Chiappori, Pierre-André; Galichon, Alfred; Salanié, Bernard

Working Paper

The Roommate Problem - Is More Stable Than You Think

CESifo Working Paper, No. 4676

Provided in Cooperation with:
Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Chiappori, Pierre-André; Galichon, Alfred; Salanié, Bernard (2014) : The Roommate Problem - Is More Stable Than You Think, CESifo Working Paper, No. 4676, Center for Economic Studies and ifo Institute (CESifo), Munich

This Version is available at:
https://hdl.handle.net/10419/93474

The international platform of Ludwig-Maximilians University's Center for Economic Studies and the Ifo Institute

Working Papers

www.cesifo.org/wp

The Roommate Problem
Is More Stable Than You Think

Pierre-André Chiappori
Alfred Galichon
Bernard Salanié

CESIFO WORKING PAPER NO. 4676
CATEGORY 12: EMPIRICAL AND THEORETICAL METHODS
FEBRUARY 2014

CESifo
Center for Economic Studies & Ifo Institute

CESifo *Working Paper No. 4676*

The Roommate Problem
Is More Stable Than You Think

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 when utility is transferable, 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.

JEL-Code: C780, J120.

Pierre-André Chiappori
Department of Economics
Columbia University
1009A International Affairs Building
420 West 118th St.
USA – New York, NY 10027
pc2167@columbia.edu

Alfred Galichon
Department of Economics
Sciences Po
27 rue Saint-Guillaume
France – 75007 Paris
alfred.galichon@sciencespo.fr

Bernard Salanié
Department of Economics
Columbia University
1132 International Affairs Building
420 West 118th Street
USA – New York, NY 10027
bsalanie@columbia.edu

February 6, 2014
We are grateful to the Editor and a referee for their comments. We also thank Yeon-Koo Che, Navin Kartik, and especially Arnaud Dupuy, Fuhito Kojima and Phil Reny for very useful input. Chiappori gratefully acknowledges financial support from the NSF (award #1124277). 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. Galichon's research has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement no 313699, and from FiME, Laboratoire de Finance des Marchés de l'Energie.

1 Introduction

Since the seminal work of Becker (1973, 1991), the economic literature has mostly modeled the marriage market as a bipartite matching game with transferable utility (TU). Each couple consists of two partners coming each from a separate subpopulation, whom we will call men and women. In the TU framework, any potential couple generates a surplus that is (endogenously) shared by its members. The resulting matching must satisfy a stability property, reflecting robustness to unilateral and bilateral deviations. The theoretical analysis of bipartite matching with transferable utility was introduced by Koopmans and Beckmann (1957). Shapley and Shubik (1971) studied the set of stable matchings under TU; 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 marriage market. 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 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 non-transferable (NTU) setting[1]. A few papers have studied the property of NTU stable matchings when they exist. Gusfield 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). Necessary and sufficient existence conditions under strict preferences have been found by Tan (1991) for complete stable matchings and by Sotomayor (2005) for stable matchings. Chung (2000) shows that a condition he calls "no odd

[1]In spite of the terminology "non-transferable utility," this branch of the literature usually makes no reference to utility, but just to ordinal preferences.

rings" is sufficient for stable matchings to exist under weak preferences.

The TU case has been less studied in the theoretical literature, in spite of its relevance in empirical applications. Chung (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. Karlander and Eriksson (2001) provide a graph-theoretic characterization of stable outcomes when they exist; and Klaus and Nichifor (2010) study their properties. Talman and Yang (2011) give a characterization in terms of integer programming.

We here prove two main results. First, a stable matching always exists when the number of individuals in each type is *even*. Second, when the number of individuals of any given type is large enough, there always exist "quasi-stable" matchings: even if a stable matching does not exist, existence can be restored with minimal policy intervention. To do this, one only needs to convince *one* individual to leave the game in each type with an odd number of individuals. If this requires a compensation to be paid, this can be done at a per capita cost that goes to zero when the population of each type goes to infinity.

