

The Roommate Problem

Is More Stable Than You Think¹

Pierre-André Chiappori²
Columbia University

Alfred Galichon³
Ecole polytechnique

Bernard Salanié⁴
Columbia University

January 24, 2012

¹We thank Yeon-Koo Che, Fuhito Kojima, and Navin Kartik for useful comments.

²Address: Department of Economics, Columbia University, 1009A International Affairs Building, 420 West 118th St., New York, NY 10027, USA. E-mail: pc2167@columbia.edu.

³Address: Economics Department, Ecole Polytechnique, 91128 Palaiseau, France. E-mail: alfred.galichon@polytechnique.edu. Parts of this paper were written while Galichon was visiting Columbia University; he thanks the Department of Economics and the Alliance Program for their hospitality. Support from Chaire EDF-Calyon “Finance and Développement Durable” and FiME (www.fime-lab.org), from Chaire Axa “Assurance et Risques Majeurs” and Chaire FDIR “Socially Responsible Investments and Value Creation” is gratefully acknowledged.

⁴Address: Department of Economics, Columbia University, 1132 International Affairs Building, 420 West 118th Street, New York, NY 10027, USA. E-mail: bsalanie@columbia.edu.

Abstract

Stable matchings may fail to exist in the roommate matching problem, both when utility is transferable and when it is not. We show that in both settings, the existence of a stable matching is restored when there is an even number of individuals of indistinguishable characteristics and tastes (types). As a consequence, when the number of individuals of any given type is large enough there always exist “quasi-stable” matchings: a stable matching can be restored with minimal policy intervention. Our results build on an analogy with an associated bipartite problem; it follows that the tools crafted in empirical studies of the marriage problem can easily be adapted to the roommate problem.

1 Introduction

Since the seminal work of Becker (1973, 1991), the economic literature has mostly modeled the marriage market as a bipartite matching game, sometimes with non transferable utility (NTU), but mostly with transferable utility (TU). Each couple consists of two partners from two separate subpopulations, whom we will call men and women. In the TU framework, any potential couple generates a surplus that is (endogenously) shared by its members, whereas in the NTU context each member receives a fixed utility from being matched with a given partner. In both cases, the resulting matching must satisfy a stability property, reflecting robustness to unilateral and bilateral deviations. The theoretical analysis of bipartite matching without transferable utility was laid down by the seminal paper of Gale and Shapley (1962). With transferable utility it was introduced by Koopmans and Beckmann (1957). Shapley and Shubik (1971) studied the set of stable matchings under transferable utility; they showed that it is both the core and the set of competitive equilibria, that stable matchings are the maximizers of aggregate surplus, and that the associated individual surpluses solve the dual imputation problem.

The applications of bipartite matching go well beyond the market for marriage. Yet the bipartite assumption is restrictive in some contexts, where a match does not have to include exactly one individual from each of two exogenously given subpopulations. Even for marriage markets, a growing number of countries or US states have authorized same-sex unions in some form. In economic relationships the buyer-seller distinction is often endogenously determined, as are positions on a presidential ticket, or roles within a team in sports. The game where agents set out to match without the bipartite requirement is classically called the *roommate matching problem*. It is well-known since Gale and Shapley (1962) that when utility is not transferable, many of the nice results of bipartite matching do not extend to the roommate problem. In particular, Gale and Shapley showed that stable matchings may not exist in the roommate problem with non-transferable utility. Since this disappointing result, the literature on the roommate problem has been very sparse; and most of it is in the NTU setting. A few papers have studied the property of stable matchings when they exist¹. Tan (1991) finds a necessary and sufficient condition for stable matchings to exist under strict preferences, and Chung (2000) shows that a condition he calls “no odd

¹Gusfield and Irving (1989) showed that the set of singles is the same in all stable matchings; Klaus and Klijn (2010) study whether any of them can be “fair”. Efficient algorithms have also been available since Irving (1985).

rings” is sufficient under weak preferences. Little appears to be known so far in the TU case, in spite of its relevance in empirical applications².

The goal of this paper is to fill some of these gaps. We first show, on a simple example, that even in the TU case a stable matching may fail to exist. We next show that there exists a close relationship between the existence of a stable, “roommate” matching and the property of the optimal matching (defined as maximizing aggregate surplus). Namely, when a stable matching exists, it maximizes aggregate surplus; conversely, a matching that maximizes aggregate surplus may be associated to a stable roommate matching if and only if a symmetry property is satisfied. A first consequence of this result is that when the framework under consideration exhibits an *even* number of individuals of indistinguishable characteristics and tastes (types), then a stable matching always exist. Moreover, this conclusion is also true in a NTU framework. As a consequence, when the number of individuals of any given type is large enough there always exist “quasi-stable” matchings: a stable matching can be restored with minimal policy intervention. Finally, our results build on an analogy with an associated bipartite problem; it follows that the tools crafted in empirical studies of the marriage problem can easily be adapted to the roommate problem.

2 A Simple Example

We start by giving the intuition of our main results on an illustrative example.

2.1 Unstable Matchings

With transferable as well as with non-transferable utility, a stable matching may not exist for the roommate problem. This is known in the NTU case from Example 3 in Gale and Shapley (1962); example 1 below is a simplified version. Recall that with NTU, a *matching* defines who is matched to whom and who remains single; and it is stable if and only if no individual would rather remain single, and no two individuals who are currently not matched prefer each other to their current match (or to remaining single if that is their current status.)

²Chung (2000) shows that when the division of surplus obeys an exogenous rule, odd rings are ruled out and the roommate problem has a stable matching; but that is clearly not an appealing assumption.

Example 1 *There are three individuals. Each of them prefers being matched to any other individual to being single. In addition, individual 1 prefers to be matched with 2 than with 3; individual 2 prefers 3 to 1; and individual 3 prefers 1 to 2. The three individuals remaining single cannot be a stable matching since it is everyone's worst outcome. Any stable matching must therefore have one pair matched and one unmatched individual. But if 1 and 2 are matched and 3 is single, both 2 and 3 would be better off by forming a pair; and by symmetry, the other two possibilities are equally unstable. Thus no stable matching exists.*

From now on we focus on the TU case; we will return to NTU in section 6. As it turns out, it is almost equally easy to construct an example of non-existence of a stable matching with transferable utility. Here a *matching* defines who is matched to whom *and* how the corresponding surplus is divided between the partners. Stability is defined exactly as in the NTU case, except that if a new match forms partners must also agree on splitting their joint surplus. Consider the following:

Example 2 *The population has three individuals. Any unmatched individual has zero utility. The joint surplus created by the matching of any two of them is given by the off-diagonal terms of the matrix*

$$\Phi = \begin{pmatrix} - & 6 & 8 \\ 6 & - & 5 \\ 8 & 5 & - \end{pmatrix} \quad (1)$$

so that individuals 1 and 2 create, if they match, a surplus of 6; 1 and 3 create a surplus of 8, etc.

