Matching with Real-Life Constraints

Thesis by:

Josué Alberto Ortega Sandoval, M.Sc.

Submitted in fulfillment of the

requirements for the Degree of:

Doctor of Philosophy in Economics

Adam Smith Business School, College of Social Science

University of Glasgow

September 20, 2017

This page is intentionally left blank.

Principal Supervisor:

Prof. Hervé Moulin

Adam Smith Business School, University of Glasgow

Second Supervisor:

Prof. Anna Bogomolnaia

Adam Smith Business School, University of Glasgow

Convener

Prof. Yiannis Vailakis

Adam Smith Business School, University of Glasgow

Internal Examiner:

Dr. Bram Driesen

Adam Smith Business School, University of Glasgow

External Examiner:

Prof. Olivier Tercieux

Paris School of Economics and CNRS

Abstract

This thesis consists of four chapters. The first chapter explains the relevance of the research that has been undertaken and it contains an overview of this research for a general audience. The second chapter studies a multi-unit assignment with endogenous quotas in a dichotomous preference domain. The main conclusion I obtain is that pseudo-market mechanisms perform poorly in this type of environment.

The third and fourth chapters use matching theory to understand segregation in matching environments ranging from integrating kidney exchanges platforms to the increase in interracial marriages after the popularization of online dating platforms. In both Chapters, using different formulations, I show under which conditions social integration can be obtained.

Contents

List of Figures

List of Tables

Acknowledgments

I thank my parents, Oliva Sandoval and Elías Ortega, for their encouragement throughout the Ph.D.

My advisor, Hervé Moulin, has provided me with extensive support to finish this thesis. His insightful comments with respect to the multi-unit assignment problem have been extremely valuable, and the corresponding article I produced has greatly benefited from his sharp advice. I also greatly benefited from attending several courses he taught on game theory, mechanism design, and fair division, and from being his teaching assistant during these three years. He has also been incredibly supportive helping me visit other research institutions. It has been an honor to be his student.

Bram Driesen has also been a terrific mentor. His constant encouragement and support has been a great asset, particularly during the second half of the Ph.D. He has also taught me a lot about the job market, and he has supported me writing way too many recommendation letters for me, like Hervé.

I am also very indebted to Olivier Tercieux for agreeing to be my external examiner. He has read all the thesis chapters with great detail and has provided me with many useful suggestions.

My research on multi-unit assignment has also greatly benefited from the comments from Anna Bogomolnaia, a real expert on the topic. I owe to Takashi Hayashi the inspiration to study social integration. My paper on integration monotonicity in matching is heavily based on his own work. Patrick Harless has given me great advice regarding my work on social networks and about its link with the monotonicity axioms in the cooperative game theory literature. Fedor Sandomirskiy, a constant visitor from HSE, has given me very valuable insights about fair division problems, in particular under specific capacity constraints.

I am also indebted to David Manlove and Baharak Rastegari from the computer science department from their comments regarding many-to-many matching, as well from their kind invitation to collaborate with them organizing the MATCH-UP workshop and to present in their internal seminar.

I spent part of my Ph.D. abroad. From my time at CERGE-EI in Prague, I had very helpful conversations with Vojtěch Bartoš and Filip Matějka. From my time at Columbia, the comments from Agustin Chaintreau, Pierre-André Chiappori, Yash Kanoria, Jacob Leshno, Jay Sethuraman, and Michael Woodford helped me in crucial parts of my research. I am particularly grateful to both Yeon-Koo Che for hosting me, and to Mark Dean, who helped me with my project on rational inattention, allowed me to take his course on behavioral economics, and supported my job market applications with recommendation letters. He definitely attracted my attention to the behavioral side of microeconomics. Thanks also to A. Chaintreau, J. Leshno, and P.A. Chiappori for inviting to present in their respective seminars.

I also want to express my gratitude to all the professors that taught me Game Theory, in particular to Antonio Jiménez-Martínez and Luis Sánchez-Mier from Guanajuato, and to Klaus Ritzberger and Jamsheed Shorish from Vienna. They all wrote recommendation letters to help me to get into the doctoral program, and without their support I just could not have started this degree. In particular, A. Jiménez-Martínez has been like an academic father to me even after getting into the Ph.D., and has provided me with invaluable advice regarding my future academic career.

My office mate Andrea Benecchi has helped in many occasions with issues from heavy coding to fixing a whiteboard. His insights regarding problems he is often unfamiliar about have given me a broader perspective about my own work. He has also constantly proofread my papers, so a sincere acknowledgment is in order. Several other colleagues from the Ph.D. have helped me to polish the work presented here, including Andrés Azqueta-Galvadón, Hualin Li, Putthi Lim, Moritz Mosenhauer, and Arsenii Paramonov.

Several of my colleagues from the master program in Vienna helped me to build some of the ideas presented in this thesis. In particular, I appreciate the suggestions from Philipp Hergovich, Fruzsina Mayer, Karol Mazur, László Tétényi and Gabriel Ziegler. I thank them for their constructive criticism and their sincere friendship since old times in Vienna. Furthermore, P. Hergovich has been an splendid coauthor who understands parts of my work better than I do.

Last but not least, I am grateful to Sarah Fox for proof-reading the introduction of this thesis and for making the process of concluding it much more bearable.

Affidavit

I declare that, except where explicit reference is made to the contribution of others, this dissertation is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Josué Alberto Ortega Sandoval

Glasgow, September 20, 2017.

Chapter 1

Introduction

Economics can be described as the systematic study of how to allocate scarce resources to a group of heterogeneous individuals. If we are interested in allocating those resources efficiently, we would give them to the persons who value them most. If we are interested in distributing them fairly instead, we should procure that every agent is relatively happy with the number of goods they obtain compared to what other persons get. Both ideas have been formalised in economics: efficiency as the maximisation of a social welfare function, and fairness as a no-envy test, that requires that each person receives a bundle of items that they value as much as the one received by any of their peers.

Luckily, when dividing scarce resources, we can generally find allocations that are efficient and fair by using prices: agents are endowed with budgets, and spend their endowments rationally. The division produced is always efficient and envy-free: agents spend their money on the goods they value most, and nobody envies the assignment obtained by someone else because they could have afforded it but instead choose another bundle in their budget set. My thesis studies how to find fair and efficient allocations whenever prices cannot be used. This type of problem is referred in the literature as a matching problem.

Matching problems have been extensively studied because economists and mathematicians realised that there are practical situations in which pricing scarce resources is either considered disgusting or unfeasible. Examples include using prices to allocate organs for transplantation, like kidneys or lungs, seats in public schools, or online dating partners.

The economics discipline understood that these assignment environments were in practice very distinct from markets, and thus deserved a systematic study on their own.

However, because of very different reasons, we as a community started studying those matching problems in the same way we studied markets and employing the same techniques. An example is the competitive equilibrium mechanism with equal incomes. The way it works is that we do not use prices but create a fake currency instead so that it is the market designer, and not the agents through their interaction, who finds the equilibrium allocations. Agents are asked what they would consume for every combination of prices, then the market designer computes an allocation, and informs each player of their bundle received at such prices. The take away from this literature is that pseudo-market methods work well: they are envy-free and Pareto efficient, and furthermore are non-manipulable whenever there are many agents interacting with each other.

The first contribution of my thesis is to show that pseudo-market mechanisms sometimes do not work as well as we may expect, producing outcomes that are neither fair nor hard to manipulate. In Chapter 2, I show that market mechanisms perform poorly in the case of multi-unit assignment, a fair division problem in which every person can receive several goods. By poorly I mean that

1. market mechanisms are unfair,

2. market mechanisms are not single-valued, and

3. market mechanisms can be manipulated by groups.

Surprisingly, we can use a Rawlsian type of mechanism that avoids all those problems present in the competitive equilibrium mechanism. Even more surprisingly, while the competitive mechanism is used in applications, I am unaware that the Rawlsian or leximin mechanism has ever been applied in practice.

Chapters 3 and 4 use matching models to understand segregation: i.e. understanding which are the real-life restrictions that make partners of the same race match among themselves in a much higher proportion than they match with people from other races.

In Chapter 3, I consider the following problem: individuals belong to different communities, and match among themselves. How many people prefer to match as a unified community instead? I am interested in this type of problem because economics has provided a rationale for behavioural patterns we encounter frequently in human interactions: envy, altruism, and cooperation are all words often used in the game-theoretic literature. What I attempt to do is to provide a theoretical framework to understand why some people may oppose social integration, and what characteristics those people theoretically share.

I show that there is no way to integrate isolated communities that guarantee that every agent is better off after integration occurs, as long as integration produces either an efficient or a Pareto optimal pairing. Remarkably, this is true even for large societies: computational simulations show that with five or fewer societies, the number of agents that become worse off after integration occurs remains around 25 percent of the society as a whole. It is also surprising that those agents who get hurt by social integration are indistinguishable in terms of expected ranking of any other agent.

Finally, Chapter 4 presents a Schelling-type model that explains how people marry. The idea behind it is simple: people want to marry other agents that have personality traits close to theirs. However, a person can only marry another person who they know, so being poorly connected to people of other races would inevitably yield a low rate of interracial marriages.

However, online dating often allows us to meet people who would otherwise be complete strangers to us, breaking an old phenomenon in networks usually referred as the strength of weak ties. But how much do those new ties through social networks will influence the decision of who do we end up marrying? We combine the tools of stable matching with those from random graph theory to answer this question with an unexpected result: even with a small number of new edges in a generalized random graph, the number of interracial marriages present in a society increases dramatically. We contrast our theory against the existing data on interracial marriages in the U.S., and find that our results are in line with current demographic trends.

In this last model, which has benefited from the collaboration with Philipp Hergovich from the University of Vienna, real life restrictions in matching strike again. Our results show that those absent ties or edges in a random graph make an important difference, and suggest that the actual ethnic composition of marriages in our societies may not be owing to intraracial preferences but could arise solely by social networks restrictions.

What should be taken away from this idea, then, is that in order to gain a proper understanding of how matching environments without money work, we need to detect the key aspects of each environment. These then need to be captured within our theoretical models, instead of forcing the use of solutions to different problems to our new environments. Matching environments are quite different from markets, and our understanding of them will be more profound as we become able to capture that difference in our research.

Chapter 2

The Random Multi-Unit Assignment Problem with Endogenous Quotas

Abstract

We study the random multi-unit assignment problem in which the number of goods to be distributed depends on players' preferences.

In this setup, the egalitarian solution is more appealing than the competitive equilibrium with equal incomes because it is Lorenz dominant, unique in utilities, and impossible to manipulate by groups when agents have dichotomous preferences. Moreover, it can be adapted to satisfy a new fairness axiom that arises naturally in this context. Both solutions are disjoint.

Two standard results disappear. The competitive solution can no longer be computed with the Eisenberg-Gale program maximizing the Nash product, and the competitive equilibrium with equal incomes is no longer unique in its corresponding utility profile.

KEYWORDS: multi-unit assignment, random assignment, endogenous quotas, dichotomous preferences, fair division, scheduling, course allocation, tennis.

JEL CLASSIFICATION: D63 (equity, justice, inequality), C78 (matching theory).

2.1 Our Problem and its Relevance

Consider a tennis club organizer who has to assign double tennis matches. He knows the players self-reported availability over the weekdays, and tries to find a reasonable schedule such that 1) no person plays on a day she is not available, 2) no person plays more than once per day, and 3) each match has exactly four players.

The previous assignment problem can be described by a matrix containing players' availability and a quota indicating how many players are required to create a game (four for our tennis example). Table 2.1 presents a real-life example from Maher (2016), in which two games of four people each can be created on both Tuesday and Thursday, one game on both Monday and Wednesday, and zero games on Friday, when less than four players are available.

Table 2.1: A tennis problem with a deterministic solution in parenthesis.

Players \ Days	Mon	Tue	Wed	Thu	Fri	Total
Barry	0	0	1 (1)	1 (1)	0	2 (2)
Tom	1 (0)	1 (1)	0	1 (1)	0	3 (2)
Peter	1 (1)	1 (1)	0	0	0	2 (2)
Colin	1 (1)	0	0	1 (1)	0	2 (2)
Mike	0	1 (1)	1 (0)	1 (1)	1 (0)	4 (2)
Keith	0	1 (0)	1 (1)	0	0	2 (1)
Alan	1 (0)	0	0	1 (1)	0	2 (1)
John	0	1 (1)	0	0	0	1 (1)
Ringo	1 (1)	0	1 (1)	0	0	2 (2)
George	1 (0)	1 (0)	1 (0)	1 (1)	0	4 (1)
Michael	0	0	1 (1)	0	0	1 (1)
Phil	0	1 (1)	0	0	0	1 (1)
Brian	1 (1)	1 (1)	0	0	0	2 (2)
Paul	0	1 (1)	0	1 (1)	0	2 (2)
Willie	0	0	0	1 (1)	0	1 (1)
Ken	0	1 (1)	0	0	0	1 (1)
Total	7 (4)	10 (8)	6 (4)	8 (8)	1 (0)	

The scheduling problem above is representative of a large class of multi-unit assignment problems where the number of available resources depends on agents' preferences. Other examples include scheduling teamwork, distributing provisions to food banks, lung transplantation or allocating courses to students. We describe them in detail in the next subsection.

For such assignment problems, we would like to have a systematic procedure to decide fairly which players should get which games, that at the same time incentivize players to reveal truthfully their availability.

Our contribution is to propose an egalitarian solution that achieves this purpose for a wide class of multi-unit assignment problems in the dichotomous preference domain.

The egalitarian solution is based on the well-known leximin principle, and performs better than the competitive equilibrium with equal incomes solution, which is theoretically appealing in similar assignment models (Hylland and Zeckhauser, 1979; Budish, 2011) and has been successfully applied in practice to allocate courses in business schools (Budish et al., 2017). By better, we mean that the egalitarian solution is, unlike the pseudo-market solution, Lorenz dominant, unique in utilities, and impossible to manipulate by groups.

Lorenz dominance is "a ranking generally accepted as the unambiguous arbiter of inequality comparison" (Foster and Ok, 1999) and is "widely accepted as embodying a set of minimal ethical judgments that should be made" (Dutta and Ray, 1989). Given two vectors of size n, the first one Lorenz dominates the second one if, when arranged in ascending order, the sum of the first $k \leq n$ elements of the first one is always greater or equal than the sum of the k first elements of the second one. A utility profile is Lorenz dominant is it Lorenz dominates any other feasible utility profile.

In our setup, that a utility profile is Lorenz dominant implies that it uniquely maximizes any strictly concave utility function representing players' preferences, like the Nash product, and is, therefore, a strong fairness property.

Uniqueness of the solution (in the utility profile obtained) is also a very desirable property, for it gives a clear recommendation of how the resources should be split. A multi-valued solution leaves the schedule designer with the complicated task of selecting a particular division among those suggested by the solution, thus raising the opportunity of rightful complaints by some agents, who may argue that there were other allocations recommended by the solution that were more beneficial to them.

It is equally interesting that the egalitarian solution is not manipulable by groups, implying that coalitions of players can never profit from misrepresenting their availability, not even when reducing the total number of resources created. The competitive solution is manipulable by groups in our setup, as in many others. Yet, it is remarkable that even in our small dichotomous preference domain, where possibilities to misreport are very limited, the pseudo-market solution can still be manipulated. Manipulating a solution can be done easily, by groups reducing their availability strategically on days where the demand for resources almost equals its supply.

The fact that the egalitarian solution satisfies these three desirable properties is a strong argument for recommending its use in this environment, instead of the pseudo-market mechanism. The egalitarian solution also satisfies these three properties in the multi-unit assignment problem with exogenous quotas, for which other solutions have been proposed in the literature, but for which the egalitarian solution has not yet been considered.

Throughout the article, we stick to the tennis jargon and denote by *generalized tennis problems* (GTPs) the set of random multi-unit assignment problems with endogenous quotas and dichotomous preferences. GTPs include several real-life problems, which we discuss below.

2.1.1 Applications

GTPs are motivated by the real-life allocation of tennis slots to players, which can be generalized to other sports (the quota required to create one game could be 2 for tennis singles, 22 for football, etcetera) but include several other problems in which the number of goods to be assigned depends on agents' preferences or characteristics. Some of them are:

1. *Scheduling team work.* Scheduling is an intuitive example that fits the dichotomous preference domain. Consider the scheduling problem faced by airlines, whose flights need a specific number of cabin crew members required by law, or the one faced by policemen who need to be in groups of certain size to patrol in some area. Other examples of this kind can be constructed.

2. *Allocation of provisions to food banks.* Food banks distribute provisions to people in need, which in turn they receive from large storage centers. A food bank may need a specific type of meal, but it is often impractical to ship a whole truckload from a distribution center to serve only a small food bank. Therefore, shipments from the storage centers can only be sent whenever a specific number of food banks request one (see Section 3.1 in Prendergast, 2017, also Aleksandrov et al., 2015).

3. *Organ Exchange.* Dichotomous preferences have been used to model whether a person is compatible or not with a particular organ for transplantation. Roth et al. (2005) write "*the experience of American surgeons suggests that preferences over kidneys can be well approximated as 0 - 1, i.e. that patients and surgeons should be more or less indifferent among kidneys from healthy donors that are blood type and immunologically compatible with the patient*".

In particular, a (living donor lobar) lung transplant requires two compatible donors to be succesfully performed, each giving a lower lung lobe to the patient (Cohen et al., 2014; Ergin et al., 2017).[1] The problem of organizing lung transplantation can be formulated as a compatibility matrix, in which rows represents hospitals, and columns denote types of compatible donors available. Note that even though the entries of the matrix can be larger than 1, the problem is equivalent to ours as it will be seen in our Examples. Each row becomes a "large" agent, whose compatibility is the sum of several individual 0-1 compatibility entries.

4. *Course Allocation.* Our problem is also similar to the real-life allocation of courses or tutorials in Universities. The number of seats available for each course is not entirely fixed, as Universities are able to open new courses if the demand for a course is significantly larger than its supply. For example, if the maximum number of students for a course is 50, and there are 125 students willing to take it, the University is likely to open two of such courses so 100 students can be served. Opening courses to fit the supply particularly fits

[1] A healthy person has five lung lobes: three in the right one, two in the left one. Given that the rows represent hospitals, it is unlikely that their constraints on the number of transplantation procedures they can perform are binding, as currently lung transplantation is rare. In the UK, only 198 were carried out in 2013-14, none of those in Scotland. Source: "Lung transplant", *NHS*, 28/06/2016.

the case of tutorials or recitation sessions, because these are usually taught by graduate students which tend to be easy to hire.[2]

There is a subtle difference, however, as students may have horizontal constraints on the maximum number of courses they can take. Including this type of constraints makes our problem much more difficult to solve, so we postpone its discussion to Section 7.

2.1.2 Related Literature

Our theoretical model is closely related to three existing problems in the literature:

1. *Single-unit random assignment with dichotomous preferences* by Bogomolnaia and Moulin (2004), henceforth BM04. Our model generalizes theirs in two regards. Firstly, in their setup agents can only get one good. Secondly, agents do not need others to obtain their desired assignment, i.e. quotas are exogenous.

They study the egalitarian and the equal income competitive solution. They show that the egalitarian solution is Lorenz dominant and can always be supported by competitive prices. Therefore, because the competitive solution is Lorenz dominant, the competitive solution can be easily computed as the maximization of the Nash product of agents' utilities. They also prove that the egalitarian solution is group strategy-proof.

Roth et al. (2005) show that the egalitarian solution is also Lorenz dominant in assignment problems on arbitrary graphs that are not necessarily bipartite. Assignment on the dichotomous domain of preferences has been further studied by Bogomolnaia et al. (2005), Katta and Sethuraman (2006), and Bouveret and Lang (2008).

2. *Shubik's bridge economy* (Shubik, 1971). He considers an economy that needs four players to create one good, eight to create two, and so on. He shows that the core of that economy may be empty. We generalize Shubik's model by considering the division of games in multiple days.

[2]For example, LSE requires undergraduate classes and graduate seminars to have a maximum of 15 students. King's College London has a maximum size of 18 students per class.

3. *Multi-unit assignment with exogenous quotas*, commonly known as the Course Allocation Problem (CAP), described by Brams and Kilgour (2001); Budish (2011); Budish and Cantillon (2012); Kominers et al. (2010); Krishna and Ünver (2008); and Sönmez and Ünver (2010), with an important difference. In CAP the number of seats available for each course (in this case game slots per day) is fixed and given exogenously, whereas in GTPs the number of seats is determined endogenously by players' preferences, representing the real possibility that the number of courses is not fully fixed in practice. This difference is important theoretically, because players may manipulate an allocation mechanism by changing the total number of seats available.

Additionally, in the combinatorial CAP version (Budish, 2011), players may have arbitrary preferences over the set of days. However, reporting combinatorial preferences is unfeasible for even few alternatives, and in practice combinatorial mechanisms never allow players to report such preferences fully, not only because such revelation would be complicated, but also because players may not know their preferences in that detail. Consequently, a new strand of theory has focused on allocation mechanisms with simpler preferences (Bogomolnaia et al., 2017; Bouveret and Lemaître, 2016), which are used successfully in modern fair division procedures in real life: see *Spliddit.com* (Goldman and Procaccia, 2015).

Although our preference domain is much smaller than those considered in CAP, it is not contained in any of those because CAP rules out indifferences.

Finally, Budish (2011) only considers deterministic assignments. We study randomized assignments instead: in practice many allocation mechanisms use some degree of randomization to achieve a higher degree of fairness.[3]

[3] Randomization is used to assign both permanent visas and housing subsidies in the US, or school places in the UK. Sources: "A one in a million chance at a better life", *The Guardian*, 2/5/2017, "Why does random chance decide who gets housing subsidies?", *NPR*, 3/5/2016, and "School admissions: is a lottery a fairer system?", *The Guardian*, 14/3/2017.

2.2 Summary of Results

We define the egalitarian and the constrained competitive solution. The egalitarian one is Lorenz dominant in the set of efficient utility profiles (Theorem 1), while the competitive one exists (Theorem 2) but is multi-valued (Example 1). The egalitarian solution is group strategy-proof, but the competitive one is not (Theorem 3). Both solutions are disjoint (Example 2).

We show that there are no competitive prices supporting the egalitarian solution, which is a stark difference between our model and BM04. As a consequence, the classical result stating that the competitive solution can be computed as the maximizer of the Nash product of utilities no longer holds: a result known as the Eisenberg-Gale program.

This result is key for algorithmic game theory as it establishes an easy method for computing economic equilibria. Its failure is important not only because leaves us with no known algorithm for computing equilibria, but also because the Eisenberg-Gale program is a rather robust result that applies to a large class of utility functions beyond the linear case (Vazirani, 2007) and to the division of goods and bads (Bogomolnaia et al., 2017).

The fact that the competitive solution is not unique is also interesting, as a unique utility profile is always obtained in Fisher markets (which is itself another consequence of solving the Eisenberg-Gale convex program, see Theorem 5.1 in Vazirani, 2007).

We show that the egalitarian solution violates a natural fairness requirement called *independence of perfect days*. We construct a refined egalitarian solution that achieves this property, while at the same time being Lorenz dominant for the set of overdemanded days. This refined solution, while appealing, violates group strategy-proofness, unlike the classical egalitarian solution (Example 3).

This article is structured as follows. Sections 3 and 4 formalize the model and the solutions we consider, respectively. Section 5 analyzes the solutions' manipulation, while Section 6 introduces the property of independence of perfect days. Section 7 discusses how our findings extend to the model in which agents face upper limits on the number of days they want to play.

We defer all proofs to the Appendix. In all of our Examples (not in the proofs), we fix the quota to 4, but it is easily seen that our arguments generalize.