The results of this paper are related to those of Azevedo, Weyl, and White (2012) who show the existence of a Walrasian equilibrium in an economy with indivisible goods, a continuum of agents and quasilinear utility. Unlike their main results, ours apply in markets with finite numbers of agents. Our methods are also original. As is well-known, in bipartite problems all feasible matchings that maximize social surplus are stable. This is not true in roommate problems; but we show how any roommate problem can be "cloned" in order to construct an associated bipartite problem. We then exploit this insight to prove existence of stable matchings in roommate problems with even numbers of agents within each type.

To the best of our knowledge, the connection between the unipartite and bipartite problems stressed in this paper is new. Importantly, this implies that the empirical tools devised for bipartite matching should carry over directly to roommate matching when the populations under consideration are large.

2 A Simple Example

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

2.1 Non-existence of Stable Matchings

That a stable matching may not exist for the roommate problem under non-transferable utility has been known since Gale and Shapley. 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 1 *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*

$$\Phi = \begin{pmatrix} - & 6 & 8 \\ 6 & - & 5 \\ 8 & 5 & - \end{pmatrix} \tag{1}$$

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. 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 matched 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 matched 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 matched 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). We conclude that no stable matching exists.*

Note that there is nothing pathological in Example 1. 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 1. 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 1; 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 a type 1 and a type 2 individuals, 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 different[2].

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

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). In the standard, bipartite framework, the adjective "feasible" refers to the fact that each individual can only be matched to one partner or stay single. Roommate matching, however, 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; since a roommate matching for which μ_{xy} and μ_{yx} differ would clearly not be feasible, it must be the case that $\mu_{xy} = \mu_{yx}$. 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 1, 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 a partner of the same type: $\mu_{xx} \equiv 0$; and neglecting symmetry, the only other feasibility constraints are

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

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; moreover, and quite interestingly, it cannot be interpreted as the outcome of randomization since it is *not* a convex combination of feasible roommate matching matrices[3].

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, $\bar{\mu}$ is *not* a solution to the maximization problem without symmetry constraint; in other words, the symmetry constraint is binding in this example. As we shall see below, this is characteristic of situations in which the roommate 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 1, 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.

[3]For 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.

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 1, 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 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.

Now if the roommate 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 that any roommate problem in which the number of individuals in each type is even has a symmetric stable match.

We now provide a formal derivation of these results.

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 we collect them in a vector $n = (n_x)$. We also denote

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

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 who are characterized by their types. An individual of any type can be matched with any individual of the same or any other type, or remain single. Let Φ_{xy} be the surplus generated by a match between partners of types x and y.

In the *roommate problem* (Φ, n), individuals match by pairs and the members of a pair share the surplus generated. Unmatched individuals obtain zero utility.

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).*

As shown in section 5, there is no loss of generality in making this assumption. Indeed, if Φ_{xy} fails to be symmetric in (x, y), then the partners will choose their roles so to maximize output; and our results hold with $\tilde{\Phi}_{xy} = \max(\Phi_{xy}, \Phi_{yx})$ (which is symmetric) instead of Φ_{xy}.

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.

We define the set of *feasible roommate matchings* as follows.

Definition 1 (Feasible roommate matching) *The set of* feasible roommate matchings*, denoted $\mathcal{P}(n)$, is defined as*

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

3.2 Stability and optimality

We define an *outcome of the roommate problem*, or *roommate outcome* (μ, u) as the specification of a feasible roommate matching μ and an associated vector of payoffs u_i to each individual i in a set $\mathcal{I}$. Define x_i as the type of individual i. Then a first remark is that, at any stable matching, individuals of the same type must have the same payoff.[4] Formally:

Lemma 1 *For any stable matching, we have*

$$\forall i, j,\ x_i = x_j \Rightarrow u_i = u_j$$

Proof. *Consider a stable outcome $(\tilde{\mu}, \tilde{u})$. Note, first, that stability implies:*

$$\tilde{u}_i = \max_{j \in \mathcal{I}} \left(\Phi_{x_i y_j} - \tilde{u}_j, 0 \right). \tag{4}$$

since $\tilde{u}_i \geq \max\left(\Phi_{x_i y_j} - \tilde{u}_j, 0 \right)$ for all j by stability, and equality is reached either at 0 (for unmatched individuals) or for the match of individual i. Next, for any type x, define

$$u_x = \max_{j \in \mathcal{I}} \left(\Phi_{x y_j} - \tilde{u}_j, 0 \right) \tag{5}$$