Assume, now, that there exists a stable matching. As in example 1, a matching in which all individuals remain single is obviously not stable; any stable matching must be such that one person remains single and the other two are matched together. Let (u_x) be the utility that individual of type $x = 1, 2, 3$ gets out of this game; stability imposes $u_x + u_y \geq \Phi_{xy}$ for all potential matches, with equality if x and y are actually matched—and $u_x \geq 0$ with equality if x is single. One can readily check, however, that no set of numbers (u_1, u_2, u_3) satisfying these relationships for all x and y exists: whichever the married pair is, one of the matched partners would increase her utility by matching with the single person.

Indeed, if the matched pair is $\{1, 2\}$, then

$$u_1 + u_2 = 6, u_3 = 0, u_2 \geq 0$$

contradicts $u_1 + u_3 \geq 8$: agent 3, being single, is willing to give up any amount smaller than 8 to be matched with 1, while the match between 1 and 2 cannot provide 1 with more than 6. Similarly, if the married pair is $\{2, 3\}$, then

$$u_2 + u_3 = 5, u_1 = 0, u_2 \geq 0, u_3 \geq 0$$

contradicts both $u_1 + u_2 \geq 6$ and $u_1 + u_3 \geq 8$ (so that 1 is willing to give more than 5 and less than 6 to agent 2 to match with her, and more than 5 and less than 8 to 3.) Finally, if the married pair is $\{1, 3\}$, then

$$u_1 + u_3 = 8, u_2 = 0, u_1 \geq 0, u_3 \geq 0$$

is incompatible with $u_1 + u_3 \geq 11$, which follows from combining $u_1 + u_2 \geq 6$ and $u_2 + u_3 \geq 5$ with $u_2 = 0$ (since agent 2 is single 1 could match with her and capture almost 6, while 3 could match with her and capture almost 5; these outside options are more attractive than anything 1 and 3 can achieve together.)

Again, no stable matching exists.

Note that there is nothing pathological in Example 2. The surpluses can easily be (locally) modified without changing the result. Also, the conclusion does not require an odd number of agents; one can readily introduce a fourth individual, who generates a small enough surplus with any roommate, without changing the non existence finding.

2.2 Cloning

However, there exists a simple modification that restores existence. Take Example 2. Let us now *duplicate* the economy by “cloning” each agent; technically, we now have three *types* $x = 1, 2, 3$ of agents, with two (identical) individuals of each type. The joint surplus created by a matching between two individuals of different types $x \neq y$ is as in Example 2; but we now also need to define the surplus generated by the matching of two clones (two individuals of the same type.) Take it to be 2 for every type—more on this later. We then have the matrix:

$$\Phi' = \begin{pmatrix} 2 & 6 & 8 \\ 6 & 2 & 5 \\ 8 & 5 & 2 \end{pmatrix} \tag{2}$$

Consider the following matching μ^* : there is one match between type 1 and type 2, one between type 1 and type 3, and one between type 2 and type

3. Assume individuals share the surplus so that each individual of type 1 gets 4.5, each individual of type 2 gets 1.5, and each individual of type 3 gets 3.5. This is clearly feasible; and it is easy to verify that it is a stable matching. Less obvious but still true is the fact (proved later on) that existence would still obtain for *any* values chosen for the diagonal of the matrix, although the stable matching pattern that would emerge may be different³. In other words, our cloning operation always restores the existence of a stable match, irrespective of the values of the joint surpluses created by matches between clones.

2.3 Surplus Maximization

Our main result is better understood when related to another, closely linked problem: finding a feasible matching that maximizes total surplus. Total surplus is simply the sum of the joint surpluses of every match (keeping to a normalized utility of zero for singles); and the adjective “feasible” refers to the fact that each individual can only be matched to one partner or stay single.

At least, that is the standard definition for the bipartite matching problem; but roommate matching introduces an additional feasibility constraint. For any two types $x \neq y$, denote μ_{xy} the number of matches between an individual of type x and an individual of type y ; a roommate matching for which μ_{xy} and μ_{yx} differ would clearly not be feasible. This additional symmetry constraint is absent from the bipartite model, where these two individuals would belong to two separate subpopulations and the number of marriages between say, a college-educated man and a woman who is a high-school graduate may well differ (and typically does) from the number of marriages between a college-educated woman and a man who is a high-school graduate.

As we will see, this symmetry constraint is the source of the difficulty in finding stable roommate matchings; and our cloning operation addresses it. To see this on our Example 2, first go back to roommate matching with one individual of each type $x = 1, 2, 3$, and neglect the symmetry constraint. Since there is only one individual of each type x , she cannot match with herself: $\mu_{xx} \equiv 0$; and neglecting symmetry, the only other feasibility constraints are

$$\text{for every } x, \sum_{y \neq x} \mu_{xy} \leq 1$$

³For instance, if the diagonal elements are large enough, the stable matching matches each individual with her clone.

and

$$\text{for every } y, \sum_{x \neq y} \mu_{xy} \leq 1.$$

The two matchings

$$\mu^1 = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \text{ and } \mu^2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

are feasible in this limited sense; and they both achieve the highest possible surplus when the symmetry conditions are disregarded. The existence of two solutions is not surprising: given the symmetric nature of the surplus matrix Φ , if a matrix μ maximizes total surplus, so does its transpose μ^t . Unfortunately, neither is symmetric, and therefore neither makes any sense in the roommate problem. For instance, μ^1 has agent 1 matched both with agent 3 (in the first row) and with agent 2 (in the first column). Also, note that a third solution to this relaxed problem is the unweighted mean of μ^1 and μ^2 ,

$$\mu^m = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{pmatrix}$$

However, while this matrix is indeed symmetric, its coefficients are not integer and thus it is not a feasible matching either.

Let us now reintroduce the symmetry constraint. The (now fully) feasible matching that maximizes total surplus can only have one matched pair and one single; and the pair that should be matched clearly consists of individuals 1 and 3:

$$\bar{\mu} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

Obviously, this implies that the symmetry constraint is binding; as we shall see below, this is characteristic of situations in which the roommate matching problem with transferable utility does not have a stable matching. Indeed, we prove in the next section that *a stable matching exists if and only if the symmetry constraint does not bind*.

Now take the “cloned” version of Example 2, in which each type x has two individuals. It is easy to see that the solution to the relaxed problem which neglects the symmetry constraint is the μ^* of section 2.2, which is symmetric; therefore the symmetry constraint does not bind, and a stable matching exists. This is a general result: we shall see below that in any

cloned roommate matching setup, at least one solution to the relaxed problem is symmetric—which implies the existence of a stable match.

