

**FORTHCOMING in Games and Economic Behavior.**

# **Ordinal and cardinal solution concepts for two-sided matching<sup>1</sup>**

**Federico Echenique<sup>2</sup>**  
California Institute of Technology

**Alfred Galichon<sup>3</sup>**  
Sciences Po, Paris

June 30, 2015

<sup>1</sup>The authors wish to thank Juan Pereyra Barreiro and Aditya Kuvalekar for many useful suggestions, and for pointing out a mistake in one of the examples in a previous version of the paper.

<sup>2</sup>Address: Division of the Humanities and Social Sciences, Mail Code 228-77, Caltech, Pasadena, CA 91125, USA. E-mail: fede@caltech.edu.

<sup>3</sup>Address: Department of Economics, Sciences Po, 75007 Paris, France. E-mail: alfred.galichon@sciences-po.fr. This research has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement n°313699.

## **Abstract**

We characterize solutions for two-sided matching, both in the transferable- and in the nontransferable-utility frameworks, using a cardinal formulation. Our approach makes the comparison of the matching models with and without transfers particularly transparent. We introduce the concept of a no-trade stable matching to study the role of transfers in matching. A no-trade stable matching is one in which the availability of transfers do not affect the outcome.

# 1 Introduction

We explore the role of transfers and cardinal utility in matching markets. Economists regularly use one- and two-sided models, with and without transfers. For example auctions allow for monetary transfers among the agents, while models of marriage, organ donation and “housing” exchanges do not. There are two-sided matching models of the labor market without transfers, such as the market for medical interns in the US; and traditional models of the labor market where salaries, and therefore transfers, are allowed. We seek to understand how and why transfers matter in markets for discrete goods.

The question is interesting to us as theorists, but it also matters greatly for one of the most important applications of matching markets, namely the medical interns market. In the market for medical interns in the US (see Roth (1984a), Roth and Sotomayor (1990), and Roth (2002)), hospitals match with applicants using a centralized clearinghouse that implements a stable matching. We always think of this market as one *without* transfers, because salaries are fixed first, before the matching is established. So at the stage in which the parties “bargain” over who is to be matched to whom, salaries are already fixed, and thus there are no transfers.

There is a priori no reason for things to be this way. Hospitals and interns could instead bargain over salaries and employment at the same time. This is arguably the normal state of affairs in most other labor markets; and it has been specifically advocated for the medical interns market in the US (see Crawford (2008)). It is therefore important to understand the impact of disallowing transfers in a matching market. Our paper is a first step towards understanding this problem.

In a two-sided matching market—for our purposes, in the Gale-Shapley marriage market—this impact is important. We consider two canonical models: the marriage market without transfers (the NTU model) and the marriage market with transfers (the TU model, also called the assignment game).

There are Pareto efficient, and even stable, matchings in the NTU model that a utilitarian social planner would never choose, regardless of how she weights agents’ utilities. Implicitly, a utilitarian social planner has access to transfers. Our results motivate an investigation into the distance between the utilitarian welfare in the presence of transfers, and the utilitarian welfare in the absence of transfers. We show that this gap can be arbitrarily large. In fact, it can grow exponentially with the size of the market. Our result should be contrasted with Lee and Yariv (2014), who show that the gap

between utilitarian welfare and stability disappears in large markets. Lee and Yariv impose certain regularity conditions on preferences; or result is a “worst case” analysis, showing that there are markets for which the difference between efficiency and stability can be arbitrarily large. Of course, our result is established in an environment in which utilities are bounded, and the bound is constant while the market grows (otherwise the exercise would be meaningless). See our Proposition 16.

From the viewpoint of the recent literature in computer science on the “price of anarchy” (see e.g. Roughgarden (2005)), Proposition 16 says that the “price of no transfers” can be arbitrarily bad, and grow super-exponentially with the size of the market. In that sense, our result avoids the main critique that economists often levy on the price of anarchy literature. Many papers on the price of anarchy compare equilibria with the utilitarian efficient outcomes, in models in which there are no transfers allowed. Our paper deals with a model in which transfer play a role: namely the assignment game, or the TU model of matching. The meaning then of our price of transfer result is that the loss in utilitarian welfare from banning transfers (and thus moving from the assignment game to the NTU matching model) is large and grows superexponentially with the size of the market. See Section 4.2 for details.

We present results characterizing Pareto efficiency and the role of transfers in marriage models. Ex-ante Pareto optimality in the model with transfers is characterized by the maximization of the weighted utilitarian sum of utilities, while Pareto optimality when there are no transfers is equivalent to a different maximization problem, one where the weighted sum of “adjusted” utilities are employed. Each of these problems, in turn, have a formulation as a system of linear inequalities. The results follow (perhaps unexpectedly) from Afriat’s theorem in the theory of revealed preference.

In order to explore the role of transfers, we study a special kind of stable matching: A *no-trade stable matching* in a marriage market is a matching that is not affected by the presence of transfers. This is the central notion in our paper. Agents are happy remaining matched as specified by the matching, even if transfers are available, and *even though they do not make use of transfers*. Transfers are available, but they are not needed to support the stable matching. There is thus a clear sense in which transfers play no role in a no-trade stable matching.

The notion of no-trade stable matching is useful for the following reason. We can think of transfers as making some agents better off at the expense of others. It is then possible to modify a market by choosing a cardinal util-

ity representation of agents preferences with the property that the matching remains stable with and without transfers (Theorem 12). Under certain circumstances, namely when the stable matchings are “isolated,” we can choose a cardinal representation that will work in this way for every stable matching. So under such a cardinal representation of preferences, any stable matching remains stable regardless of the presence of transfers. Finally (Example 14), we cannot replicate the role of transfers by re-weighting agents’ utilities. In general, to instate a no-trade stable matching, we need the full freedom of choosing alternative cardinal representations.

It is easy to generate examples of stable matchings that cannot be sustained when transfers are allowed, and of stable matchings that can be sustained with transfers (in the sense of being utilitarian-efficient), but where transfers are actually used to sustain stability. We present conditions under which a market has a cardinal utility representation for which stable matchings are no trade matchings.

In sum, the notion of a no-trade stable matching captures both TU and NTU stability: a no-trade stable matching is also a TU and NTU stable matching. TU stability is, on the other hand, strictly stronger than ex-ante Pareto efficiency, which is strictly stronger than ex-post Pareto efficiency. NTU stability is strictly stronger than ex-post Pareto efficiency.<sup>1</sup>

The model without transfers was introduced by Gale and Shapley (1962). The model with transfers is due to Shapley and Shubik (1971). Kelso and Crawford (1982) extended the models further, and in some sense Kelso and Crawford’s is the first paper to investigate the effects of adding transfers to the Gale-Shapley marriage model. Roth (1984b) and Hatfield and Milgrom (2005) extended the model to allow for more complicated contracts, not only transfers (see Hatfield and Kojima (2010) and Echenique (2012) for a discussion of the added generality of contracts). We are apparently the first to consider the effect of transfers on a given market, with specified cardinal utilities, and the first to study the notion of a no-trade stable matching.

---

<sup>1</sup>TU and NTU stability are not comparable in this sense. Empirically, though, they are comparable, with TU stability having strictly more testable implications than NTU stability (Echenique, Lee, Shum, and Yenmez, 2013).

## 2 The Marriage Problem

### 2.1 The model

Let  $M$  and  $W$  be finite and disjoint sets of, respectively, men and women, which are assumed to be in equal number;  $M \cup W$  comprise the *agents* in our model. We can formalize the marriage “market” of  $M$  and  $W$  in two ways, depending on whether we assume that agents preferences have cardinal content, or that they are purely ordinal. For our results, it will be crucial to keep in mind the difference between the two frameworks.