From (4) and (5), we must have $\tilde{u}_i = u_{x_i}$, and the conclusion follows. ■

An important implication of this Lemma is that, since we are only interested in stable matchings, we may restrict to roommate outcomes (μ, u) such that the associated vector of payoffs u_{x_i} is defined exclusively as a function of the individual's type.

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 is[5]

$$S_R(\mu, \Phi) = \sum_x \mu_{xx} \Phi_{xx} + \sum_{x \neq y} \mu_{xy} \frac{\Phi_{xy}}{2}. \tag{6}$$

This leads to the following definition of a feasible outcome:

[4] We thank an anonymous referee for suggesting that we prove Lemma 1.

[5] Note 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.

Definition 2 (Feasible roommate outcome) *A roommate outcome* (μ, u) *is* feasible *if* μ *is a feasible roommate matching and*

$$\sum_{x \in \mathcal{X}} n_x u_x = \sum_x \mu_{xx} \Phi_{xx} + \sum_{x \neq y} \mu_{xy} \frac{\Phi_{xy}}{2}. \tag{7}$$

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:

Definition 3 (Stable roommate outcome) *A roommate outcome* (μ, u) *is* stable *if* (μ, u) *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 roommate matching* μ *is called stable if there exists a payoff vector* (u_x) *such that the outcome* (μ, u) *is stable.*

In bipartite matching, stability is equivalent to *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

$$\begin{aligned} \mathcal{W}_{\mathcal{P}}(\Phi, n) \quad &= \quad \max_{\mu \geq 0} \sum_x \mu_{xx} \Phi_{xx} + \sum_{x \neq y} \mu_{xy} \frac{\Phi_{xy}}{2} \\ s.t. \qquad & \quad 2\mu_{xx} + \sum_{y \neq x} \mu_{xy} \leq n_x \\ & \quad \mu_{xy} = \mu_{yx} \\ & \quad \mu_{xy} \in \mathbb{Z}_+. \end{aligned} \tag{8}$$

While no stable roommate matching may actually achieve this value, it plays an important role in our argument.

3.3 The Associated Bipartite Matching Problem

We shall now see that to every roommate problem we can associate a bipartite matching problem, and that the comparison between these problems provides crucial insights on the existence of a stable matching. We start with a quick reminder of standard, *bipartite matching problems.* These are defined by *two* populations, one described by a vector of type numbers $(n_x)_{x \in \mathcal{X}}$ and the second by a vector of type numbers $(m_y)_{y \in \mathcal{X}}$. Each agent in the first population should be matched with an agent in the second population

or remain single; similarly, each agent in the second population should be matched with an agent in the first population or remain single. Two individuals from the same population cannot match[6]. If an individual of type $x \in \mathcal{X}$ from the first population matches with an individual of type $y \in \mathcal{X}$, they share surplus ϕ_{xy} between them. As before, individuals who remained unmatched get a surplus that can be normalized to zero.

A bipartite matching is a vector $(\nu_{xy})_{x,y\in\mathcal{X}}$ such that ν_{xy} is the number of matches between individuals of type x in the first population and inviduals of type y in the second population. The set of feasible bipartite matchings, denoted $\mathcal{B}(n,m)$, is formally defined as follows:

Definition 4 (Feasible bipartite matching) *The set of* feasible bipartite matchings *in the bipartite problem, denoted $\mathcal{B}(n,m)$, is defined by*

$$\mathcal{B}(n,m) = \left\{ \nu = (\nu_{xy}) : \begin{pmatrix} \sum_y \nu_{xy} \leq n_x \\ \sum_x \nu_{xy} \leq m_y \\ \nu_{xy} \in \mathbb{Z}_+ \end{pmatrix} \right\}. \tag{9}$$