2.3 Model

Let R be a $n \times m$ binary matrix containing the availability of each person $i \in N$ about playing on day $k \in M$. The entry $r_{ik} = 1$ if person i is available to play on day k, and 0 otherwise. Abusing notation slightly, R_{iM} (resp. R_{Nk}) denotes both the i-th row (k-th column) of R and the set of days $k \in M$ (resp. players $i \in N$) for which $r_{ik} = 1$.

Let $q \geq 2$ be the number of people required for making a game. The notation $\lfloor x \rfloor$ denotes the floor function applied to x, i.e. $\lfloor 3.2 \rfloor = 3$. For each day $k \in M$, there are $\delta(k)$ identical slots to assign, where δ is given by

$$\delta(k) = q \cdot \left\lfloor \frac{|R_{Nk}|}{q} \right\rfloor \tag{2.1}$$

and $\delta = (\delta(1), \ldots, \delta(m))$ is the vector of available slots. The set of slots for a day k is denoted by S_k, and S represents the set of all slots, i.e. $S = \bigcup_M S_k$. The pair (R, q) is called a **generalized tennis problem** or GTP. The matching size of a GTP is denoted by $\nu(R, q) = \sum_{k \in M} \delta(k)$.

A random assignment is a probability distribution over allocations of slots to players such that no player receives more than one slot per day. It can be represented by a **random allocation matrix** (RAM) Z, which entry z_{ik} denotes the probability of person i playing on day k.

$\mathcal{F}(R, q)$ denotes the set of all RAMs for the GTP (R, q). To describe it, we need to define individually rational (IR) matrices first, i.e. those that assign positive probabilities only to days in which a player is available. Formally, the matrix X is IR for R if they are of same size and, $\forall i \in N, k \in M$, $x_{ik} > 0$ only if $r_{ik} > 0$. Then

$$\mathcal{F}(R, q) = \{Z \in [0,1]^{n \times m} \mid Z \text{ is IR for } R \text{ and } \forall k \in M, \sum_{i \in N} z_{ik} = \delta(k)\} \tag{2.2}$$

As before, the notation Z_{iM} (resp. Z_{Nk}) denotes both the i-th row (k-th column) of Z and the set of days $k \in M$ (resp. players $i \in N$) for which $z_{ik} = 1$.

Several random assignments can have the same corresponding RAM. Theorem 1 in Budish et al. (2013) implies[4] that

Lemma 1. *Any RAM can be decomposed into a convex combination of deterministic binary RAMs, and thus can be implemented.*

We assume that players are indifferent between when and with whom they play, as long as they do it on an available day. The canonical utility function representing those preferences is

$$u_i(Z) = \sum_{k \in M} z_{ik} = \sum_{k \in M} r_{ik} \cdot z_{ik} \tag{2.3}$$

for an arbitrary agent $i \in N$ and an arbitrary Z that is IR for R. This function is clearly not unique but it is convenient to work with. The preference relation represented by it is a complete order over all feasible and individually rational random assignments.

The preference relation represented by the utility function above implies that a RAM Z is Pareto optimal in a GTP (R, q) if and only if $Z \in \mathcal{F}(R, q)$.

The set of efficient utility profiles $\mathcal{U}(R, q)$ can be described as

$$\mathcal{U}(R, q) = \{U \in \mathbb{R}^n \mid \exists\, Z \in \mathcal{F}(R, q) : U_i = \sum_{k \in M} r_{ik} z_{ik},\ \forall i \in N\} \tag{2.4}$$

We do not distinguish between ex-ante and ex-post efficiency because in our preference domain they coincide. This equivalence occurs because in all efficient assignments the sum of utilities is constant and equal to the matching size of the problem $\nu(R, q)$.[5] In our setup, efficiency simply requires that no game slot is wasted.

A **welfarist solution** is a mapping Φ from (R, q) to a set of efficient utility profiles in $\mathcal{U}(R, q)$, and hence it only focuses on the expected number of slots received by an agent

[4]The implication follows because the set of vertical constraints on any RAM is a hierarchy. Hierarchies are also known as laminar families in combinatorial optimization.

[5]A RAM is ex-post efficient if it can be written as a convex combination of deterministic Pareto optimal RAMs, and ex-ante efficient if it is optimal with respect to agents' preferences over lotteries. Both notions are equivalent in assignment problems with dichotomous preferences (BM04, Roth et al., 2005).

and not on the exact probability distribution over deterministic assignments. Whenever a solution is single-valued we use the notation ϕ instead.

For each GTP (R, q) there exists a corresponding **course allocation problem** (CAP), defined as a tuple (N, M, δ, R) in which N is the set of students, M is the set of courses, δ is the vector of exogenous capacities for each course, and R contains the preferences of each student over the set of courses. Therefore, any statement we make about the efficiency and fairness for GTP solutions also applies to the corresponding CAP.[6]

2.3.1 Reductions and Decompositions

Any day in which there are less than q players available is irrelevant and can be deleted. Players who are always unavailable or that are only available on irrelevant days are inconsequential too and are also removed. Henceforth we work with the corresponding **irreducible problem** of any GTP, which satisfies

$$\forall i \in N, \qquad R_{iM} \neq \emptyset \tag{2.5}$$

$$\forall k \in M, \qquad |R_{Nk}| \geq q \tag{2.6}$$

Furthermore, for any GTP (R, q), we can partition the corresponding set of days M into two subsets $\mathcal{P}(R, q)$ and $\mathcal{O}(R, q)$, which are called **perfect** and **overdemanded** respectively.[7] The set of perfect days is defined as

$$\mathcal{P}(R, q) = \{k \in M : |R_{Nk}| = \delta(k)\} \tag{2.7}$$

Given a GTP (R, q), a perfect complement for player i represents adding an arbitrary perfect day in which i can play. Formally, a perfect complement for player i in a GTP (R, q) is a pair $(k', R_{Nk'})$ such that $k' \notin M$, $r_{ik'} = 1$, and $k' \in \mathcal{P}([R\, R_{Nk'}], q)$, where $[R\, R_{Nk'}]$

[6]We emphasize again that the equivalence between GTPs and CAPs only holds for CAP without horizontal constraints, i.e. without limits on how many days each agent can play.

[7]These categories can be thought of as the Gallai-Edmonds decomposition of the bipartite graph $G = ((N, S_k), R_{Nk})$ associated with the matching problem in day k.

denotes the $n \times (m+1)$ juxtaposition of the two matrices. The GTP $([R\, R_{Nk'}], q)$ is a **perfect extension** of the original problem for player i.

2.4 Three Efficient Solutions

2.4.1 The Egalitarian Solution

An intuitive solution equalizes players' utilities as much as possible respecting efficiency and individual rationality: this is the well-known leximin solution. We refer to it as the **Egalitarian Solution** (ES), proposed theoretically by BM04, and applied to the exchange of live donor kidneys for transplant by Roth et al. (2005) and Yılmaz (2011).

To define it formally, let $\succ^l$ be the well-known lexicographic order.[8] For each $U \in \mathbb{R}^n$, let $\gamma(U) \in \mathbb{R}^n$ be the vector containing the same elements as U but sorted in ascending order, i.e. $\gamma_1(U) \leq \ldots \leq \gamma_n(U)$. The leximin order $\succ^{LM}$ is defined by $U \succ^{LM} U'$ if and only if $\gamma(U) \succ^l \gamma(U')$. The ES is defined by

$$\phi^{\mathrm{ES}}(R, q) = \arg \max_{\succ^{LM}} \mathcal{U}(R, q) \tag{2.8}$$

The ES satisfies a strong fairness notion called **Lorenz dominance**, defined as follows. Define the order $\succ^{ld}$ on $\mathbb{R}^n$ so that for any two vectors U and U', $U \succ^{ld} U'$ only if $\sum_{i=1}^{t} U_i \geq \sum_{i=1}^{t} U_i' \quad \forall t \leq n$, with strict inequality for some t. We say that U Lorenz dominates U', written $U \succ^{LD} U'$, if $\gamma(U) \succ^{ld} \gamma(U')$. A vector $U \in \mathcal{U}(R, q)$ is Lorenz dominant for a GTP (R, q) if it Lorenz dominates any other vector in $\mathcal{U}(R, q)$.

Lorenz dominance is a partial order in $\mathcal{U}(R, q)$ and therefore a Lorenz dominant utility profile need not exist. Nevertheless, the ES solution is Lorenz dominant.

Theorem 1. *The ES solution is Lorenz dominant in the set of efficient utility profiles.*

We prove Theorem 1 using Theorem 3 in Dutta and Ray (1989), which states that the core of every supermodular cooperative game has a Lorenz dominant element. We postpone to the Appendix the construction of the corresponding cooperative game.

[8] So that for any two vectors $U, U' \in \mathbb{R}^n$, $U \succ^l U'$ only if $U_t > U_t'$ for some integer $t \leq n$, and $U_p = U_p'$ for any positive integer $p \leq t$.

The reader may think that Theorem 1 extends to the case when players may have arbitrary preferences that are not dichotomous. This is not the case. Consider a problem with $n = 7$, $m = 3$, $q = 4$, and players utilities given by subtable 2.2a. The players give a 2 to a day in which they really like to play on, a 1 on a day they are available but do not prefer as much as a day to which they gave a 2, and 0 to a day in which they are not available at all.

Table 2.2: The limits of Theorem 1

$N \setminus M$	M	T	W
a	2	1	1
b	1	2	0
c	1	2	0
d	0	2	2
e	1	0	2
f	1	0	2

(a) Corresponding R matrix.

M	T	W
4/9	1	1
8/9	1	0
8/9	1	0
0	1	1
8/9	0	1
8/9	0	1

(b) RAM supporting ES.

M	T	W
1	1	1
0	1	0
1	1	0
0	1	1
1	0	1
1	0	1

(c) Alternative RAM.

The egalitarian solution suggests the utility profile (2.89, 2.89, 2.89, 4, 2.89, 2.89), which produces a total of 18.44 units of utility. However, consider the RAM in subtable 2.2c, suggesting the utility profile (4, 2, 3, 4, 3, 3). The ES utility profile does not Lorenz dominate this feasible utility profile, showing that Theorem 1 only works with dichotomous preferences.

2.4.2 The Constrained Competitive Equilibrium with Equal Incomes

A second solution, substantially more complicated, requires to find an equilibrium between supply and demand of slots when players are endowed with equal budgets. These equal budgets are often normalized to one currency unit, a normalization that we also use. This solution is known as the **Competitive Equilibrium with Equal Incomes** or CEEI (Varian, 1974; Hylland and Zeckhauser, 1979). In our tennis problem, each agent can consume at most one slot per day, hence having particular constraints on their consumption set that play a major role. We use the term **Constrained Competitive Equilibrium** (CCE, still with equal incomes) from now on to make this distinction obvious.[9]

The CCE solution is different from the CEEI as defined in Hylland and Zeckhauser (1979) in that they impose no constraints in the goods to be consumed: in our case agents never partially consume goods that have different prices, see their Table 1 in their paper. This difference justifies the different terminology of CCE.

Definition 1. A CCE for a GTP (R, q) is a pair of a RAM Z^* and a non-negative price vector p^* such that, $\forall i \in N$, agents maximize their utilities

$$Z^*_{iM} \in \arg \max_{Z_{iM} \in \beta_i(p^*)} u_i(Z_{iM}) \tag{2.9}$$

where $\beta_i(p)$ is the budget set defined as $\beta_i(p) = \{Z_{iM} \mid \sum_{k \in M} z_{ik} \leq |R_{iM}|\,; p \cdot Z_{iM} \leq 1\}$, and the market clears, so that

$$Z^* \in \mathcal{F}(R, q) \tag{2.10}$$

As we shall see in Theorem 2, the set of CCE is never empty but may be large. The optimality conditions of CCE imply

$$\begin{aligned} k \notin \mathcal{P}(R, q) \quad &\Longrightarrow \quad p^*_k > 0 && (2.11)\\ z^*_{ik}, z^*_{ik'} \in (0,1) \quad &\Longrightarrow \quad p^*_k = p^*_{k'} && (2.12)\\ [p^*_k < p^*_{k'}] \wedge [0 < z^*_{ik'}] \quad &\Longrightarrow \quad z^*_{ik} = 1 && (2.13)\\ \sum_k z^*_{ik} < |R_{iM}| \quad &\Longrightarrow \quad \sum_k p^*_k \cdot z^*_{ik} = 1 && (2.14) \end{aligned}$$

These are the equivalent of the Fisher equations in our model, see Brainard and Scarf (2005). Condition (2.11) allows a zero price only for perfect days, while expression (2.12) forces the same marginal benefit for every good in which the agents plays with a strictly positive probability but not with certainty.

[9] We stress that the CCE is a standard competitive equilibrium with restricted preferences.

The CCE is in general multivalued. Given a GTP, we denote the set of pairs (Z^*, p^*) as $\mathcal{C}(R, q)$. The CCE solution is defined by

$$\Phi^{CCE}(R, q) = \{u(Z') \mid \exists p' : (Z', p') \in \mathcal{C}(R, q)\} \tag{2.15}$$

2.4.3 The Naive Egalitarian per Day

Finally, a naive and most intuitive solution (that we use as a benchmark only) breaks up the allocation problem into m sub-problems of assigning S_k into R_{Nk}, giving an equal share of the slots in day k among all players available on that day. We call this solution **Egalitarian Per Day** (EPD). This is, given a GTP (R, q), the EPD solution assigns to each player

$$\phi_i^{EPD}(R, q) = \sum_{k \in M} r_{ik} \cdot \frac{\delta(k)}{|R_{Nk}|} \tag{2.16}$$

We note that, in our preference domain, EPD is equivalent to the well-known random priority mechanism, aka random serial dictatorship.[10] We do not consider EPD an appropriate solution for GTPs because it ignores the interaction between the m fair division problems of each day.

EPD also fails the following basic fairness property: if $n-1$ players get at least 1 utility unit, the n-th player also gets at least 1 utility unit too; see Example 1 for an illustration.

2.4.4 Two Examples Showing that All the Solutions Differ

Example 1 (Multivalued CCE differs from EPD)**.** Table 2.3 shows the different outcomes these three solutions produce for a problem with $n = 6$, $m = 3$, $q = 4$, and R given in subtable 2.3a. The CCE utilities are written in brackets in subtable 2.3b because there are CCE that support utility profiles between $(2.4, 1.4, 1)$ and $(2.25, 2, 1)$ with $0 \leq p_W \leq \frac{4}{9}$. This multiplicity is interesting: the competitive solution is always unique in the corresponding utility profile in Fisher markets, and also in the more general Eisenberg-

[10] EPD would not be efficient in a more general domain of preferences. The equivalence with random priority would also disappear.

Gale markets; see for example Theorem 5.1 in Vazirani (2007) or p. 87 in Jain and Vazirani (2010). It is also problematic, as there is no obvious selection from the CCE.

Table 2.3: Example 1

$N\backslash M$	Mon	Tue	Wed	Total
$a:d$	1	1	1	3
e	1	1	0	2
f	1	0	0	1
Total	6	5	4	

(a) Corresponding R matrix.

$N\backslash$Solution	ES	CCE	EPD
a:d	2.25	[2.25 - 2.4]	2.47
e	2	[1.4 - 2]	1.47
f	1	1	0.67

(b) Utility profiles for each solution.

Any CCE in example 1 gives a slot with probability one to player f. This implies that there are no CCE prices that support the EPD outcome, and thus is a strong argument against this solution, as competitive equilibria are considered "essentially the description of perfect justice" (Arnsperger, 1994), and the base of Dworkin's "*equality of resources*" (Dworkin, 1981).

The EPD solution is therefore not ideal, as expected. But interestingly, the ES solution can also produce outcomes that cannot be supported as a CCE.

Example 2 (ES differs from CCE)**.** We show it using a GTP with $n = 9$, $m = 6$, $q = 4$, and R given in subtable 2.4a. Note that in the single-unit case (Theorem 1 in BM04), the ES is always supported by competitive prices.

Table 2.4: Example 2.

$N\backslash M$	M	T	W:Th	F:S	Total
$a:c$	1	1	0	0	2
d	0	1	1	0	3
e	0	1	0	1	3
$f:i$	1	0	1	1	5
Total	7	5	5	5	

(a) Corresponding R matrix.

M	T	W:Th	F:S	Total
1	0.97	0	0	1.97
0	0.54	1	0	2.54
0	0.54	0	1	2.54
0.25	0	0.75	0.75	3.25
4	4	4	4	

(b) Corresponding Z^*.

If the ES solution (2, 2.5, 2.5, 3.25) could be supported as a CCE, then $p_M = p_W = p_{Th} = p_F = p_S$ because agents f:i play with positive probability in those days. Furthermore, players $d{:}i$ must spend their whole budget, implying prices $p_M = \frac{4}{13}$ and $p_T = \frac{10}{13}$. However, at such prices, the ES utility for players $a{:}c$ is unaffordable.[11]

The fact that ES and CCE do not coincide is interesting: in the non constrained context, the competitive solution can be computed by maximizing the Nash product, solving what is known as the Eisenberg-Gale program (Eisenberg, 1961; Eisenberg and Gale, 1959; Chipman, 1974, see chapter 7 in Moulin (2003) for a textbook treatment or Sobel (2009) for a brief overview). That the competitive solution cannot be computed solving the Eisenberg-Gale program implies that we lack an algorithm for computing the competitive equilibrium, which can be a hard task (Uzawa, 1962; Othman et al., 2010, 2014).

The Eisenberg-Gale program is otherwise a rather robust result since it extends to a large family of utility functions beyond the linear case (Jain and Vazirani, 2010), as well as to the mixed division of goods and bads (Bogomolnaia et al., 2017).

The multiplicity of the competitive solution and its non-equivalence with the egalitarian outcome justify the new terminology of CCE. For any GTP, the set of CCE is non-empty, a result we prove in the Appendix using a classical fixed point argument with a small twist. We summarize our findings in Theorem 2.

Theorem 2. *For generalized tennis problems, the ES solution is well-defined and single-valued, and the CCE solution exists. Their intersection can be empty.*

2.4.5 Minimal Fairness Guarantees

It is easy to see that both the ES and CCE solutions achieve minimal fairness guarantees existing in the literature: namely equal treatment of equals and envy-freeness.

A solution ϕ **treats equals equally** if, for any GTP (R, q) that has players i and j such that $R_{iM} = R_{jM}$, $\phi_i(R, q) = \phi_j(R, q)$. A solution ϕ is **envy-free** if, for any GTP (R, q) with players i and j such that $R_{iM} \subseteq R_{jM}$, $\phi_i(R, q) \leq \phi_j(R, q)$. Clearly, envy-freeness implies equal treatment of equals. For a multi-valued solution, both properties hold if they hold for any selection of it.

[11] We do not consider the possibility that prices are defined over bundles, an interpretation which is not very intuitive in our model, but which is often used in combinatorial auctions.

Lemma 2. *ES and CCE are envy-free, and hence treat equals equally.*

We postpone an easy proof. Note that there is no efficient solution that is strongly envy-free, i.e. that for any GTP (R, q) with players i and j such that $|R_{iM}| < |R_{jM}|$, $\phi_i(R, q) \leq \phi_j(R, q)$, see Theorem 1 in Ortega (2016).

2.5 Manipulation by a Group of Players

We consider players' manipulation in the direct revelation mechanism associated with each solution. To do so, we need to know exactly how the tennis slots are assigned. A detailed solution ψ maps every GTP (R, q) into a RAM $Z \in \mathcal{F}(R, q)$, specifying which agents play in which day, whereas a welfarist solution ϕ maps every GTP into a utility profile $U \in \mathcal{U}(R, q)$ and only tells us the expected number of games received by each player. Every detailed solution ψ projects onto the welfarist solution $\phi(R, q) = u(\psi(R, q))$.

The direct revelation mechanism associated with a detailed solution ψ is such that all players reveal their preferences R_{iM}, and then ψ is applied to the corresponding irreducible problem (R, q), implementing the RAM $\psi(R, q) = Z$.

We assume that player i with true preferences R_{iM} can only misrepresent her preferences by declaring a profile $R'_{iM} \subset R_{iM}$. The intuition is that, declaring to be available on days players are not, would be strongly punished by the schedule designer in case of a game cancellation. Such assumption has already been imposed in scheduling problems in the context of algorithmic mechanism design (Koutsoupias, 2014). We say then that R'_{iM} is IR for R_{iM} (Section 3).

Considering manipulations in which players can exaggerate their availability is complicated for a number of reasons. First, to define it properly we would need to specify how the players substitute between goods and bads, i.e. in our tennis example, how many good days is a bad day worth. Secondly, if we allow transfers of days, so that a player obtains a game in a day she is unavailable, she can transfer it to another player, it is very likely that no rule, even a dictatorship, would be non-manipulable.

A detailed solution ψ is **group strategy-proof**[12] if for every GTP (R, q) and every coalition $S \subset N$, $\nexists$ R' satisfying i) $R'_{jM} = R_{jM}$ $\forall j \notin S$, and ii) R'_{SM} is IR for R_{SM}, such that

$$\forall i \in S, \quad u_i(\psi(R', q)) \geq u_i(\psi(R, q)) \tag{2.17}$$

with strict inequality for at least one player in S. A welfarist solution ϕ is group strategy-proof only if every detailed solution ψ projecting onto ϕ is group strategy-proof.

BM04 show that any deterministic solution fails group strategy-proofness for single-unit assignment, including priority solutions, i.e. those in which players choose sequentially their most preferred available bundle according to a specific order. The reason is that the player with the highest priority could change his report and still receive one acceptable alternative, leaving his utility unchanged, and at the same time benefiting a player with low priority: a property known as bossiness.

The argument does not extend to GTPs. Because agents can play on multiple days, the player with higher priority can belong to a manipulating coalition only by claiming fewer days. But since she has the highest priority, it is immediate that such manipulation would always give her strictly less utility, so she cannot be in the coalition. The same argument applies to all remaining players and, consequently,

Lemma 3. *Any deterministic priority solution is group strategy-proof.*

The previous Lemma shows that group strategy-proofness is relatively easy to achieve for GTPs in the dichotomous domain, in fact we show below that the ES solution also satisfies it. Is CCE also group strategy-proof? There are two extensions of our group strategy-proofness definition to set valued solutions.

One requires that for every GTP (R, q), there is no equilibrium of the manipulated GTP (R', q) that is weakly better than every equilibria of the original problem (R, q), for every member of the manipulating coalition S. A stronger extension is that there is at least one equilibrium of (R, q) that yields a weakly higher utility than some equilibrium of

[12]We note again that our definition corresponds to the one of *partial* group strategy-proofness, as we do not consider manipulations in which players exaggerate their availability.

(R', q), with strict inequality for at least one member of the deviating coalition S. It turns out that CCE violates both conditions.[13] The reason is that a group can coordinate to make several days perfect, and thus price them at 0.

Theorem 3. *ES is group strategy-proof but CCE is not.*

We postpone the proof of ES being group strategy-proof to the Appendix, but we show, using a simple example, that CCE is unambiguously manipulable by groups.

Example 3 (CCE not group strategy-proof)**.** Let $n = 7$, $m = 4$, $q = 4$, and R given by Table 2.5.

Table 2.5: Example 3.

$N \backslash M$	M	T	W	Th	Φ^{CCE}
$\boldsymbol{a}$	1	**1**	1	1	**2.5**
$\boldsymbol{b}$	1	1	**1**	1	**2.5**
$\boldsymbol{c}$	1	1	1	**1**	**2.5**
d	1	0	1	1	2.5
e	1	1	0	1	2.5
f	1	1	1	0	2.5
g	1	0	0	0	1
Total	7	5	5	5	

(a) True preferences R.