2.4 A Bipartite Interpretation

The relaxed problem, in turn, has a natural interpretation in terms of bipartite matching. Start from the three-agent Example 2, and define an *associated bipartite matching problem* as follows: clone the population again, but this time assign a label (such as “man” or “woman”) to each of the two subpopulations. Then consider the bipartite matching problem between these the subpopulations of “men” and “women”, with the joint surplus matrix given by Φ' in (2).

By standard results, there always exists a stable matching in this associated bipartite matching problem; and it maximizes the associated total surplus. In our example, μ^1 and μ^2 are the two stable matchings. Any convex combination such as μ^m can be interpreted as a randomization between these two matchings; it is natural to focus on μ^m since it is the only symmetric one and feasible roommate matchings must be symmetric. As remarked above, In the original roommate problem μ^m cannot be stable, since it has non-integer elements; and it cannot be interpreted as the product of randomization since it is *not* a convex combination of feasible roommate matching matrices⁴.

Now if the roommate matching problem is cloned we can proceed as in the above paragraph, except that with twice the number of individuals we should work with $2\mu^m$. As an integer symmetric matrix, reinterpreted in the cloned roommate matching setup, it defines a feasible roommate matching which is stable—in fact it is the stable matching μ^* of section 2.2. This construction is general: we shall see below than any roommate matching problem in which the number of individuals in each type is even has a symmetric stable match. In addition, when the initial population is “large” enough, while a stable match may still fail to exist, there exists allocations that are “almost” stable.

We now provide a formal derivation of these results.

⁴For any stable roommate matching matrix, the sum of coefficients equals 2, reflecting the fact that one agent must remain single. This property is preserved by convex combination; however, the sum of coefficients of μ^m equals 3.

3 The Formal Setting

We consider a population of individuals who belong to a finite set of types \mathcal{X} . Individuals of the same type are indistinguishable. We denote n_x the number of individuals of type $x \in \mathcal{X}$, and

$$N = \sum_{x \in \mathcal{X}} n_x$$

the total size of the population.

Without loss of generality, we normalize the utilities of singles to be zero throughout.

3.1 Roommate Matching

A match consists of two partners of types x and y . An individual of any type can be matched with any individual of the same or any other type, or remain single. In particular, there is no restriction that matches only involve two partners of different “genders.”

Let a match $\{x, y\}$ generate a surplus Φ_{xy} . In principle the two partners could play different roles. In sections 3 and 4 we will assume that they are in fact symmetric within a match, so that Φ_{xy} is assumed to be a symmetric function of (x, y) :

Assumption 1 *The surplus Φ_{xy} is symmetric in (x, y) .*

We show in section 5 that, surprising as it may seem, there is in fact no loss of generality in making this assumption. The intuition is simple: if Φ_{xy} fails to be symmetric in (x, y) , so that the partners’ roles are not exchangeable, then they should choose their roles so to maximize output. This boils down to replacing Φ_{xy} with the symmetric $\max(\Phi_{xy}, \Phi_{yx})$. Thus our results extend easily when we do not impose Assumption 1; but it is easier to start from the symmetric case.

A matching can be described by a matrix of numbers (μ_{xy}) indexed by $x, y \in \mathcal{X}$, such that

- μ_{x0} is the number of singles of type x
- when $y \neq 0$, μ_{xy} is the number of matches between types x and y .

The numbers μ_{xy} should be integers; given Assumption 1, they should be symmetric in (x, y) ; and they should satisfy the scarcity constraints. More precisely, the number of individuals of type x must equal the number μ_{x0} of singles of type x , plus the number of pairs in which only one partner has type x , plus twice the number of pairs in which the two partners are of type x —since such a same-type pair has two individuals of type x .

Finally, the set of *feasible roommate matchings* is

$$\mathcal{P}(n) = \left\{ \mu = (\mu_{xy}) : \begin{pmatrix} 2\mu_{xx} + \sum_{y \neq x} \mu_{xy} \leq n_x \\ \mu_{xy} = \mu_{yx} \\ \mu_{xy} \in \mathbb{N} \end{pmatrix} \right\} \quad (3)$$

3.2 TU stability and optimality

We define an *outcome* (μ, u) as the specification of a feasible roommate matching μ and an associated vector of payoffs u_x to each individual of type x . These payoffs have to be feasible: that is, the sum of payoffs across the population has to be equal to the total output under the matching μ . Now in a roommate matching μ , the total surplus created is⁵

$$\sum_x \mu_{xx} \Phi_{xx} + \sum_{x \neq y} \mu_{xy} \frac{\Phi_{xy}}{2}. \quad (4)$$

This leads to the following definition of a feasible outcome: an outcome (μ, u) is *feasible* if μ is a feasible roommate matching and

$$\sum_{x \in \mathcal{X}} n_x u_x = \sum_{x \in \mathcal{X}} \mu_{xx} \Phi_{xx} + \sum_{x \neq y} \mu_{xy} \frac{\Phi_{xy}}{2}. \quad (5)$$

We define stability as in Gale and Shapley (1962): an outcome (μ, u) is stable if it cannot be blocked by an individual or by a pair of individuals. More precisely, an outcome (μ, u) is *stable* if it is feasible, and if for any two types $x, y \in \mathcal{X}$, (i) $u_x \geq 0$, and (ii) $u_x + u_y \geq \Phi_{xy}$. By extension, a matching μ is called stable if there exists a payoff vector (u_x) such that the outcome (μ, u) is stable.

⁵Note that in the second sum operator the pair $\{x, y\}$ appears twice, one time as (x, y) and another time as (y, x) ; but the joint surplus Φ_{xy} it creates must only be counted once, hence the division by 2.

In bipartite matching the problem of stability is equivalent to the problem of *optimality*: stable matchings maximize total surplus. Things are obviously more complicated in roommate matchings—there always exist surplus-maximizing matchings, but they may not be stable. The maximum of the aggregate surplus over the set of feasible roommate matchings $\mathcal{P}(n)$ is

$$\begin{aligned} \mathcal{W}_{\mathcal{P}}(n, \Phi) &= \max \left(\sum_x \mu_{xx} \Phi_{xx} + \sum_{x \neq y} \mu_{xy} \frac{\Phi_{xy}}{2} \right) & (6) \\ \text{s.t.} \quad & 2\mu_{xx} + \sum_{y \neq x} \mu_{xy} \leq n_x \\ & \mu_{xy} = \mu_{yx} \\ & \mu_{xy} \in \mathbb{N}. \end{aligned}$$

It may not be attainable in a decentralized manner, but it plays an important role in our argument.