Let $\mathcal{W}_{\mathcal{B}}(\phi,n,m)$ be the maximal total surplus attainable in the bipartite matching problem. We formally define

$$\mathcal{W}_{\mathcal{B}}(\phi,n,m) = \max_{\nu \in \mathcal{B}(n,m)} \sum_{x,y\in\mathcal{X}} \nu_{xy}\phi_{xy}. \tag{10}$$

We define a *bipartite outcome* as a triple (ν,u,v) where ν_{xy} is the number of individuals of type x matched with individuals of type y, u_x is the payoff of any individual of type x, and v_y is the payoff of any individual of type y. A feasible bipartite outcome is defined by:

Definition 5 (Feasible bipartite outcome) *A bipartite outcome (ν,u,v) is* feasible *if $\nu \in B(n,m)$, and*

$$\sum_{x\in\mathcal{X}} n_x u_x + \sum_{y\in\mathcal{X}} m_y v_y = \sum_{x,y\in\mathcal{X}} \nu_{xy}\phi_{xy}.$$

We also define stability of a bipartite outcome (ν,u,v) in the usual way:

[6]The set of types $\mathcal{X}$ is the same for agents of the two populations, but these populations are distinct. For instance, $\mathcal{X}$ is the set of possible ages, and the two populations are men and women.

Definition 6 (Stable bipartite outcome) *A bipartite outcome* (ν, u, v) *is* stable *if* (ν, u, v) *is feasible, and if for any two types* $x, y \in \mathcal{X}$*, (i)* $u_x \geq 0$*, (ii)* $v_y \geq 0$*, and (iii)* $u_x + v_y \geq \phi_{xy}$*. By extension, a bipartite matching* ν *is called stable if there exist payoff vectors* (u, v) *such that bipartite outcome* (ν, u, v) *is stable.*

By classical results of Shapley and Shubik (1971), there exist stable bipartite matchings ν, and they coincide with the solutions of (10). Moreover, the associated payoffs (u, v) solve the dual program; that is, they minimize $\sum_{x \in \mathcal{X}} n_x u_x + \sum_{y \in \mathcal{X}} m_y v_y$ over $u \geq 0$ and $v \geq 0$ subject to constraint $u_x + v_y \geq \phi_{xy}$. Finally, for any stable matching, $\mu_{xy} > 0$ implies $u_x + v_y = \phi_{xy}$, and $\mu_{x0} > 0$ implies $u_x = 0$.

Note that the marriage problem obviously is a particular case of the roommate problem: if in a roommate problem $\Phi_{xy} = -\infty$ whenever x and y have the same gender, and $\Phi_{xy} = \phi_{xy}$ otherwise, then any optimal or stable matching will be heterosexual.

Finally, to any roommate problem we associate a bipartite matching in the following way:

Definition 7 (Associated bipartite matching problem) *For each roommate problem* (Φ, n)*, the associated bipartite matching problem is the bipartite matching problem* (ϕ, n, n) *with*

$$\phi_{xy} = \Phi_{xy}/2.$$

Note that in the associated bipartite matching problem, the first population and the second population have the same distributions of types.

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

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

Theorem 1 *Under Assumption 1,*

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

where

$$\overline{\Phi} = \max_{x,y \in \mathcal{X}} \Phi_{xy}.$$

and $|\mathcal{X}|$ is the cardinality of the set $\mathcal{X}$, i.e. the number of types in the population.
Proof. *See appendix.* ■

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

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

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

Proof. *See appendix.* ■

3.3.2 Stable Roommate Matchings

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

Theorem 2 *Under Assumption 1,*
(i) There exist stable roommate matchings if and only if

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

(ii) Whenever they exist, stable roommate matchings achieve the maximal aggregate surplus $\mathcal{W}_{\mathcal{P}}(\Phi, n)$ in (8).
(iii) Whenever a stable roommate matching exists, individual utilities at equilibrium (u_x) solve the following, dual program:

$$\begin{aligned} \min_{u,A} \quad & \sum_x u_x n_x \\ s.t. \quad & u_x \geq 0 \\ & u_x + u_y \geq \Phi_{xy} + A_{xy} \\ & A_{xy} = -A_{yx} \end{aligned} \tag{11}$$

Proof. See appendix. ■

Note that while the characterization of the existence of a stable matching in terms of equality between an integer program and a linear program is a well-known problem in the literature on matching (see Talman and Yang (2011) in the case of the roommate problem), the link with a bipartite matching problem is new.