M	T	W	Th	Φ^{CCE}
1	**0**	1	1	**[2.5 - 2.57]**
1	1	**0**	1	**[2.5 - 2.57]**
1	1	0	**1**	**[2.5 - 2.57]**
1	0	1	0	[2.5 - 2.57]
1	1	0	1	[2.5 - 2.57]
1	1	1	0	[2.5 - 2.57]
1	0	0	0	[0.57 - 1]
7	5	5	5	

(b) Misreport R' for $S = \{e, f, g\}$.

Consider the coalition $S = \{a, b, c\}$. When players submit their real preferences, there exists a unique CCE that supports the ES solution: players a, b, c obtain 2.5 expected tennis games. By changing their report each on a different day, as in subtable 2.5b, they make Tuesday, Wednesday, and Thursday perfect days, thus enlarging the set of CCE solutions, which includes utilities that are always weakly above 2.5 and up to 2.57. By misrepresenting and creating artificially perfect days, they allow those days to be priced at 0, weakly increasing the number of expected slots received in any equilibria of (R', q), at the expense of players with limited availability, in this case g.

Given that ES is impossible to manipulate, unique, and Lorenz dominant, we suggest its use as a solution for GTPs. The competitive solution lacks these three properties.

[13] A weaker notion of group strategy-proofness is satisfied if there is a selection of the CCE solution in which no coalition of agents can weakly benefit, with one making positive gains. We are unsure whether the CCE satisfies this condition, we conjecture that it does.

Two remarks on the manipulation of our solutions. First, we do not discuss strategy-proofness (manipulation by individuals on their own) as it is immediate that ES (and EPD) satisfy it. For CCE, we can construct a selection of it that is strategy-proof, as reducing the total availability for a day either reduces the day's price, relatively increasing the price of other days, or it leaves the day's price unchanged.

Secondly, even though ES is group strategy-proof, it may offer weak incentives for truthful preference revelation for some players, so that they may misreport without affecting the solution outcome. This is a concern only inasmuch as the designer cares to perfectly capture players' availability. Players who may misreport never affect the number of slots available, so this lack of truthful revelation has no effect on the solution outcome.

Efficiency, fairness, and non-manipulability are standard goals in the design of resource allocation mechanisms. Now we consider a new goal that arises naturally for GTPs.

2.6 Independence of Perfect Days

Some solutions do not depend on the number of perfect days on which a player is available. If an agent is available on an extra perfect day we could expect that she would always receive one extra expected day in full. This is what our following property captures.

A solution ϕ is **independent of perfect days** (IPD) if, for every GTP, every $i \in N$ and for any of its perfect extensions $([R\ R_{Nk'}], q)$,

$$\phi_i(R, q) + 1 = \phi_i([R\ R_{Nk'}], q) \tag{2.18}$$

IPD is a desirable property because of two reasons. Firstly, perfect days belong unambiguously to players available on them, so they can argue that they should obtain them fully, irrespectively of the share they obtain from overdemanded days. Secondly, if the clearinghouse used a solution that was not IPD, the set of players who are available on perfect days could avoid reporting their availability for perfect days and organize a game on perfect days outside the centralized mechanism. That way, they would obtain a better share from the overdemanded days while fully receiving the benefits of perfect days.

Only one of our solutions (partially) satisfies this requirement.

Lemma 4. *Although ES is not IPD, there exists a selection of CCE that satisfies IPD.*

Lemma 4 highlights that CCE can always assign a zero price to all perfect days: this is the how we construct the selection of CCE that satisfies IPD. But it may also assign a zero price to some perfect days only, or to no perfect day at all. The designer has a high flexibility choosing the equilibrium prices.

The selection problem extends to Budish (2011) competitive mechanism for CAP in which students reveal their preferences to a centralized clearinghouse who announces a corresponding equilibrium allocation. Budish argues that this mechanism is transparent, meaning that students can verify that the allocation is an equilibrium. But the mechanism can be "manipulated from the inside", assigning selectively zero prices to hand-picked courses, while at the same time rightly arguing that it produces a competitive allocation.

If IPD must be achieved (a decision depending on the context and the designer's objectives), we would like to have a solution that, at the same time, avoids the multiplicity problem of the CCE, while being envy-free and as egalitarian as possible.

Such solution exists: we call it the **refined egalitarian solution** or ES*. To define it, we use the partition of M into $\mathcal{P}(R,q)$ and $\mathcal{O}(R,q)$, and split the original GTP (R,q) into two independent problems $(R_{N\mathcal{P}(R,q)},q)$ and $(R_{N\mathcal{O}(R,q)},q)$, which correspond to the independent GTPs with perfect and the overdemanded days, respectively. ES* is given by

$$\phi_i^{\mathrm{ES}^*}(R,q) = \phi^{\mathrm{ES}}(R_{N\mathcal{O}(R,q)},q) + \left|R_{i\,\mathcal{P}(R,q)}\right| \tag{2.19}$$

ES* takes the egalitarian solution for the GTP with overdemanded days only, and adds the number of perfect days in which a player is available. ES* is close to a suggestion in Budish (2011). Budish, recognizing that some courses may be in excess supply, informally proposes to run the allocation mechanism only on the set of overdemanded courses: "*if some courses are known to be in substantial excess supply, we can reformulate the problem as one of allocating only the potential scarce courses*". ES* does exactly that, making precise what "substantial" means. It also satisfies several desiderata.

Lemma 5. *The ES* solution is well-defined and single-valued, efficient, IPD, envy-free, and Lorenz dominant for the problem* $(R_{N\mathcal{O}(R,q)}, q)$.

It is immediate that ES* is single-valued, efficient and IPD. The remaining properties are straightforward modifications of the proofs of Lemmas 1 and 2 and Theorem 1.

Unfortunately, the properties in Lemma 4 come at a cost: **ES* is not group strategy-proof**.[14] ES* can be manipulated by groups reducing their availability so to make certain days perfect. Therefore, the members of the manipulating coalition obtain those days fully, while obtaining also an egalitarian fraction of the remaining overdemanded problem.

Group strategy-proofness and IPD are compatible: priority solutions like EPD satisfy them both. However, their poor performance with respect to fairness make them inappropriate for the problems we have considered, as argued in subsection 4.3.

Before concluding the article, we discuss the complexity of solving GTPs with upper limits on the number of games an agent can play. We refer to those as horizontal quotas.

2.7 Adding Horizontal Quotas

An intuitive generalization of GTPs is to add upper limits or quotas on the number of games a player is willing to participate in, e.g. an agent that is available on 5 days but wants to play on at most 3. We did not present the results using this more general framework because our results do not extend to this setup.

The formalization of this generalized problem is similar to the one of a GTP defined in Section 3, with minor notational changes. A GTP with horizontal quotas (GTPQ) is a triple (R, q, κ), where (R, q) is a GTP with n agents, and $\kappa = (\kappa_1, \ldots, \kappa_n)$ is a vector of positive integers such that, $\forall i \in N$, $\kappa_i \leq |R_{iM}|$.

The set of all canonical RAMs for a GTPQ (R, q, κ) is defined as

$$\begin{aligned}\tilde{\mathcal{F}}(R,q,\kappa) \;=\; & \{Z \in [0,1]^{n\times m} \mid Z \text{ is IR for } R, \text{ and } \forall i \in N, \sum_{k\in M} z_{ik} \leq \kappa_i \\ & \text{and } \forall k \in M, \sum_{i \in N} z_{ik} \mod q = 0\} \end{aligned} \tag{2.20}$$

[14] For a manipulation example, use the GTP and manipulation R' illustrated in Table 4.

Given a GTPQ (R, q, κ), the set of feasible RAMs[15] is

$$\mathcal{F}(R, q, \kappa) = \{Y \in [0,1]^{n \times m} \mid Y = \sum_l \alpha_l Z_l\} \quad (2.21)$$

where $0 < \alpha_l \leq 1$, $\sum_l \alpha_l = 1$, and every $Z_l \in \tilde{\mathcal{F}}(R, q, \kappa)$.

As before, the set of utility profiles is only defined over the set $\mathcal{F}(R, q, \kappa)$. In the direct revelation mechanisms for GTPQs, players reveal (R_{iM}, κ_i), and then a detailed solution ϕ is applied to the corresponding irreducible problem (R, q, κ).

Solving a GTPQ is substantially more difficult than solving a GTP. First of all, the matching size is not constant across Pareto optimal assignments, as we illustrate in Example 4. Furthermore, we can use the same GTPQ to show that the ES solution is no longer Lorenz dominant nor group strategy-proof.

Example 4. (ES not Lorenz dominant nor group strategy-proof for GTPQs) Consider a GTPQ with $n = 13$, $m = 5$, $q = 4$, and (R, κ) given by subtable 2.6a. Two RAMs for this GTPQ are given in subtables 2.6b and 2.6c. While both are Pareto optimal, their matching size is 12 and 20, respectively. This is a first stark difference with the structure of Pareto optimal RAMs in GTPs.

Table 2.6: Example 4

$N \backslash M$	M	T	W	Th	F	κ
a	1	0	0	0	0	1
b	1	1	0	0	0	1
c	1	0	1	0	0	1
d	1	0	0	1	0	1
$e:g$	0	1	0	0	1	2
$h:j$	0	0	1	0	1	2
$k:m$	0	0	0	1	1	2
Total	4	4	4	4	9	

(a) (R, κ)

M	T	W	Th	F
1	0	0	0	0
1	0	0	0	0
1	0	0	0	0
1	0	0	0	0
0	0	0	0	8/9
0	0	0	0	8/9
0	0	0	0	8/9
4	0	0	0	8

(b) Z

M	T	W	Th	F
0	0	0	0	0
0	1	0	0	0
0	0	1	0	0
0	0	0	1	0
0	1	0	0	8/9
0	0	1	0	8/9
0	0	0	1	8/9
0	4	4	4	8

(c) Z'

The egalitarian solution for this problem, $U^{\text{ES}} = (U_a^{\text{ES}}, U_{b:d}^{\text{ES}}, U_{e:m}^{\text{ES}}) = (1, 1, 1)$, is obtained by randomizing between RAM Z with probability $\frac{17}{18}$, and RAM Z', with probability $\frac{1}{18}$.

[15] Defining $\mathcal{F}(R, q, \kappa)$ more succinctly is not possible because the matching size of Pareto optimal RAMs is not constant. Note also that $\tilde{\mathcal{F}}(R, q, \kappa) \subsetneq \mathcal{F}(R, q, \kappa)$, as the latter contains RAMs with columns whose sum is not mod $q = 0$, that can be obtained from randomization between RAMs in the former.

The ES solution does not Lorenz dominate the feasible utility profile corresponding to the RAM Z', which is $U' = (U'_a, U'_{b:d}, U'_{e:m}) = (0, 1, \frac{17}{9})$. Moreover, the ES solution is not even efficient here because more slots can be created by imposing a zero utility for agent a.

The egalitarian solution can be manipulated by groups in several ways. One manipulating coalition is $S = \{b, e\}$, with agent b reporting that he is only available on Tuesday (and agent e reporting her true availability). It is straightforward that now the ES solution to the new GTPQ (R', q, κ) is U', which benefits agent e (and 8 other agents $g : m$) while leaving the utility of agent b unchanged.

Which solution should we use for GTPQs? The problems of CCE that we have discussed obviously remain, so CCE is as at least as bad as ES, and EPD is not even defined for GTPQs. Finding a fair, efficient, and non-manipulable solution for GTPs with horizontal quotas remains an open question that we leave for further research.

2.8 Conclusion

We introduced a novel assignment problem, which differs from the previous literature in that the number of the goods to be shared is endogenously determined by players' preferences. Our problem is inspired by scheduling, but can be applied to several other matching problems in which the number of resources to be assigned is not fixed.

The egalitarian solution is single-valued, Lorenz dominant, and impossible to manipulate. For these reasons, we recommend its use as a solution in the dichotomous domain. If the market designer is interested in satisfying independence of perfect days, the refined egalitarian solution becomes an appealing alternative.

Two open questions are 1) whether the CCE is single-valued for GTPs with no perfect days, and 2) whether there are efficient, fair, and non-manipulable solutions for GTPs with horizontal quotas. Both are hard questions.

Appendix: Omitted Proofs

Theorem 1 *The ES solution is Lorenz dominant in the set of efficient utility profiles.*

Proof. Fix a GTP (R, q). Consider the concave cooperative game (N, μ) where $\mu : 2^N \to \mathbb{R}$ is a function that assigns, to each subset of players, the maximum number of slots they can obtain together, fixing the total number of slots available at $\nu(R, q)$. To formalize this intuitive function, given a coalition $S \subset N$, let us partition the set of days M into $M^+(S)$ and $M^-(S)$, defined as

$$M^+(S) = \{k \in M : |R_{Sk}| \leq \delta(k)\} \tag{2.22}$$

The function μ is given by

$$\mu(S) = \sum_{k \in M^+(S)} \sum_{i \in S} |r_{ik}| + \sum_{k \in M^-(S)} \delta(k) \tag{2.23}$$

This function is clearly submodular, i.e. for any two subsets $T, S \subset N$

$$\mu(S) + \mu(T) \geq \mu(S \cup T) + \mu(S \cap T) \tag{2.24}$$

The "core from above" is defined as the following set of profiles

$$C(R, q) = \{x \in \mathbb{R}^n \mid \sum_{i \in N} x = \nu(R, q) \text{ and } \nexists\, S \subset N : \sum_{s} x_i > \mu(S)\} \tag{2.25}$$

It follows from Theorem 3 in Dutta and Ray (1989) that the set $C(R, q)$ has a Lorenz dominant element and is the egalitarian solution. But by construction of the "core from above", $\mathcal{U}(R, q) \subset C(R, q)$, the ES solution is also Lorenz dominant in the set of efficient utility profiles $\mathcal{U}(R, q)$. □

Theorem 2 *For generalized tennis problems, the ES solution is well-defined and single-valued, and the CCE solution exists. Their intersection can be empty.*

Proof. Fix a GTP (R, q). Let $p \in \mathbb{R}^m_+$ be an arbitrary price vector such that $p \cdot \delta = n$, and use the notation $y_i = R_{iM}$ to denote the optimal consumption bundle for player $i \in N$, and $y_N = (|R_{N1}|, \ldots, |R_{Nm}|)$. Note that

$$p \cdot y_N \geq p \cdot \delta \tag{2.26}$$

Define the vector $\vec{\lambda}$ as

$$\vec{\lambda}(p) = (\lambda_1, \ldots, \lambda_n) = \text{UNIF}\{p \cdot y_i; n\} \tag{2.27}$$

where UNIF denotes the uniform rationing rule: a mapping that gives to every player the money needed to buy her preferred schedule as long as it is less than λ, chosen so that $p \cdot \vec{\lambda} = n$. Define the sets of satiated and non-satiated players

$$N_0(p) = \{i \in N \mid \lambda_i = p \cdot y_i\} \tag{2.28}$$

$$N_+(p) = \{i \in N \mid \lambda_i < p \cdot y_i\} \tag{2.29}$$

So that $\lambda_i = \lambda \, \forall i \in N_+$. Define the demand correspondence $d_i(p)$ as

$$d_i(p) = \arg \max_{Z_{iM} \in \mathcal{I}(R_{iM})} \{p \cdot Z_{iM} \leq \lambda_i\} \tag{2.30}$$

where $\mathcal{I}(R_{iM})$ denotes the set of individually rational assignments for R_{iM}. Note that $d_i(p) = \{y_i\}$ for every $i \in N_0(p)$, while for agents in $N_+(p)$, any vector $z_i \in d_i(p)$ satisfies $p \cdot z_i = \lambda$. By Berge's maximum theorem, the demand correspondence is upper hemi-continuous and convex valued. The excess demand correspondence for the whole society, which inherits the properties of d_i, is given by

$$e(p) = d_N(p) - \delta \tag{2.31}$$

where $d_N(p)$ denotes the aggregate demand correspondence for each day. Using the Gale-Nikaido-Debreu theorem (Theorem 7 in pp. 716-718 of Debreu (1982)), we know that there exists both a price vector $p^* \in R_+$ and an excess demand vector $x^* \in e(p^*)$ for which the following two conditions are satisfied

$$x^* = \vec{0} \tag{2.32}$$

$$p^* \cdot x^* = 0 \tag{2.33}$$

Where Walras' law in equation (2.33) holds by construction of $\vec{\lambda}$ and d. Finally, $\forall i \in N$

$$Z^*_{iM} = d_i(p^*) \tag{2.34}$$

so that the corresponding $Z^* \in \mathcal{F}(R, q)$ by equation (2.32), concluding the proof of existence of CCE. That ES is single-valued follows from Theorem 1. We have shown in Example 2 that for some GTP there do not exist prices that support the ES as a CCE. □

Lemma 2 *ES and CCE are envy-free, and hence treat equals equally.*

Proof. For an arbitrary GTP, let $\phi^{\text{ES}}(R, q) = (U_1, \dots, U_i, U_j, \dots, U_n)$, and assume player i is envious of j, which means that $R_{jM} \subseteq R_{iM}$ and that there exists a Pigou-Dalton transfer ϵ so that the utility profile $U' = (U_1, \dots, U_i + \epsilon, U_j - \epsilon, \dots, U_n) \in \mathcal{U}(R, q)$. But U' Lorenz dominates $\phi^{\text{ES}}(R, q)$, so $\phi^{\text{ES}}(R, q)$ was not the ES solution, a contradiction.

Any selection of the CCE solution is envy-free because of the standard argument: if there is any player who is envious, she could afford the schedule of the player she envies. □

Theorem 3 *ES is group strategy-proof but CCE is not.*

That CCE is not group strategy-proof was shown in the main text. To show that ES is group strategy-proof, we start with a few preliminaries. Let $\mathcal{Z}$ denote the set of all feasible RAMs supporting the egalitarian solution, i.e.

$$\mathcal{Z} = \{Z \in \mathcal{F}(R, q) \mid \forall i \in N : \sum_{k \in M} z_{ik} = \phi^{\text{ES}}_i(R, q)\} \tag{2.35}$$

As we mentioned in the main text, a rule is non-bossy if no player can change anyone's else utility without changing his own. This is, a solution ϕ is **non-bossy** if, for every GTP (R, q), $\forall i \in N$, and any manipulation R' such that 1) $\forall j \neq i, R_{jM} = R'_{jM}$, and 2)

$R'_{iM} \subsetneq R_{iM}$, we have

$$\phi_i(R, q) = \phi_i(R', q) \quad \text{only if} \quad \phi(R, q) = \phi(R', q) \tag{2.36}$$

We show that ES is non-bossy now.

Proof. We proceed by way of contradiction. Let R' be as specified in the previous definition. The manipulation may come from a reduction of availability in three types of days:

1. $k \in \mathcal{P}(R, q)$, but if player i reduces the number of perfect days, she always reduces the utility she obtains (we postpone this proof), so her utility is not constant and she cannot be bossy.

2. $k \in \mathcal{O}(R, q)$ and $\{k \in M \mid \exists Z \in \mathcal{Z} : z_{ik} = 0\}$, and hence there is a way to implement the ES solution even when player i misreported, so her change in availability is inconsequential and all utilities remain the same, so player i cannot be bossy.

3. $k \in \mathcal{O}(R, q)$ and $\{k \in M \mid \forall Z \in \mathcal{Z} : z_{ik} > 0\}$, so clearly player i's utility changes, so she cannot be bossy.

Now we prove our postponed claim: reducing the number of perfect days in which player i is available always strictly reduces her utility. The certain loss of the perfect day(s) must be (at least) exactly compensated by an increase of the shares she gets from all overdemanded days, which is constant in any $Z \in \mathcal{Z}$. Player i was not getting full shares on those day (as otherwise we obtain a contradiction) so another player(s) j must be obtaining shares those days, implying $\phi_j^{\mathrm{ES}}(R, q) \leq \phi_i^{\mathrm{ES}}(R, q)$. Moreover,

$$\phi_i^{\mathrm{ES}}(R, q) - 1 < \phi_j^{\mathrm{ES}}(R, q) \leq \phi_i^{\mathrm{ES}}(R, q) \tag{2.37}$$

as otherwise j does not transfer any shares to i when i reduces the number of perfect days. Let γ be the Pigou-Dalton transfer from j to i required so that the utility of i is kept constant. We have

$$\phi_i^{\mathrm{ES}}(R', q) = \phi_i^{\mathrm{ES}}(R, q) - 1 + \gamma = \phi_j^{\mathrm{ES}}(R, q) - \gamma < \phi_i^{\mathrm{ES}}(R, j) \tag{2.38}$$

showing that indeed reducing the number of perfect days always yields lower utility, and thus concluding the proof that ES is non-bossy. $\square$

We are now ready to prove that ES is group strategy-proof. We will do it by showing that nobody can join a manipulating coalition.

Proof. By way of contradiction, assume there exists a GTP (R, q), a coalition $S \subsetneq N$, and a manipulation R' such that, for all $i \in S$ $\phi_i^{\text{ES}}(R', q) \geq \phi_i^{\text{ES}}(R, q)$, and for some $j \in S$ $\phi_j^{\text{ES}}(R', q) > \phi_j^{\text{ES}}(R, q)$.

Let $\phi^{\text{ES}}(R, q) = U^{\text{ES}}$ and order the players such that $U_1^{\text{ES}} \leq \ldots \leq U_n^{\text{ES}}$. We will show by induction on the order of players the following property

$$i \notin S \tag{2.39}$$

There are two cases in which an agent i can be in S. Case 1) either he gets more utility, $\phi_i^{\text{ES}}(R', q) > \phi_i^{\text{ES}}(R, q)$, or case 2) he gets the same utility but he changes his reported preferences to help another member of S. This is ruled out by non-bossiness of ES so we focus on case 1) only.

We prove it for $i = 1$ first, i.e. the player with lowest utility. Player 1 gets a strictly higher number of slots with the new profile R', which must come from a set of days $K \subseteq \mathcal{O}(R, q)$ in which he was not playing with certainty ($K = \{k \in M \mid \exists Z \in \mathcal{Z} : 0 < z_{ik} < 1\}$), for which players $2, \ldots, q, \ldots, t$ are also available and $U_1^{\text{ES}} = U_2^{\text{ES}} = \ldots = U_t^{\text{ES}}$. Those players exhaust $\delta(k)$ entirely; i.e. $\forall k \in K, \forall Z \in \mathcal{Z}, \sum_1^t z_{ik} = \delta(k)$.

Let $T = \{1, \ldots, t\} \cap S$. For any availability matrix R'_{TM} that is individually rational for R_{TM}, the days $\{k \in K \mid R_{Nk} \neq R'_{Nk}\}$ become less overdemanded for players $\{1, \ldots, t\} \setminus T$, and therefore the players in T get less games as a whole. Therefore there must be at least one player in T who is worst off, and the coalition S is not viable. Therefore $1 \notin S$.

Now we assume that $i \notin S$ for player $i = h - 1$ and we show it holds for player h. We must have that $U_h^{\text{ES}} < |R_{hM}|$. We assume $\phi_1^{\text{ES}}(R, q)_1^{\text{ES}} < \phi_h^{\text{ES}}(R, q)$ as otherwise our argument for player 1 works exactly the same.

If player $h \in S$, it must be that there exists a manipulation R' so that $\phi_h(R', q) > \phi_h(R, q)$. The increase in her utility must come from more game shares on overdemanded days in which she was not playing with certainty, i.e. $K^h = \{k \in M \mid \exists Z \in \mathcal{Z} : 0 < z_{hk} < 1\}$. Some of these days are exhausted by players $1, \ldots, h-1$. There is no way player h could get more shares in any of those days because $\{1, \ldots, h-1\} \cap S = \emptyset$ by our induction step.