3.3 The Associated Bipartite Matching Problem

We shall now see that to every roommate matching problem we can associate a bipartite matching problem which generates almost the same level of aggregate surplus. More precisely, we will prove that for every vector of populations of types $n = (n_x)$ and every symmetric surplus function $\Phi = (\Phi_{xy})$, the highest possible surplus in the roommate matching problem is “close to” that achieved in a bipartite problem with mirror populations of men and women and half the surplus function:

$$\mathcal{W}_{\mathcal{P}}(n, \Phi) \simeq \mathcal{W}_{\mathcal{B}}(n, n, \Phi/2).$$

where $\mathcal{W}_{\mathcal{B}}(n, n, \Phi/2)$ is defined as the maximal surplus of the bipartite matching problem:

$$\mathcal{W}_{\mathcal{B}}(n, n, \Phi/2) = \max_{\nu \in \mathcal{B}(n, n)} \sum_{x, y \in \mathcal{X}} \nu_{xy} \frac{\Phi_{xy}}{2} \quad (7)$$

with $\mathcal{B}(n, n)$ the set of feasible matchings in the bipartite problem:

$$\mathcal{B}(n, n) = \left\{ \nu = (\nu_{xy}) : \begin{pmatrix} \sum_y \nu_{xy} \leq n_x \\ \sum_x \nu_{xy} \leq n_y \\ \nu_{xy} \in \mathbb{N} \end{pmatrix} \right\} \quad (8)$$

We also define stability for a feasible bipartite matching (ν_{xy}) in the usual way: there must exist payoffs (u_x, v_y) such that

$$\begin{aligned} \sum_{x,y \in \mathcal{X}} \nu_{xy} \frac{\Phi_{xy}}{2} &= \sum_{x \in \mathcal{X}} n_x u_x + \sum_{y \in \mathcal{X}} n_y v_y & (9) \\ u_x + v_y &\geq \frac{\Phi_{xy}}{2} \\ u_x \geq 0, & \quad v_y \geq 0 \end{aligned}$$

By classical results of Shapley and Shubik (1971), there exist stable matchings ν , and they coincide with the solutions of (7). Moreover, the associated payoffs (u, v) solve the dual program; that is, they minimize $\sum_{x \in \mathcal{X}} n_x u_x + \sum_{y \in \mathcal{X}} n_y v_y$ over the feasible set of program (9). Finally, for any stable matching, $\mu_{xy} > 0$ implies $u_x + v_y = \Phi_{xy}/2$, and $\mu_{x0} > 0$ implies $u_x = 0$.

3.3.1 Links Between $\mathcal{W}_{\mathcal{P}}$ and $\mathcal{W}_{\mathcal{B}}$

It is not hard to see that $\mathcal{W}_{\mathcal{P}}(n, \Phi) \leq \mathcal{W}_{\mathcal{B}}(n, n, \Phi/2)$. In fact, we can bound the difference between these two values:

Theorem 1 *Under Assumption 1,*

$$\mathcal{W}_{\mathcal{P}}(n, \Phi) \leq \mathcal{W}_{\mathcal{B}}(n, n, \Phi/2) \leq \mathcal{W}_{\mathcal{P}}(n, \Phi) + |\mathcal{X}|^2 \bar{\Phi}$$

where

$$\bar{\Phi} = \sup_{x,y \in \mathcal{X}} \Phi_{xy}.$$

and $|\mathcal{X}|$ is the cardinal of the set \mathcal{X} , i.e. the number of types in the population.

Proof. See appendix. ■

In some cases, $\mathcal{W}_{\mathcal{P}}(n, \Phi)$ and $\mathcal{W}_{\mathcal{B}}(n, n, \Phi/2)$ actually coincide. For instance:

Proposition 2 *If n_x is even for each $x \in \mathcal{X}$, then under Assumption 1,*

$$\mathcal{W}_{\mathcal{P}}(n, \Phi) = \mathcal{W}_{\mathcal{B}}(n, n, \Phi/2).$$

Proof. See appendix. ■

3.3.2 Stable Roommate Matchings

The existence of stable roommate matchings is directly related to the divergence of $\mathcal{W}_{\mathcal{P}}(n, \Phi)$ and $\mathcal{W}_{\mathcal{B}}(n, n, \Phi/2)$. Indeed, our main result is the following:

Theorem 3 *Under Assumption 1,*

(i) *There exist stable roommate matchings if and only if*

$$\mathcal{W}_{\mathcal{P}}(n, \Phi) = \mathcal{W}_{\mathcal{B}}(n, n, \Phi/2).$$

(ii) *Whenever they exist, stable roommate matchings achieve the maximal aggregate surplus $\mathcal{W}_{\mathcal{P}}(n, \Phi)$ in (6).*

(iii) *Whenever a stable roommate matching exists, individual utilities at equilibrium (u_x) solve the following, dual program:*

$$\begin{aligned} \min_{u, A} \quad & \sum_x u_x n_x & (10) \\ \text{s.t.} \quad & u_x \geq 0 \\ & u_x + u_y \geq \Phi_{xy} + A_{xy} \\ & A_{xy} = -A_{yx} \end{aligned}$$

Proof. See appendix. ■

In program (10), the antisymmetric matrix A has a natural interpretation: A_{xy} is the Lagrange multiplier of the symmetry constraints $\mu_{xy} = \mu_{yx}$ in the initial program (6). Our proof shows that if $\mu_{xy} > 0$ in a stable roommate matching, then the corresponding A_{xy} must be non-positive; but since $\mu_{yx} = \mu_{xy}$ the multiplier A_{yx} must also be non-positive, so that both must be zero. The lack of existence of a stable roommate matching is therefore intimately linked to a binding symmetry constraint.

Given Proposition 2, Theorem 3 has an immediate corollary: with an even number of individuals per type, there must exist a stable roommate matching. Formally:

Corollary 3.1 *If n_x is even for each $x \in \mathcal{X}$, then under Assumption 1, there exists a stable roommate matching.*

In particular, for any roommate matching problem, its “cloned” version, in which each agent has been replaced with a couple of clones, has a stable matching; and this holds irrespective of the surplus generated by the matching of two identical individuals. Of course, in general much less than full

cloning is needed to restore existence. Take any roommate matching problem, consider the categories in which the number of individuals is odd, and add exactly one individual in each of them; by Corollary 3.1, the resulting roommate matching problem has a stable matching.

4 Matching in Large Numbers

We now consider the case of a “large” game, in which there are “many” agents *of each type*. Intuitively, even though an odd number of agents in any type may result in non existence of a stable roommate matching, the resulting game becomes “close” to one in which a stable matching exists (just add or remove one agent in each odd type.) We now flesh out this intuition by providing a formal analysis. We first consider the surplus maximization problems, then we turn to stability issues.

4.1 Surplus Maximization

We start with a formal definition of a large game. For that purpose, we consider a sequence of games with the same number of types and the same surplus matrix, but with increasing populations in each type. If n_x^k denotes the population of type x in game k and $N^k = \sum_x n_x^k$ is the total population of that game, then we consider situations in which, when $k \rightarrow \infty$:

$$N^k \rightarrow \infty \text{ and } n_x^k/N^k \rightarrow f_x$$

As the population gets larger, aggregate surplus increases proportionally; it is therefore natural to consider the *average surplus*, computed by dividing aggregate surplus by the size of the population. We also extend the definition of \mathcal{W}_B in program (7) to non-integers in the obvious way so as to define the limit average bipartite problem $\mathcal{W}_B(f, f, \Phi/2)$. Note that the linearity of the program implies

$$\mathcal{W}_B(cn, cm, \Phi/2) = c\mathcal{W}_B(n, m, \Phi/2)$$

for any $c > 0$.