An *ordinal marriage market* is a tuple  $(M, W, P)$ , where  $P$  is a *preference profile*: a list of preferences  $>_i$  for every man  $i$  and  $>_j$  for every woman  $j$ . Each  $>_i$  is a linear order over  $W$ , and each  $>_j$  is a linear order over  $M$ . Here, agents always prefer being matched with anyone rather than being unmatched. The weak order associated with  $>_s$  is denoted by  $\geq_s$  for any  $s \in M \cup W$ .<sup>2</sup>

We often specify a preference profile by describing instead utility functions for all the agents. A *cardinal marriage market* is a tuple  $(M, W, U, V)$ , where  $U$  and  $V$  define the agents’ utility functions:  $U(i, j)$  (resp.  $V(i, j)$ ) is the amount utility derived by man  $i$  (resp. woman  $j$ ) out of his match with woman  $j$  (resp. man  $i$ ). The utility functions  $U$  and  $V$  *represent*  $P$  if, for any  $i$  and  $i'$  in  $M$ , and  $j$  and  $j'$  in  $W$ ,

$$\begin{aligned} U(i, j) > U(i, j') &\iff j >_i j', \text{ and} \\ V(i, j) > V(i', j) &\iff i >_j i'. \end{aligned}$$

We say that  $U$  and  $V$  are a *cardinal representation of*  $P$ . Clearly, for any cardinal marriage market  $(M, W, U, V)$  there is a corresponding ordinal market.

A one-to-one function  $\sigma : M \rightarrow W$  is called a *matching*. When  $w = \sigma(m)$  we say that  $m$  and  $w$  are matched, or married, under  $\sigma$ . In our setting, under a given matching, each man or woman is married to one and only one partner of the opposite sex. We shall denote by  $\mathcal{A}$  the set of matchings. We shall assume that  $M$  and  $W$  have the same number of elements, so that  $\mathcal{A}$  is non-empty.

In our definition of matching is that agents are always married: we do not allow for the possibility of singles.

---

<sup>2</sup>A linear order is a binary relation that is complete, transitive and antisymmetric. The weak order  $\geq_s$  is defined as  $a \geq_s b$  if  $a = b$  or if  $a >_s b$ .

Under our assumptions, we can write  $M = \{m_1, \dots, m_n\}$  and  $W = \{w_1, \dots, w_n\}$ . For notational convenience, we often identify  $m_i$  and  $w_j$  with the numbers  $i$  and  $j$ , respectively. So when we write  $j = \sigma(i)$  we mean that woman  $w_j$  and man  $m_i$  are matched under  $\sigma$ . This usage is a bit different than what it standard notation in matching theory, but it makes the exposition of our results a lot simpler.

We shall often fix an arbitrary matching, and without loss of generality let this matching be the *identity matching*, denoted by  $\sigma_0$ . That is,

$$\sigma_0(i) = i.$$

For a matching  $\sigma$ , let  $u_\sigma(i) = U(i, \sigma(i))$  and  $v_\sigma(j) = V(\sigma^{-1}(j), j)$ . When  $\sigma = \sigma_0$ , we shall often omit it as a subscript and just use the notation  $u$  and  $v$ .

One final concept relates to random matchings. We consider the possibility that matching is chosen according to a lottery: a *fractional matching* is a matrix  $\pi = (\pi_{i,j})$  such that  $\pi_{ij} \geq 0$  and letting  $\pi_{ij}$  the probability that individuals  $i$  and  $j$  get matched, the constraints on  $\pi$  are

$$1 = \sum_{i'=1}^n \pi_{i'j} = \sum_{j'=1}^n \pi_{ij'} \quad \forall i, j \in \{1, \dots, n\}$$

(i.e.  $\pi$  is a *bistochastic matrix*). It is a celebrated result (the Birkhoff von-Neumann Theorem) that such matrices result from a lottery over matchings. Let  $\mathcal{B}$  denote the set of all fractional matchings.

## 2.2 Solution concepts

We describe here some commonly used solution concepts. The first solutions capture the notion of Pareto efficiency. In second place, we turn to notion of core stability for matching markets. For simplicity of exposition, we write these definitions for the specific matching  $\sigma_0$ . Of course by relabeling we can express the same definitions for an arbitrary matching.

A solution concept singles out certain matchings as immune to certain alternative outcomes that could be better for the agents. If we view such alternatives as arising ex-post, after any uncertainty over which matching arises has been resolved, then we obtain a different solution concept than if we view the alternatives in an ex-ante sense.

### 2.2.1 NTU Pareto efficiency

Matching  $\sigma_0(i) = i$  is *ex-post NTU Pareto efficient*, or simply *ex-post Pareto efficient* if there is no matching  $\sigma$  that is at least as good as  $\sigma_0$  for all agents, and strictly better for some agents. That is, such that the inequalities  $U(i, \sigma(i)) \geq U(i, i)$  and  $V(\sigma^{-1}(j), j) \geq V(j, j)$  cannot simultaneously hold with at least one strict inequality.

In considering alternative matchings, it is easy to see that one can restrict oneself to cycles. The resulting formulation of efficiency is very useful, as it allows us to relate efficiency with standard notions in the literature on revealed preference.

Hence matching  $\sigma_0(i) = i$  is ex-post Pareto efficient if and only if for every cycle<sup>3</sup>  $i_1, \dots, i_{p+1} = i_1$ , inequalities  $U(i_k, i_{k+1}) \geq U(i_k, i_k)$  and  $V(i_k, i_{k+1}) \geq V(i_k, i_k)$  cannot hold simultaneously unless they all are equalities. In other words:

**Observation 1** *Matching  $\sigma_0(i) = i$  is ex-post Pareto efficient if for every cycle  $i_1, \dots, i_{p+1} = i_1$ , and for all  $k$ , inequalities*

$$U(i_k, i_{k+1}) \geq U(i_k, i_k) \text{ and } V(i_k, i_{k+1}) \geq V(i_k, i_k),$$

*cannot hold simultaneously unless they are all equalities.*

In an ex-ante setting, we can think of probabilistic alternatives to  $\sigma_0$ . As a result, we obtain the notion of ex-ante Pareto efficiency. To define this notion, we require not only that there is no other matching which is preferred by every individual, but also that there is no lottery over matchings that would be preferred.

Formally: Matching  $\sigma_0(i) = i$  is *ex-ante NTU Pareto efficient* or simply *ex-ante Pareto efficient*, if for any  $\pi \in \mathcal{B}$ , and for all  $i$  and  $j$ , inequalities

$$\sum_j \pi_{ij} U(i, j) \geq U(i, i) \text{ and } \sum_i \pi_{ij} V(i, j) \geq V(j, j)$$

cannot hold simultaneously unless they are all equalities.

Note that the problem of ex-post efficiency is purely ordinal, as ex-post efficiency of some outcome only depends on the rank order preferences, not

---

<sup>3</sup>A cycle  $i_1, \dots, i_p, i_{p+1} = i_1$  is a permutation  $\sigma$  such that  $\sigma(i_1) = i_2$ ,  $\sigma(i_2) = i_3, \dots, \sigma(i_{p-1}) = i_p$ , and  $\sigma(i_p) = i_1$ .



on the particular cardinal representation of it. In contrast, the problem of ex-ante efficiency is cardinal, as we are adding and comparing utility levels across states of the world. The concept of NTU Pareto efficiency is interesting for example in the context of school choice problems, where the assignment is often not deterministic, and transfers are not permitted. In this context, this concept provides an adequate assessment of welfare.

### 2.2.2 TU Pareto efficiency

We now assume that utility is transferable across individuals. In this case, a matching is Pareto efficient if no other matching produces a higher welfare, accounted for as the sum of individual cardinal utilities. It is a direct consequence of the Birkhoff-von Neumann theorem that if a fractional matching produces a higher welfare, then some deterministic matching also produces a higher welfare. As a result, the notions of ex-ante and ex-post TU Pareto efficiency coincide, and we do not need to distinguish between them.

Matching  $\sigma_0(i) = i$  is *TU Pareto efficient* if there is no matching  $\sigma$  for which

$$\sum_{i=1}^n U(i, \sigma(i)) + V(i, \sigma(i)) > \sum_{i=1}^n (U(i, i) + V(i, i)).$$

**Observation 2** *Matching  $\sigma_0(i) = i$  is TU Pareto efficient if for every cycle  $i_1, \dots, i_{p+1} = i_1$ , and for all  $k$ , inequalities*

$$\sum_{k=1}^p U(i_k, i_{k+1}) + V(i_k, i_{k+1}) \geq \sum_{k=1}^p U(i_k, i_k) + V(i_{k+1}, i_{k+1})$$

*cannot hold simultaneously unless they are all equalities.*

In the previous definitions, transfers are allowed across any individuals. One may have considered the possibility of transfers only between matched individuals. It is however well known since Shapley and Shubik (1971) that this apparently more restrictive setting leads in fact to the same notion of efficiency. As we recall below, TU Pareto efficiency is equivalent to TU stability, so to avoid confusions, we shall systematically refer to “TU stability” instead of “TU Pareto efficiency,” and, in the sequel, reserve the notion of efficiency for NTU efficiency.

