**(1), 53–76.

Lee, SangMok, & Yariv, Leeat. 2018. On the efficiency of stable matchings in large markets. *Unpublished*.

Lee, Soohyung. 2009. Marriage and online mate-search services: Evidence from South Korea. *University of Maryland, Working Paper*.

Lundberg, Shelly, & Pollak, Robert A. 1993. Separate spheres bargaining and the marriage market. *Journal of political Economy*, **101**(6), 988–1010.

Lundberg, Shelly J, Pollak, Robert A, & Wales, Terence J. 1997. Do husbands and wives pool their resources? Evidence from the United Kingdom child benefit. *Journal of Human Resources*, **32**(3), 463–481.

Mas-Colell, Andreu, Whinston, Michael Dennis, Green, Jerry R, *et al.* 1995. *Microeconomic theory*. Vol. 1. Oxford university press New York.

Nedoluzhko, Lesia, & Agadjanian, Victor. 2015. Between tradition and modernity: Marriage dynamics in Kyrgyzstan. *Demography*, **52**(3), 861–882.

Pollak, Robert A. 2019. How Bargaining in Marriage Drives Marriage Market Equilibrium. *Journal of Labor Economics*, **37**(1), 297–321.

Roth, Alvin E, & Sotomayor, Marilda A Oliveira. 1990. *Two-sided matching, Volume 18 of Econometric Society Monographs*.

Shapley, Lloyd S, & Shubik, Martin. 1971. The assignment game I: The core. *International Journal of game theory*, **1**(1), 111–130.

Voena, Alessandra. 2015. Yours, Mine, and Ours: Do Divorce Laws Affect the Intertemporal Behavior of Married Couples? *American Economic Review*, **105**(8), 2295–2332.

Appendices

A Example: Top Trading Cycles is not strategy-proof for men

The example below illustrates that *top trading cycles* algorithm is not strategy proof for men.

Suppose there are three men and three women and that their true preferences are as follows:

$m_1 : w_1 \succ w_3 \succ w_2$

$m_2 : w_2 \succ w_1 \succ w_3$

$m_3 : w_2 \succ w_1 \succ w_3$

$w_1 : m_2 \succ m_1 \succ m_3$

$w_2 : m_1 \succ m_2 \succ m_3$

$w_3 : m_1 \succ m_2 \succ m_3$

If everyone reveals her/his true preference, the *top trading cycles* mechanism converges to the following assignment:

$m_1 \rightarrow w_2, \quad m_2 \rightarrow w_1, \quad m_3 \rightarrow w_3$

Notice that man 1 is matched with the woman ranked lowest according to his preference ordering.

Suppose man 1, instead of revealing his true preferences, reveals the following: $m_1^{false} : w_3 \succ w_1 \succ w_2$.

Suppose further that all other individuals in the economy state their true preferences. In this case the top trading cysles mechanism converges to the following assignment.

$m_1 \rightarrow w_3, \quad m_2 \rightarrow w_2, \quad m_3 \rightarrow w_1$

Notice that man 1 is now matched with the woman ranked second according to his true preference ordering. Thus, truth-telling is not a dominant strategy for man 1.

B Example 1 satisfies CONDITION A

Table 5: Tabulation of sum of Utilities from all Possible Assignments

(m, w) pairs	$\sum_{m\in\mathcal{M}} U_m$	$\sum_{w\in\mathcal{W}} U_w$
(1,1), (2,2), (3,3)	2	10
(1,2), (2,3), (3,1)	0.5	12.5
(1,3), (2,1), (3,2)	2	13
(1,1), (2,3), (3,2)	2	7.5
(1,2), (2,1), (3,3)	0.5	20
(1,3), (2,2), (3,1)	2	8
Maximum	2	20

Here, $U^*_{\mathcal{M}} = 2$, $U^*_{\mathcal{W}} = 20$, $U^{*,(-1)}_{\mathcal{W}} = 13$

$\therefore U^*_{\mathcal{W}} - U^{*,(-1)}_{\mathcal{W}} = 7 > 2 = U^*_{\mathcal{M}}$.

Hence, CONDITION A holds.