Theorem 4 *In the large population limit, under assumption 1, the average surplus in the roommate matching problem converges to the limit average surplus in the related bipartite matching problem. That is,*

$$\lim_{N \rightarrow \infty} \frac{\mathcal{W}_P(n, \Phi)}{N} = \lim_{N \rightarrow \infty} \frac{\mathcal{W}_B(n, n, \Phi/2)}{N} = \mathcal{W}_B(f, f, \Phi/2).$$

Proof. See appendix. ■

The marriage problem obviously is a particular case of the roommate problem: if in a roommate matching problem $\Phi_{xy} = -\infty$ whenever x and y have the same gender, then each realized match will be heterosexual. Our results show that given a symmetric surplus⁶, the converse approximately holds—and it holds exactly in the large population limit.

4.2 Stability

The next result shows that one can restore the existence of a stable matching by removing a bounded number of individuals from a population; if these individuals have to be compensated for leaving the game, this can be done at a bounded total cost (so that the per capita cost goes to zero):

Theorem 5 (Approximate stability) *Under Assumption 1, in a population of N individuals, there exists a subpopulation of at least $N - |\mathcal{X}|$ individuals among which there exist a stable matching, where $|\mathcal{X}|$ is the number of types. The total cost for the regulator to compensate the individuals left aside is bounded above by $|\mathcal{X}|\bar{\Phi}$.*

Proof. See appendix. ■

As a direct consequence, consider the large population limit of section 4.1. When the number of individuals becomes much larger than the number of types,

Theorem 6 *In the large population limit and under Assumption 1,*

(i) one may remove a subpopulation of asymptotically negligible size in order to restore the existence of stable matchings.

(ii) the average cost per individual of restoring the existence of stable matchings tends to zero.

Proof. See appendix. ■

Our approximation results crucially rely on the number of types becoming small relative to the total number of individuals. By definition, two individuals of the same type are indistinguishable in our formulation, both

⁶Section 5 extends our results to non-symmetric surplus.

in their preferences and in the way potential partners evaluate them. This may seem rather strong; however, a closer look at the proof of Theorem 4 shows that our bound can easily be refined. In particular, we conjecture that with a continuum of types Theorem 4 would hold exactly.

5 The Nonexchangeable Roommate Problem⁷

We now investigate what happens when the surplus Φ_{xy} is not necessarily symmetric. This will arise when the roles played by the partners are not exchangeable. For instance, a pilot and a copilot on a commercial airplane have dissymmetric roles, but may be both chosen from the same population. Hence, in this section, we shall assume away Assumption 1, and we refer to the “nonexchangeable roommate problem”; it contains the exchangeable problem as a special case.

As it turns out, this can be very easily recast in the terms of an equivalent symmetric roommate problem. Indeed if $\Phi_{xy} > \Phi_{yx}$, then any match of an (ordered) 2-uple (y, x) will be dominated by a matching of a (x, y) 2-uple, and the partners may switch the roles they play and generate more surplus. Therefore, in any optimal (or stable) solution there cannot be such a (y, x) 2-uple. As a consequence, the nonexchangeable roommate problem is equivalent to an exchangeable problem where the surplus function is equal to the maximum joint surplus x and y may generate together, that is

$$\Phi'_{xy} = \max(\Phi_{xy}, \Phi_{yx});$$

and since this is symmetric our previous results apply almost directly. Denoting π_{xy} the number of (x, y) pairs (in that order), one has

$$\begin{aligned}\mu_{xy} &= \pi_{xy} + \pi_{yx}, \quad x \neq y \\ \mu_{xx} &= \pi_{xx}\end{aligned}$$

and obviously, π_{xy} need not equal π_{yx} . The population count equation is

$$n_x = \sum_{y \in \mathcal{X}} (\pi_{xy} + \pi_{yx}), \quad \forall x \in \mathcal{X}$$

and the social surplus from a matching π is

$$\sum_{x, y \in \mathcal{X}} \pi_{xy} \Phi_{xy}.$$

⁷We are grateful to Arnaud Dupuy for correcting a mistake in a preliminary version of the paper.

so that the optimal surplus in the nonexchangeable problem is

$$\begin{aligned} \mathcal{W}'_{\mathcal{P}}(n, \Phi) &= \max \sum_{x,y \in \mathcal{X}} \pi_{xy} \Phi_{xy} \\ \text{s.t. } n_x &= \sum_{y \in \mathcal{X}} (\pi_{xy} + \pi_{yx}), \quad \forall x \in \mathcal{X}. \end{aligned}$$

The following result extends our previous analysis to the nonexchangeable setting:

Theorem 7 *The nonexchangeable roommate matching problem is solved by considering the surplus function*

$$\Phi'_{xy} = \max(\Phi_{xy}, \Phi_{yx})$$

which satisfies Assumption 1. Call optimized symmetric problem the problem with surplus Φ'_{xy} and population count n_x . Then:

(i) the optimal surplus in the nonexchangeable roommate problem coincides with the optimal surplus in the corresponding optimized symmetric problem, namely

$$\mathcal{W}'_{\mathcal{P}}(n, \Phi) = \mathcal{W}_{\mathcal{P}}(n, \Phi')$$

(ii) the nonexchangeable roommate problem has a stable matching if and only if the optimized symmetric problem has a stable matching.

Given Theorem 7, all results in Sections 3 and 4 hold in the general (nonexchangeable) case. In particular:

- Theorem 1 extends to the general case: the social surplus in the roommate problem with asymmetric surplus Φ_{xy} is approximated by a bipartite problem with surplus function $\Phi'_{xy} = \max(\Phi_{xy}, \Phi_{yx})/2$, or more formally:

$$\mathcal{W}'_{\mathcal{P}}(n, \Phi) \leq \mathcal{W}_{\mathcal{B}}(n, n, \Phi'/2) \leq \mathcal{W}'_{\mathcal{P}}(n, \Phi) + |\mathcal{X}|^2 \bar{\Phi},$$

and as an extension of Proposition 2, equality holds in particular when the number of individuals in each types are all even.

- Theorem 3 extends as well: there is a stable matching in the roommate problem with asymmetric surplus Φ_{xy} if and only if there is equality in the first equality above, that is:

$$\mathcal{W}'_{\mathcal{P}}(n, \Phi) = \mathcal{W}_{\mathcal{B}}(n, n, \Phi'/2).$$

- All the asymptotic results in Section 4 hold true: in the asymmetric roommate problem, there is approximate stability and the optimal matching solves a linear programming problem.