### 2.2.3 NTU Stability

We now review notions of stability. Instead of focusing on the existence of a matching which would be an improvement for everyone (as in Pareto efficiency), we focus on matchings which would be an improvement for a newly matched pair of man and woman. Thus we obtain two solution concepts, depending on whether we allow for transferable utility.

Our definitions are classical and trace back to Gale and Shapley (1962) and Shapley and Shubik (1971). See Roth and Sotomayor (1990) for an exposition of the relevant theory.

Matching  $\sigma_0(i) = i$  is *stable in the nontransferable utility matching market*, or *NTU stable* if there is no “blocking pair”  $(i, j)$ , i.e. a pair  $(i, j)$  such that  $U(i, j) > U(i, i)$  and  $V(i, j) > V(j, j)$  simultaneously hold.

Hence, using our assumptions on utility, we obtain the following:

**Definition 3** *Matching  $\sigma_0(i) = i$  is NTU stable if*

$$\forall i, j : \min(U(i, j) - U(i, i), V(i, j) - V(j, j)) \leq 0.$$

Of course, this notion is an ordinal notion and should not depend on the cardinal representation of men and women’s preferences, only on the underlying ordinal matching market.

### 2.2.4 TU Stability

Utility is transferable across pair  $(i, j)$  if there is the possibility of a utility transfer  $t$  (of either sign) from  $j$  to  $i$  such that the utility of  $i$  becomes  $U(i, j) + t$ , and utility of  $j$  becomes  $V(i, j) - t$ . When we assume that utility is transferable, in contrast, we must allow blocking pairs to use transfers. Then a couple  $(i, j)$  can share, using transfers, the “surplus”  $U(i, j) + V(i, j)$ . Thus we obtain the definition:

**Definition 4** *Matching  $\sigma_0(i) = i$  is TU stable, if there are vectors  $\tilde{u}(i)$  and  $\tilde{v}(j)$  such that for each  $i$  and  $j$ ,*

$$\tilde{u}(i) + \tilde{v}(j) \geq U(i, j) + V(i, j)$$

*must hold with equality for  $i = j$ .*

By a celebrated result of Shapley and Shubik (1971), this notion is equivalent to the notion of TU Pareto efficiency. Note that there may be multiple vectors  $\tilde{u}$  and  $\tilde{v}$  for the given matching  $\sigma_0$ .

### 2.3 No-Trade stability

The notions of TU and NTU stability have been known and studied for a very long time. Here, we seek to better understand the effect that the possibility of transfers has on a matching market. We introduce a solution concept that is meant to relate the two notions.

Note that if matching  $\sigma_0(i) = i$  is TU stable, then there are transfers between the matched partners, say from woman  $i$  to man  $i$ , equal to

$$T_i = \tilde{u}(i) - U(i, i) = V(i, i) - \tilde{v}(i) \quad (2.1)$$

where the payoffs  $\tilde{u}(i)$  and  $\tilde{v}(j)$  are those of Definition 4. We want to understand the situations when matching  $\sigma_0(i) = i$  is TU stable but when no actual transfers are made “in equilibrium.” As a result, the matching  $\sigma_0$  is NTU stable as well as it is TU stable.

We motivate the notion of a No-Trade stable matching with an example. We present a matching market with a matching which is both the unique TU stable matching and also the unique NTU stable matching. In order for agents to *accept it*, however, transfers are needed.

**Example 5** *In this and other examples, we write the payoffs  $U$  and  $V$  in matrix form. In the matrices, the payoff in row  $i$  and column  $j$  is the utility  $U(i, j)$  for man  $i$  in matrix  $U$ , and utility  $V(i, j)$  for woman  $j$  in matrix  $V$ .*

*Consider the following utilities*

$$U = \begin{pmatrix} 0 & 2 & 1 \\ 1 & 2 & 0 \\ 1 & 0 & 2 \end{pmatrix}, \quad V = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 1 & 2 \end{pmatrix}$$

*Note that the matching  $\sigma_0(i) = i$  is the unique NTU stable matching, and is also the unique TU stable matching. To sustain it in the TU game, however, requires transfers. Indeed,  $u(i_1) = 0$  and  $v(j_2) = 2$  cannot be TU stable payoffs as*

$$2 = u(i_1) + v(j_2) < U(i_1, j_2) + V(i_1, j_2) = 3$$

*contradicts Definition 4. Intuitively, one needs to compensate agent  $i = 1$  in order for him to remain matched with  $j = 1$ . Hence, even though  $\sigma_0$  is NTU-stable and TU-stable, transfers between the agents are required to sustain it as TU-stable. Anticipating the definition to follow, this means this matching is not a No-trade stable matching.*

Matching  $\sigma_0(i) = i$  is *no-trade stable* when it is TU stable and there are no actual transfers between partners at equilibrium. In other words, Equation (2.1) should hold with  $T_i = 0$ . That is,  $U(i, i) = u(i)$ ,  $V(j, j) = v(j)$ , and so:

**Definition 6 (No-Trade Matching)** *Matching  $\sigma_0(i) = i$  is no-trade stable if and only if for all  $i$  and  $j$ ,*

$$U(i, j) + V(i, j) \leq U(i, i) + V(j, j).$$

Therefore in a no-trade stable matching, two matched individuals would have the opportunity to operate monetary transfers, but they choose not to do so. To put this in different terms, in a no-trade stable matching, spouses are “uncorrupted” because no monetary transfer actually takes place between them, but they are not “incorruptible”, because the rules of the game would allow for it.

### 3 Cardinal characterizations

We now present simple characterizations of the solution concepts described in Section 2.2.

Our characterizations involve cardinal notions, even for the solutions that are purely ordinal in nature. The point is to characterize all solutions using similar concepts, so it is easier to understand how the solutions differ. It will also help us understand the role of transfers in matching markets.

Define  $\sigma_0$  as the matching such that  $\sigma_0(i) = i$ . We need to introduce the following notation:

$$\begin{aligned} R_{ij} &= U(i, i) - U(i, j) \\ S_{ij} &= V(j, j) - V(i, j), \end{aligned}$$

defined for each  $i \in M$  and  $j \in W$ . Note that  $R_{ij}$  measures how much  $i$  prefers his current partner to  $j$ , and  $S_{ij}$  measures how much  $j$  prefers her current partner to  $i$ . These two concepts are dependent on the matching  $\sigma_0$ , which we take as fixed in the following result.

**Theorem 7** *Matching  $\sigma_0(i) = i$  is:*

(a) No-trade stable iff for all  $i$  and  $j$  in  $\{1, \dots, n\}$

$$0 \leq R_{ij} + S_{ij} \quad (3.1)$$

(b) NTU stable iff for all  $i$  and  $j$  in  $\{1, \dots, n\}$

$$0 \leq \max(R_{ij}, S_{ij}) \quad (3.2)$$

(c) TU stable iff there exists  $T \in \mathbb{R}^n$  such that for all  $i$  and  $j$  in  $\{1, \dots, n\}$

$$T_j - T_i \leq R_{ij} + S_{ij} \quad (3.3)$$

(d) Ex-ante Pareto efficient iff there exist  $v_i$  and  $\lambda_i, \mu_j > 0$  such that for all  $i$  and  $j$  in  $\{1, \dots, n\}$

$$v_j - v_i \leq \lambda_i R_{ij} + \mu_j S_{ij} \quad (3.4)$$

(e) Ex-post Pareto iff there exist  $v_i$  and  $\lambda_i > 0$  such that for all  $i$  and  $j$  in  $\{1, \dots, n\}$

$$v_j - v_i \leq \lambda_i \max(R_{ij}, S_{ij}). \quad (3.5)$$

Observe that (3.4) and (3.5) are “Afriat inequalities,” using the terminology in revealed preference theory.

As a consequence of the previous characterizations, it is straightforward to list the chains of implications between the various solution concepts.

**Theorem 8** (i) *The two following chains of implications always hold:*

- No-trade Stable implies TU Stable, TU Stable implies Ex-ante Pareto Efficient, Ex-ante Pareto Efficient implies Ex-post Pareto Efficient, and
- No-trade Stable implies NTU stable, NTU stable implies Ex-post Pareto Efficient.