C Proof of Lemma 1

Suppose **CONDITION A** holds, $A^* \in A^*_{\mathcal{W}}$ but $A^* \notin \underset{A\in\mathcal{A}}{\arg\max} \left[\sum_{m\in\mathcal{M}} \widetilde{U}_m(A) + \sum_{w\in\mathcal{W}} \widetilde{U}_w(A)\right]$. Then, $\exists\, A' \in \mathcal{A}$, $A' \neq A^*$ such that

$$\left[\sum_{m\in\mathcal{M}} \widetilde{U}_m(A') + \sum_{w\in\mathcal{W}} \widetilde{U}_w(A')\right] > \left[\sum_{m\in\mathcal{M}} \widetilde{U}_m(A^*) + \sum_{w\in\mathcal{W}} \widetilde{U}_w(A^*)\right]$$

$$\implies \left[\sum_{m\in\mathcal{M}} \widetilde{U}_m(A') + \sum_{w\in\mathcal{W}} \widetilde{U}_w(A')\right] > \left[\sum_{m\in\mathcal{M}} \widetilde{U}_m(A^*) + U^*_{\mathcal{W}}\right] \quad [\because A^* \in A^*_{\mathcal{W}}]$$

Rearranging the above inequality and applying **CONDITION A**, we have

$$\left[\sum_{m\in\mathcal{M}} \widetilde{U}_m(A') - \sum_{m\in\mathcal{M}} \widetilde{U}_m(A^*)\right] > \left[U^*_{\mathcal{W}} - \sum_{w\in\mathcal{W}} \widetilde{U}_w(A')\right] \geq U^*_{\mathcal{W}} - U^{*,(-1)}_{\mathcal{W}} > U^*_{\mathcal{M}} \tag{8}$$

$$\text{But} \quad U^*_{\mathcal{M}} \geq \sum_{m\in\mathcal{M}} \widetilde{U}_m(A') \geq \left[\sum_{m\in\mathcal{M}} \widetilde{U}_m(A') - \sum_{m\in\mathcal{M}} \widetilde{U}_m(A^*)\right] \quad \forall A' \in \mathcal{A} \tag{9}$$

From (8) and (9), $U^*_{\mathcal{M}} > U^*_{\mathcal{M}}$ which is a contradiction.■

D Proof of Proposition 1

Suppose the top-trading cycles algorithm produces an element $A^* \in A^*_{\mathcal{W}}$ and **CONDITION A** holds. With transferable utility, the set of equilibrium assignments under **BAMM** is given by $\mathcal{A}^*_{BAMM} = \{A \in \mathcal{A} : A \in \arg\max_{A \in \mathcal{A}} \left[\sum_{m \in \mathcal{M}} \widetilde{U}_m(A) + \sum_{w \in \mathcal{W}} \widetilde{U}_w(A)\right]\}$. Thus, if **CONDITION A** is true, from **Lemma 1** we conclude that $A^* \in \mathcal{A}^*_{BAMM}$. ■

E Proof of Proposition 2

Step 1

Definition: Nested Cycle

A cycle $C = (m_1, w_1, ..., m_n, w_n)$ is nested within level k, $k \leq K$ if and only if for any j, $j = 1, 2, ..., n$, such that $m_j, w_j \in C$, m_j and w_j are both in level k.

Claim: *At Step 1 of the TTC, all cycles are nested within the top level, ie, level 1.*

Proof:

Suppose not. Then there is at least one cycle not nested within level 1. First, notice that all individuals below level 1 are pointing at someone in level 1. So any cycle has to include at least one person from level 1. Suppose such a cycle is not nested within level 1. Then there is at least one man or woman in level 1 who is pointing at a woman or man at level k, $k > 1$. But that implies she or he prefers a partner below her or his level to all partners at her/his level, which violates Assumption 2.■

Step 2

Claim: *Each man and woman in level 1 is part of some cycle at Step 1 of the top trading cycles (TTC).*

Proof:

Suppose there are some women and men in level 1 who are not part of any cycle at Step 1. Note that there is at least one cycle. Since each man and woman has a unique and distinct most preferred mate, no individual who does not belong to any cycle is pointing to any individual who is part of a cycle. Further, there are as many men as women who do not belong to any cycle. Let each man point to his most preferred woman and each woman point to her most preferred man. Let such men and women form the following ordered list: $(m_1^c, w_1^c, ..., m_l^c, w_l^c)$. Then, w_l^c must be pointing back at some man in the ordered list, thus forming a cycle, and contradicting the initial claim that no man or woman in $(m_1^c, w_1^c, ..., m_l^c, w_l^c)$ belongs to a cycle.■

Step 3

Claim: *Each man and woman in level 1 is matched at Step 1 of the TTC.*

Proof:

From Step 2 of this proof, each man and each woman at level 1 is part of some cycle. By construction of the TTC, each woman is matched with the man she points to at *Step 1* of the TTC. Hence, each man and woman in level 1 is matched at *Step 1* of the TTC.■

Step 4

Claim: *At any subsequent Step k of the TTC, all cycles are nested within level k. All individuals at level k are matched in Step k.*

Proof:

By induction on k.

Suppose the statement is true for some $k = m$, $m \leq K - 1$. We will show that the statement is true for $k = m + 1$. Note that by Step $m + 1$ of the TTC, all individuals at or above level m are already matched. (This holds by the induction hypothesis.) By the same argument as in Step 1 of this proof, all cycles at Step $m + 1$ of the TTC are nested within level $m+1$. By the same argument as in Step 2 of this proof, each man and woman at level $m+1$ is part of some cycle at Step $m+1$ of the TTC, and are, therefore, matched at Step $m+1$ of the TTC.■

Step 5

Claim: *The top trading cycles algorithm produces a matching A^* where A^* is given by:*

$A^*(w_1) = m_j$ s.t. $U_{w_1}(m_j) = max\{U_{w_1}(m_1), U_{w_1}(m_2), ..., U_{w_1}(m_N)\}$

For $j = \{2, 3, ..., N\}$, $A^*(w_j) = m_l$

s.t. $U_{w_j}(m_l) = max\Big\{\{U_{w_j}(m_1), U_{w_j}(m_2), ..., U_{w_j}(m_N)\} \setminus \underset{i=1,..,j-1}{\cup} A(w_i)\Big\}$

Proof:

Consider women at level 1, ie $w_1, .., w_{K_1}$. Each woman at level 1 has a unique and distinct most preferred man. Thus, $m_l^* \neq m_{l'}^*$ $\forall$ $l \neq l'$, $w_l, w_{l'} \in$ level 1, where

$m_l^* := argmax\{U_{w_l}(m_1), U_{w_l}(m_2), ..., U_{w_l}(m_N)\}$

Hence, for all women at level 1, A^* satisfies the following property:

$A^*(w_1) = m_j$ s.t. $U_{w_1}(m_j) = max\{U_{w_1}(m_1), U_{w_1}(m_2), ..., U_{w_1}(m_N)\}$, and

for $j = \{2, 3, ..., K_1\}$, $A^*(w_j) = m_l^*$

s.t. $U_{w_j}(m_l^*) = max\Big\{\{U_{w_j}(m_1), U_{w_j}(m_2), ..., U_{w_j}(m_N)\} \setminus \underset{i=1,..,j-1}{\cup} A(w_i)\Big\}$

All men and women at level 1 are matched in Step 1 of the TTC. Thus, when the TTC proceeds to Step 2, the most preferred men of all women at level 2 have already been eliminated at Step 1 of the TTC. Thus, each woman at level 2 has a unique and distinct most preferred man from amongst the set of unmatched men. Hence, the argument in the above paragraph can be applied repeatedly to establish the claim. ■

Step 6

Claim: *A^* is woman-Pareto optimal.*

Proof:

Suppose A^* is not woman-Pareto optimal. Then $\exists$ an assignment $A' \neq A^*$ such that $\widetilde{U}_{w_i}(A') \geq \widetilde{U}_{w_i}(A^*)$ $\forall w_i \in \mathcal{W}$ and $\exists$ $w_{i'} \in \mathcal{W}$ such that $\widetilde{U}_{w_{i'}}(A') > \widetilde{U}_{w_{i'}}(A^*)$. Define $i'_{\min} := \min\Big\{i' \in \{1, 2, ..., N\} | \widetilde{U}_{w_{i'}}(A') > \widetilde{U}_{w_{i'}}(A^*)\Big\}$. By definition, A^* assigns w_1 to her most preferred man. Hence, $i'_{\min} \neq 1$. Further, by construction of A^*, the following holds: If the preference of $w_{i'_{min}}$ clashes with the preference of a woman with a higher index, $w_{i'_{min}}$'s preferences are given priority. Since preferences over men are strict, it follows that if assignment A' matches $w_{i'_{min}}$ with the partner that some woman with a higher index had under assignment A^*, $w_{i'_{min}}$ would be worse off under assignment A' than under assignment A^*. But that cannot be the case since that would violate the definition of i'_{min}. Formally, $\nexists$ i, $i \geq i'_{\min}$, such that $A'(w_{i'_{\min}}) = A^*(w_i)$. So, $w_{i'_{min}}$'s partner under A' must have been the partner of a woman with a lower index under A^*. Formally, $\exists$ $w_i \in \mathcal{W}, i \in \{2, ..., i'_{\min} - 1\}$ such that $A'(w_{i'_{\min}}) = A^*(w_i)$. But then A' must match w_i with a man who, under A^*, was the partner of a woman with an index (weakly) higher than i'_{min}. So it must be the case that $A'(w_i) = A^*(w_{\widetilde{i}})$ where $\widetilde{i} \in \{i'_{\min}, ..., N\}$. However, by construction of A^*, $U_{w_i}(A^*(w_i)) > U_{w_i}(A^*(w_{\widetilde{i}})) = U_{w_i}(A'(w_i)) \implies \widetilde{U}_{w_i}(A') < \widetilde{U}_{w_i}(A^*)$ which contradicts the definition of A'. ■

Step 7

$A \in A^*_{\mathcal{W}} \implies A$ is woman-Pareto optimal

Proof:

Suppose A is not woman Pareto optimal. Then there is an assignment $A' \neq A$ such that $\widetilde{U}_{w_i}(A') \geq \widetilde{U}_{w_i}(A)$ $\forall w_i \in \mathcal{W}$ and $\exists$ $w_{i'} \in \mathcal{W}$ such that $\widetilde{U}_{w_{i'}}(A') > \widetilde{U}_{w_{i'}}(A)$.

Let $\mathcal{W}_b := \{w \in \mathcal{W} | \widetilde{U}_w(A') > \widetilde{U}_w(A)\}$.

$\sum_{w\in\mathcal{W}} \widetilde{U}_w(A') = \sum_{w\in\mathcal{W}_b} \widetilde{U}_w(A') + \sum_{w\in\mathcal{W}\setminus\mathcal{W}_b} \widetilde{U}_w(A') > \sum_{w\in\mathcal{W}} \widetilde{U}_w(A)$.

Hence, $A \notin A^*_{\mathcal{W}}$. ■

Step 8

Claim: *If* Assumption 1 *holds*, A^* *is the unique element in* $A^*_{\mathcal{W}}$.

Proof:

From Step 7 above, it suffices to demonstrate that $\sum_{w\in\mathcal{W}} \widetilde{U}_w(A^*) > \sum_{w\in\mathcal{W}} \widetilde{U}_w(A')$ where $A' \neq A^*$ and A' is an arbitrary woman Pareto optimal assignment.