6 Stability under Non-transferable Utility

Finally, and for the sake of completeness, we briefly consider the NTU case. Our “cloning” technique still works: any “cloned” NTU roommate matching game has a stable roommate matching. Take Example 1 for instance. Clone types so that there are two individuals of each type, with preferences over types as in the example; and assume for simplicity that clones of each type prefer all other types to each other. Then the following is one among several stable roommate matchings in the cloned game:

- one individual of type 1 matches with a type 2
- the other individual of type 1 matches with a type 3
- the two remaining individuals (a type 2 and a type 3) match.

This is an instance of a much more general result. Formally:

Theorem 8 *Consider \mathcal{X} a set of individuals. Let $\bar{\mathcal{X}}$ be the cloned set, where there are now two identical copies of individuals in \mathcal{X} , with identical preferences over other types. Then whatever the preferences of each individuals with respect to her identical twin, there is a stable roommate matching in $\bar{\mathcal{X}}$.*

Proof. See appendix. ■

Once again, this result would be trivial if we assumed that each individual prefers her clone to all other individuals. What is interesting is that these preferences can be set arbitrarily; in particular, each individual may regard her clone as an unacceptable partner.

7 Conclusion

Some roommate problems involve extensions to situations where more than two partners can form a match; but the two-partner case is a good place to start the analysis. Here, we have shown that when the population is large enough with respect to the number of observable types, the structure of the roommate problem is the same as the structure of the bipartite matching

problem. As a result, the empirical tools developed in the bipartite setting, especially for the analysis of the marriage markets (such as Choo and Siow (2006), Chiappori, Salanié, and Weiss (2010), Fox (2010), Galichon and Salanié (2011), to cite only a few⁸) can be extended to other contexts where the bipartite constraint is relaxed. These include marriage markets incorporating single-sex households, tickets in US presidential elections, team jobs such as pilot/copilot, team sports and many others.

References

- BALINSKI, M. (1970): “On maximum matchings, minimum coverings, and their connections,” in *Proc. of the Princeton Symposium on Mathematical Programming*, ed. by H. Kuhn. Princeton University Press.
- BECKER, G. (1973): “A theory of marriage, part I,” *Journal of Political Economy*, 81, 813–846.
- (1991): *A Treatise on the Family*. Harvard University Press.
- CHIAPPORI, P.-A., B. SALANIÉ, AND Y. WEISS (2010): “Partner Choice and the Marital College Premium,” mimeo Columbia University.
- CHOO, E., AND A. SIOW (2006): “Who Marries Whom and Why,” *Journal of Political Economy*, 114, 175–201.
- CHUNG, K.-S. (2000): “On the Existence of Stable Roommate Matchings,” *Games and Economic Behavior*, 33, 206–230.
- FOX, J. (2010): “Identification in Matching Games,” *Quantitative Economics*, 1, 203–254.
- GALE, D., AND L. SHAPLEY (1962): “College Admissions and the Stability of Marriage,” *American Mathematical Monthly*, 69, 9–14.
- GALICHON, A., AND B. SALANIÉ (2011): “Cupid’s Invisible Hand: Social Surplus and Identification in Matching Models,” mimeo Columbia University.
- GRAHAM, B. (2011): “Econometric Methods for the Analysis of Assignment Problems in the Presence of Complementarity and Social Spillovers,” in *Handbook of Social Economics*, ed. by J. Benhabib, A. Bisin, and M. Jackson. Elsevier.

⁸Graham (2011) has a good discussion of this burgeoning literature.

- GUSFIELD, D., AND R. IRVING (1989): *The Stable Marriage Problem: Structure and Algorithms*. MIT Press.
- IRVING, R. (1985): “An Efficient Algorithm for the ‘Stable Roommates’ Problem,” *Journal of Algorithms*, 6, 577–595.
- KLAUS, B., AND F. KLIJN (2010): “Smith and Rawls Share a Room: Stability and Medians,” *Social Choice and Welfare*, 35, 647–667.
- KOOPMANS, T. C., AND M. BECKMANN (1957): “Assignment Problems and the Location of Economic Activities,” *Econometrica*, 25, 53–76.
- SHAPLEY, L., AND M. SHUBIK (1971): “The assignment game I: the core,” *International Journal of Game Theory*, 1, 111–130.
- TAN, J. (1991): “A Necessary and Sufficient Condition for the Existence of a Complete Stable Matching,” *Journal of Algorithms*, 12, 154–178.

A Appendix: Proofs

Our proofs use an auxiliary object: the highest possible surplus for a *fractional* roommate matching, namely

$$\mathcal{W}_{\mathcal{F}}(n, \Phi) = \max_{\mu \in \mathcal{F}(n)} \left(\sum_x \mu_{xx} \Phi_{xx} + \sum_{x \neq y} \mu_{xy} \frac{\Phi_{xy}}{2} \right). \quad (11)$$

where $\mathcal{F}(n)$ is the set of *fractional (roommate) matchings*, which relaxes the integrality constraint on μ :

$$\mathcal{F}(n) = \left\{ (\mu_{xy}) : \begin{pmatrix} 2\mu_{xx} + \sum_{y \neq x} \mu_{xy} \leq n_x \\ \mu_{xy} = \mu_{yx} \\ \mu_{xy} \geq 0 \end{pmatrix} \right\}. \quad (12)$$

The program (11) has no immediate economic interpretation since fractional roommate matchings are infeasible in the real world; and while obviously $\mathcal{W}_{\mathcal{P}}(n, \Phi) \leq \mathcal{W}_{\mathcal{F}}(n, \Phi)$, the inequality in general is strict. We are going to show, however, that the difference between the two programs vanishes when the population becomes large. Moreover, we will establish a link between (11) and the surplus at the optimum of the associated bipartite matching problem.

We start by proving:

Lemma A.1

$$\mathcal{W}_{\mathcal{F}}(n, \Phi) = \mathcal{W}_{\mathcal{B}}(n, n, \Phi/2). \quad (13)$$

Moreover, problem (11) has a half-integral solution.