(ii) *Assume that there are two agents on each side of the market. Then two additional implications hold:*

- Ex-Post Pareto Efficient implies (and thus is equivalent to) Ex-Ante Pareto Efficient, and
- NTU stable implies Ex-Ante Pareto Efficient.

*Any further implication which does not logically follow from those written is false. See Figure (1).*

(iii) *Assume that there are at least three agents on each side of the market. Then any implication that does not logically follow from the ones stated in part (i) of Theorem 7 above are false. See Figure (2).*

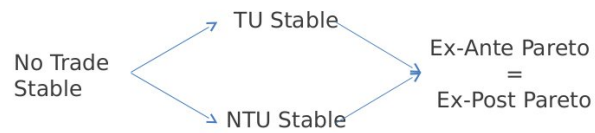


Figure 1: Summary of the implications in Theorem 8, part (ii) when there are only two agents on each side of the market.

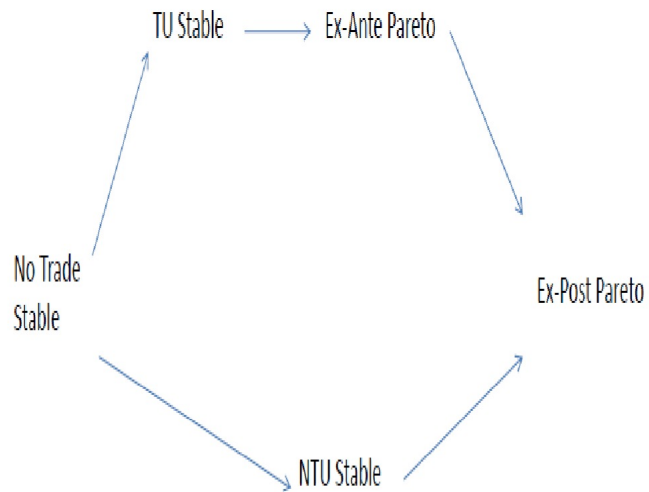


Figure 2: Summary of the implication in Theorem 8, part (iii) when there are three agents or more on each side of the market.

The implications in Theorem 8 are illustrated in Figures (1) and (2).

The proof of Theorem 8 is given in the Appendix. It relies on the following counterexamples.

**Example 9** Consider

$$U = \begin{pmatrix} 0 & -2 \\ 1 & 0 \end{pmatrix} \text{ and } V = \begin{pmatrix} 0 & -2 \\ 1 & 0 \end{pmatrix} \quad (3.6)$$

Then  $\sigma_0$  is TU stable (and thus Ex-Ante Pareto Efficient and Ex-Post Pareto Efficient), but not NTU stable, and not No-trade stable.

**Example 10** Consider

$$U = \begin{pmatrix} 0 & 2 & -1 \\ -1 & 0 & 2 \\ 2 & -1 & 0 \end{pmatrix}, \quad V = \begin{pmatrix} 0 & -1 & 2 \\ 2 & 0 & -1 \\ -1 & 2 & 0 \end{pmatrix}$$

Then  $\sigma_0$  is NTU stable (hence Ex-Post Pareto efficient). But it is not Ex-Ante Pareto efficient (hence neither No-trade stable nor TU stable). Indeed consider the fair lottery over the 6 existing pure assignments. Under this lottery, each agent achieves a payoff of  $1/3$ , hence this lottery is ex-ante preferred by each agent to  $\sigma_0$ .

**Example 11** Consider now

$$U = \begin{pmatrix} 0 & 1 \\ -2 & 0 \end{pmatrix} \text{ and } V = \begin{pmatrix} 0 & 1 \\ 3 & 0 \end{pmatrix} \quad (3.7)$$

Then  $\sigma_0$  is ex-ante Pareto efficient, and it is ex-post Pareto efficient, but it is not TU stable, and it is not NTU stable. Ex-ante Pareto efficiency follows from  $\lambda_i = 5$ ,  $\mu_j = 1$  and  $v_2 = 2$ ,  $v_1 = 1$ . Ex-post Pareto efficiency is clear. It is easily seen that  $\sigma_0$  is not TU stable. It is also clear that the matching is not NTU stable, as  $i = 1$ ,  $j = 2$  form a blocking pair.

## 4 NTU Stability and No-trade stable matchings

As we explained above, we use no-trade stability to shed light on the role of transfers. Given a stable NTU matching, one may ask if there is a cardinal representation of the agents' utility such that the stable matching is no-trade stable.

As we shall see, the answer is yes if we are allowed to tailor the cardinal representation to the given stable matching. If we instead want a representation that works for all stable matchings in the market, we shall resort to a regularity condition: that the matchings be isolated.

Finally, some statements in Theorem 7 involve the rescaling of utilities: they show how optimality can be understood through the existence of *weights* on the agents that satisfy certain properties. We can similarly imagine finding, not an arbitrary cardinal representation of preferences, but a restricted rescaling of utilities that captures the role of transfers. That is to say, a rescaling of utilities that ensures that the matching is no-trade stable. We shall present an example to the effect that such a rescaling is not possible.

### 4.1 Linking models with and without transfers

Our first question is whether for a given NTU stable matching, there is a cardinal representation of preferences under which the same matching is No-trade stable. The answer is yes.

**Theorem 12** *Let  $(M, W, P)$  be an ordinal matching market. If  $\sigma$  is a NTU stable matching, then there is a cardinal representation of  $P$  such that  $\sigma$  is a no-trade stable matching in the corresponding cardinal market.*

It is natural to try to strengthen this result in two directions. First, we could expect to choose the cardinal representation of preferences as a linear rescaling of a given cardinal representation of the preferences. Given what we know about optimality being characterized by choosing appropriate utility weights, it makes sense to ask whether any stable matching can be obtained as a No-Trade Stable Matching if one only weights agents in the right way. Namely:



**Problem 13** *Is it the case that matching  $\sigma_0$  is NTU stable if and only if there is  $\lambda_i, \mu_j > 0$  such that for all  $i$  and  $j$ ,*

$$0 \leq \lambda_i R_{ij} + \mu_j S_{ij}?$$

After all, transfers favor some agents over others, and utility weights play a similar role. Our next example shows that this is impossible. It exhibits a stable matching that is not No-Trade for any choice of utility weights.

**Example 14** *Consider a marriage market defined as follows. The sets of men and women are:  $M = \{m_1, m_2, m_3, m_4\}$  and  $W = \{w_1, w_2, w_3, w_4\}$ . Agents' preferences are defined through the following utility functions:*

$$U = \begin{pmatrix} 1.01 & 0 & \underline{1/2} & -1 \\ 0 & 1 & -1 & \underline{1/2} \\ \underline{1/2} & 1/5 & 1/3 & 1/4 \\ 1/5 & \underline{1/2} & 1/3 & 1/4 \end{pmatrix}, \quad V = \begin{pmatrix} 0 & 1 & \underline{1/2} & -1 \\ 1 & 0 & -1 & \underline{1/2} \\ \underline{1/2} & 1/5 & 1/3 & 1/4 \\ 1/5 & \underline{1/2} & 1/3 & 1/4 \end{pmatrix}$$

*The unique stable matching is underlined. Uniqueness is readily verified by running the Gale-Shapley algorithm. So  $u_i = v_j = 1/2$  for all  $i$  and  $j$ . Yet it is shown in the appendix that there are no  $\lambda_i, \mu_j > 0$  such that for all  $(i, j)$ ,*

$$\lambda_i(u_i - U_{ij}) + \mu_j(v_j - V_{ij}) \geq 0.$$

*This example has indifference in payoffs. It is simple to perturb the payoffs so that there are no more indifference, and the conclusion still holds. This is explained in the proof in the appendix.*

Example 14 has the following implication (which also follows from Example 5).

**Corollary 1** *There are cardinal matching markets that do not possess a no-trade matchings.*

Given Example 14, it is clear that No-Trade can only be achieved by appropriate choice of agents' utility functions. Our next question deals with the existence of cardinal utilities such that the set of No-trade stable matchings and NTU stable matchings will coincide for *all* stable matchings in a market. We show that if the stable matchings are isolated then one can choose cardinal utilities such that all stable matchings are No-Trade.

Let  $S(P)$  denote the set of all stable matchings. A matching  $\sigma \in S(P)$  is *isolated* if  $\sigma'(a) \neq \sigma(a)$  for all  $a \in M \cup W$  and all  $\sigma' \in S(P) \setminus \{\sigma\}$ .