Therefore, the increase must come from days that are not exhausted by $\{1, \ldots, h-1\}$. Those days become less overdemanded for $\{h, \ldots, n\} \setminus S$, and therefore players in S get less game shares as a whole. It follows that there must be a player in S who gets less utility, so coalition S is not viable. Therefore $h \notin S$, and this concludes the proof.

As a technical remark, in some assignment problems strategy-proofness plus non-bossiness implies group strategy-proofness. This is not the case for GTPs: see for example the refined egalitarian solution ES*, which is strategy-proof and non-bossy, and yet fails group strategy-proofness. $\square$

Lemma 4 *Although ES is not IPD, there exists a selection of CCE that satisfies IPD.*

Proof. It is straightforward to show that ES is not IPD. Let $n = 5, M = \{\text{Mon}\}, q = 4$, and $R^\top = [1\ 1\ 1\ 1\ 1]$. $\phi_i^{\text{ES}}(R, 4) = 0.8$ for any player, but adding a perfect day k' for any player i changes $\phi_i^{\text{ES}}([R\, R_{Nk'}], 4) = 1.75 \neq 2$.

To show that there is a selection of Φ^{CCE} that is IPD, let (Z^*, p^*) be a CCE of (R, q). Then fix $p^*_{k'} = 0$ and, for every $i \in N$ let $z^*_{ik'} = 1$ if $r_{ik'} = 1$, and 0 otherwise. The pair $([Z^*\, Z^*_{Nk'}], (p^*_1, \ldots, p^*_n, 0))$ is a CCE of the new problem $([R\, R_{Nk'}], q)$, because everybody interested in the perfect day is able to afford it, and the demand for k' equals its supply, because the new day is perfect. $\square$

Chapter 3

Can Everybody Benefit from Social Integration?

Abstract

There is no matching mechanism that satisfies integration monotonicity and stability. If we require integration monotonicity, we cannot even achieve Pareto optimality: the only option is to remain segregated.

A weaker monotonicity condition is compatible with Pareto optimality but not with path independence, which implies that the dynamics of social integration matter.

If the outcome of integration is stable, integration is always approved by majority voting, but a non-vanishing fraction of agents always oppose segregation. The side who receives the proposals in the deferred acceptance algorithm suffers significant welfare losses, which nevertheless become negligible when societies grow large.

KEYWORDS: integration, desegregation, integration monotonicity, interracial marriage.

JEL CLASSIFICATION: DC78 (matching theory), J12 (marriage and marital dissolution)

3.1 Our Problem and its Relevance

Imagine several completely isolated communities that match within themselves, but that could expand their boundaries to merge and match as a unified community instead. The question I ask is whether every person would prefer that all communities integrate as one, provided that the matching outcome is either stable or efficient. Some examples that motivate my research question are:

1. **Interracial marriage**. Three infamous cases of societies that banned interracial marriage are i) the U.S. before the *Virginia vs Loving* case in 1967 (Arrow, 1998; Fryer, 2007), ii) Nazi Germany, where the marriage between arians and non-arians was forbidden (Caestecker and Fraser, 2008), and iii) South Africa during the apartheid era, when the *the Prohibition of Mixed Marriages Act* was established (Hyslop, 1995).

   In all cases, social integration occurred only after complicated social movements, and without unanimous approval.

2. **Centralized Kidney Exchange**. After 2000, kidney exchanges began to take place internally in hospitals around the U.S. Few years after, centralized programs started to conduct regional kidney exchanges by asking hospitals to share their donor-patient pairs. Using a centralized procedure would always weakly increase the number of transplants, yet it has been noticed that some hospitals may not have incentives to integrate into to the central clearinghouse, preferring to conduct exchanges only internally. The aforementioned rejection to integrate to a centralized clearinghouse has been documented in practice (Ashlagi and Roth, 2014).

3. **School Desegregation** In 1954, school segregation was declared illegal in the U.S. following the *Brown vs Board of Education* case. Although the desegregation ruling was widely acknowledged as a major accomplishment, it was not well-received by some. A shameful example is the resistance by the governor of Arkansas, who tried to prevent a few Black students from attending a newly desegregated school. The

students were able to enter the school only when they were escorted by federal forces (U.S. Commission of Civil Rights, 1977).

All these examples show instances on which social integration was complicated to achieve in matching environments. I formalize these environments with an extended version of the Gale and Shapley (1962) matching problem with non-transferable utility. The idea of the model is simple: we take several classical Gale-Shapley problems and find their segregated women-optimal stable partner. Then we put them together and compute the integrated women-optimal stable matching, and compare who prefers the integrated matching to the segregated one. Stability is natural requirement to ask for, because decentralized matching environments produce outcomes close to those predicted by stability c, and because centralized mechanisms that produce stable outcomes are widely regarded as successful (Roth, 2002).

I derive several impossibilities showing that social integration cannot benefit everyone whenever the matching outcome is stable or efficient.[1] The impossibility disappears when only weak integration monotonicity and Pareto efficiency are required: an observation which may be obvious, but the reader is reminded that weak monotonicity is still incompatible with stability. Weak integration monotonicity only requires that when the complete society merges, everybody is better off than remaining in disjoint segregated communities. It says nothing about the process of partial integration of all communities. I also prove that social integration is always approved by a weak majority of the population.

Interestingly, such majority rarely surpasses 80% of the population when considering random instances of matching problems, emphasizing the complicated dynamics of desegregation, and making evident that a non-vanishing minority may oppose social integration because they foresee that they will be worse off belonging to an integrated community. Also interestingly, the welfare losses for those who oppose integration become negligible with respect to the size of the grand society as communities grow large, but in small societies the side who receives the proposals in the deferred acceptance algorithm suffers

[1]The impossibilities occur when we require that for every union of disjoint communities, the resulting matching is weakly better off for each agent.

significant welfare losses. Finally, it is also surprising that those who oppose integration are indistinguishable in terms of expected ranking from those who prefer integration.

I use a one-to-one matching framework, but all the impossibility results obviously extend to the many-to-one matching case. Since the interracial marriage example is the one closer to one-to-one matching, I will present the model in those terms, but the reader should keep in mind that the results apply to general matching problems.

After reviewing the related literature in Section 2, Section 3 introduces the model and defines the integration monotonicity property. Section 4 presents several impossibility theorems regarding the existence of integration monotonic matchings that have also any degree of efficiency, emphasizing the rigid structure that monotonicity imposes in matching problems.

Section 5 details the limits of our impossibility results, showing that although some agents oppose social integration, they are always a minority. It also describes the size of such group when the societies become large. Section 6 explores the properties of those who oppose integration and their welfare loss. Finally, section 7 concludes.

3.2 Related Literature

3.2.1 Comparative Statics in Matching Problems

Within the matching literature, there is a body of work that studies how the set of stable outcomes changes when a new agent joins an existing society. The main result in this literature is that, when a new man joins a stable matching problem, every women weakly improves, while every man becomes weakly worse off. This result is robust to various formulations of the problem such as many-to-one extensions and preferences determined by choice functions: see theorem 5 in Kelso and Crawford (1982), theorems 2.25 and 2.26 in Roth and Sotomayor (1992), theorems 1 and 2 in Crawford (1991), and theorem 2 in Chambers and Yenmez (2017).

The aforementioned welfare loss that men suffer when a new men joins the problem has been recently quantified, by assigning agents with independent random preferences over all their partners. Pittel (1989) shows that, in expectation, the side of the society that proposes in the deferred acceptance algorithm, say men, gets matched to a woman ranked $\log(n)$ in their preference lists, whereas women get in expectation a man in the $\frac{n}{\log(n)}$ position of theirs. The partner ranked first is the best possible partner, and so on, and n is the number of potential partners for each agent.

Using the same probabilistic framework as Pittel, Ashlagi et al. (2017) find that just by adding an additional man, men receive a partner ranked $\frac{n}{\log(n)}$ with high probability (in the men-optimal matching), whereas women will receive someone close to $\log(n)$. Interestingly, all stable marriages are similar whenever societies are even slightly unbalanced in their ratio between men and women.

In all cases, the discussion centers on what happens when adding an individual alone to a society and not when merging isolated societies of same size. This is my main contribution with respect to the surveyed literature.

3.2.2 Integration and Population Monotonicity Elsewhere

Chambers and Hayashi (2017) introduced integration monotonicity and derived similar results for economic integration. They consider several exchange economies, in which each agent has an initial endowment, that integrate as one. They find that there is no path-independent exchange mechanism that is integration monotonic and Pareto efficient. If the integration mechanism is Pareto efficient and satisfies the additional property of equal treatment of equals, it must necessarily harm one third of all agents in the economy.

Sprumont (1990) considers population monotonic schemes in cooperative games with transferable utility. An allocation scheme is population monotonic if each time an agents joins an existing problem, the payoff for every existing member increases. He shows that every convex game admits a population monotonic allocation scheme, and provides a tighter characterization using linear combinations of games with veto control. His work deals with transferable utility games only.

Related population monotonicity concepts in cooperative games are widely used in different environments, based on the seminal work of Moulin and Thomson (Thomson

(1983), Moulin (1990, 1992); Moulin and Thomson (1988)). This paper's title is inspired on the last of those articles. As in the matching literature, all these monotonicity concepts deal with adding an agent to a problem instead of merging problems of the same size. Sprumont (2008) presents a detailed review of the work in this area.

3.3 Model

Let S^k be a **society** of race k that consist of n men M^k and n women W^k. I refer to man i (woman j) that belongs to society k by m_i^k (w_j^k). When I refer to an agent of arbitrary gender I use x_i^k; I omit the subindices when $n = 1$. There are $r \geq 2$ races and $R = \{1, \ldots, r\}$. For any subset $T \subseteq R$, let $M^T = \bigcup_{k \in T} M^k$, $W^T = \bigcup_{k \in T} W^k$, and $S^T = M^T \cup W^T$. S^R is called the grand society.

Each man (woman) has strict preferences over the entire set of women W^R (men M^R) and not only over those belonging to her own race. I represent the preferences of an arbitrary person x_i^k by $P(x_i^k)$. $F(x_i^k)$ denotes the weak preference relation associated to $P(x_i^k)$ so that for any two agents y_j^l and z_g^e, $y_j^l \; F(x_i^k) \; z_g^e$ if and only if $y_j^l \; P(x_i^k) \; z_g^e$ or $y_j^l = z_g^e$. I assume that every person prefers matching with any potential partner of the opposite gender than remaining alone.

P^T will denote the preference lists of every person in $S^T \subseteq S^R$. The pair (S^R, P^R) is an **interracial matching problem** (IMP).

A **matching** $\mu : M^R \cup W^R \times 2^R \rightarrow M^R \cup W^R$ is a mapping such that, $\forall T \subseteq R$,

$$\forall m_i^k \in M^T, \qquad \mu(m_i^k, T) \in W^T \tag{3.1}$$

$$\forall w_i^k \in W^T, \qquad \mu(w_i^k, T) \in M^T \tag{3.2}$$

$$\forall x_i^k \in S^T, \qquad \mu(\mu(x_i^k, T), T) = x_i^k \tag{3.3}$$

so that every man is married to a woman in the specified society S^T and vice versa. The function μ indicates who marries whom under every union of races. Naturally, $\mu(x_i^k, T)$ is only defined whenever $x_i^k \in T$.

In the majority of matching literature, a matching is defined instead as a mapping $\mu' : M \cup W \to M \cup W$. My definition of μ corresponds to the one of an allocation scheme of the matching μ', as defined by Sprumont (1990), which specifies a matching μ' for each subset of R. For convenience, I just refer to such allocation scheme as a matching.

Let T and Q be an arbitrary partition of R, with $T = \{a, b, \ldots, r'\}$. Let the **colorblind equivalent** of S^T be denoted by $S^{\overline{T}}$, in which every agent $x_i^k \in S^T$ becomes of a new race $\overline{ab \ldots r'}$, and in which the preferences of each agent in S^R remain the same up to the renaming in agents' race.

We define the following properties of interest for an arbitrary matching μ, given an IMP.

Pareto Optimality There is no different matching μ' such that, for all $T \subseteq R$ and all $x_i^k \in S^T$

$$\mu'(x_i^k, T)\ F(x_i^k)\ \mu(x_i^k, T), \tag{3.4}$$

and for some $Q \subseteq R$ and some $y_j^l \in S^R$,

$$\mu'(y_j^l, Q)\ P(y_j^l)\ \mu(y_j^l, Q) \tag{3.5}$$

Pareto optimality is a classical requirement and a basic efficiency concern. It can be strengthened to the stronger efficiency concept of stability.

Stability For every subset $T \subseteq R$, and for every $m_i^k, w_j^l \in S^T$, such that

$$m_i^k \notin \mu(w_j^l, T) \quad \text{and} \quad w_j^l \notin \mu(m_i^k, T) \tag{3.6}$$

either

$$\mu(m_i^k, T)\ P(m_i^k)\ w_j^l \quad \text{or} \quad \mu(w_j^l, T)\ P(w_j^l)\ m_i^k \tag{3.7}$$

Stability is an important requirement because it closely predicts realized outcomes in decentralized environments and because if the final outcome was not stable, it would be unlikely that it lasted long from a game-theoretical perspective.[2] Now we turn to integration monotonicity.

Integration Monotonicity For all disjoint subsets $T, Q \subseteq R$, and for every $x_i^k \in S^T$

$$\mu(x_i^k, T \cup Q)\ F(x_i^k)\ \mu(x_i^k, T) \tag{3.8}$$

Note that integration monotonicity not only requires that the matching obtained when all races have integrated is better than the one obtained with a society alone. It requires that anytime another race joins, it always benefits every agent in the existing societies. We will relax this requirement in Section 5. Finally, we define path independence.

Path Independence For all disjoint subsets $T, Q \subseteq R$, and for every $x_i^k \in S^R$

$$\mu(x_i^k, T \cup Q) = \mu(x_i^k, \overline{T} \cup Q) \tag{3.9}$$

where $\overline{T}$ denotes the colorblind equivalent of T.

Path independence is a more technical requirement, but nevertheless relevant because if a matching violates path independence, the dynamics of integration would play a role in determining the final pairings.

From our four properties, only stability and Pareto optimality are related.[3]

Lemma 6. *Every stable matching is Pareto optimal.*

Proof. Let μ be stable. Therefore in any alternative matching μ' that is better for person x_i^k at $T \subseteq R$, we have that $\mu(\mu'(x_i^k, T), T)\ P(\mu'(x_i^k, T))\ x_i^k$, which means the new partner of x_i^k prefers matching μ to μ' at T, and hence μ' is not a Pareto improvement. $\square$

The converse statement is clearly not true.

[2] For example, realized romantic pairings are similar to those predicted by stability, see Hitsch et al. (2010) and Banerjee et al. (2013). Stability is also related to the successful operation of centralized matching mechanisms such as kidney exchanges programs and school choice (Roth, 2002).

[3] In exchange economies integration monotonicity and efficiency imply core stability (Lemma 2 in Chambers and Hayashi (2017)). A similar conclusion applies in the housing model of Shapley and Scarf (1974). For two-sided matching that relationship does not hold.

3.4 Results

Unfortunately, stability and integration monotonicity are not compatible even with just two societies with two persons each.[4]

Proposition 1. *Not every IMP admits a matching that satisfies stability and integration monotonicity.*

Proof. *(Example 1)* Let $R = \{a, b\}$ and $n = 1$, and let agents' preferences be

$$\boldsymbol{m^a} : w^a \ P(m^a) \ w^b \qquad \boldsymbol{w^a} : m^b \ P(w^a) \ m^a$$

$$\boldsymbol{m^b} : w^a \ P(m^b) \ w^b \qquad \boldsymbol{w^b} : m^b \ P(w^b) \ m^a$$

The unique stable matching has $\mu(m^k, \{k\}) = w^k$ for $k \in \{a, b\}$ but $\mu(w^a, R) = m^b$ and $\mu(w^b, R) = m^a$. Yet μ violates integration monotonicity for m^a because $\mu(m^a, \{a\}) \ P(m^a) \ \mu(m^a, R)$. The same occurs for w^b. □

Given that stability and integration monotonicity are incompatible, an obvious question is whether we can weaken any of those two properties to avoid the impossibility. To address it, let us define a particular matching, called the **segregated matching**.

Let λ be a matching such that $\lambda(x_i^k, T)$ assigns to each agent x_i^k the women-optimal stable matching[5] in the matching problem $(M^T, W^T; P^T)$ for each $T \subseteq R$. The segregated matching σ is defined as

$$\forall T \subseteq R, \forall x_i^k \in S^R, \quad \sigma(x_i^k, T) = \sigma(x_i^k) = \lambda(x_i^k, \{k\}) \tag{3.10}$$

so that for any subset T, it assigns to each individual the women-optimal matching obtained when matching each race alone. The segregated matching is clearly integration monotonic,

[4] Sprumont (1990) proves a similar result: any assignment game with two men and two women lacks a population monotonic assignment scheme. His result does not imply any of mines because his works deals with transferable utility games only.

[5] I always pick the women-optimal stable matching to have a consistent selection from the set of stable matchings. We could consider the men-optimal one as well. The selection problem is not a big issue, as in large societies there is a unique stable matching whenever agents have short preferences or the societies are unbalanced in their ratio between men and women (Immorlica and Mahdian, 2005; Kojima and Pathak, 2009; Ashlagi et al., 2017).

but it fails to be stable when aggregating the individual societies. The segregated matching even fails Pareto optimality, as the preferences in Example 2 shows.

Example 2: The segregated matching is not Pareto optimal.

$$\boldsymbol{m^a} : w^b \ P(m^a) \ w^a \qquad \boldsymbol{w^a} : m^b \ P(w^a) \ m^a$$
$$\boldsymbol{m^b} : w^a \ P(m^b) \ w^b \qquad \boldsymbol{w^b} : m^a \ P(w^b) \ m^b$$

I start by weakening stability and requiring Pareto optimality only. Can we obtain always a matching that is Pareto optimal and integration monotonic? The answer is that not even such weakening of optimality is enough.

Proposition 2. *Not every IMP admits a Pareto optimal and integration monotonic matching.*

Proof. (Example 3) Let $R = \{a, b, c\}$ and $n = 3$, and let agents' preferences be

$$\boldsymbol{m^a} : w^b \ P(m^a) \ w^c \ P(m^a) \ w^a \qquad \boldsymbol{w^a} : m^b \ P(w^a) \ m^c \ P(w^a) \ m^a$$
$$\boldsymbol{m^b} : w^c \ P(m^b) \ w^a \ P(m^b) \ w^b \qquad \boldsymbol{w^b} : m^c \ P(w^b) \ m^a \ P(w^b) \ m^b$$
$$\boldsymbol{m^c} : w^a \ P(m^c) \ w^b \ P(m^c) \ w^c \qquad \boldsymbol{w^c} : m^a \ P(w^c) \ m^b \ P(w^c) \ m^c$$

Any Pareto optimal matching μ has $\mu(w^a, \{a, b\}) = m^b$ and $\mu(w^a, \{a, c\}) = m^c$. Therefore $\mu(w^a, R) = m^b$ by integration monotonicity. But exactly the same argument for m^b shows that he gets $\mu(m^b, R) = w^c$. Therefore, no Pareto optimal matching satisfies integration monotonicity. □

Note that Chambers and Hayashi (2017) are always able to find a mechanism that is Pareto optimal and integration monotonic, although not path-independent. Therefore, the impossibility we obtain is stronger. A first conclusion is that *achieving complete social integration is more difficult than obtaining complete economic integration.*

An immediate corollary follows, showing that if one is to pursue integration monotonicity, there is no room for efficiency even in its weakest form.

Corollary 1. *The only matching that satisfies integration monotonicity in every IMP is the segregated matching.*

Given the negative result obtained, let us focus on matchings that satisfy a more flexible monotonicity condition, defined below.

Weak Integration Monotonicity For any race $k \in R$, and for every x_i^k,

$$\mu(x_i^k, R)\ F(x_i^k)\ \sigma(x_i^k) \tag{3.11}$$

Weak integration monotonicity only requires that the corresponding matching when all races have integrated is better than the segregated matching when all societies are segregated. It says nothing about the relationship between matchings obtained under partial integration.

This mild monotonicity is still inconsistent with stability, as our previous Example 1 shows. It can be combined with optimality, yet not without consistency problems.

Proposition 3. *Every IMP admits a matching that is weakly integration monotonic and Pareto efficient. If we add path-independence, we obtain an impossibility.*

Proof. For any $T \subseteq R$, implement the matching σ. If the segregated matching is not Pareto optimal, then implement a Pareto optimal matching μ that dominates σ, and so on. Trivially, every agent is better off. Note that every agent has a veto power over stable matchings that benefits others but hurt her/him.

To show that path-independence cannot be added, consider the society in Example 3. Let us merge societies into their colorblind equivalents: a and b into $\overline{ab}$, and a and c into $\overline{ac}$. The unique Pareto optimal matching μ is such that $\mu(m_1^a, \{\overline{ab}, c\}) = w_1^b$, but $\mu(m_1^a, \{\overline{ac}, b\}) = w_1^c$. $\square$

3.5 The Limits of Segregation

How many people prefer segregation over complete integration, provided that the (women-optimal) stable matching will realize when societies merge? If there will be a referendum asking whether all individual societies should merge, could it be that segregation would obtain a majority of votes?

Let us assume that everybody who does not get hurt by integration votes in favor of it. In Example 1, half of the society votes against integration. Can it be more? The answer is no.

Proposition 4. *For any IMP (S^R, P^R), at most $\lfloor rn \rfloor$ agents prefer segregation. The bound is tight.*

Proof. Let us partition S^R into three sets A, B^+ and B^-, defined as

$$A = \{x_i^k \in S^R \mid \lambda(x_i^k, R) = \sigma(x_i^k)\} \tag{3.12}$$

$$B^+ = \{x_i^k \in S^R \mid \lambda(x_i^k, R)\ P(x_i^k)\ \sigma(x_i^k)\} \tag{3.13}$$

So A is the set of people who keep the same partner after integration, B^+ are those who prefer their "integrated" partner, and B^- are those who prefer the "segregated" partner.

Now consider the directed graph which contains all women from B^+ and B^-, in which every woman points to the woman from whom she "stole" her new husband in the integrated society, i.e. this is each woman w_i^k points towards $\sigma(\lambda(w_i^k, R))$. A cycle always forms whenever $A \neq S^R$.

Now consider an arbitrary woman in B^+, which must exist if $A \neq S^R$ because λ produces the women-optimal stable matching. She points to a woman w_j^l, who can either be in B^+ or in B^-. If she is in the latter, it must be that $\sigma(w_j^l)$ is in B^+, because she proposed to him at some point in the woman-proposing deferred acceptance algorithm but he rejected her. This goes on for any woman who is worst off after integration: her previous partner must necessarily be better off after integration, because he rejected her when she proposed to him in the deferred acceptance algorithm.

It follows that $|B^+| \geq |B^-|$, and thus $|A| + |B^+| \geq |B^-|$, which implies that always at least half of the society supports integration, completing the proof. $\square$

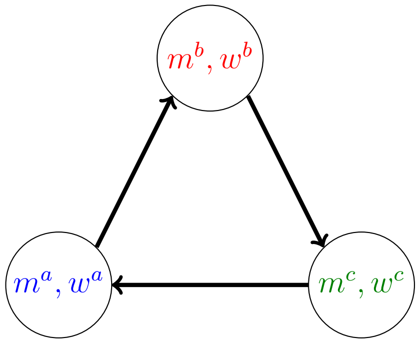


Figure 3.1: The procedure in the proof of Proposition 4 applied to Example 3.