Proof of Lemma A.1. First consider some fractional roommate matching $\mu \in \mathcal{F}(n)$, and define

$$\begin{aligned} \nu_{xy} &= \mu_{xy} \text{ if } x \neq y \\ \nu_{xx} &= 2\mu_{xx}. \end{aligned}$$

As a (possibly fractional) bipartite matching, clearly $\nu \in \mathcal{B}(n, n)$; and

$$\sum_x \mu_{xx} \Phi_{xx} + \sum_{x \neq y} \mu_{xy} \frac{\Phi_{xy}}{2} = \frac{1}{2} \sum_{x, y \in \mathcal{X}} \nu_{xy} \Phi_{xy}.$$

Now the right-hand side is the aggregate surplus achieved by ν in the bipartite matching problem with margins (n, n) and surplus function $\Phi/2$. It follows that

$$\mathcal{W}_{\mathcal{F}}(n, \Phi) \leq \mathcal{W}_{\mathcal{B}}(n, n, \Phi/2). \quad (14)$$

Conversely, let (ν_{xy}) maximize aggregate surplus over $\mathcal{B}(n, n)$ with surplus $\Phi/2$. By symmetry of Φ , (ν_{yx}) also is a maximizer; and since (7) is a linear program, $\nu'_{xy} = \frac{\nu_{xy} + \nu_{yx}}{2}$ also maximizes it. Define

$$\begin{aligned}\mu'_{xy} &= \nu'_{xy} \text{ if } x \neq y \\ \mu'_{xx} &= \frac{\nu_{xx}}{2}.\end{aligned}$$

Then

$$\begin{aligned}2\mu'_{xx} + \sum_{y \neq x} \mu'_{xy} &= \nu_{xx} + \frac{1}{2} \sum_{y \neq x} (\nu_{xy} + \nu_{yx}) \\ &= \frac{1}{2} (\nu_{xx} + \sum_{y \neq x} \nu_{xy}) \\ &\quad + \frac{1}{2} (\nu_{xx} + \sum_{y \neq x} \nu_{yx}).\end{aligned}$$

Now $\nu_{xx} + \sum_{y \neq x} \nu_{xy} \leq n_x$ by the scarcity constraint of “men” of type x , and $\nu_{xx} + \sum_{y \neq x} \nu_{yx} \leq n_x$ by the scarcity constraint of “women” of type x . It follows that $\mu' \in \mathcal{F}(n)$, and

$$\sum_x \mu'_{xx} \Phi_{xx} + \sum_{x \neq y} \mu'_{xy} \frac{\Phi_{xy}}{2} = \frac{1}{2} \sum_{x, y \in \mathcal{X}} \nu_{xy} \Phi_{xy}.$$

Therefore the values of the two programs coincide.

Half-integrality follows from the Birkhoff-von Neumann theorem: there always exists an integral solution ν of the associated bipartite matching problem, and the construction of μ' makes it half-integral⁹. ■

Given Lemma A.1, we can now prove Theorem 1.

Proof of Theorem 1. The first inequality simply follows from the fact that $\mathcal{P}(n) \subset \mathcal{F}(n)$. Let us now show the second inequality. Lemma A.1 proved that $\mathcal{W}_{\mathcal{F}}(n, \Phi) = \mathcal{W}_{\mathcal{B}}(n, n, \Phi/2)$. Let μ achieve the maximum in $\mathcal{W}_{\mathcal{F}}(n, \Phi)$, so that

$$\mathcal{W}_{\mathcal{F}}(n, \Phi) = \sum_x \mu_{xx} \Phi_{xx} + \sum_{x \neq y} \mu_{xy} \frac{\Phi_{xy}}{2}.$$

⁹The half-integrality of the solution of problem (11) also follows from a general theorem of Balinski (1970); but the proof presented here is self-contained.

Let $\lfloor x \rfloor$ denote the floor rounding of x ; by definition, $x < \lfloor x \rfloor + 1$, so that

$$\mathcal{W}_{\mathcal{F}}(n, \Phi) < \sum_x \lfloor \mu_{xx} \rfloor \Phi_{xx} + \sum_{x \neq y} \lfloor \mu_{xy} \rfloor \frac{\Phi_{xy}}{2} + \sum_x \Phi_{xx} + \sum_{x \neq y} \frac{\Phi_{xy}}{2}.$$

The right-hand side can also be rewritten as

$$\sum_{x,y} \lfloor \mu_{xy} \rfloor \Phi_{xy} + \sum_{x,y} \Phi_{xy}.$$

But $\lfloor \mu \rfloor$ is in $\mathcal{B}(n, n)$, and is integer by construction; therefore

$$\sum_{x,y \in \mathcal{X}} \lfloor \mu_{xy} \rfloor \Phi_{xy} \leq \mathcal{W}_{\mathcal{P}}(n, \Phi).$$

Finally,

$$\sum_{x,y \in \mathcal{X}} \Phi_{xy} \leq |\mathcal{X}|^2 \bar{\Phi}$$

so that

$$\mathcal{W}_{\mathcal{F}}(n, \Phi) \leq \mathcal{W}_{\mathcal{P}}(n, \Phi) + |\mathcal{X}|^2 \bar{\Phi}.$$

■

A.1 Proof of Proposition 2

Proof. Let $n'_x = \frac{n_x}{2}$. By Lemma A.1, problem $\mathcal{W}_{\mathcal{F}}(n', \Phi)$ has an half-integral solution μ' ; therefore problem $\mathcal{W}_{\mathcal{F}}(n, \Phi)$ has an integral solution $2\mu'$, which must also solve (7). It follows that

$$\mathcal{W}_{\mathcal{P}}(n, \Phi) = \mathcal{W}_{\mathcal{F}}(n, \Phi).$$

■

A.2 Proof of Theorem 3

Proof. By Theorem A.1, Problem (11) coincides with a bipartite matching problem between marginal (n_x) and itself. By well-known results on bipartite matching, there exist vectors (v_x) and (w_y) such that

$$\begin{aligned} v_x &\geq 0, & w_y &\geq 0 \\ v_x + w_y &\geq \Phi_{xy} \end{aligned}$$

and the latter inequality is an equality when $\mu_{xy} > 0$. Setting

$$u_x = \frac{v_x + w_x}{2}$$

the symmetry of Φ implies

$$\begin{aligned} u_x &\geq 0 \\ u_x + u_y &\geq \Phi_{xy} \end{aligned}$$

and

$$\sum_{x \in \mathcal{X}} n_x u_x = \sum_{x \in \mathcal{X}} \mu_{xx} \Phi_{xx} + \sum_{x \neq y} \mu_{xy} \frac{\Phi_{xy}}{2}$$

so that the outcome (μ, u) is stable.

Conversely, assume that μ is a stable roommate matching. Then by definition, there is a vector (u_x) such that

$$\begin{aligned} u_x &\geq 0 \\ u_x + u_y &\geq \Phi_{xy} \end{aligned}$$

and

$$\sum_{x \in \mathcal{X}} n_x u_x = \sum_{x \in \mathcal{X}} \mu_{xx} \Phi_{xx} + \sum_{x \neq y} \mu_{xy} \frac{\Phi_{xy}}{2}.$$

Therefore $(u, A = 0)$ are Lagrange multipliers for the linear programming problem (11), and μ is an optimal solution of (11); finally, μ is integral since it is a feasible roommate matching. QED.