**Theorem 15** *There is a representation  $(U, V)$  of  $P$  such that for all  $\sigma \in S(P)$ , if  $\sigma$  is isolated then  $\sigma$  is no trade stable for  $(U, V)$ .*

The question whether the conclusion holds without the assumption that the matching is isolated remains open to investigation.

## 4.2 Price of no transfers

The logic of the previous subsection can be pushed further, to obtain a “Price of Anarchy,” in the spirit of the recent literature in computer science (Roughgarden (2005)). We *quantify* the cost in social surplus (sum of agents’ utilities) that results from NTU stability: we can think of this cost as an efficiency gap inherent in the notion of NTU stable matching. The result is that the gap can be arbitrarily large, and that it grows “super exponentially” in the size of the market (i.e. it grows at a faster rate than  $n^g$ , for any  $g$ , where  $n$  is the size of the market).

Let  $\Delta^\epsilon$  denote the subset of the simplex in  $\mathbf{R}^{2n}$  in which every component is at least  $\epsilon$ :  $\Delta^\epsilon = \{((\alpha(i))_{i \in M}, (\beta(j))_{j \in W}) : \forall i \in M \alpha(i) \geq \epsilon, \forall j \in W, \beta(j) \geq \epsilon\}$ . Let  $S(M, W, U, V)$  denote the set of stable matchings in the cardinal matching market  $(M, W, U, V)$ . In the statement of the results below, we write the matchings in  $S(M, W, U, V)$  as fractional matchings  $\pi$  in which every entry in  $\pi$  is either 0 or 1.

**Proposition 16** *For every  $\epsilon > 0$ ,  $n$ ,  $g$ , and  $K > 0$  There is a cardinal marriage market  $(M, W, U, V)$ , with  $n$  men and women, and where utilities  $U$  and  $V$  are bounded by  $K$  such that*

$$\frac{\min_{(\alpha, \beta) \in \Delta^\epsilon} \left\{ \max_{\pi \in \Pi} \sum_{i=1}^n \alpha(i) \sum_{j'=1}^n \pi_{i,j'} U(i, j') + \sum_{j=1}^n \beta(j) \sum_{i'=1}^n \pi_{i',j} V(i', j) \right\}}{\max_{(\alpha, \beta) \in \Delta^\epsilon} \left\{ \max_{\pi \in S(M, W, U, V)} \sum_{i=1}^n \alpha(i) \sum_{j'=1}^n \pi_{i,j'} U(i, j') + \sum_{j=1}^n \beta(j) \sum_{i'=1}^n \pi_{i',j} V(i', j) \right\}} \text{ is } \Omega(n^g K)$$

Proposition 16 shows that the gap in the sum of utilities, between the maximizing (probabilistic) matchings, and the stable matchings, is large and grows with the size of the market at a rate that is arbitrarily large. Moreover, the gap is large regardless of how one weighs agents’ utilities.

Of course, the interpretation of Proposition 16 is not completely straightforward. It does not seem right to compare the sum of utilities in a model in

which transfers are not allowed with the sum of utilities in the TU model.<sup>4</sup> Nevertheless, we hope that Proposition 16 sheds additional light on the role of transfers in matching markets.

## References

- CRAWFORD, V. (2008): “The flexible-salary match: a proposal to increase the salary flexibility of the national resident matching program,” *Journal of Economic Behavior & Organization*, 66(2), 149–160.
- ECHENIQUE, F. (2012): “Contracts versus Salaries in Matching,” *American Economic Review*, 102(1), 594–601.
- ECHENIQUE, F., S. LEE, M. SHUM, AND M. B. YENMEZ (2013): “The Revealed Preference Theory of Stable and Extremal Stable Matchings,” *Econometrica*, 81(1), 153–171.
- EKELAND, I., AND A. GALICHON (2012): “The Housing Problem and Revealed Preference Theory: Duality and an Application,” forthcoming, *Economic Theory*.
- FOSTEL, A., H. E. SCARF, AND M. J. TODD (2004): “Two new proofs of Afriat’s theorem,” *Economic Theory*, 24(1), 211–219.
- GALE, D., AND L. S. SHAPLEY (1962): “College Admissions and the Stability of Marriage,” *The American Mathematical Monthly*, 69(1), 9–15.
- HATFIELD, J., AND F. KOJIMA (2010): “Substitutes and stability for matching with contracts,” *Journal of Economic Theory*, 145(5), 1704 – 1723.
- HATFIELD, J., AND P. MILGROM (2005): “Matching with Contracts,” *American Economic Review*, 95(4), 913–935.
- KELSO, A. S., AND V. P. CRAWFORD (1982): “Job Matching, Coalition Formation, and Gross Substitutes,” *Econometrica*, 50, 1483–1504.
- LEE, S., AND L. YARIV (2014): “On the efficiency of stable matchings in large markets,” mimeo: Caltech.

---

<sup>4</sup>This problem of interpretation is present throughout the literature on the price of anarchy.

- ROTH, A., AND M. SOTOMAYOR (1990): *Two-sided Matching: A Study in Game-Theoretic Modelling and Analysis*, vol. 18 of *Econometric Society Monographs*. Cambridge University Press, Cambridge England.
- ROTH, A. E. (1984a): “The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory,” *The Journal of Political Economy*, 92(6), 991–1016.
- (1984b): “Stability and Polarization of Interests in Job Matching,” *Econometrica*, 52(1), 47–57.
- (2002): “The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics,” *Econometrica*, 70(4), pp. 1341–1378.
- ROUGHGARDEN, T. (2005): *Selfish routing and the price of anarchy*. MIT press.
- SHAPLEY, L., AND M. SHUBIK (1971): “The assignment game I: The core,” *International Journal of Game Theory*, 1(1), 111–130.

# A Appendix: Proofs

## A.1 Proof of Theorem 7

**Proof.** (i) (a) Characterization of No-trade stable matchings as in (3.1) follows directly from the definition.

(b) Characterization (3.2) follows directly from the definition.

(c) For characterization of TU stability in terms of (3.3), recall that according to the definition, matching  $\sigma_0(i) = i$  is TU Stable if there are vectors  $u(i)$  and  $v(j)$  such that for each  $i$  and  $j$ ,

$$u(i) + v(j) \geq U(i, j) + V(i, j)$$

with equality for  $i = j$ . Hence, there exists a monetary transfer  $T_i$  (of either sign) from man  $i$  to woman  $i$  at equilibrium given by

$$T_i = u(i) - U(i, i) = V(i, i) - v(i).$$

The stability condition rewrites as  $T_j - T_i \leq R_{ij} + S_{ij}$ , thus, one is led to characterization (3.3).

(d) For characterization of Ex-ante Pareto efficient matchings in terms of (3.4), the proof is an extension of the proof by Fostel, Scarf, and Todd (2004), which give in full. Assume  $\sigma_0$  is Ex-ante efficient. Then the Linear Programming problem

$$\begin{aligned} & \max \sum_i x_i + \sum_j y_j \\ & \text{s.t.} \\ & x_i = - \sum_j \pi_{ij} R_{ij} \text{ and } y_j = - \sum_i \pi_{ij} S_{ij} \\ & \sum_k \pi_{ik} = \sum_k \pi_{ki} \text{ and } \sum_k \pi_{ik} = 1 \\ & x_i \geq 0, y_j \geq 0, \pi_{ij} \geq 0. \end{aligned}$$

is feasible and its value is zero. Thus it coincides with the value of its dual, which is

$$\begin{aligned} & \min - \sum_i \phi_i \\ & \text{s.t.} \\ & v_j - v_i \leq \lambda_i R_{ij} + \mu_j S_{ij} + \phi_i \\ & \lambda_i \geq 1 \text{ and } \mu_j \geq 1 \end{aligned}$$

where variables  $\lambda_i$ ,  $\mu_j$ ,  $v_i$  and  $\phi_i$  in the dual problem are the Lagrange multipliers associated to the four constraints in the primal problem, and variables  $\pi_{ij}$ ,  $x_i$ , and  $y_j$  in the primal problem are the Lagrange multipliers associated to the three constraints in the dual problem. Hence the dual program is feasible, and there exist vectors  $\lambda$ ,  $\mu$ , and  $\phi$ , such that

$$\begin{aligned} v_j - v_i &\leq \lambda_i R_{ij} + \mu_j S_{ij} + \phi_i & (A.1) \\ \lambda_i &\geq 1 \text{ and } \mu_j \geq 1 \\ \sum_i \phi_i &= 0 \end{aligned}$$

but setting  $j = i$  in inequality (A.1) implies (because  $R_{ii} = S_{ii} = 0$ ) that  $\phi_i \geq 0$ , hence as  $\sum_i \phi_i = 0$ , thus  $\phi_i = 0$ . Therefore it exist vectors  $\lambda_i > 0$  and  $\mu_j > 0$ , such that

$$v_j - v_i \leq \lambda_i R_{ij} + \mu_j S_{ij}.$$

(e) For characterization of Ex-post Pareto efficient matchings in terms of (3.5), assume  $\sigma_0$  is Ex-post Pareto efficient, and let

$$Q_{ij} = \max(R_{ij}, S_{ij}),$$

so that by definition, matrix  $Q_{ij}$  satisfies “**cyclical consistency**”: for any cycle  $i_1, \dots, i_{p+1} = i_1$ ,

$$\forall k, Q_{i_k i_{k+1}} \leq 0 \text{ implies } \forall k, Q_{i_k i_{k+1}} = 0, \quad (A.2)$$