Also note that in program (11), 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 (8). 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 1, Theorem 2 has an immediate corollary: with an even number of individuals per type, there must exist a stable roommate matching. Formally:

Corollary 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 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; we give this statement a precise meaning in the next paragraph.

Our next result shows that one can restore the existence of a stable matching by removing at most one individual of each type from the population; if these individuals have to be compensated for leaving the game, this can be done at limited total cost:

Theorem 3 (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 $|\mathcal{X}|\,\overline{\Phi}$.*

Proof. See appendix. ■

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. We now flesh out this intuition by providing a formal analysis.

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 $k \to \infty$:

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

where f_x are constant numbers.

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 $\mathcal{W}_{\mathcal{B}}$ to non-integers vectors of type distributions in the obvious way so as to define the limit average bipartite problem $\mathcal{W}_{\mathcal{B}}(\Phi/2, f, f)$. Note that the linearity of the program implies

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

for any $c > 0$.

Proposition 2 *In the large population limit, under Assumption 1, the average surplus in the roommate problem converges to the limit average surplus in the related bipartite matching problem. That is,*

$$\lim_{k\to\infty} \frac{\mathcal{W}_{\mathcal{P}}\left(\Phi, n^k\right)}{N^k} = \lim_{N^k\to\infty} \frac{\mathcal{W}_{\mathcal{B}}\left(\Phi/2, n^k, n^k\right)}{N^k} = \mathcal{W}_{\mathcal{B}}\left(\Phi/2, f, f\right).$$

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 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 2 shows that our bound can easily be refined. In particular, we conjecture that with a continuum of types, Theorem 2 would hold exactly.

A related effect of the number of individuals becoming much larger than the number of types is that the costs of the policy to restore stability in Theorem 3 become negligible:

Proposition 3 *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. ■

In particular, in the case of a continuum of individuals (that is, when there is a finite number of types and an infinite number of individuals of each type), we recover the results of Azevedo, Weyl, and White (2012) (hereafter, AWW). To make the connection with this paper, the partner types in our setting is replaced by goods in AWW's. The social welfare in our setting translates into the utility u of a single consumer in AWW. u is such that $u(C) = \Phi(\{x, y\})$ for $C = \{x, y\}$, $u(\{x\}) = 0$, and $u = -\infty$ elsewhere (or very negative). Then it can be shown without difficulty that the existence of a TU stable matching in our setting is equivalent to the existence of a Walrasian equilibrium in the AWW setting. Thus existence and TU stability in the case of a continuum of individuals follow from the Theorem and the Proposition in AWW.

A question worthy of interest is whether there is a way for the social planner to restore stability in the roommate problem at no cost[7]. This is obviously possible if we allow a constant per capita tax which applies both to matched and unmatched individuals. If only matched individuals can be taxed, this is a more delicate matter. Indeed, assume a constant tax is levied per matched invidividual. In that case, the incentive to match is distorted, and hence the endogenous surplus share will be likely to be altered. In that case, more agents may have an incentive in exiting the market, and some agents who were intended by the social planner to remain matched may actually prefer to break away from their match and receive the monetary compensation for exiting the market. In this case, it is difficult to qualify the outcome as "stable". As a result, we can think of a new matching problem, where the social planner is required to break even, and it is impossible to levy tax on unmatched individuals.