Suppose $A' \neq A^*$ and A' is an arbitrary woman Pareto optimal assignment. Define the set of losers $L := \{w \in \mathcal{W} | \widetilde{U}_w(A') < \widetilde{U}_w(A^*)\}$ and the set of gainers $G := \{w \in \mathcal{W} | \widetilde{U}_w(A') > \widetilde{U}_w(A^*)\}$

For $I = L, G$, define $I_{min} := \min\Big\{i \in \{1, 2, ..., N\} | w_i \in I\Big\}$ and $I_{max} := \max\Big\{i \in \{1, 2, ..., N\} | w_i \in I\Big\}$. Notice, $L_{min} < G_{min}$. To see why this holds, assume, for the sake of contradiction, that $L_{min} > G_{min}$[8]. Note that $G_{min} \neq 1$, because A^* matches w_1 with her most preferred man. Further, by construction of A^*, the following holds: Under assignment A', any woman $w_i \in G$ would have to be matched with a man, who, under A^*, was partnered with a woman $w_{i'}$, where $i' < i$. In particular, this holds for G_{min}. Thus, $\exists\, i \in \{1, .., G_{min} - 1\}$ such that under A', G_{min} gets i's partner under A^*. But then, woman i has a different partner under A' than under A^*. Since preferences are strict, woman i is not indifferent between A' and A^*. She has a lower index than G_{min}, so $w_i \notin G$. Therefore, $w_i \in L$, which contradicts the supposition that $L_{min} > G_{min}$.

Hence, $\exists$ at least one woman, $w_{L_{min}} \in L$, who is more sensitive (or equivalently, has a lower index) than any woman in G.

Now, $\forall\, w \in G$, $\widetilde{U}_w(A') - \widetilde{U}_w(A^*) \leq \big(\widetilde{\mathcal{U}}_w^{(N)} - \widetilde{\mathcal{U}}_w^{(1)}\big) \leq \big(\widetilde{\mathcal{U}}_{w_{G_{min}}}^{(N)} - \widetilde{\mathcal{U}}_{w_{G_{min}}}^{(1)}\big)$ where the second inequality above follows from observing the fact that any $w \in G$ is weakly less sensitive than $w_{G_{min}}$ and by applying Assumption 1[9].

[8] $L_{min} = G_{min}$ is not a possibility because the same woman cannot be both a loser and a gainer, i.e., she cannot be in both sets L and G.

[9] From Assumption 1 it follows that for any $i < i'$,
$\big(\widetilde{\mathcal{U}}_{w_{i'}}^{(N)} - \widetilde{\mathcal{U}}_{w_{i'}}^{(1)}\big) < \frac{1}{N}\big(\widetilde{\mathcal{U}}_{w_i}^{(j+1)} - \widetilde{\mathcal{U}}_{w_i}^{(j)}\big) < \frac{1}{N}\big(\widetilde{\mathcal{U}}_{w_i}^{(N)} - \widetilde{\mathcal{U}}_{w_i}^{(1)}\big) < \big(\widetilde{\mathcal{U}}_{w_i}^{(N)} - \widetilde{\mathcal{U}}_{w_i}^{(1)}\big)$ where $j \in \{1, 2, ..., N-1\}$

Hence, we have

$$\sum_{w\in G}\left[\widetilde{U}_w(A')-\widetilde{U}_w(A^*)\right]\leq |G|.\big(\widetilde{\mathcal{U}}^{(N)}_{w_{G_{min}}}-\widetilde{\mathcal{U}}^{(1)}_{w_{G_{min}}}\big)<N.\big(\widetilde{\mathcal{U}}^{(N)}_{w_{G_{min}}}-\widetilde{\mathcal{U}}^{(1)}_{w_{G_{min}}}\big) \tag{10}$$

Since $L_{min} < G_{min}$, by Assumption 1,

$$\begin{aligned}&\left[\widetilde{U}_{w_{L_{min}}}(A^*)-\widetilde{U}_{w_{L_{min}}}(A')\right]>N.\big(\widetilde{\mathcal{U}}^{(N)}_{w_{G_{min}}}-\widetilde{\mathcal{U}}^{(1)}_{w_{G_{min}}}\big)\\ \implies &\sum_{w\in L}\left[\widetilde{U}_w(A^*)-\widetilde{U}_w(A')\right]>N.\big(\widetilde{\mathcal{U}}^{(N)}_{w_{G_{min}}}-\widetilde{\mathcal{U}}^{(1)}_{w_{G_{min}}}\big)\end{aligned} \tag{11}$$

Now, $\sum_{w\in\mathcal{W}}\widetilde{U}_w(A^*)-\sum_{w\in\mathcal{W}}\widetilde{U}_w(A')$