Proposition 4 only applies to one-to-one matching. To extend it to many-to-one matching one needs to be careful to define: 1) the structure of the preferences, which may exhibit substitutes and complements; and 2) how do we count colleges. It could be that either each college counts as one, or that each college counts for as many seats it has, i.e. its quota. It is well-known that if preferences are *responsive*, each many-to-one matching has a corresponding one-to-one matching, in which each college with capacity or quota q_c is replaced with q_c copies of itself. Using this equivalence, our Proposition 4 extends to many-to-one matchings too.

Proposition 4 is interesting because it tells us that a referendum for integration will always be accepted by a *weak* majority. However, it could be that the voting rule we need is a *supermajority*, that implements integration only if the number of agents that get hurt from integration are at most a fraction ϵ of the population.[6]

A natural conjecture is that, for large societies, integration is always approved in any ϵ-supermajority, for any arbitrarily small ϵ. The conjecture is natural because, when the number of agents grows, agents win a larger pool of potential partners when social integration realizes. Stated differently, that the fraction of people that reject social integration is vanishingly small when the societies become large. Yet, this conjecture

[6] Chambers and Hayashi (2017) use the equivalent concept of *integration monotonicity under P-vetoes*, in which at most a group of people of size $|P|$ may oppose integration. Both concepts are mild versions of integration monotonicity.

appears to be false for small values of r. Looking at what happens when r is small is particularly interesting because in reality we have only a few races.

Let (S^R, P^R) be an IMP in which each agents' preferences are chosen independently and uniformly at random from the set of possible strict preferences. Let $\Omega_r(n)$ denote the expected number of agents who prefer the segregated matching over the women-optimal stable matching in the grand society.

Conjecture 1. *For $r \leq 5$,*

$$\lim_{n\to\infty} \frac{\Omega_r(n)}{2rn} \neq 0 \tag{3.14}$$

Our conjecture that $\frac{\Omega_r(n)}{2rn}$ does not vanish is supported by Monte Carlo simulations using Matlab presented in Table 1.[7] The code used is available from my webpage. I stopped the simulations at $2n = 1000$ because it already took three days to run in a high performance computing facility ($2n$ is the number of agents of each race). It is clear from Table 1 that convergence occurs in all cases.

Table 3.1: How many people (in percentage) prefer segregation?

$r \backslash 2n$	100	200	1000
2	25.42	25.76	25.47
	(0.04)	(0.02)	(0.01)
3	25.43	25.83	25.99
	(0.03)	(0.01)	(0.01)
4	24.84	25.14	25.57
	(0.02)	(0.01)	(0.01)
5	24.30	24.60	25.05
	(0.02)	(0.01)	(0.01)

Average over a thousand simulations with preferences drawn uniformly at random. Standard errors in parenthesis.

Our $\frac{\Omega_r(n)}{2rn} \approx .25$ is related to theorem 2 in Chambers and Hayashi (2017). Their result states that when societies merge, the fraction of people who oppose *economic* integration is always above one third under equal treatment of equals. Their result, looking at a worst case scenario, is obtained in a very different fashion. The comparison of our results suggest

[7] The code was run on the high performance computing facilities of the University of Glasgow, and it uses the Matlab package to compute the women-optimal stable marriage, developed by S. Gopalakrishnan.

a second conclusion: *it is easier to achieve partial social integration than partial economic integration.*

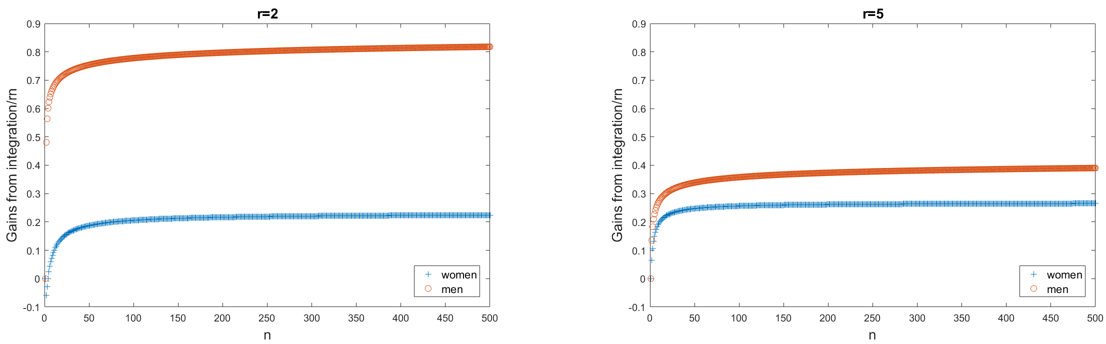
Using the same probabilistic IMP with random preferences, we can find the expected welfare gains derived from integration. Applying the well-known result from Pittel (1989) about expected rankings of partners in random matching problems, it is easy to see that women get a higher ranked partner in expectation after integration occurs, because

$$\underbrace{\log(n)\left(\frac{rn+1}{n+1}\right)}_{\text{exp. ranking w. segregation}} - \underbrace{\log(rn)}_{\text{exp. ranking w. integration}} = \frac{n(r-1)}{n+1}\log(n) - \log(r) \tag{3.15}$$

which is positive for all sensible values of r and n, meaning women get a partner that appears earlier on their preference lists. Similarly, men get a better partner after integration for sensible values of r and n, because

$$\underbrace{\frac{n}{\log(n)}\left(\frac{rn+1}{n+1}\right)}_{\text{exp. ranking w. segregation}} - \underbrace{\frac{rn}{\log(rn)}}_{\text{exp. ranking w. integration}} = A[(rn+1)\log(r) - (r-1)\log(n)] \tag{3.16}$$

where $A = n/[(n+1)\log(n)\log(rn)]$. The gains from integration in an IMP with random uncorrelated preferences are given by the sum of the previous expressions multiplied by rn. The normalized gains from integration for a man and a woman are depicted in Figure 2.



(a) $r = 2$ (b) $r = 5$

Figure 3.2: Individual gains from integration divided by rn, by gender

3.6 Who Prefers Segregation?

First let us look at the expected relative number of people who keep the same partner after integration. Since the preferences are drawn uniformly, everybody has the same probability of matching an agent from their own race: this is $1/r$. A natural guess is that, among those, $1/2$ of them do not change their marriage, which provides a good intuition of the real numbers described in Table 2.

Table 3.2: How many people (in percentage) keep the same partner after integration?

$r \backslash 2n$	100	200	1000	$1/2r$
2	27.26	27.14	27.02	25
	(0.12)	(0.07)	(0.02)	
3	18.39	18.23	17.89	16.66
	(0.06)	(0.03)	(0.01)	
4	13.93	13.62	13.41	12.5
	(0.03)	(0.02)	(0.01)	
5	11.31	11.21	10.84	10
	(0.02)	(0.1)	(0.01)	

Average over a thousand simulations with preferences drawn uniformly at random. Standard errors in parenthesis.

Table 2 shows that, as r grows, the number of people who are indifferent between integration and segregation becomes smaller. Since the proportion of people who oppose social integration keeps relatively constant as described in Table 1, the number of people who strongly prefer integration does grow, although as we saw it rarely goes over four-fifths of the entire society.

Another natural conjecture is that the people who oppose social integration have a lower expected desirability than those who do not. In other words, they are usually ranked lower in the preference lists of the potential partners. And this new conjecture is false too. Table 3 describes the expected rank of people who prefer segregation: it is immediate that those who prefer segregation have the same expected ranking as a random person, showing that people who prefer segregation are not particularly undesirable agents, they are just like anybody else.[8]

[8] This result depends crucially on the assumption of independent preferences, as it can be noted in the next Chapter.

Table 3.3: Average rank of people who prefer segregation, by gender.

$r\backslash 2n$	100		200		1000	
	women	men	women	men	women	men
2	50.7	50.5	100.6	100.5	500.5	500.5
	(0.30)	(0.07)	(0.22)	(0.10)	(0.21)	(0.09)
3	75.7	75.5	150.7	150.5	750.6	750.5
	(0.25)	(0.09)	(0.24)	(0.08)	(0.25)	(0.08)
4	100.6	100.5	200.6	200.5	1000.7	1000.5
	(0.25)	(0.09)	(0.27)	(0.07)	(0.24)	(0.08)
5	125.6	125.5	250.6	250.5	1250.6	1250.5
	(0.29)	(0.08)	(0.29)	(0.08)	(0.28)	(0.08)

Average over a thousand simulations with preferences drawn uniformly at random. Standard errors in parenthesis.

Finally, we look at the welfare losses suffered by those who prefer segregation when integration realizes, in terms of ranking of their current partner. If their loss was relatively small it would be a strong argument for saying that the impossibilities described in Section 3 are basically irrelevant. Table 4 summarizes an interesting result: the side of the society who does not propose, in this case men, get severely hurt by integration for moderate values of n. Women, the proposing side, suffer a moderate hurt at most.

Table 3.4: Average welfare loss by people who prefer segregation, by gender.

$r\backslash 2n$	100		200		1000	
	women	men	women	men	women	men
2	4.9	19.7	5.7	34.9	7.4	136.6
	(0.91)	(9.47)	(0.95)	(22.91)	(1.04)	(246.83)
3	5.4	27.4	6.2	49.2	7.9	193.8
	(1.01)	(14.57)	(0.96)	(39.22)	(0.83)	(409.86)
4	5.7	35	6.5	62.1	8.1	250.8
	(1.08)	(21.67)	(1.03)	(60.48)	(0.84)	(623.95)
5	6	41.8	6.8	74.9	8.4	303.3
	(1.07)	(28.26)	(1.14)	(84.86)	(0.88)	(709.15)

Average over a thousand simulations. Welfare loss measured in difference in ranking of partners. Standard errors in parenthesis.

Table 4 reveals that the welfare loss becomes smaller with respect to the size of the grand society as n grows, suggesting that they become negligible in the limit. This finding suggests that integration could be more easily implemented in large societies.

3.7 Integration with Correlated Preferences

The numbers in Table 4 should be understood as a lower bound for the welfare losses, which would increase when agents are endowed with correlated preferences. Correlation in preferences is evident for certain matching environments, like school choice and marriage. We also assume that preferences are independent from race, an ideal scenario but probably not the case in reality for many matching environments, see Fisman et al. (2008) and Garcia (2008) for evidence of racial preferences for dating and school choice, respectively.

In this section, we modify the assumption that preferences are independent, and see how our results change. The theoretical results of course apply, but what about our simulation results?

How we introduce correlation in preferences is by defining a random status quo in preferences for both men and women. This is, an order over all possible partners. Then, each agent's preferences is identical to the status quo, except for c positions. The expected correlation coefficient between each person preferences and the status quo equals $\rho = 1 - \frac{c}{nr}$. Note that in all of our previous results the expected correlation coefficient is $\rho = 0$.

For example, if $nr = 6$, $c = 2$, and the status quo is 1, 2, 3, 4, 5, 6, a person preferences could be 1, 2, 6, 4, 5, 3. In two places of her preference list there were changes. The changes are chosen randomly, so it could be as well that a person's preferences are the status quo itself.

The results we find is that the fraction of people against integration remains similar around 25% independently of the correlation in preferences, and so does the fraction of agents who keep the same partner. The expected ranking of those against integration increases, as expected, implying that those who oppose integration seem to be less likable by their peers. Also, not surprisingly, the welfare losses of men and women become similar, because it is well-known that with highly correlated preferences the set of stable matchings becomes a singleton. We summarize our observations in Table 3.5

Table 3.5: Statistics for correlated preferences, $n = 100$, $r = 2$.

ρ	%	%	Exp. rank		Welfare loss	
	worse	same	women	men	women	men
0.9	24.61	21.49	105.55	93.92	30.53	31.90
	(0.02)	(0.06)	(59.22)	(50.78)	(17.75)	(24.36)
0.7	26.06	23.13	114.29	89.15	17.22	34.72
	(0.02)	(0.06)	(25.51)	(20.07)	(8.45)	(25.87)
0.5	26.12	25.55	109.67	93.37	10.86	34.81
	(0.02)	(0.06)	(15.81)	(11.85)	(3.46)	(22.99)
0.3	25.89	26.92	104.39	97.46	7.79	34.12
	(0.02)	(0.07)	(6.46)	(4.43)	(1.83)	(21.68)
0.1	25.81	27.18	101.11	100.13	6.14	34.11
	(0.02)	(0.07)	(1.05)	(0.62)	(1.24)	(22.35)

Average over a thousand simulations, $2n = 200$. Columns (2) and (3) refer to the expected number of people who get a worse partner and those who keep the same partner after integration occurs, respectively. All the other columns refer to statistics of those who get a worse partner under integration. Welfare loss measured in difference in ranking of partners. Standard errors in parenthesis.

3.8 Conclusion

When two or more communities integrate to match, there are always some people that become worse off. If the final matching pattern is stable, integration is always approved by a majority of agents, but the fraction of those that oppose social integration does not vanish, even when communities grow large. The welfare losses of those hurt by integration become negligible with respect to the size of the grand society when communities grow large, suggesting that social integration is easier to achieve in sizable communities.

Two interesting questions remain open. The first one is studying the limits of social segregation in many to one matching. The impossibility results carry over, but the question on whether more or less people get hurt by integration, and the exact magnitude of the welfare losses, remains open.

Secondly, there is a recent literature that studies matching in the large using cardinal utilities: e.g. Che and Tercieux (2015) and Lee (2017). Their formulation of preferences makes it easier to introduce correlation, and can provide cardinal measures on the welfare loss of agents that get hurt by integration. Although I conjecture one would obtain similar results using their type of formulation, this remains to be shown formally.

3.9 Appendix: Matlab Code

```
% Matlab Code by JOSUE ORTEGA, University of Glasgow, used in the article "Can everyone benefit from social integration?"
%The stable marriage package is needed, available at
%https://uk.mathworks.com/matlabcentral/fileexchangegale-shapley-stable-marriage-algorithm
%The function march10 computes integrated and segregated matchings, and welfare gains and losses
%The function integration10, below, runs the MonteCarlo simulations to obtain the average over many random instances
function [x,y,SMP,SWP,LOSSM,LOSSW] = march10(n,r);
%n-> Number of males (females) in each race, balanced societies
%r-> Number of races
%x-> Number of people who dislike integration
%y-> 1+Number of people indifferent to integration
%SM(SW)-> Index of males (females) who prefer segregation
%SMP(SWP)-> Average raking of segregated males (females)
%LOSSM(LOSSW)-> Average loss of SM (SW) by integrating
A = zeros(n*r); %Men preferences
B = zeros(n*r); %Women preferences
for i=1:n*r
     A(i,:) = randperm(n*r);%Generates random preferences
     B(i,:) = randperm(n*r);
end
for i=1:r
    aa{i} = transpose(A);
    bb{i} = transpose(B);
```

```
    indicesA{i} = find(aa{i}>i*n |  aa{i}<(i-1)*n+1); %Restricting preferences to their own race
    indicesB{i} = find(bb{i}>i*n |  bb{i}<(i-1)*n+1); %aa{1} contains preferences of men of race 1 over women of race 1
    aa{i}(indicesA{i}) = [];
    bb{i}(indicesB{i}) = [];
    aa{i} = transpose(reshape(aa{i},[n,n*r]));
    bb{i} = transpose(reshape(bb{i},[n,n*r]));
    aa{i}=aa{i}((i-1)*n+1:i*n,:);
    bb{i}=bb{i}((i-1)*n+1:i*n,:);
end
int=galeshapley(n*r,A,B);%Computes integrated marriage
seg=zeros(n,r);
for i=1:r
      seg(:,i)=((i-1)*n)+galeshapley(n,aa{i}-n*(i-1),bb{i}-n*(i-1));%Computes segregated marriage
end
seg=reshape(seg,[r*n,1]);
match=cat(2,int,seg);%Who marries whom in both scenarios,
%Women 1 marries guy in the 1st position, and so on
if match(:,1)==match(:,2)
        x=0;
        y=1
        return;
end%Just in case everybody keeps their same match
indices=find((match(:,1)~=match(:,2)));%Women with different partner
y=1+mean(match(:,1)==match(:,2));%This gives the value of y straight away
```

```
sum=0;
SW=zeros(size(indices));%Women who dislike integration
for i=1:size(indices);
    if find(B(indices(i),:)==int(indices(i)))>find(B(indices(i),:)==seg(indices(i)));
            sum=sum+1;
            SW(i)=indices(i);
    end
end
xx=seg(indices);%Men with different partners
SM=zeros(size(xx));%Men who dislike integration
for i=1:size(xx)
    if find(A(xx(i),:)==find(int==xx(i)))>find(A(xx(i),:)==find(seg==xx(i)));
        SM(i)=xx(i);
        sum=sum+1;
    end
end
SW(SW==0) = []; SM(SM==0) = [];
match;
x=sum/(2*n*r);
SMP=zeros(size(SM,1),1);SWP=zeros(size(SW,1),1);%Average ranking of those who prefer segregation
LOSSM=zeros(size(SM,1),1);LOSSW=zeros(size(SW,1),1);%Average loss of those who prefer segregation
for i=1:size(SM,1)
    SMP(i)=mean(find(B.'==SM(i))-(0:(r*n):((r*n)^2)-1).')    ;
    j=SM(i);
```

```
        LOSSM(i)=find(A(j,:)==find(int==j))-find(A(j,:)==find(seg==j
        ));
end
for i=1:size(SW,1)
        SWP(i)=mean(find(B.'==SW(i))-(0:(r*n):((r*n)^2)-1).')        ;
        j=SW(i);
        LOSSW(i)=find(B(j,:)==int(j))-find(B(j,:)==seg(j));
end
SMP=mean(SMP); SWP=mean(SWP);
LOSSM=mean(LOSSM); LOSSW=mean(LOSSW);
end

----

function [e] = integration10(t,n,r)
v=zeros(t,6);
parfor i=1:t
        [x y SMP SWP LOSSM LOSSW]=march10(n,r);
        v(i,:)=[x y SMP SWP LOSSM LOSSW];
end
e=sum(v,1)/t;
end
```

Chapter 4

The Strength of the Weak Ties: Online Integration via Online Dating

Abstract

We used to marry people to which we are somehow connected to: friends of friends, coworkers, or colleagues from school. Since we are much more connected to people that are like us, this implies that we were likely to marry someone that share our own characteristics, in particular our race.

The irruption of online dating platforms have changed this pattern: other online daters are very likely to be complete stranger to us. Given that one third of modern marriages start online, we investigate theoretically the effects of those previously absent ties in the diversity of modern societies.

We find that when a society benefits from previously absent ties, that we interpret as online dating contacts, social integration occurs rapidly, even if the number of partners met online is arbitrarily small. Our findings are consistent with the sharp increase in interracial marriages in the U.S. after the popularization of online dating platforms.

KEYWORDS: Online dating, Tinder, interracial marriages, segregation.

JEL CLASSIFICATION: J12 (marriage), D85 (networks), C78 (matching).

In the most cited article on social networks,[1] Granovetter (1973) argued that the most important connections we have may not be our close friends but our acquaintances: people that are not very close to us, either physically or emotionally, help us to relate to groups that otherwise we would not be linked to. It is from acquaintances, for example, that we are more likely to hear about job offers. Those weak ties serve as bridges between our group of close friends and other clustered groups, hence allowing us to connect to the global community in several ways.[2]

Interestingly, the process of how we meet our romantic partners in at least the last hundred years closely resembles this phenomenon. We would probably not marry our best friends, but we are likely to end up marrying a friend of a friend or someone we coincided with in the past. Rosenfeld and Thomas (2012) show how Americans meet their partners in the last decades, listed by importance: through mutual friends, in bars, at work, in educational institutions, at church, through their families, or because they became neighbors. This is nothing but the weak ties phenomenon in action.[3]

But in the last two decades, the way in which we meet our romantic partners has changed dramatically. Online dating has become the second most popular way to meet a spouse for most U.S. residents, as can be observed in Figure 4.1, taken from Rosenfeld and Thomas (2012).[4]

The aforementioned article explains: *" the Internet increasingly allows Americans to meet and form relationships with perfect strangers, that is, people with whom they had no previous social tie"*. We suspect that this is happening not only to Americans, but is a consistent global phenomenon. If the reader needs another example, Figure 4.2 shows one of the author's Facebook friends graph. The yellow triangles reveal previous relationships that started in offline venues. It can easily be seen that those ex-partners had several mutual friends with the author; in the corresponding graph, their edge had

[1] "What are the most-cited publications in the social sciences according to Google?", *LSE Blog*, 12/05/2016.

[2] Strong ties are also valuable, and in the case of job search, they may outweigh weak ones (Kramarz and Skans, 2014; Gee et al., 2017).

[3] Backstrom and Kleinberg (2014) reinforce the previous point: given the social network of a Facebook user who is in a romantic relationship, the node which has the highest chances to be his romantic partner is, perhaps surprisingly, not the one who has more friends in common with him.

[4] We thank Michael Rosenfeld for allowing us to use his figure.

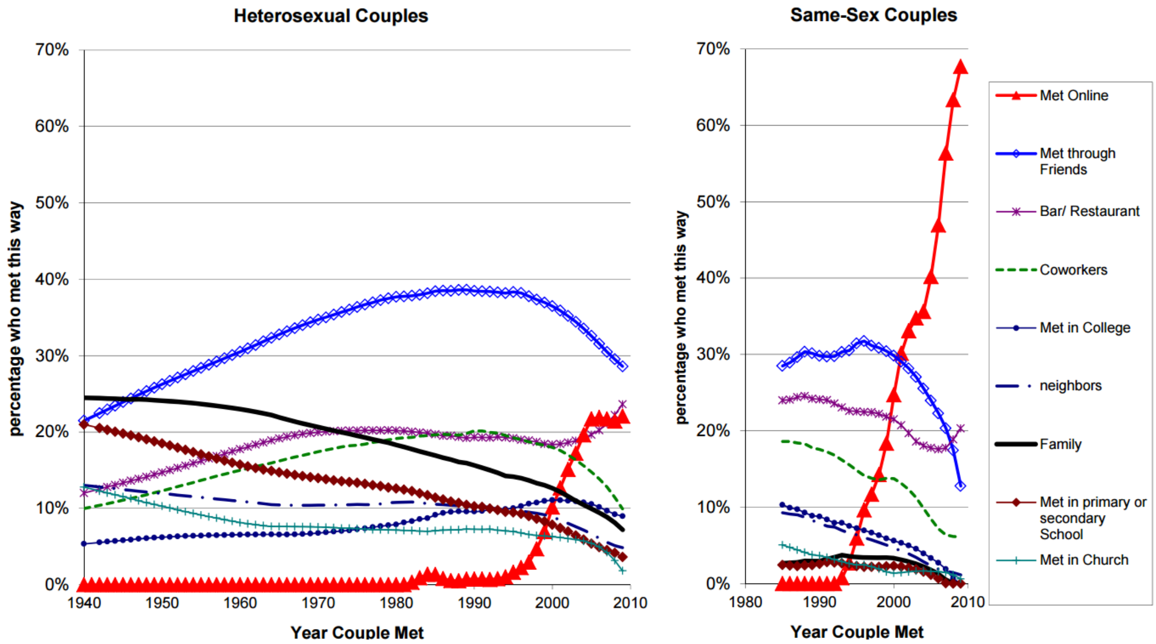


Figure 4.1: How we met our partners in the last decades.

a high embeddedness in graph theoretical jargon. In contrast, nodes appearing as red stars represent partners he met through online dating. It is easily seen that those have no contacts in common with him, and thus it is likely that, if it would not have been for online dating, those persons would have never interacted with him.