In this problem, two groups of agents are formed by the social planner. A first group of individuals are compensated by the social planner to leave the market; their compensation may depend on their type. Individuals from a second group participate in a standard roommate matching, and are levied

[7]We thank an anonymous Referee for suggesting the essence of this discussion.

a per capita tax if they are matched. The social planner faces a balanced budget constraint: payments to agents in the first group must be financed by payments from agents in the second group. An outcome specifies (i) which group each agent belongs to; (ii) who is unmatched, and which partner each matched agent has; and (iii) what payoff each agent obtains. An outcome is stable if (a) the roommate problem in the second group is stable in the sense developed in this paper, and (b) no agent from either group would strictly prefer to belong to the other group. The existence of stable outcomes in this problem is an interesting question, which we leave for future exploration.

5 The Asymmetric Roommate Problem[8]

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 dissymmetrical 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 "asymmetric roommate problem"; it contains the symmetric 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-tuple (y, x) will be dominated by a matching of a (x, y) 2-tuple, 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-tuple. As a consequence, the asymmetric roommate problem is equivalent to a symmetric problem where the surplus function is equal to the maximum joint surplus x and y may generate together, that is

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

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

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

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

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

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

and the social surplus from a matching π is

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

so that the optimal surplus in the asymmetric problem is

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

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

Theorem 4 *The asymmetric roommate 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 asymmetric roommate problem coincides with the optimal surplus in the corresponding optimized symmetric problem, namely

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

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

Given Theorem 4, all results in Sections 3 and 4 hold in the general (asymmetric) 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:

  $$\mathcal{W}'_{\mathcal{P}}(\Phi,n) \leq \mathcal{W}_{\mathcal{B}}\left(\phi',n,n\right) \leq \mathcal{W}'_{\mathcal{P}}(\Phi,n) + |\mathcal{X}|^2\,\overline{\Phi},$$

  and as an extension of Proposition 1, equality holds in particular when the number of individuals in each types are all even.

- Theorem 2 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:

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

- 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 Conclusion

While our analysis has been conducted in the discrete case, an interesting extension of our results is to the case where there is an infinite number of agents with a continuum of types. The relevant theory here is the theory of optimal transportation, see Villani (2003) and McCann and Guillen (2010). For the precise connection between matching models and optimal transportation theory, see Ekeland (2010), Gretsky, Ostroy, and Zame (1999) and Chiappori, McCann, and Nesheim (2010). We believe these results could be easily extended to the unipartite matching. It is also worth mentioning that recent papers by Ghoussoub and Moameni (2013) and Galichon and Ghoussoub (2013) use the same type of mathematical structure for the purpose of the study of "2-monotone" and "N-monotone" operators.

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. Most empirical applications of matching models under TU use a framework as in this paper in order to understand, depending on the context, how the sorting on a given matching market depends on age, education or income, but also height, BMI, caste, etc.[9]. As a consequence of our results, the empirical tools developed in the bipartite setting, especially for the analysis of the marriage markets (see Choo and Siow (2006), Chiappori, Salanié, and Weiss (2010), Fox (2010), Galichon and Salanié (2011), to cite only a few[10]) can be extended to other contexts where the bipar-

[9] See for instance Choo and Siow (2006), Hitsch, Hortacsu, and Ariely (2010), Banerjee, Duflo, Ghatak, and Lafortune (2009), Chiappori, Oreffice, and Quintana-Domeque (2009) among many others.

[10] Graham (2011) has a good discussion of this burgeoning literature.

tite constraint is relaxed. These include marriage markets incorporating same-sex households, tickets in US presidential elections, team jobs such as pilot/copilot, team sports and many others.

Finally, it seems natural to apply our "cloning technique" when utility is not transferable. One may think of assigning arbitrarily genders to both clones of each type, and considering a bipartite stable matching between the two genders. Such a matching will be stable in the roommate matching framework if the bipartite matching of the cloned populations is symmetric. However, such a symmetric stable bipartite matching of the cloned population may not exist. Therefore, the usefulness of cloning to restore stability in the non-transferable utility version of the roommate problem is an open question.

References

Azevedo, E., E. Weyl, and A. White (2012): "Walrasian Equilibrium in Large, Quasilinear Markets," *Theoretical Economics*, forthcoming.