By the Linear Programming proof of Afriat’s theorem in Fostel, Scarf, and Todd (2004)<sup>5</sup>, see implication (i) implies (ii) in Ekeland and Galichon (2012), there are scalars  $\lambda_i > 0$  and  $v_i$  such that (3.5) holds. ■

## A.2 Proof of Theorem 8

**Proof.** (i) No trade stable implies TU Stable is obtained by taking with  $T_i = 0$  in (3.3).

TU Stable implies Ex-ante Pareto is obtained by taking  $v_i = T_i$  and  $\lambda_i = \mu_j = 1$  in (3.4).

---

<sup>5</sup>The link between Afriat’s theorem and the characterization of efficiency in the housing problem was first made in Ekeland and Galichon (2012).

To show that Ex-ante Pareto implies Ex-post Pareto, assume there exist  $v_i$  and  $\lambda_i, \mu_j > 0$  such that  $v_j - v_i \leq \lambda_i R_{ij} + \mu_j S_{ij}$ . Now assume  $\max(R_{ij}, S_{ij}) \leq 0$ . Then  $v_j - v_i \leq 0$ , and the same implication holds with strict inequalities. By implication (iii) implies (ii) in Ekeland and Galichon (2012), there exist scalars  $v'_i$  and  $\lambda'_i$  such that  $v'_j - v'_i \leq \lambda'_i \max(R_{ij}, S_{ij})$ .

No-trade stable implies NTU stable follows from  $R_{ij} + S_{ij} \leq 2 \max(R_{ij}, S_{ij})$ .

NTU Stable implies Ex-post Pareto is obtained by taking  $\lambda_i = 1$  and  $v_i = 0$  in (3.5).

Part (i) of the result is proved using a series of counterexample, which for the most part only require two agents (one can incorporate a third neutral agents, which has zero utility regardless of the outcome).

We show point (iii) before point (ii). In order to show (iii), it is enough to show the following claims, proved in Examples 9 to 11:

- TU Stable does not imply NTU Stable – cf. example 9
- NTU Stable does not imply Ex-Ante Pareto – cf. example 10
- Ex-Ante Pareto does not imply TU Stable – cf. example 11
- Ex-Ante Pareto does not imply NTU Stable – cf. example 11
- Ex-Post Pareto does not imply Ex-Ante Pareto – cf. example 10
- Ex-Post Pareto does not imply NTU stable – cf. example 11.

To prove part (ii), we note that in the proof of part (i), the only instance where we needed three agents was to disprove that NTU stable implies Ex-Ante Pareto efficient and to disprove that Ex-Post Pareto efficient implies Ex-Ante Pareto efficient. We will show that these implications actually hold when there are only two agents. Indeed, when there are two agents,  $\sigma_0$  is Ex-Ante Pareto efficient if there are positive scalars  $\lambda_1, \lambda_2, \mu_1$  and  $\mu_2$  such that

$$0 \leq \lambda_1 R_{12} + \lambda_2 R_{21} + \mu_1 S_{12} + \mu_2 S_{21}$$

which is equivalent to

$$0 \leq \max(R_{12}, R_{21}, S_{12}, S_{21}). \tag{A.3}$$

Therefore, if  $\sigma_0$  is Ex-Post Pareto efficient, then  $v_j - v_i \leq \lambda_i \max(R_{ij}, S_{ij})$ . But either  $v_1 - v_2$  or  $v_2 - v_1$  is nonnegative, thus (A.3) holds, and Ex-Post Pareto efficient implies Ex-Ante Pareto efficient. ■

### A.3 Proof of Theorem 12

Assume  $\mu_0(i) = i$  (this is w.l.o.g. as can always relabel individuals). Take  $R_{ij} = U(i, j) - U(i, i)$  and  $S_{ij} = V(i, j) - V(j, j)$ .  $\mu_0$  is Stable iff  $\min(R_{ij}, S_{ij}) \leq 0$  for all  $i$  and  $j$ , with strict inequality for  $j \neq i$ . Consider

$$\begin{aligned}\bar{U}(i, j) &= \frac{1}{2} - e^{-tR_{ij}} \text{ for } i \neq j \\ \bar{U}(i, i) &= 0\end{aligned}$$

one has:

- $\bar{U}(i, j) > 0$  if and only if  $\frac{1}{2} > e^{-tR_{ij}}$  that is  $-\log 2 > -tR_{ij}$  that is  $tR_{ij} > \log 2$  hence  $R_{ij} > 0$ .
- $\bar{U}(i, j) < 0$  if and only if  $tR_{ij} < \log 2$  hence  $R_{ij} < 0$ .

Take

$$t > \max_{i \neq j} \left( \left| \frac{\log 2}{R_{ij}} \right|, \left| \frac{\log 2}{S_{ij}} \right| \right)$$

and let

$$\begin{aligned}\bar{V}(i, j) &= \frac{1}{2} - e^{-tS_{ij}} \text{ for } i \neq j \\ \bar{V}(i, i) &= 0\end{aligned}$$

Then  $\bar{U}(i, j) \leq 0$  and  $\bar{V}(i, j) \leq 0$ , thus

$$\bar{U}(i, j) + \bar{V}(i, j) \leq 0 = \bar{U}(i, i) + \bar{V}(j, j).$$

Thus  $\mu_0$  is a No-Trade stable Matching associated to utilities  $\bar{U}$  and  $\bar{V}$ . ■

### A.4 Claim in Example 14

Rephrasing, we want to know if there are  $\alpha(m) > 0$  and  $\beta(w) > 0$  such that, for all  $(m, w)$ ,

$$\alpha(m)(u(m) - U(m, w)) + \beta(w)(v(w) - V(m, w)) \geq 0$$

Consider the matrix  $A$  which has one column for each  $m$  and each  $w$ , and one row for each pair  $(m, w) \in M \times W$ . The matrix  $A = (a_{(m,w),a})_{(m,w) \in M \times W, a \in M \cup W}$



is defined as follows. The row corresponding to  $(m, w)$  has zeroes in all its entries except in the columns corresponding to  $m$  and  $w$ . It has  $u(m) - U(m, w)$  in  $m$ 's column and  $v(w) - V(m, w)$  in  $w$ 's column.