Because one-third of modern marriages start online (Cacioppo et al., 2013), and up to 70% of homosexual relationships, the way we match online with potential partners shapes the demography of our communities, in particular its racial diversity. Meeting complete strangers online can intuitively increase the number of interracial marriages in our societies, which is remarkably low: only 6.3% and 9% of the total number of marriages are interracial in the U.S. and the U.K., respectively.[5] The low rates of interracial marriage are expected, given that in the U.S. it was illegal in 16 states 50 years ago, until the Supreme Court ruled out anti-miscegenation laws in the famous *Loving vs. Virginia* case (Arrow, 1998; Fryer, 2007).[6]

The research question that motivates us is to understand how many more interracial marriages, if any, will occur after online dating becomes available in a society, and what

[5] "Interracial marriage: Who is marrying out", *Pew Research Center*, 12/6/2015; and "What does the 2011 census tell us about inter-ethnic relationships?", *UK Office for National Statistics*, 3/7/2014.

[6] Interracial marriage in the U.S. has increased considerably from 1970, but it is still rare (Kalmijn, 1998; Fryer, 2007; Furtado, 2015). Interracial marriage occurs far less frequently than interfaith marriages (Qian, 1997).

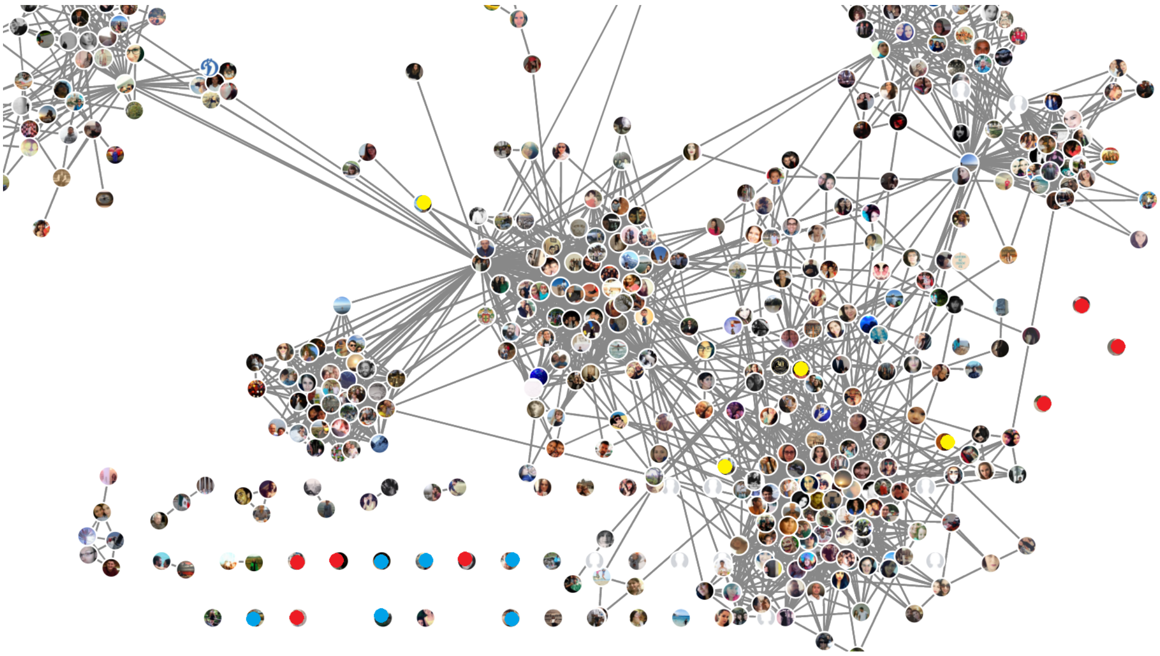


Figure 4.2: How one of us met his partners in the last decade.

drives this increase. In addition, we are also interested in whether marriages created online are any different from those that existed before.

Understanding the evolution of interracial marriage is a relevant problem, for intermarriage is widely considered a measure of social distance in our societies (Wong, 2003; Furtado, 2015), just like residential or school segregation. In the words of Fryer (2007), "*social intimacy is a way of measuring whether or not a majority group views a minority group on equal footing*".

Moreover, the number of interracial marriages in a society has important economic implications. Interracial marriage is known to affect the employment status[7] (Meng and Gregory, 2005; Goel and Lang, 2009; Furtado and Theodoropoulos, 2010) and the social identity (Bisin and Verdier, 2000; Duncan and Trejo, 2011) of those engaging into it, as well as the education levels of their offspring (Furtado, 2012).

4.0.1 Overview of Results

This article builds a theoretical framework to explain how many more interracial marriages occur after the popularization of online dating. Our model builds an intuitive combination

[7] Intermarriage affects the probability of finding a job, but surprisingly, not the average wage earned (Kantarevic, 2004).

of non-transferable utility[8] matching *à la* Gale and Shapley (1962) in random graphs, first studied by Erdős and Rényi (1959) and Gilbert (1959). Our theoretical framework is easy to grasp and has an intuitive graphical visualization.

We take several disjoint Gale-Shapley marriage problems, with agents randomly located on the unit square. Agents want to marry the person who is closest to them, but they can only marry people who they know, i.e. to whom they are connected. As in real life, agents are highly connected with agents of their own race, but poorly so with people from other races. Also, as it seems to be the case in real life,[9] we assume that the marriages that occur in our society are are those predicted by game-theoretic stability.

Then, to model online dating, we introduce absent ties, by slightly increasing the probability that any two agents of different races are connected, and compare how many more interracial marriages we observe now in the expanded society. We also keep an eye on the characteristics of those newly formed marriages. In particular, we focus on the average distance between partners before and after the introduction of online dating. Assuming that marriages between partners who are closer to each other are stronger, given that they are less susceptible to break up when new agents arrive, we can also measure whether marriages created after online dating are more or less likely to divorce.

The graphical interpretation of our model is similar to the one used by the mathematics literature in matching of Poisson point processes (Holroyd et al., 2009; Holroyd, 2011; Amir et al., 2016), from which we borrow useful technical results (see the proof of Proposition 1). Our model also roughly resembles the graphical model of residential segregation of Schelling (1969, 1971, 1972). However, unlike the famous Schelling model, our model predicts that nearly complete racial integration occurs when online dating emerges, even if the number of partners that individuals meet from newly formed ties is small.

We contrast our model with empirical data from U.S. and find that, as predicted, the number of interracial marriages substantially increases after the popularization of online

[8]Most of the literature studying marriage with matching models uses transferable utility, following the seminal work of Becker (1973, 1974, 1981). A review of that literature appears in Browning et al. (2014). Although our model departs substantially from this literature, we point out similarities with particular papers in this field when we detail the model in Section 4.1.

[9]See Banerjee et al. (2013), also Hitsch et al. (2010) for the case of online dates.

dating. We discuss how the observed sharp increase cannot be purely due to changes in the composition of the U.S. population.

Our result contributes to clarify the relationship between social networks and interracial marriage. In a related paper, Furtado and Theodoropoulos (2010) find that immigrants who intermarry have a higher chance of finding employment than those who marry within their own ethnic group. Interestingly, most of this effect is due to the valuable social networks that immigrants gain by marrying a local (and not because an easier chance to get a visa). In their model, intermarriage creates social networks. In ours, social networks generate intermarriage, by creating previously absent ties within races via online dating. This increase is not due to changes in agents' preferences.

Our model also predicts that marriages created in a society with online dating should be stronger, another feature that has been documented empirically.

4.0.2 Structure of the Article

We present our model in Section 1, and discuss the welfare measures we consider in Section 2. Sections 3 and 4 analyze how our welfare measures change when societies become more connected using theory and computations, respectively.

Section 5 contrasts our model predictions with observed demographic trends from the U.S. Section 6 concludes and details on other applications of our theoretical framework, which is a general model of matching under network constraints. Those applications include social integration after student participate in exchange programs and collaboration between interdisciplinary researchers, among others.

4.1 Marriages in a Network

4.1.1 Agents

There are r races or communities, each with n heterosexual agents. Each race is assigned a particular color. Each agent i is identified by a pair of coordinates $(x_i, y_i) \in [0, 1]^2$, that can be understood as measures of agents' social and political opinions,[10] to which we refer as *personality traits*. Both coordinates are drawn uniformly and independently for all agents.[11]

Each agent is either male or female. Female agents are plotted as stars and males as dots. Each race is balanced in its ratio between men and women.

4.1.2 Edges

Agents are connected to others of their own race with probability p: these edges are represented as solid lines and occur independently of each other. Agents are connected to others of different race with probability q: these interracial edges appear as dotted lines and are also independent. We present an example in Figure 4.3.

Our model is a generalization of the random graph model (Erdős and Rényi, 1959; Gilbert, 1959; for a textbook reference, see Bollobás, 2001), in which there are r random graphs with parameter p and n nodes, interacting across graphs with probability q. The intuition in our model is that two agents are connected if they know each other. In expectation, each agent is connected to $n(r-1)q + (n-1)p$ persons.

A *society* S is a realization from a generalized random graph model, defined by a four-tuple (n, r, p, q). A society S has a corresponding bipartite graph $S = (M, W; E)$ where M and W are the set of men and women, respectively, and E is the set of edges. We use the notation $E(i, j) = 1$ if there is an edge between agents i and j, and 0 otherwise. We denote such edge by either (ij) or (ji).

[10]For a real-life representation using a 2-dimensional plane see `www.politicalcompass.org`. A similar interpretation appears in Chiappori et al. (2012) and in Chiappori et al. (2016), in which the traits include age, education, race, religion, weight or height.

[11]Another way to understand how agents' personality traits are drawn is to consider a Poisson point process (PPP) defined on the unit square with intensity $\lambda = n$. In a PPP the number of agents is not fixed but drawn from a Poisson distribution, although there are n in expectation. In our case, the number of agents is fixed throughout.

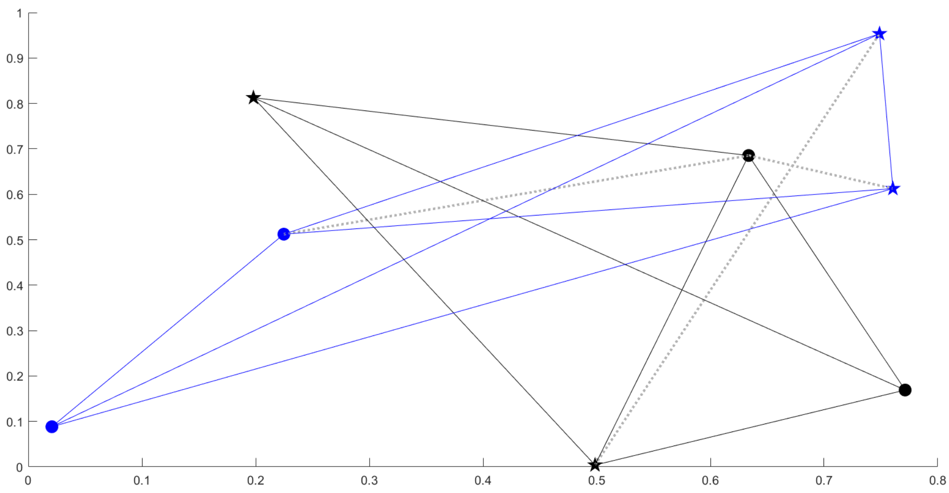


Figure 4.3: 4 agents, 2 races, linked with $p = 1$ and $q = 0.2$.

4.1.3 Agents' Preferences

All agents are heterosexual and prefer marrying anyone of different gender instead of remaining alone.[12] We denote by P_i the set of potential partners for i. The preferences of agent i are given by a function $\delta_i : P_i \to \mathbb{R}_+$ that has a distance interpretation.[13] An agent i prefers agent j over agent k if $\delta_i(i, j) \leq \delta_i(i, k)$. The intuition is that agents like potential partners that are close to them.

The function δ_i could be arbitrary, or could be the same for agents of the same race. It could also be weighted to account for strong intraracial preferences that are often observed in reality (Wong, 2003; Fisman et al., 2008; Hitsch et al., 2010; Rudder, 2014; Potarca and Mills, 2015; McGrath et al., 2016).[14] Inter or intraracial preferences can easily be incorporated into the model, as in equation (4.3) below, but for ease of exposition and

[12]Heterosexuality is assumed for convenience, because it is well-known that in one-sided matching there may be no stable pairings.

[13]Although δ can be generalized to include functions that violate the symmetry ($\delta(x, y) \neq \delta(y, x)$) and identity ($\delta(x, x) = 0$) characteristic properties of mathematical distances.

[14]It is not clear whether the declared intraracial preferences show an intrinsic intraracial predilection or capture external biases, which, when removed, leave the partner indifferent to match across races. Evidence supporting the latter hypothesis includes: Fryer (2007) documents that White and Black U.S. veterans have had higher intermarry rates after serving with mixed communities. Fisman et al. (2008) finds that people do not find partners of their own race more attractive. Rudder (2009) shows that online daters have a roughly equal user compatibility. Lewis (2013) finds that users are more willing to engage on interracial dating if they interacted earlier with a dater from another race.

mathematical convenience (see Proposition 5) we only consider two intuitive and simple functions that do not incorporate homophily.

The first one is the Euclidean distance for all agents, so that for any agent i and every potential partner $j \neq i$,

$$\delta^E(i,j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{4.1}$$

and $\delta^E(i,i) = \sqrt{2}\ \forall i \in M \cup W$. Euclidean preferences are intuitive and have been widely used in social science (Bogomolnaia and Laslier, 2007). The indifference curves associated with Euclidean preferences can be described by balls around each point.

The second preferences we consider are such that every agent prefers a partner close to them in personality trait x, but they all agree on which is the best in personality trait y. The intuition is that the y-coordinate indicates something like attractiveness, wealth, or any other attribute usually considered desirable by all partners. We call these preferences *assortative*,[15] so that for any agent i and every potential partner $j \neq i$,

$$\delta^A(i,j) = |x_i - x_j| + (1 - y_j) \tag{4.2}$$

and $\delta^E(i,i) = 2\ \forall i \in M \cup W$. The indifferences curves of assortative preferences are depicted in Figure 4.4.

Both Euclidean and assortative preferences can be generalized by weighting them by specific constants β_{ij}, such that

$$\delta_i' = \beta_{ij}\ \delta(i,j) \tag{4.3}$$

The constant β_{ij} captures intraracial specific preferences whenever it is constant for all pairs i, j who belong to the same race. Similarly, it can capture specific reluctance to match with agents from specific races whenever above 1.

A society in which all agents have either all Euclidean or all assortative preferences will be called Euclidean or assortative, respectively. We focus on these two cases. In both cases

[15]If we keep the x-axis fixed, so that agents only care about the y-axis, we get full assortative mating as a particular case.

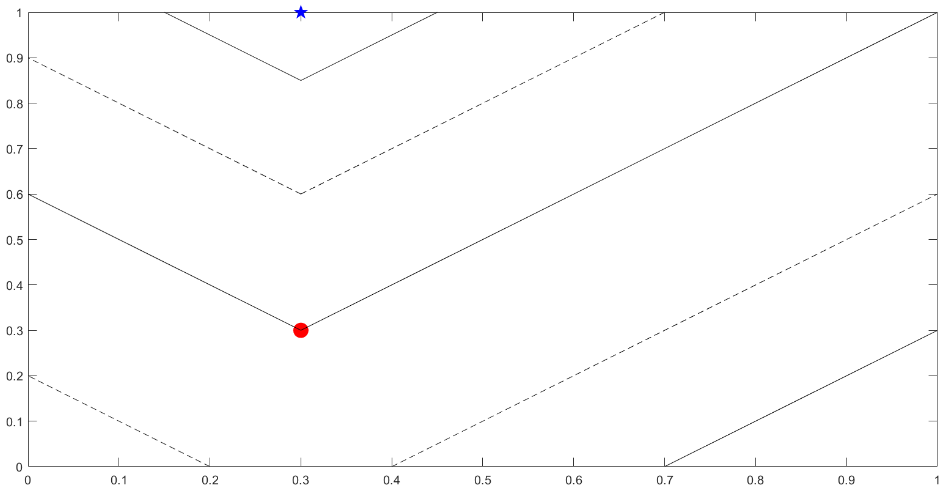


Figure 4.4: Indifference curves for assortative preferences (the blue star is the best partner for the red dot).

agents' preferences are strict because of our assumption on the distribution of personality traits.

4.1.4 Marriages

Agents can only marry potential partners that they know: i.e. if between them and their partner there exists a path of length at most k between them in the society graph.[16] We consider two types of marriages:

1. Direct marriages: $k = 1$. Agents can only marry if they know each other.
2. Long marriages: $k = 2$. Agents can only marry if they know each other or if they have a mutual friend in common.

To formalize the previous marriage notion, let $\rho_k(i, j) = 1$ if there is a path of at most length k between i and j, with the convention $\rho_1(i, i) = 1$. A marriage $\mu : M \cup W \to M \cup W$

[16] A path from node i to t is a set of edges $(ij), (jk), \ldots, (st)$. The length of the path is the number of such pairs.

of length k is a function that satisfies

$$\forall m \in M \qquad \mu(m) \in W \cup \{m\} \tag{4.4}$$

$$\forall w \in W \qquad \mu(w) \in M \cup \{w\} \tag{4.5}$$

$$\forall i \in M \cup W \qquad \mu(\mu(i)) = i \tag{4.6}$$

$$\forall i \in M \cup W \qquad \mu(i) = j \text{ only if } \rho_k(i, j) = 1 \tag{4.7}$$

We use the convention that agents that remain unmarried are matched to themselves. We use $M^* = \{m \in M \mid \mu(m) \in W\}$ to denote the set of all married men.

Because realized romantic pairings are close to those predicted by stability (Hitsch et al., 2010; Banerjee et al., 2013), we assume that marriages that occur in each society are *stable*. A marriage μ is k-stable if there is no man-woman pair (m, w) who are not married to each other such that

$$\rho_k(m, w) = 1 \tag{4.8}$$

$$\delta(m, w) < \delta(m, \mu(m)) \tag{4.9}$$

$$\delta(w, m) < \delta(w, \mu(w)) \tag{4.10}$$

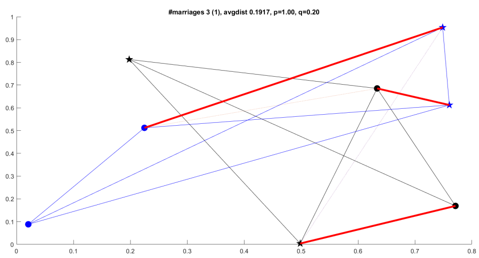
Condition (4.8) is the only non-standard one that ensures that a pair of agents cannot block a direct marriage if they are not connected in the corresponding graph, even if they prefer each other to their respective partner. Given our assumptions regarding agents' preferences,

Proposition 5. *For any positive integer k, every Euclidean or assortative society has a unique k-stable marriage.*

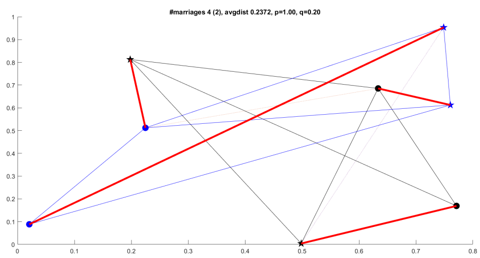
Proof. For the Euclidean society, an easy algorithm computes the unique k-stable marriage. Let every person point to their preferred partner to whom they are connected to by a path of length at most k. In case two people point to each other, marry them and remove them from the graph. Let everybody point to their new preferred partner to which they are connected to among those still left. Again, marry those that choose each other, and repeat the procedure until no mutual pointing occurs. The procedure ends after at most rn iterations. A similar algorithm has been suggested by Holroyd et al. (2009, Proposition 9) for 1-stable matchings.[17]

For the assortative society, assume by contradiction that there are two k-stable matchings μ and μ' such that for two men m_1 and m_2, and two women w_1 and w_2, $\mu(w_1) = w_1$ and $\mu(w_2) = w_2$, but $\mu'(w_1) = w_2$ and $\mu'(w_2) = w_1$.[18] The fact that both marriages are k-stable implies, without loss of generality, that for $i, j \in \{1, 2\}$ and $i \neq j$, $\delta(m_i, w_i) - \delta(m_i, w_j) < 0$ and $\delta(w_i, m_j) - \delta(w_i, m_i) < 0$. Adding up those four inequalities, one obtains $0 < 0$, a contradiction. ■ □

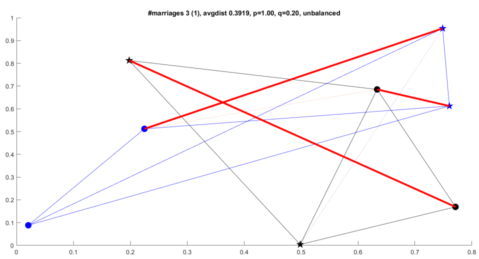
Figure 4.5 shows the direct and long stable marriages for the Euclidean and assortative societies depicted in Figure 4.3.



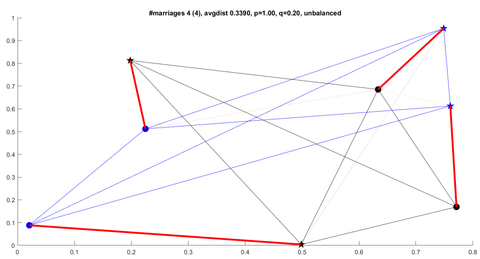
(a) Direct marriage, Euclidean pref.



(b) Long marriage, Euclidean pref.



(c) Direct marriage, assortative pref.



(d) Long marriage, assortative pref.

Figure 4.5: Direct and long stable marriages for the assortative society in Fig. 4.3.

[17] Holroyd et al. (2009) require two additional properties: non-equidistance and no descending chains. The first one is equivalent to strict preferences, the second one is trivially satisfied. In their algorithm, agents point to the closest agent, independently if they are connected to them.

[18] It could be the case that in the two matchings there are no four people who change partner, but that the swap involves more agents. The argument readily generalizes.

4.1.5 Online Dating on Networks

We model online dating in a society S by increasing the number of interracial edges. Given the bipartite graph $S = (M, W; E)$, we create new interracial edges between every pair that is disconnected with a probability ϵ.[19,20]

S_ϵ denotes a society that results after online dating has occurred in society S. S_ϵ has exactly the same nodes as S, and all its edges, but possibly more. We say that the society S_ϵ is an *expansion* of the society S.

4.2 Welfare Indicators

We want to understand how the welfare of a society changes after online dating becomes available, i.e. after it becomes more interracially connected. There are three clear indicators of agents' welfare in a given society, namely its

1. **Size**, i.e. the total number of marriages in a society. Formally,

$$sz(S) = |M^*| \tag{4.11}$$

2. **Diversity**, i.e. how close is the society to having the marriages produced in a completely connected and colorblind society. We normalize this measure so that 0 indicates a society with no interracial marriages, and 1 indicates a society in which $\frac{r-1}{r}$ of the marriages are interracial. Note that it may well be the case that diversity is above 1.

Let $\mathcal{R}$ be a function that maps each agent to their race. Then

$$dv(S) = \frac{|\{m \in M^* \mid \mathcal{R}(m) \neq \mathcal{R}(\mu(m))\}|}{sz(S)} \cdot \frac{r}{r-1} \tag{4.12}$$

[19] Online dating is likely to also increase the number of edges inside each race, but since we assume that each race is already highly connected, these new edges play no role in the results of the model. We perform robustness checks in Appendix 4.6.2, increasing both p and q but keeping its ratio fixed.