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.

Banerjee, A., E. Duflo, M. Ghatak, and J. Lafortune (2009): "Marry for What? Caste and Mate Selection in Modern India," Discussion Paper 14958, NBER.

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., R. McCann, and L. Nesheim (2010): "Hedonic Price Equilibria, Stable Matching, and Optimal Transport: Equivalence, Topology, and Uniqueness," *Economic Theory*, pp. 1–49.

Chiappori, P., S. Oreffice, and C. Quintana-Domeque (2009): "Fatter Attraction: Anthropometric and Socioeconomic Characteristics in the Marriage Market," DP 4594, IZA.

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.

Ekeland, I. (2010): "Notes on Optimal Transportation," *Economic Theory*, 42, 437–459.

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 N. Ghoussoub (2013): "Variational representations for N-cyclically monotone vector fields," *Pacific Journal of Mathematics*, forthcoming.

Galichon, A., and B. Salanié (2011): "Cupid's Invisible Hand: Social Surplus and Identification in Matching Models," mimeo.

Ghoussoub, N., and A. Moameni (2013): "A Self-dual Polar Factorization for Vector Fields," *Comm. Pure. Applied. Math.*, 66, 905–933.

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.

Gretsky, N., J. Ostroy, and W. Zame (1999): "Perfect competition in the continuous assignment model," *Journal of Economic Theory*, 88, 60–118.

Gusfield, D., and R. Irving (1989): *The Stable Marriage Problem: Structure and Algorithms*. MIT Press.

Hitsch, G., A. Hortacsu, and D. Ariely (2010): "Matching and Sorting in Online Dating," *American Economic Review*, 100, 130–163.

Irving, R. (1985): "An Efficient Algorithm for the 'Stable Roommates' Problem," *Journal of Algorithms*, 6, 577–595.

Karlander, J., and K. Eriksson (2001): "Stable outcomes of the roommate game with transferable utility," *International Journal of Game Theory*, 29, 555–569.

Klaus, B., and F. Klijn (2010): "Smith and Rawls Share a Room: Stability and Medians," *Social Choice and Welfare*, 35, 647–667.

Klaus, B., and A. Nichifor (2010): "Consistency in one-sided assignment problems," *Social Choice and Welfare*, 35, 415–433.

Koopmans, T. C., and M. Beckmann (1957): "Assignment Problems and the Location of Economic Activities," *Econometrica*, 25, 53–76.

McCann, R., and N. Guillen (2010): "Five Lectures on Optimal Transportation: Geometry, Regularity and Applications," Discussion paper, University of Toronto.

Shapley, L., and M. Shubik (1971): "The assignment game I: the core," *International Journal of Game Theory*, 1, 111–130.

Sotomayor, M. (2005): "The Roommate Problem Revisited," mimeo U. São Paulo.

Talman, D., and Z. Yang (2011): "A model of partnership formation," *Journal of Mathematical Economics*, 47, 206–212.

Tan, J. (1991): "A Necessary and Sufficient Condition for the Existence of a Complete Stable Matching," *Journal of Algorithms*, 12, 154—178.

Villani, C. (2003): *Topics in Optimal Transportation*. American Mathematical Society.

A Appendix: Proofs

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

$$\mathcal{W}_{\mathcal{F}}(\Phi, n) = \max_{\mu \in \mathcal{F}(n)} \left(\sum_x \mu_{xx} \Phi_{xx} + \sum_{x \neq y} \mu_{xy} \frac{\Phi_{xy}}{2} \right). \tag{12}$$

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

$$\mathcal{F}(n) = \left\{ (\mu_{xy}) : \begin{pmatrix} 2\mu_{xx} + \sum_{y \neq x} \mu_{xy} \leq n_x \\ \mu_{xy} = \mu_{yx} \\ \mu_{xy} \geq 0 \end{pmatrix} \right\}. \tag{13}$$

The program (12) has no immediate economic interpretation since fractional roommate matchings are infeasible in the real world; and while obviously $\mathcal{W}_{\mathcal{P}}(\Phi, n) \leq \mathcal{W}_{\mathcal{F}}(\Phi, n)$, 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 (12) and the surplus at the optimum of the associated bipartite matching problem.