The problem is to find  $x \gg 0$  such that  $A \cdot x \geq 0$ . We introduce the matrix  $B$  such that the  $i$ 'th row of  $B$  is the vector  $e_i = (0, \dots, 1, \dots, 0)$  with a 1 only in entry  $i$ . Then we want to find a vector  $x \in \mathbf{R}^n$  such that  $A \cdot x \geq 0$  and  $B \cdot x \gg 0$ . By Motzkin's Theorem of the Alternative, such a vector  $x$  exists iff there is no  $(y, z)$ , with  $z > 0$  (meaning  $z \geq 0$  and  $z \neq 0$ ) such that

$$y \cdot A + z \cdot B = 0.$$

$$\begin{array}{cccc|cccc} i & i' & i_0 & i_1 & j & j' & j_0 & j_1 \\ \hline j & j' & \underline{j} & \underline{w'} & i' & i & \underline{i} & \underline{i'} \\ \hline \underline{j_0} & \underline{j_1} & \underline{j_0} & \underline{j_0} & i_0 & i_1 & i_0 & i_0 \\ \hline j' & \underline{j} & j_1 & j_1 & i & i' & i_1 & i_1 \\ \hline j_1 & j_0 & j' & j & i_1 & i_0 & i' & i \end{array}$$

Utilities are:

$$\begin{array}{cccc|cccc} i & i' & i_0 & i_1 & j & j' & j_0 & j_1 \\ \hline 1 + \delta & 1 & 1/2 & 1/2 & 1 & 1 & 1/2 & 1/2 \\ 1/2 & 1/2 & 1/3 & 1/3 & 1/2 & 1/2 & 1/3 & 1/3 \\ 0 & 0 & 1/4 & 1/4 & 0 & 0 & 1/4 & 1/4 \\ -1 & -1 & 1/5 & 1/5 & -1 & -1 & 1/5 & 1/5 \end{array}$$

The upper 4 rows of  $A$  are:

$$\begin{array}{c|cccccccc} & i & i' & i_0 & i_1 & j & j' & j_0 & j_1 \\ \hline i, j & 1/2 - (1 + \delta) & 0 & 0 & 0 & 1/2 & 0 & 0 & 0 \\ i, j' & 1/2 & 0 & 0 & 0 & 0 & 1/2 - 1 & & \\ i', j & 0 & 1/2 & 0 & 0 & -1/2 & 0 & 0 & 0 \\ i', j' & 0 & -1/2 & 0 & 0 & 0 & 1/2 & 0 & 0 \end{array}$$

So the sum of the first four rows of  $A$  is  $(-\delta, 0, 0, 0, 0, 0, 0, 0, 0)$ .

Notice that we have other rows: for example the row corresponding to  $(i, j_1)$  is:

$$\begin{array}{c|cccccccc} & i & i' & i_0 & i_1 & j & j' & j_0 & j_1 \\ \hline i, j_1 & 1/2 + 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1/2 - 1/5, \end{array}$$

but these rows will get weight zero in the linear combination below.

So  $y = (1, 1, 1, 1, 0, \dots, 0)$  and  $z = (\delta, 0, \dots, 0)$  exhibit a solution to the alternative system as

$$y \cdot A + z \cdot B = (-\delta, 0, 0, 0, 0, 0, 0, 0) + \delta(1, 0, \dots, 0) = 0$$

■

Observe that in the construction of a solution to the dual system above, we could perturb utilities by adding payoffs to each entry, in such a way that we obtain the matrix  $A+A'$  instead of  $A$  above. By choosing the perturbation so that  $y \cdot A' = 0$  the result goes through.

## A.5 Proof of Theorem 15

Let  $S(P)$  be the set of stable matchings in the ordinal matching market  $(M, W, P)$ . Suppose that there are  $N$  stable matchings, and enumerate them, so  $S(P) = \{\mu^1, \dots, \mu^K\}$ .

To prove the proposition we first establish some simple lemmas.

**Lemma 2** *For any  $i \in M$  and  $j \in W$ ,*

$$|\{k : j >_i \mu^k(i)\}| + |\{k : i >_j \mu^k(j)\}| \leq K - |\{k : j = \mu^k(i)\}|$$

**Proof.** Let  $j >_i \mu^k(i)$ ; then for  $\mu^k$  to be stable we need that  $\mu^k(j) >_j i$ . So  $|\{k : j >_i \mu^k(i)\}| \leq |\{k : \mu^k(j) >_j i\}|$ .

Then,

$$\begin{aligned} |\{k : i >_j \mu^k(j)\}| &= K - |\{k : \mu^k(j) \geq_j i\}| \\ &\leq K - |\{k : j >_i \mu^k(i)\}| - |\{k : j = \mu^k(i)\}|, \end{aligned}$$

where the last inequality follows from the previous paragraph and the fact that preferences  $>_j$  are strict. ■

Let  $\hat{U}(i, j) = |\{k : j \geq_i \mu^k(i)\}|$  and  $\hat{V}(i, j) = |\{k : i >_j \mu^k(j)\}|$ . By the previous lemma,  $\hat{U}(i, j) + \hat{V}(i, j) \leq K$  for all  $i$  and  $j$ .

**Lemma 3** *If  $\mu$  is an isolated stable matching,  $\mu'$  is a stable matching, and  $i, \hat{i} \in M$ , then  $\mu(i) >_i \mu'(i)$  iff  $\mu(\hat{i}) >_i \mu'(\hat{i})$ .*

**Proof.** Suppose (reasoning by contradiction) that  $\mu(i) >_i \mu'(i)$  while  $\mu'(\hat{i}) \geq_i \mu(\hat{i})$ . Since  $\mu$  is isolated and preferences are strict, we have  $\mu'(\hat{i}) >_i \mu(\hat{i})$ . Now let  $\hat{\mu} = \mu \vee \mu'$ , using the join operator in the lattice of stable matchings (see Roth and Sotomayor (1990)). Then  $\hat{\mu}(i) = \mu(i)$  and  $\hat{\mu}(\hat{i}) = \mu'(\hat{i})$ . So  $\hat{\mu} \in S(P)$ ,  $\hat{\mu}(i) = \mu(i)$ , and  $\hat{\mu} \neq \mu$ ; a contradiction of the hypothesis that  $\mu$  is isolated. ■

**Lemma 4** *If  $\mu$  is an isolated stable matching then*

$$\hat{U}(i, \mu(i)) + \hat{V}(\mu(j), j) = K.$$

**Proof.** We prove that

$$\{k : \mu \neq \mu^k \text{ and } \mu(i) \geq_i \mu^k(i)\} = \{k : \mu \neq \mu^k \text{ and } \mu^k(j) \geq_j \mu(j)\}.$$

The lemma follows then because

$$\begin{aligned} \hat{U}(i, \mu(i)) + \hat{V}(\mu(j), j) &= |\{n : \mu(i) \geq_i \mu^k(i)\}| + |\{k : \mu^k(j) >_j \mu(j)\}| \\ &= 1 + |\{k : \mu \neq \mu^k \text{ and } \mu(i) \geq_i \mu^k(i)\}| \\ &\quad + (K - |\{k : \mu \neq \mu^k \text{ and } \mu^k(j) \geq_j \mu(j)\}| - 1). \end{aligned}$$

Let  $\mu(i) \geq_i \mu^k(i)$  and let  $i = \mu(j)$ . Since  $\mu \neq \mu^k$  is isolated and preferences are strict,  $\mu(i) >_i \mu^k(i)$ . Then by Lemma 3,  $\mu(i) >_i \mu^k(i)$ ; so  $j = \mu(i)$  implies that  $\mu^k(j) >_j \mu(j)$ . Similarly, if  $\mu^k(j) >_j \mu(j)$  then  $\mu(i) >_i \mu^k(i)$ . So  $\mu(i) >_i \mu^k(i)$ . ■

We are now in a position to prove the proposition.

Define a representation  $U$  and  $V$  of  $P$  as follows. Fix  $\delta$  such that  $0 < \delta < 1/2$ . Let  $U(i, j) = \hat{U}(i, j)$  and  $V(i, j) = \hat{V}(i, j)$  if there is  $\mu \in S(P)$  such that  $j = \mu(i)$ . Otherwise, if  $j$  is worse than  $i$ 's partner in any stable matching, let  $U(i, j) < 0$  (and chosen to respect representation of  $P$ ); and if there is  $\mu \in S(P)$  such that  $j >_i \mu(i)$ , let  $\mu^0$  be the best such matching for  $i$ , and choose  $U(i, j)$  such that  $U(i, j) - U(i, \mu^0(i)) < \delta$ . Choose  $V$  similarly.

Let  $\mu$  be an isolated matching. Fix a pair  $(i, j)$  and suppose, wlog that  $u_\mu(i) - U(i, j) < 0$  and  $v_\mu(i) - V(i, j) \geq 0$  (if  $u_\mu(i) - U(i, j) \geq 0$  and  $v_\mu(i) - V(i, j) \geq 0$  then there is nothing to prove; and they cannot both be  $< 0$  or  $(i, j)$  would constitute a blocking pair).

First, if  $i$  and  $j$  are matched in some matching  $\mu' \in S(P)$  then  $u_\mu(i) - U(i, j) + v_\mu(i) - V(i, j) = u_\mu(i) - \hat{U}(i, j) + v_\mu(i) - \hat{V}(i, j)$  so it follows that

$u_\mu(i) - U(i, j) + v_\mu(i) - V(i, j) \geq 0$  by Lemmas 2, 4, and the definition of  $\hat{U}(i, j)$  and  $\hat{V}(i, j)$ .