[20] We could assume that particular persons are more likely than others to use online dating, e.g. younger people. Data shows that, from 2013 to 2015, the percentage of people who use online dating has increased for people of all ages. See: "5 facts about online dating", *Pew Research Center*, 29/2/2016. While this occurs at a different rate, to obtain our main result we only need an infinitesimal increase in the probability of interconnection for each agent.

3. **Strength**, defined as $\sqrt{2}$ minus the average Euclidean distance between each married couple, denoted as $ds(S)$. A marriage with a small distance is better than one with a large one because is less susceptible to break up when random agents appear on the unit square, and the new outcome is to be k-stable too. The previous observation holds for assortative societies too.

The above indicator is divided by $\sqrt{2}$ (or the maximal distance possible) to normalize it between 0 and 1.

Formally,

$$ds(S) = \frac{\sum_{m \in M^*} \delta^E(m, \mu(m))}{sz(S)} \tag{4.13}$$

$$st(S) = \frac{\sqrt{2} - ds(S)}{\sqrt{2}} \tag{4.14}$$

If every married agent gets paired with her perfect match, then $st(S) = 1$.

4.3 Edge Monotonicity of Welfare Indicators

Given a society S, the first question is whether the welfare indicators of a society grow when its number of interracial edges grow, i.e. when online dating becomes available. We refer to this property as *edge monotonicity*.[21]

Definition 2. A welfare indicator w is edge monotonic if, for any society S, and any of its extensions S_ϵ, we have

$$w(S_\epsilon) \geq w(S) \tag{4.15}$$

If a welfare indicator is edge monotonic it means that a society unambiguously becomes better off from becoming more interracially connected. Unfortunately,

Proposition 6. *Diversity, strength, and size are all not edge monotonic.*

[21] Edge monotonicity is different from node monotonicity, in which one node, with all its corresponding edges, is added to the matching problem. It is well-known that when a new man joins a stable matching problem, every woman weakly improves, while every man becomes weakly worse off (Theorems 5 in Kelso and Crawford, 1982, 2.25 and 2.26 in Roth and Sotomayor, 1992, and 1 and 2 in Crawford, 1991).

Before proving Proposition 6, let us build some intuition about it. It may be surprising that the number of interracial marriages can decrease when more interracial edges are formed. The intuition behind it is that an interracial edge may create one interracial marriage at the cost of destroying two existing ones, and the left-alone partners may now marry partners of their own race.

An interracial edge may similarly increase the average distance between couples if it provides a link between very desirable partners, i.e. those in the center for the case of Euclidean preferences. Those desirable partners drop their current spouses, which now have to match with partners that have been dropped too. That their new partner has been previously dropped implies it is far from the center, and thus the marriage between dropped partners may marry people in the corners of the unit square.

Finally, size may be reduced if the new interracial edge links people who were already highly connected in the society, making them leave partners who are poorly so. The left-alone partners may now become unable to find a partner.

We present now a formal proof for Euclidean societies with direct marriages.

Proof. To show that size is not edge monotonic, consider the society in Figure 4.3 and its direct stable matching in Figure 4.5a. Remove all interracial edges: it is immediate that in the unique stable matching there are 4 couples now, one more than when interracial edges are present.

For the case of strength, consider a simple society in which all nodes share the same y-coordinate, as the one depicted in Figure 4.6. There are two intraracial marriages and the average Euclidean distance is 3.5. When we add the interracial edge between the two central nodes, the closest nodes marry and the two far away nodes marry too. The average Euclidean distance in the expanded society increases to 4.5, hence reducing its strength.

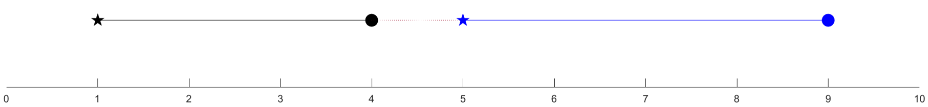
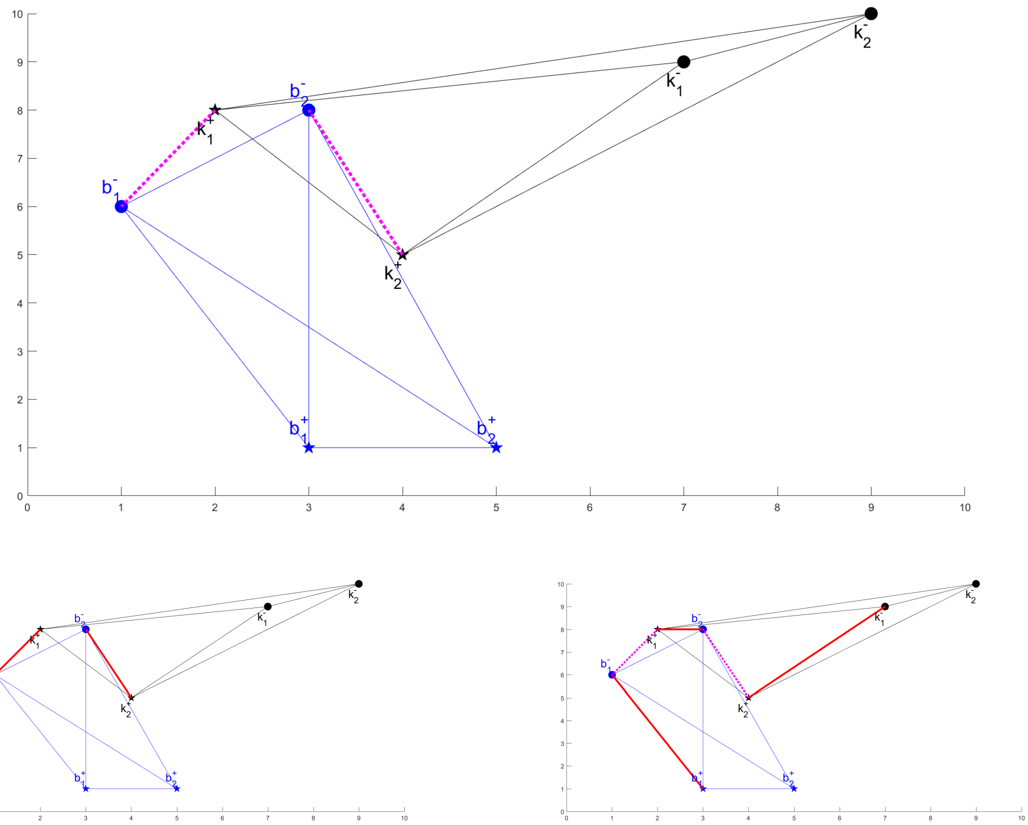


Figure 4.6: Strength is not edge monotonic.

To show that diversity is not edge monotonic, consider Figure 4.7. There are two men and two women of each of two races a and b. Each gender is represented with the superscript $^+$ or $^-$.



(a) $dv(S) = 2$

(b) $dv(S_\epsilon) = 1/3$

Figure 4.7: Diversity is not edge monotonic.

Stability requires that $\mu(b_1^-) = a_1^+$ and $\mu(b_2^+) = a_2^-$, and everyone else is unmarried. However, when we add the edge $(a_1^+ b_2^-)$, the married couples become $\mu(b_1^-) = b_1^+$, $\mu(a_2^+) = a_1^-$, and $\mu(a_1^+) = b_2^-$. In this extended society, there is just one interracial marriage, out of a total of three, when before we had two out of two. Therefore diversity reduces after adding the edge $(a_1^+ b_2^-)$. ■ □

The failure of edge monotonicity by our three welfare measures makes evident that to evaluate welfare changes in societies, we need to understand how welfare varies on an average society after introducing new interracial edges. We develop this comparison in the next Section.

A final comment on edge monotonicity. The fact that the size of a society is not edge monotonic, as shown in Proposition 6, implies that adding interracial edges may not lead to a Pareto improvement of the society, i.e. some agents may become worse off after the society becomes more connected. Nevertheless, the fraction that becomes worse off after adding an extra edge never more than one-half of the society. Ortega (2017) discusses this phenomenon in detail and the associated welfare losses of those hurt by integration.

4.4 Average Welfare Indicators

In the last Section we found that our three welfare indicators may increase or decrease after adding interracial edges. Therefore, we need to analyze what happens in an average case: i.e. what is expected to happen to the diversity, strength and size of a society when agents become more connected.

There are two ways to answer this question. The first one is to provide analytical expressions for the expected welfare indicators as a function of the number of interracial edges. However, providing analytical solutions is incredibly complicated, if not impossible. Already solving the expected average distance in a society with just one race containing only one man and one woman requires a complicated computation (which equals to $\frac{2+\sqrt{2}+5\ln(\sqrt{2}+1)}{15} \approx 0.52$).[22]

The second way to approach the problem is to simulate several random societies and observe how their welfare change when they become more connected. This is the route we follow. We create ten thousand random societies, and increase the expected number of interracial edges by increasing the parameter q. In the following subsections, we describe the changes of our welfare indicators for different values of q.

In all cases we fix $n = 50$ and $p = 1$.[23] We consider the following four scenarios:

1. Two races and direct marriages, appears in blue with diamond markers ♦.

2. Five races and direct marriages, appears in grey with square markers ■.

[22] The detailed computation appears in "Distance between two random points in a square", *Mind your Decisions*, 3/6/2016.

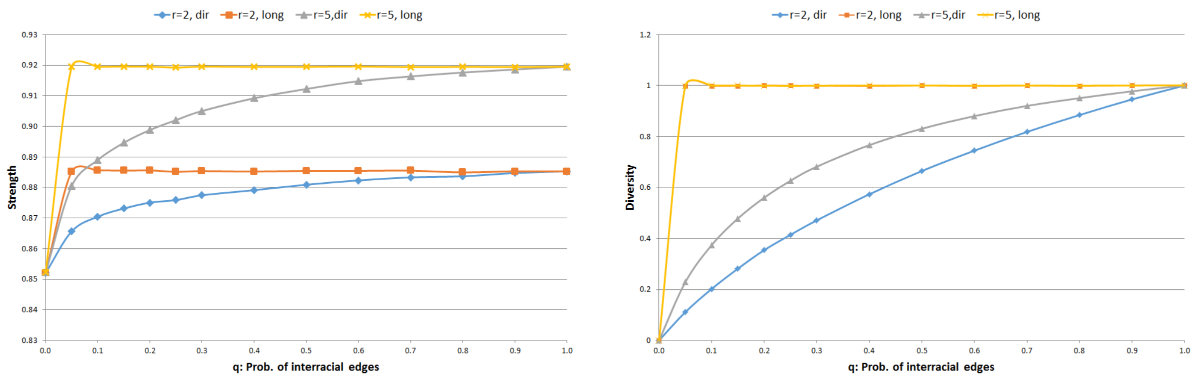
[23] We limit ourselves to $n = 50$ and ten thousand replications because of computational limitations, even though we used the high performance computing facilities at the University of Glasgow.

3. Two races and long marriages, appears in orange with triangle markers ▲.

4. Five races and long marriages, appears in yellow with cross markers ×.

4.4.1 Diversity

In the case of long marriages, even the smallest increase in the probability of interracial connections (in this case of 0.05) achieves perfect social integration with either two or five races: diversity is exactly one. For the cases with direct marriages, the increase in diversity is slower but still fast: an increase of q from 0 to 0.1 increases diversity to 0.19 for $r = 2$, and from 0 to 0.37 with $r = 5$.[24]



(a) Euclidean society. (b) Assortative society

Figure 4.8: Average diversity (y-axis) of a random society for different values of q. The yellow and orange curves are indistinguishable in this plot because they are identical. Exact values and standard errors (which are in the order of 1.0e-04) provided in Appendix 4.6.2.

Figure 4.8 summarizes our main result, namely

Result 1. *Diversity is fully achieved with long marriages, even if the increase in interracial connections is arbitrarily small.*

With direct marriages, diversity is achieved partially but still substantially, so that an increase in q always yields an increase in diversity of a larger size, i.e. diversity is a concave function of q.

[24]Empirical evidence suggests that q is close to zero. Echenique and Fryer (2007) find that the typical American public school student has 0.7 friends of another race. It is also a sensible assumption that p is large, given the clear residential segregation patterns in the U.S. (Cutler et al., 1999) and that around 90% of people who attend religious services do so with others from their same race (Fryer, 2007).

The intuition behind full diversity for the case of long marriages, in which agents are allowed to marry any person with whom they have a friend in common, is that once an agent obtains just one edge to any other race, he gains $\frac{n}{2}$ potential partners, i.e. just one edge to a person of different race gives access to that person's complete race.

The reader may think that the full diversity result heavily depends on each race being fully connected, i.e. $p = 1$. This is not the case. Full diversity is also for many other values of p, as we present in Appendix 4.6.2. When same-race agents are less interconnected within themselves, agents gain fewer connections once an interracial edge is created, but those fewer connections are relatively more valuable, because the agent had himself less potential partners before the creation of new interracial edges.

Result 1 implies that, assuming long marriages are formed, very few interracial links can lead a society to almost complete racial integration, and leads to very optimistic views on the role that dating platforms can play in the reduction of racial segregation in our society. Our result is in sharp contrast to the one of Schelling (1969, 1971) in its well-known models of residential segregation, in which a society always gets completely segregated.

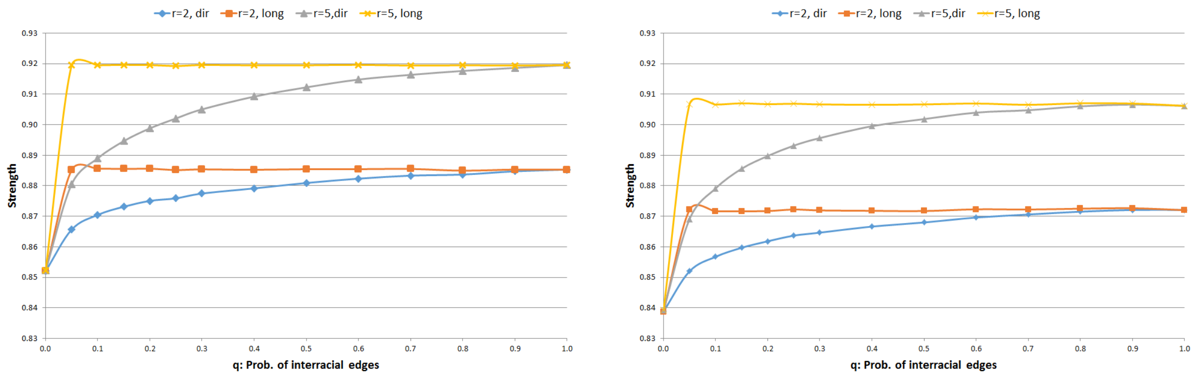
What is the ingredient in our model that allows us to get predictions so different from the ones derived from the celebrated Schelling's model? It is not specific preferences, which can be accommodated in our model so that an agent will slightly prefer her own race as in Schelling's (for example as a tie-breaker, an event of probability zero). It is not the parameters of the society, which in both cases can be easily generalized (think for example of nodes represented in $\mathbb{R}^n$ instead of $\mathbb{R}^2$). The whole difference, it would appear, is the formation of romantic links. In our model, agents' utility only depends on the location of agent they marry, instead of depending on all their neighbors.

We pose this finding as the first testable hypothesis of our model

Hypothesis 1. The number of interracial marriages should increase after the popularization of online dating.

4.4.2 Strength

A second observation, less pronounced that the increase in diversity, is that the strength of the society goes up when increasing q. For an illustration, see Figure 4.9, which considers the same four cases as before in both Euclidean and assortative societies.



(a) Euclidean society. (b) Assortative society

Figure 4.9: Average strength (y-axis) of a random society for different values of q.
Exact values and standard errors (which are in the order of 1.0e-04) provided in Appendix 4.6.2.

It is clear that, for all combinations of parameters (see Appendix 4.6.2 for further robustness checks), there is a consistent trend downwards in the average distance of partners after adding new interracial edges, and thus a consistent increase in strength of the societies. We present this observation as our second result.

Result 2. *Strength increases after the number of interracial edges increases. The increase is faster whenever the society has more races, and converges to a higher level with long marriages.*

Assuming that marriages with a higher average distance have a higher chance to end up divorcing, because they are more susceptible to break up when new nodes are added to the society graph, we can reformulate our result as our second hypothesis.

Hypothesis 2. Marriages created in societies with online dating should have a lower divorce rate.

Finally, our last welfare indicator, size, keeps constant for most of our simulations, so we do not discuss it further. The detailed data behind Figures 4.8 and 4.9, with its standard errors, appear in Appendix 4.6.2.

Our analysis of the expected changes in welfare gives us with two testable hypotheses. In the next Section, we contrast them against data on of interracial marriage in the U.S, and the quality of the marriages created through online dating.

4.5 Hypotheses and Data

4.5.1 Hypothesis 1: More Interracial Marriages

What does the data reveal? Is our model consistent with observed demographic trends? Figure 4.10 presents the evolution of interracial marriages among newlyweds in the U.S. from 1967 to 2015, based on the 2008-2015 American Community Survey and 1980, 1990 and 2000 decennial censuses (IPUMS). In this Figure, interracial marriages include those between white, black, Hispanic, Asian, American Indian or multiracial persons.[25,26]

In the data, we observe that the number of interracial marriages has consistently increased in the last 50 years, as it has been documented by several other authors (Kalmijn, 1998; Fryer, 2007; Furtado, 2015). However, it is intriguing that, shortly after the introduction of the first dating websites in 1995, like Match.com, the percentage of new marriages created by interracial couples increased rapidly. The increase becomes steeper around 2004, when online dating became more popular: it is then when well-known platforms such like OKCupid emerged. During the 2000's decade, the percentage of new marriages that are interracial changed from 10.68% to 15.54%, a huge increase of nearly 5%.

After the 2009 increase, the proportion of new interracial marriage jumps again in 2014 to 17.24%, remaining above 17% in 2015 too. Again, it is interesting that this increase occurs shortly after the creation of Tinder, considered the most popular online dating app.

[25] We are grateful to Gretchen Livingston from the Pew Research Center for providing us with the data. Data prior to 1980 are estimates. The methodology on how the data was collected is described in Livingston and Brown (2017).

[26] Although Hispanic is not a race, Hispanics do not associate with other races. In the 2010 U.S. census, over 19 million of Latinos selected to be of "some other race". See "For many Latinos, racial identity is more culture than color", *New York Times*, 13/1/2012.

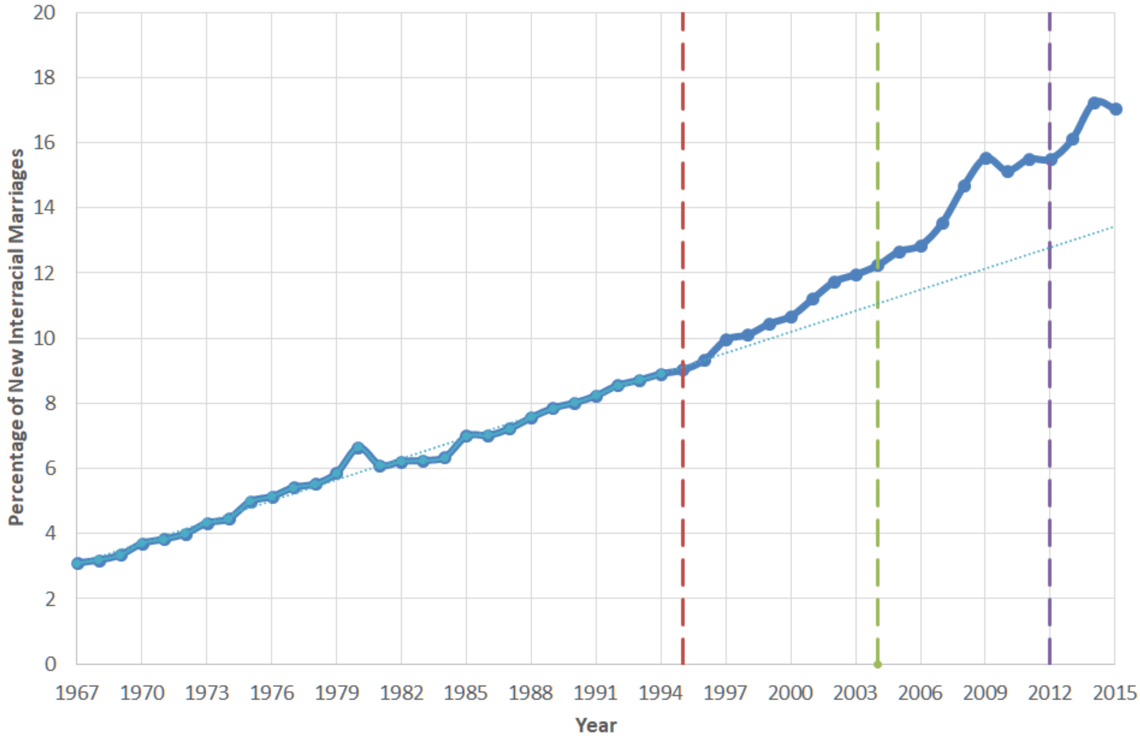


Figure 4.10: Percentage of interracial marriages among newlyweds in the U.S.
Source: Pew Research Center analysis of 2008-2015 American Community Survey and 1980, 1990 and 2000 decennial censuses (IPUMS). The red, green, and purple lines represent the creation of Match.com, OKCupid, and Tinder, three of the largest dating websites. The blue line represents a prediction for 1996 – 2015 using the data from 1967 to 1995.

Tinder, created in 2012, has approximately 50 million users and produces more than 12 million matches per day.[27]

We do not claim that the increase in the share of new marriages that are interracial in the last 20 years is a direct consequence of the emergence of online dating in the same period, but this finding is in line with Hypothesis 1 in our model.

Another cause for the steep increase described could be that the U.S. population is more interracial now than 20 years ago. The reduction of the percentage of Americans who are white, falling from 80.3% to 72.4% from 1990 to 2010,[28] combined with the fact that white people are the ones who show higher reluctance to intermarriage (Livingston and Brown, 2017, and even to date interracially, see Rudder, 2009), provides an alternative explanation.

[27] "Tinder, the fast-growing dating app, taps an age-old truth", *New York Times*, 29/10/2014. The company claims that 36% of Facebook users have had an account on their platform.
[28] "Demographic trends in the 20th century", and "The White Population: 2010 Census Briefs", U.S. Census Bureau.

However, this explanation is inconsistent with the empirical observation that white people are intermarrying more. By 1980, only 4% of the interracial newlyweds involved white persons, while the percentage raised to 11% in 2015 (Livingston and Brown, 2017).

4.5.2 Hypothesis 2: Marriages Created Online Are Less Likely to Divorce

Cacioppo et al. (2013) find that marriages created online were less likely to break up and reveal a higher marital satisfaction, using a sample of 19,131 Americans who married between 2005 and 2012. They write: *"What is clear from this research is that a surprising number of Americans now meet their spouse on-line, meeting a spouse on-line is on average associated with slightly higher marital satisfaction and lower rates of marital break-up than meeting a spouse through traditional (off-line) venues"*.

The findings of Cacioppo and his coauthors show that our model predictions closely match the observed properties of marriages created online, and its strength compared to marriages created on other, more traditional venues.

Our model predicts that, on average, marriages created when online dating becomes available last longer than those created in societies without this technology. Yet, it is silent regarding comparisons between the strength of interracial and intraracial marriages. There is empirical evidence showing that interracial marriages are more likely to end up in divorce (Bratter and King, 2008; Zhang and Van Hook, 2009).