We start by proving:

Lemma 2

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

Moreover, problem (12) has a half-integral solution.

Proof of Lemma 2. 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

$$\mathcal{W}_{\mathcal{F}}(\Phi, n) \leq \mathcal{W}_{\mathcal{B}}(\Phi/2, n, n). \tag{15}$$

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 (10) 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

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

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. Finally, it follows from the Birkhoff-von Neumann theorem that $\mathcal{B}(n,n)$ is the set of extreme points of the set $\left\{\nu \geq 0 : \sum_y \nu_{xy} = n_x; \sum_x \nu_{xy} = n_y\right\}$; thus, $\mathcal{W}_{\mathcal{B}}(\Phi/2, n, n)$ coincides with

$$\max_{\nu\geq 0}\left\{\frac{1}{2}\sum_{x,y\in\mathcal{X}}\nu_{xy}\Phi_{xy} : \sum_y \nu_{xy} = n_x, \ \sum_x \nu_{xy} = n_y\right\}$$

thus there always exists an integral solution ν of the latter program, and the construction of μ' shows that μ' can be taken half-integral[11]. ■

Given Lemma 2, 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 2 states

[11]The half-integrality of the solution of problem (12) also follows from a general theorem of Balinski (1970); but the proof presented here is self-contained.

that $\mathcal{W}_{\mathcal{F}}(\Phi, n) = \mathcal{W}_{\mathcal{B}}(\Phi/2, n, n)$. Let μ achieve the maximum in $\mathcal{W}_{\mathcal{F}}(\Phi, n)$, so that

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

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

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

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 $\mathcal{P}(n)$, and is integer by construction; therefore

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

Finally,

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

so that

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

■

A.1 Proof of Proposition 1

Proof. Let $n'_x = \frac{n_x}{2}$. By Lemma 2, problem $\mathcal{W}_{\mathcal{F}}(\Phi, n')$ has an half-integral solution μ'; therefore problem $\mathcal{W}_{\mathcal{F}}(\Phi, n)$ has an integral solution $2\mu'$, which must also solve (10). It follows that

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

■

A.2 Proof of Theorem 2

Proof. By Lemma 2, Problem (12) coincides with a bipartite matching problem between marginal (n_x) and itself. It follows from Shapley and Shubik (1971) that there exist vectors (v_x) and (w_y) such that

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

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

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

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

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

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 (12), and μ is an optimal solution of (12); finally, μ is integral since it is a feasible roommate matching.

(i), (ii) and (iii) follow, as there exist integral solutions of (12) if and only if

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

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

A.3 Proof of Theorem 3

Proof. For each type x, remove one individual of type x from 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 1 there exists a stable matching.

Each individual so picked can be compensated with u_x, the payoff of individuals of the same type who have not been removed. Since $u_x \leq \overline{\Phi}$, the total cost of compensating at most one individual of each type is bounded from above by $|\mathcal{X}|\,\overline{\Phi}$. ■

A.4 Proof of Proposition 2

Proof. By Theorem 1, in the large population limit

$$\lim_{k\to\infty} \frac{\mathcal{W}_{\mathcal{P}}\left(\Phi, n^k\right)}{N^k} = \mathcal{W}_{\mathcal{F}}\left(\Phi, f\right)$$

and Lemma 2 yields the conclusion. ■

A.5 Proof of Proposition 3

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

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

■

A.6 Proof of Theorem 4

Proof. (i) Consider an optimal solution μ_{xy} to $\mathcal{W}_{\mathcal{P}}\left(\Phi', n\right)$. 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

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

Conversely, consider π_{xy} an optimal solution to $\mathcal{W}'_{\mathcal{P}}(\Phi, n)$. 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

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

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

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

(ii) Assume there is a stable matching π_{xy} in the asymmetric 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

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

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 asymmetric roommate problem. ■