$=\sum_{w\in L}\left[\widetilde{U}_w(A^*)-\widetilde{U}_w(A')\right]-\sum_{w\in G}\left[\widetilde{U}_w(A')-\widetilde{U}_w(A^*)\right].$

From (10) and (11), $\sum_{w\in\mathcal{W}}\widetilde{U}_w(A^*)-\sum_{w\in\mathcal{W}}\widetilde{U}_w(A')>0$ ■

Step 9

Claim: *Under* Assumption 1 *and* CONDITION A, $A^* = A^*_{BAMM}$

Proof:

From Step 8 above, A^* is the unique element in $A^*_{\mathcal{W}}$ if Assumption 1 holds. From Proposition 1, if CONDITION A holds, $A^* = A^*_{BAMM}$ ■

F Proof of Proposition 3

For any level k, $k \leq K$, each woman at level k gets her most preferred man within level k. So a woman from level k is not interested in deviating to any man at level k.■

G Proof of Proposition 4

Suppose $\mathcal{A}^*_{BAMM} = A^*_{TTC} = A^*_{\mathcal{W}}$ and $\mathcal{A}^*_{BAMM}$ is a singleton. Then,

$U^*_{\mathcal{W}}+\sum_{m\in\mathcal{M}}\widetilde{U}_m(\mathcal{A}^*_{BAMM})>\sum_{w\in\mathcal{W}}\widetilde{U}_w(A)+\sum_{m\in\mathcal{M}}\widetilde{U}_m(A)\quad\forall A\neq\mathcal{A}^*_{BAMM}.$

In particular, for $A = A^*_{\mathcal{M}}$, we have

$U^*_{\mathcal{W}} + \sum_{m\in\mathcal{M}} \widetilde{U}_m(\mathcal{A}^*_{BAMM}) > \sum_{w\in\mathcal{W}} \widetilde{U}_w(A^*_{\mathcal{M}}) + U^*_{\mathcal{M}}$

$\implies U^*_{\mathcal{W}} - \sum_{w\in\mathcal{W}} \widetilde{U}_w(A^*_{\mathcal{M}}) > U^*_{\mathcal{M}} - \sum_{m\in\mathcal{M}} \widetilde{U}_m(\mathcal{A}^*_{BAMM}) > U^*_{\mathcal{M}}$

$\implies$ **CONDITION B.**$\blacksquare$

H Details relevant to Example 4

H.1 Example 4 satisfies CONDITION A

Table 6: Tabulation of sum of Utilities from all Possible Assignments

(m, w) pairs	$\sum_{m\in\mathcal{M}} U_m$	$\sum_{w\in\mathcal{W}} U_w$
(1,1), (2,2), (3,3)	10.1	550.1
(1,2), (2,3), (3,1)	1.2	30.1
(1,3), (2,1), (3,2)	7	1006.1
(1,1), (2,3), (3,2)	10.1	505.1
(1,2), (2,1), (3,3)	1.2	1025.1
(1,3), (2,2), (3,1)	7	56.1
Maximum	10.1	1025.1

Here, $U^*_{\mathcal{M}} = 10.1$, $U^*_{\mathcal{W}} = 1025.1$, $U^{*,(-1)}_{\mathcal{W}} = 1006.1$

$\therefore U^*_{\mathcal{W}} - U^{*,(-1)}_{\mathcal{W}} = 19 > 10.1 = U^*_{\mathcal{M}}$.

Further, the matching $\{(1,2),(2,1),(3,3)\}$ uniquely maximizes $\sum_{w\in\mathcal{W}} U_w$.

Hence, CONDITION A holds.

H.2 Example 4 satisfies Assumption 1

First, observe that each woman's utility from her worst possible match is 0.1. Next, notice that w_1 is more sensitive than w_2. By moving from her first-best to her second-best match, w_1 loses 500 utils while she loses 499.9 utils by moving from her second-best to her third-best. Both these number are more than 149.7 utils[10]. Note that w_2's gain by moving from her best to her worst partner equals 49.9. Similarly, w_2 is more sensitive than w_3. By moving from her first to her second-best w_2 loses 25

[10] 49.9 X 3 = 149.7

utils while she loses 24.9 utils. Both these numbers are larger than 17.7 utils, which is what w_3 gains by moving from her worst to her best choice[11].