Second, let us assume that  $i$  and  $j$  are not matched in any matching in  $S(P)$ . Since  $u_\mu(i) - U(i, j) < 0$  we know that there is a matching that is worse for  $i$  than  $j$ . Let  $\mu^0$  be such that  $j >_i \mu^0(i)$  implies that  $\mu^0(i) \geq_i \mu^0(i)$ . Thus  $u_{\mu^0}(i) - U(i, j) > -\delta$  by definition of  $U(i, j)$ . Since  $j >_i \mu^0(i)$ , we also have  $\mu^0(j) >_j i$ , or  $\mu^0$  would not be stable. Then, letting  $\mu^1$  be the best matching in  $S(P)$  for  $j$ , out of those that are worse than  $i$ , we have  $v_{\mu^0}(j) - V(i, j) = v_{\mu^0}(j) - v_{\mu^1}(j) + v_{\mu^1}(j) - V(i, j) > 1 - \delta$ , as  $\mu^0(j) >_j \mu^1(j)$  implies that  $v_{\mu^0}(j) - v_{\mu^1}(j) \geq 1$  and the definition of  $V(i, j)$  implies that  $v_{\mu^1}(j) - V(i, j) > -\delta$ .

Finally,

$$\begin{aligned} u_\mu(i) - U(i, j) + v_\mu(i) - V(i, j) &= u_\mu(i) - u_{\mu^0}(i) + u_{\mu^0}(i) - U(i, j) \\ &\quad + v_\mu(i) - v_{\mu^0}(j) + v_{\mu^0}(j) - V(i, j) \\ &= (u_\mu(i) - u_{\mu^0}(i) + v_\mu(i) - v_{\mu^0}(j)) \\ &\quad + (u_{\mu^0}(i) - U(i, j)) + (v_{\mu^0}(j) - V(i, j)) \\ &\geq 0 + (-\delta) + (1 - \delta) > 0, \end{aligned}$$

where the first inequality follows from the remarks in the previous paragraphs, and from the fact that  $K = u_\mu(i) + v_\mu(i) \geq u_{\mu^0}(i) + v_{\mu^0}(j)$  by Lemmas 2 and 4. The second inequality follows because  $\delta < 1/2$ . This proves the proposition.

## A.6 Proof of Proposition 16

Let  $n$  be an even positive number. Let  $(M, W, U, V)$  be a marriage market with  $n$  men and  $n$  women, defined as follows. The agents ordinal preferences are defined in the following tables:

$i_1$	$i_2$	$i_3$	$\cdots$	$i_{n-2}$	$i_{n-1}$	$i_n$
$j_1$	$j_2$	$j_3$	$\cdots$	$j_{n-2}$	$j_{n-1}$	$j_{n-1}$
$j_2$	$j_3$	$j_4$	$\cdots$	$j_{n-1}$	$j_1$	$j_1$
$j_3$	$j_4$	$j_5$	$\cdots$	$j_{n-2}$	$j_1$	$j_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$j_{n/2}$	$j_{n/2+1}$	$j_{n/2+1}$	$\cdots$	$j_{n/2-2}$	$j_{n/2-1}$	$j_{n/2-1}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$j_n$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$j_{n-1}$	$j_n$	$j_1$	$\cdots$	$\vdots$	$\vdots$	$\vdots$
$j_n$	$j_1$	$j_2$	$\cdots$	$j_{n-3}$	$j_{n-2}$	$\vdots$

The table means that  $j_1$  is the most preferred partner for  $i_1$ , followed by  $j_2$ , and so on. The women's' preferences are as follows.

$j_1$	$j_2$	$j_3$	$\cdots$	$j_{n-2}$	$j_{n-1}$	$j_n$
$i_2$	$i_3$	$i_4$	$\cdots$	$i_{n-1}$	$i_1$	$i_1$
$i_3$	$i_4$	$i_5$	$\cdots$	$i_{n-1}$	$i_2$	$i_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$i_{n/2}$	$i_{n/2+1}$	$i_{n/2+2}$	$\cdots$	$i_{n/2-2}$	$i_{n/2-1}$	$i_{n/2-1}$
$i_{n/2+1}$	$i_{n/2+2}$	$i_{n/2+3}$	$\cdots$	$i_{n/2}$	$i_{n/2}$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$i_{n-1}$	$i_n$	$i_1$	$\cdots$	$i_{n-4}$	$i_{n-3}$	$\vdots$
$i_n$	$i_1$	$i_2$	$\cdots$	$i_{n-3}$	$i_{n-1}$	$\vdots$
$i_1$	$i_2$	$i_3$	$\cdots$	$i_{n-2}$	$i_n$	$\vdots$
$i_1$	$i_2$	$i_3$	$\cdots$	$i_{n-2}$	$i_{n-1}$	$i_n$

It is a routine matter to verify that there is a unique stable matching in this market. It has  $i_1$  matched to  $j_{n/2}$ ,  $i_2$  matched to  $j_{n/2+1}$ , and so on, until we obtain that  $i_{n-1}$  is matched to  $j_{n/2-1}$ . We have  $i_n$  matched to  $j_n$ . (The logic of this example is that  $i_n$  creates cycles in the man-proposing algorithm which pushes the men down in their proposals until reaching the matching in the “middle” of their preferences;  $j_n$  plays the same role in the woman proposing version of the algorithm).

Define agents' cardinal preferences as follows. Let

$$U(i, j) = [n - r_m(w)] \frac{1}{n^g} + \max\{0, n - 1 - r_i(j)\} \left(K - \frac{n - 1}{n^g}\right),$$

where  $r_i(j)$  is the rank of woman  $j$  in  $i$ ' preferences. Similarly define  $V(i, j)$ , replacing  $r_i(j)$  with  $r_j(i)$ . Then, given the preferences defined above, the agents utilities at the unique stable matching satisfy:

$$u(i_l) = v(j_l) = \frac{1}{2n^{g-1}}, l = 1, \dots, n-1 \text{ and } u(i_n) = v(j_n) = (n/2 - 1) \frac{1}{n^g}.$$

So that the sum of all agents utilities at the unique stable matching is:

$$2(n-1) \left( \frac{1}{2n^{g-1}} \right) + 2(n/2 - 1) \frac{1}{n^g},$$

and agents' weighted sum of utilities is at most

$$\max \left\{ \frac{1}{2n^{g-1}}, (n/2 - 1) \frac{1}{n^g} \right\}.$$

Consider the matchings  $\mu^*(i_l) = j_l, l = 1, \dots, n$ , and  $\hat{\mu}(j_1) = i_2, \dots, \hat{\mu}(j_{n-2}) = i_{n-1}, \hat{\mu}(j_{n-1}) = i_1, \hat{\mu}(j_n) = i_n$ . Let  $\pi$  be the random matching that results from choosing  $\mu^*$  and  $\hat{\mu}$  with equal probability. Then, for all  $i \neq i_n$  and  $j \neq j_n$  we have that

$$\sum_{j'} \pi_{i,j'} U(i, j') = \sum_{i'} \pi_{j,i'} V(i', j) = K/2,$$

while

$$\sum_{j'} \pi_{i_n,j'} U(i_n, j') = \sum_{i'} \pi_{j,i'} V(i', j_n) = (n/2 - 1) \frac{1}{n^3}.$$

Then

$$\sum_{i \in M} \alpha(i) \sum_{j' \in W} \pi_{i,j'} U(i, j') + \sum_{j \in W} \beta(j) \sum_{i' \in M} \pi_{j,i'} V(i', j) \geq \epsilon n K / 2.$$

So, regardless of the values of  $\alpha$  and  $\beta$  in  $\Delta^\epsilon$ , the fraction

$$\frac{\sum_{i \in M} \alpha(i) \sum_{j' \in W} \pi_{i,j'} U(i, j') + \sum_{j \in W} \beta(j) \sum_{i' \in M} \pi_{j,i'} V(i', j)}{\sum_{i \in M} \alpha(i) u(i) + \sum_{j \in W} \beta(j) v(j)}$$

is bounded below by

$$\frac{\epsilon n K / 2}{\max \left\{ \frac{1}{2n^{g-1}}, (n/2 - 1) \frac{1}{n^g} \right\}},$$

which is  $\Omega(K n^g)$ . ■