(i), (ii) and (iii) follow, as there exist integral solutions of (11) if and only if

$$\mathcal{W}_{\mathcal{P}}(n, \Phi) = \mathcal{W}_{\mathcal{F}}(n, \Phi),$$

and $\mathcal{W}_{\mathcal{F}}(n, \Phi)$ coincides with $\mathcal{W}_{\mathcal{B}}(n, n, \Phi/2)$ from Lemma A.1. ■

A.3 Proof of Theorem 4

Proof. By Theorem 1, in the large population limit

$$\lim_{N \rightarrow \infty} \frac{\mathcal{W}_{\mathcal{P}}(n, \Phi)}{N} = \mathcal{W}_{\mathcal{F}}(f, \Phi)$$

and Lemma A.1 yields the conclusion. ■

A.4 Proof of Theorem 5

Proof. For each type x , remove one individual of type x to the population if n_x is odd. The resulting subpopulation differs from the previous one by at most $|\mathcal{X}|$ individuals, and there is an even number of individuals of each type; hence by Proposition 3.1 there exists a stable matching.

Each individual so picked can be compensated with his payoff u_x . Since $u_x \leq \bar{\Phi}$, the total cost of compensating at most one individual of each type is bounded from above by $|\mathcal{X}|\bar{\Phi}$. ■

A.5 Proof of Theorem 6

Proof. (i) The number of individuals to be removed is bounded from above by $|\mathcal{X}|$, hence its frequency tends to zero as $|\mathcal{X}|/N \rightarrow 0$. (ii) follows from the fact that

$$\frac{\mathcal{W}_{\mathcal{F}}(n, \Phi) - \mathcal{W}_{\mathcal{P}}(n, \Phi)}{N} \rightarrow 0.$$

■

A.6 Proof of Theorem 7

Proof. (i) Consider an optimal solution μ_{xy} to $\mathcal{W}_{\mathcal{P}}(n, \Phi')$. For any pair $x \neq y$ such that $\Phi_{xy} > \Phi_{yx}$, set $\pi_{xy} = \mu_{xy}$, and $\pi_{xy} = 0$ if $\Phi_{xy} < \Phi_{yx}$. If $\Phi_{xy} = \Phi_{yx}$, set π_{xy} and π_{yx} arbitrarily nonnegative integers such that $\pi_{xy} + \pi_{yx} = \mu_{xy}$; set $\pi_{xx} = \mu_{xx}$. Then π is feasible for the optimized symmetric problem, and one has

$$\sum_{x \in \mathcal{X}} \mu_{xx} \Phi'_{xx} + \sum_{x \neq y} \mu_{xy} \frac{\Phi'_{xy}}{2} = \sum_{x, y \in \mathcal{X}} \pi_{xy} \Phi_{xy}$$

so that

$$\mathcal{W}_{\mathcal{P}}(n, \Phi') \leq \mathcal{W}'_{\mathcal{P}}(n, \Phi).$$

Conversely, consider π_{xy} an optimal solution to $\mathcal{W}'_{\mathcal{P}}(n, \Phi)$. First observe that if $\Phi_{xy} < \Phi_{yx}$ then $\pi_{xy} = 0$; otherwise subtracting one from μ_{xy} and adding one to π_{yx} would lead to an improving feasible solution, contradicting the optimality of π . Set

$$\begin{aligned} \mu_{xy} &= \pi_{xy} + \pi_{yx}, \quad x \neq y \\ \mu_{xx} &= \pi_{xx} \end{aligned}$$

so that

$$\sum_{x \in \mathcal{X}} \mu_{xx} \Phi'_{xx} + \sum_{x \neq y} \mu_{xy} \frac{\Phi'_{xy}}{2} = \sum_{x, y \in \mathcal{X}} \pi_{xy} \Phi_{xy}$$

and hence

$$\mathcal{W}'_{\mathcal{P}}(n, \Phi) \leq \mathcal{W}_{\mathcal{P}}(n, \Phi').$$

(ii) Assume there is a stable matching π_{xy} in the nonexchangeable roommate problem. Then if there is a matched pair (x, y) in that order, one cannot have $\Phi_{yx} > \Phi_{xy}$; otherwise the coalition (y, x) would be blocking. Hence one can define

$$\begin{aligned} \mu_{xy} &= \pi_{xy} + \pi_{yx}, \quad x \neq y \\ \mu_{xx} &= \pi_{xx} \end{aligned}$$

and the matching μ is stable in the optimized symmetric problem. Conversely, assume that the matching μ is stable in the optimized symmetric problem. Then it is not hard to see that, defining π from μ as in the first part of (i) above, the matching π is stable in the nonexchangeable roommate problem. ■

A.7 Proof of Theorem 8

Proof. For each individual x in \mathcal{X} , we create two identical copies in $\bar{\mathcal{X}}$ by assigning them a “male” and “female” gender, which we denote (x, m) and (x, f) . One has

$$\bar{\mathcal{X}} = \bigcup_{x \in \mathcal{X}} \{(x, m), (x, f)\}.$$

Consider

$$\begin{aligned} \mathcal{M} &= \bigcup_{x \in \mathcal{X}} \{(x, m)\} \\ \mathcal{F} &= \bigcup_{x \in \mathcal{X}} \{(x, f)\} \end{aligned}$$

the set of “males” and “females”, so that $\bar{\mathcal{X}} = \mathcal{M} \cup \mathcal{F}$. Consider σ the Gale-Shapley “men propose first” stable bipartite matching between \mathcal{M} and \mathcal{F} . This can be seen as a roommate matching on $\bar{\mathcal{X}}$. Let us show that it is stable. Take a pair of individuals in $\bar{\mathcal{X}}$. Then either these individuals have different genders, and they cannot be a blocking pair, as the matching σ

is stable in the bipartite sense. The remaining case to consider is the case when these individuals have the same gender, call them w.l.o.g. (x, m) and (x', m) . Assume that they form a blocking pair. Let $(y, f) = \sigma(x, m)$ be the partner of (x, m) , and $(y', f) = \sigma(x', m)$ be the partner of (x', m) , with the convention that y (resp. y') is \emptyset if x (resp. x') is single. Then one has

$$x' >_x y \text{ and } x >_{x'} y'$$

which implies that $(x', f) \neq (y, f)$, thus (x, m) and (x', f) are not matched. By the stability of the bipartite matching, this implies that (x, m) and (x', f) is not a blocking pair, thus if (y'', m) is the partner of (x', f)

$$\text{either } y \geq_x x' \text{ or } y'' \geq_{x'} x$$

and as the first statement is ruled out, one gets

$$y'' \geq_{x'} x$$

hence

$$y'' >_{x'} y'.$$

Now consider σ^T the bipartite matching that consists in inverting the genders: m becomes f and f becomes m . This is the Gale and Shapley “women propose first” stable matching. One has $\sigma(x', m) = (y', f)$ and $\sigma^T(x', m) = (y'', f)$. As men weakly prefer their partners under the “men propose first” matching, one has

$$y' \geq_{x'} y''$$

which is a contradiction. Hence there can be no blocking pair in the roommate matching problem, and the roommate matching is stable. ■