H.3 Example 4 violates Assumption 2

We must show that there is no partition of $\mathcal{M} \cup \mathcal{W}$ such that the conditions of Assumption 2 are satisfied. We can consider all possible partitions of $\mathcal{M} \cup \mathcal{W}$ below.

Case 1

$K = 1$, $P_1 = \mathcal{M} \cup \mathcal{W}$

$\mathcal{M} \cup \mathcal{W}$ does not satisfy Assumption 2. w_1 and w_2 both have m_2 as their stated preference, which is a violation of Assumption 2.

Case 2

$K = 2$. There are two sub-cases of this case.

1. Consider a partition in which w_1 and w_2 are at the same level and w_3 is at a lower level. Both w_1 and w_2 both have m_2 as their stated preference, which is a violation of Assumption 2. To see why, notice that m_2 can either belong to the same level as w_2 or the lower level. In the first case, Assumption 2 is violated because the most preferred man (in the same level) for two women are not distinct. In the second case, Assumption 2 is violated because the most preferred man of both w_1 and w_2 belong to a lower level.

2. Consider a partition in which w_1 is at the highest level and w_2 and w_3 both belong to the lower level. m_2 can belong to the higher or to the lower level. If m_2 belongs to the lower level, m_1 and m_3 belong to a higher level. But w_1 prefers m_2, who is at a lower level over m_1, who is at a higher level. This is a violation of Assumption 2. Alternatively, if m_2 belongs to the higher level, m_1 and m_3 must belong to the lower level. Then, w_3's preference ordering violates

[11] 5.9 X 3 = 17.7

Assumption 2, because she prefers m_1, who is at a lower level, over m_2, who is in a higher level.

Case 3

If $K = 3$, we might have the following sub-cases:

Sub-case 1

$P_1 = \{w_1, m_1\}$, $P_2 = \{w_2, m_2\}$, $P_3 = \{w_3, m_3\}$

Notice, w_1 prefers m_2, who is in a lower level over m_1, who is in a higher level. This is a violation of Assumption 2.

Sub-case 2

$P_1 = \{w_1, m_1\}$, $P_2 = \{w_2, m_3\}$, $P_3 = \{w_3, m_2\}$

Notice, w_1 prefers m_2, who is in a lower level over m_1, who is in a higher level. This is a violation of Assumption 2.

Sub-case 3

$P_1 = \{w_1, m_2\}$, $P_2 = \{w_2, m_1\}$, $P_3 = \{w_3, m_3\}$

Notice, m_1 prefers w_3, in level 3, over w_2, who is in level 2, thereby violating Assumption 2.

Sub-case 4

$P_1 = \{w_1, m_3\}$, $P_2 = \{w_2, m_1\}$, $P_3 = \{w_3, m_2\}$

Notice, w_1 prefers m_2 in level 2 over m_3 in level 1, thereby violating Assumption 2.

Sub-case 5

$P_1 = \{w_1, m_2\}$, $P_2 = \{w_2, m_3\}$, $P_3 = \{w_3, m_1\}$

Notice, w_3 prefers m_1 in level 3 over m_3 in level 1, thus violating Assumption 2.

Sub-case 6

$P_1 = \{w_1, m_3\}$, $P_2 = \{w_2, m_2\}$, $P_3 = \{w_3, m_1\}$

Notice, w_1 prefers m_2 in level 2 over m_3 in level 1, thus violating Assumption 2.

H.4 Top Trading Cycles (TTC) on Example 4

In Step 1 of the TTC, M_2 and W_2 are the only two agents that are part of a cycle. They are matched in Step 1. The algorithm proceeds to Step 2. M_1 and W_1 are the only two agents that are part of a cycle. They are matched in Step 2. In Step 3, M_3 and W_3 point to one another, and are matched. Thus, the TTC algorithm produces the following assignment:

$$W_1 \to M_1, \quad W_2 \to M_2, \quad W_3 \to M_3$$