Our model is also silent on why some intraracial marriages from a particular race last longer than intraracial marriages from another race (e.g. Stevenson and Wolfers, 2007 show that Blacks who divorce spend more time in their marriage than their White counterparts).

4.6 Final Remarks

4.6.1 Further Applications

The theoretical model we present discusses a general matching problem under network constraints, and hence it can be useful to study other social phenomena besides interracial marriage. The races or communities in our model can be understood as arbitrary groups of highly clustered agents. Agents can be clustered by race, but also by ethnicity, education, socieconomic status, religion, etcetera. Thus, our theoretical model can be also applied to study interfaith marriages, or marriages between people of different social status.

The role of connecting highly clustered groups is also not only linked to online dating. Another example is the European student exchange program "Erasmus", which helped more than 3 million students and over 350 thousand academics and staff members to spend time at a University abroad.[29]

The matching of agents also goes beyond marriage. Think of nodes being researchers at a University, races being academic departments, and edges represting who knows whom. Matchings indicate academic collaboration in articles or grants. The Euclidean distance interpretation makes sense, as a microeconomist in a business school may be better off partnering with a game theorist at the biology department rather with an econometrician in his own business school. Diversity in a University would be then a measure for interdisciplinary research, often encouraged by higher education institutions and funding bodies. Interdiscplinary seminars, for example, could take the role of creating links between academics in different departments.

It would be interesting to test our model against in this other scenarios. We leave this task for further research.

[29] "ERASMUS: Facts, figures and trends.", *European Comission*, 10/6/2014.

4.6.2 Conclusion

We introduce a simple theoretical model which tries to explain the complex process of deciding whom to marry in the times of online dating. As any model, ours has limitations. It categorizes every individual with only two characteristics, it assumes a very simple structure inside each race, it poses restrictions on agents' preferences. Furthermore, it fails to capture many of the complex features of romance in social networks, like love. There are multiple ways to enrich and complicate the model with more parameters.

However, the simplicity of our model is its main strength. With a basic structure, it can generate very strong predictions: the diversity of societies, measured by the number of interracial marriages in it, should increase drastically after the introduction of online dating. And societies with online dating available should produce marriages that are less likely to break up. Both predictions are consistent with observed demographic trends.

Simple models are great tools to convey an idea. Schelling's segregation model clearly does not capture many important components of how people decide where to live. It could have been enhanced by introducing thousands of parameters. Yet it has broadened the way how we understand racial segregation, and has been widely influential: according to Google Scholar, it has been quoted 3,258 times by articles coming from sociology to mathematics. It has provided us a way to think about an ubiquitous phenomenon.

Our model goes in the same direction.

Appendix A: Simulation Results

Table 1: Supporting data for Figures 4.8 and 4.9

q	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00
Panel A: Welfare on Euclidean societies														
$r = 2$, direct marriages														
Dv	0.00	0.10	0.19	0.27	0.34	0.41	0.47	0.57	0.66	0.75	0.82	0.89	0.94	1.00
St	0.85	0.87	0.87	0.87	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.89
Sz	1.00	0.98	0.98	0.98	0.98	0.99	0.99	0.99	0.99	0.99	1.00	1.00	1.00	1.00
$r = 2$, long marriages														
Dv	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
St	0.85	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.88	0.89	0.89
Sz	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$r = 5$,direct marriages														
Dv	0.00	0.22	0.37	0.47	0.56	0.62	0.68	0.77	0.83	0.88	0.92	0.95	0.98	1.00
St	0.85	0.88	0.89	0.89	0.90	0.90	0.91	0.91	0.91	0.91	0.92	0.92	0.92	0.92
Sz	1.00	0.97	0.98	0.98	0.99	0.99	0.99	0.99	0.99	1.00	1.00	1.00	1.00	1.00
$r = 5$, long marriages														
Dv	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
St	0.85	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92
Sz	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Panel B: Welfare on assortative societies														
$r = 2$, direct marriages														
Dv	0.00	0.11	0.20	0.28	0.35	0.41	0.47	0.57	0.66	0.75	0.82	0.88	0.95	1.00
St	0.84	0.85	0.86	0.86	0.86	0.86	0.86	0.87	0.87	0.87	0.87	0.87	0.87	0.87
Sz	1.00	0.98	0.98	0.98	0.98	0.99	0.99	0.99	0.99	0.99	1.00	1.00	1.00	1.00
$r = 2$, long marriages														
Dv	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
St	0.84	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87
Sz	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$r = 5$,direct marriages														
Dv	0.00	0.23	0.37	0.48	0.56	0.63	0.68	0.77	0.83	0.88	0.92	0.95	0.98	1.00
St	0.84	0.87	0.88	0.89	0.89	0.89	0.90	0.90	0.90	0.90	0.90	0.91	0.91	0.91
Sz	1.00	0.97	0.98	0.98	0.99	0.99	0.99	0.99	0.99	1.00	1.00	1.00	1.00	1.00
$r = 5$, long marriages														
Dv	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
St	0.84	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91
Sz	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

*Average over 10,000 random simulations, $n = 50$, $p = 1$.
Sz equals the percentage of agents married.
Standard errors in the order of 1.0e-04, so we do not present them.

Appendix B: Robustness Checks

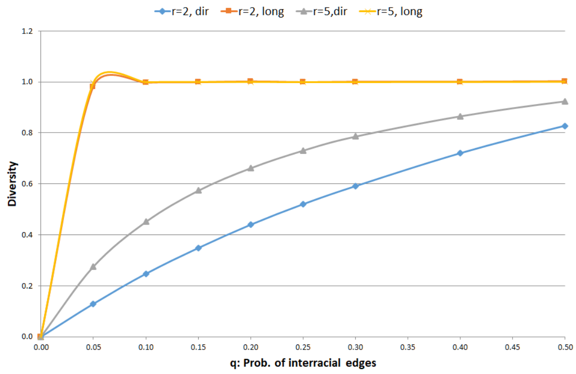
In this Appendix we conduct several robustness checks to show that the fast increase in the diversity of societies that we described in Result 1 occurs for many combinations of parameters of the model.

The first exercise we perform is to conduct again 10,000 simulations as those supporting Figures 4.8 and 4.9, but now changing the probability of intraracial connectivity p to 0.7, 0.5, and 0.3. We allow q to vary within 0 and p, as we have explained in the text that $q \leq p$, since persons tend to be more connected to people from their own race.

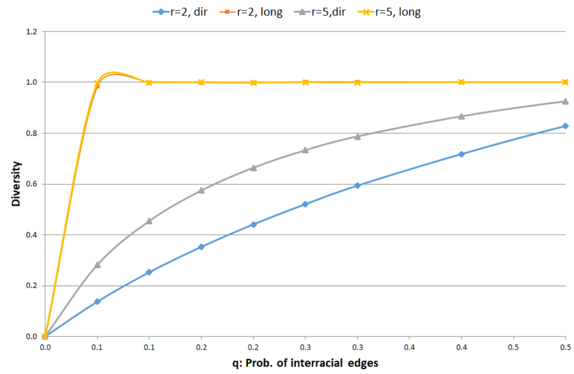
With respect to diversity, with long marriages we always observe an almost immediate increase to 1, meaning complete social integration. This increase appears in Figure 11. As expected, the increase becomes steeper as p increases.

With respect to strength, we also observe minor variations, which appear in Figure 12. As expected, a smaller p makes agents less connected to potential partners, and thus the strength of resulting marriages becomes weaker when agents are poorly connected. With long marriages, strength converges quite quickly to its optimal value, around 0.9, which again, is smaller in societies with low values of p and q.

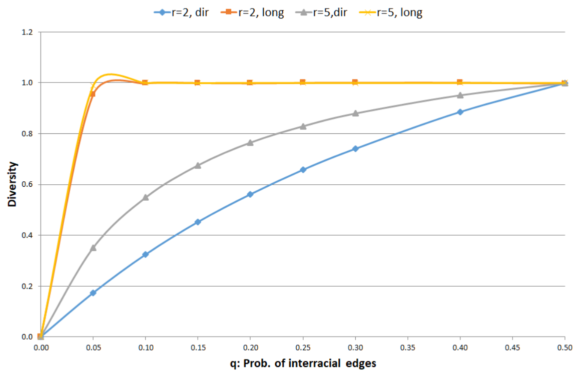
The detailed results of our simulations with p equal to 0.7, 0.5, and 0.3 appear in Tables 2, 3 and 4 at the end of this Appendix.



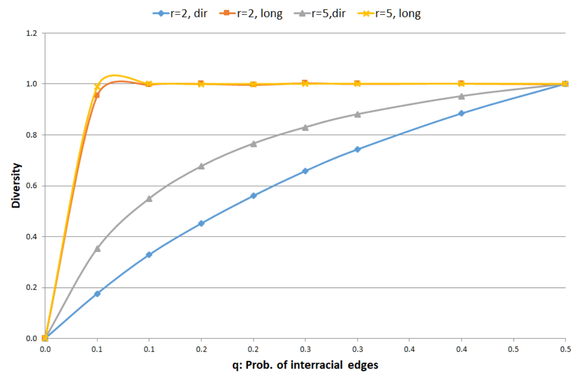
(a) Euclidean society, $p = .7$.



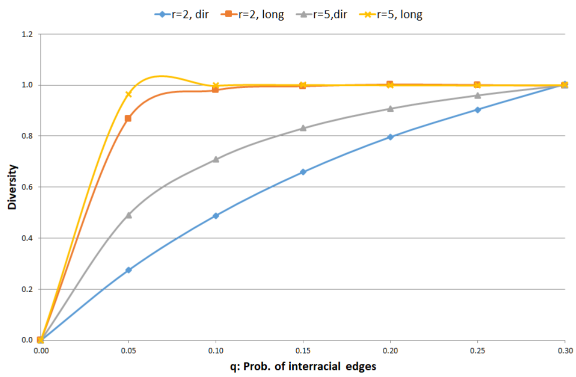
(b) Assortative society, $p = .7$



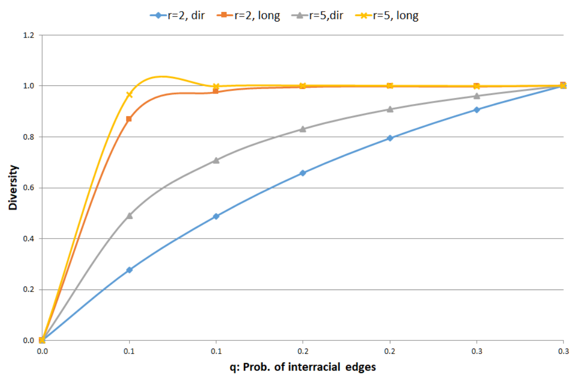
(c) Euclidean society, $p = .5$.



(d) Assortative society, $p = .5$

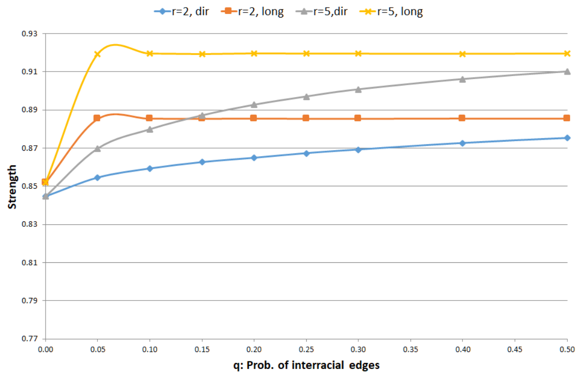


(e) Euclidean society, $p = .3$.

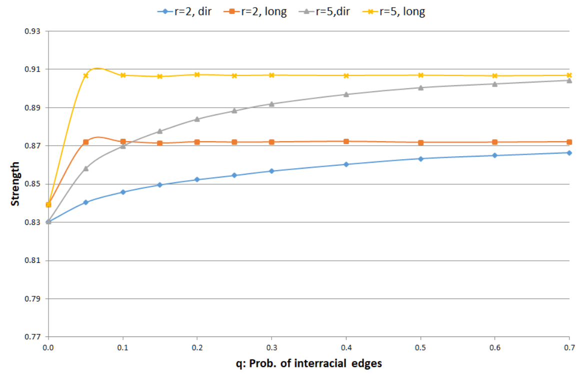


(f) Assortative society, $p = .3$

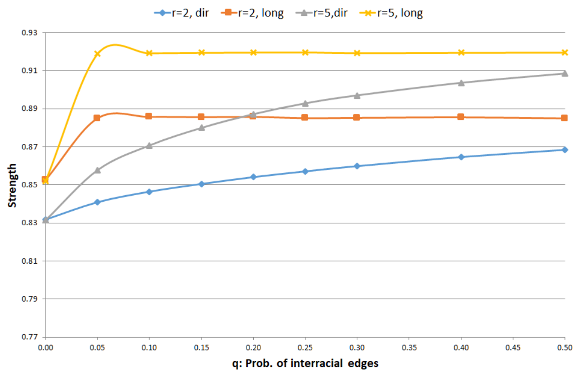
Figure 11: Average diversity (y-axis) of a random society for several values of p.



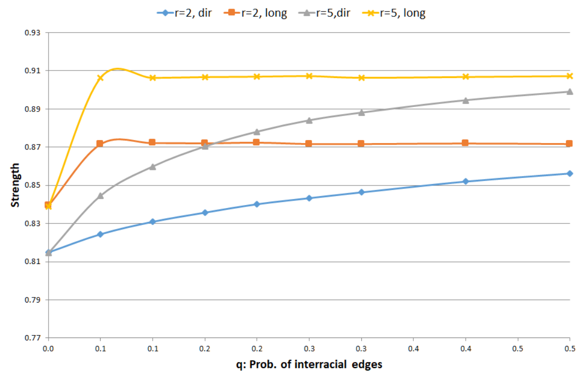
(a) Euclidean society, $p = .7$.



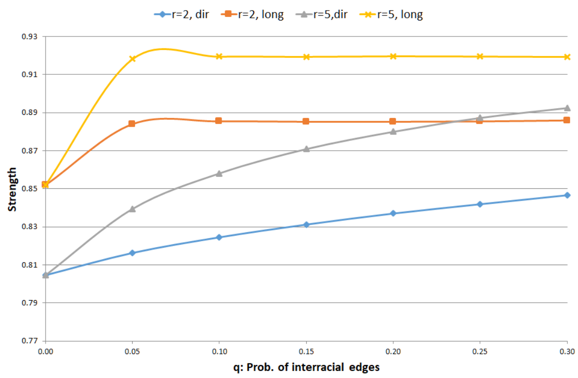
(b) Assortative society, $p = .7$



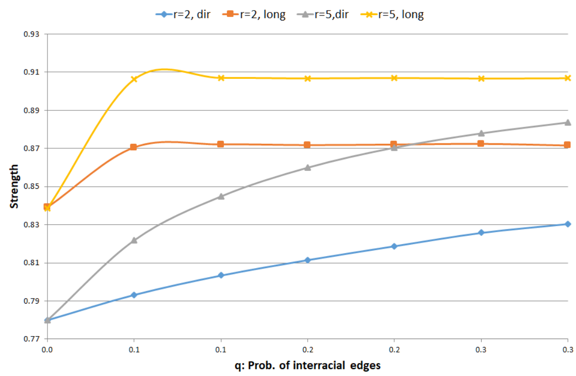
(c) Euclidean society, $p = .5$.



(d) Assortative society, $p = .5$



(e) Euclidean society, $p = .3$.



(f) Assortative society, $p = .3$

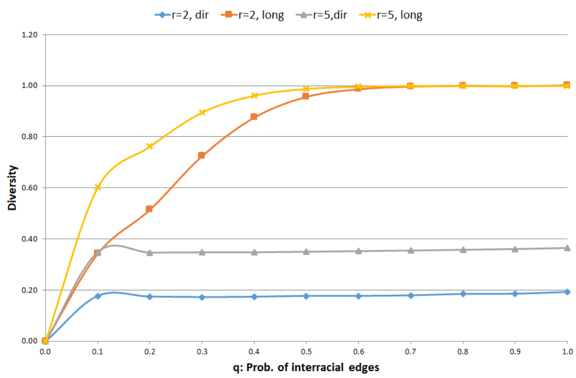
Figure 12: Average stregth (y-axis) of a random society for several values of p.

The second robustness test we perform is to vary p and q simultaneously but keeping its ratio fixed. Both parameters indicate how connected is a person to people of his race compared to people of other races.

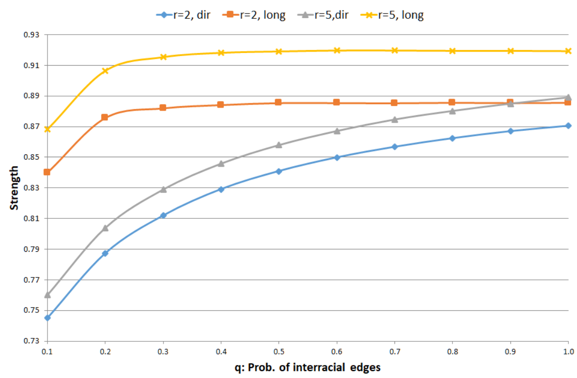
To find a good estimate of the ratio $\frac{p}{q}$, we use data from the American Values Survey by the Public Religion Research Institute (PRRI), a nonpartisan, independent research organization. The data is well described in the following article from the Washington Post: "Three quarters of whites dont have any non-white friends".

The PRRI data shows that, if a White American had 100 friends, 91 are expected to be of his own race, and 1 Black, 1 Latino, and 1 Asian (the rest are multiracial or of unknown race). Black Americans are more interracially connected, with 83 friends expected to be of his own race, 8 Whites, 2 Latinos, and and no Asians.

Based on this data, we use the ratio $p/q = 10$, based on the ratio between the expected number of Black and White friends for Black people. We vary p from 0 to 1. We present the results for Euclidean societies only (as we have seen that Euclidean and assortative societies produce almost identical results).



(a) Diversity.



(b) Strength.

Figure 13: Average diversity and strength of a random society for $p \in [0, 1]$.

A first conclusion we obtain is that, with long marriages, we again obtain complete integration. However, this time is not as fast as with an increase of q alone. With direct marriages the increase is again very fast but full integration is not obtained, only around 20% and 40% of it in societies with 2 and 5 races, respectively.

We could say that the diversity achieved when agents intra and interracial circles both grow is much lower, compared to the results shown in the main text. But this lecture is not accurate, because we are comparing the diversity to the one that obtains in a fully connected society, i.e. a complete graph. Therefore, the diversity obtained already is 20% and 40% of the diversity in a complete graph, and that is a very high percentage of interracial marriages, because we are fixing that agents are 10 times more connected to its own race.

Finally, the strength levels we observe with direct marriages are the lowest we have found so far, which is not a surprise given the small number of potential partners that agents have when p is small. It is equally expected to observe that the strength of a society increases when p grows.

Table 2: Welfare with $p = 0.7$

q	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.40	0.50	0.60	0.70
Panel A: Welfare on Euclidean societies											
$r = 2$, direct marriages											
Dv	0.00	0.13	0.25	0.35	0.44	0.52	0.59	0.72	0.83	0.92	1.00
St	0.84	0.85	0.86	0.86	0.86	0.87	0.87	0.87	0.88	0.88	0.88
Sz	0.98	0.97	0.97	0.97	0.97	0.98	0.98	0.98	0.99	0.99	0.99
$r = 2$, long marriages											
Dv	0.00	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
St	0.85	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89
Sz	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$r = 5$, direct marriages											
Dv	0.00	0.28	0.45	0.57	0.66	0.73	0.79	0.87	0.92	0.97	1.00
St	0.84	0.87	0.88	0.89	0.89	0.90	0.90	0.91	0.91	0.91	0.92
Sz	0.98	0.97	0.97	0.98	0.98	0.98	0.99	0.99	0.99	0.99	1.00
$r = 5$, long marriages											
Dv	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
St	0.85	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92
Sz	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Panel B: Welfare on assortative societies											
$r = 2$, direct marriages											
Dv	0.00	0.14	0.25	0.35	0.44	0.52	0.59	0.72	0.83	0.92	1.00
St	0.83	0.84	0.85	0.85	0.85	0.85	0.86	0.86	0.86	0.86	0.87
Sz	0.98	0.97	0.97	0.97	0.98	0.98	0.98	0.98	0.99	0.99	0.99
$r = 2$, long marriages											
Dv	0.00	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
St	0.84	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87
Sz	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$r = 5$, direct marriages											
Dv	0.00	0.28	0.45	0.57	0.66	0.73	0.79	0.87	0.93	0.97	1.00
St	0.83	0.86	0.87	0.88	0.88	0.89	0.89	0.90	0.90	0.90	0.90
Sz	0.98	0.97	0.97	0.98	0.98	0.99	0.99	0.99	0.99	1.00	1.00
$r = 5$, long marriages											
Dv	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
St	0.84	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91
Sz	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

*Average over 10,000 random simulations, $n = 50$.
Sz equals the percentage of agents married.
Standard errors in the order of 1.0e-04.

Table 3: Welfare with $p = 0.5$

q	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.40	0.50
Panel A: Welfare on Euclidean societies									
$r = 2$, direct marriages									
Dv	0.00	0.17	0.32	0.45	0.56	0.66	0.74	0.89	1.00
St	0.83	0.84	0.85	0.85	0.85	0.86	0.86	0.86	0.87
Sz	0.96	0.95	0.96	0.96	0.96	0.97	0.97	0.97	0.98
$r = 2$, long marriages									
Dv	0.00	0.95	1.00	1.00	1.00	1.00	1.00	1.00	1.00
St	0.85	0.88	0.89	0.89	0.89	0.89	0.89	0.89	0.88
Sz	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$r = 5$, direct marriages									
Dv	0.00	0.35	0.55	0.68	0.77	0.83	0.88	0.95	1.00
St	0.83	0.86	0.87	0.88	0.89	0.89	0.90	0.90	0.91
Sz	0.96	0.96	0.97	0.97	0.98	0.98	0.98	0.99	0.99
$r = 5$, long marriages									
Dv	0.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
St	0.85	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92
Sz	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Panel B: Welfare on assortative societies									
$r = 2$, direct marriages									
Dv	0.00	0.18	0.33	0.45	0.56	0.66	0.74	0.88	1.00
St	0.58	0.58	0.59	0.59	0.59	0.60	0.60	0.60	0.61
Sz	0.96	0.96	0.96	0.96	0.97	0.97	0.97	0.98	0.98
$r = 2$, long marriages									
Dv	0.00	0.95	1.00	1.00	1.00	1.00	1.00	1.00	1.00
St	0.84	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87
Sz	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$r = 5$, direct marriages									
Dv	0.00	0.35	0.55	0.68	0.77	0.83	0.88	0.95	1.00
St	0.58	0.60	0.61	0.62	0.62	0.63	0.63	0.63	0.64
Sz	0.96	0.96	0.97	0.98	0.98	0.98	0.99	0.99	0.99
$r = 5$, long marriages									
Dv	0.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00
St	0.84	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91
Sz	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

*Average over 10,000 random simulations, $n = 50$.

Sz equals the percentage of agents married.

Standard errors in the order of 1.0e-04.

Table 4: Welfare with $p = 0.3$

q	0	0.05	1	0.15	0.2	0.25	0.3
Panel A: Welfare on Euclidean societies							
$r = 2$, direct marriages							
Dv	0.00	0.27	0.49	0.66	0.80	0.90	1.00
St	0.80	0.82	0.82	0.83	0.84	0.84	0.85
Sz	0.91	0.92	0.93	0.93	0.94	0.95	0.95
$r = 2$, long marriages							
Dv	0.00	0.87	0.98	1.00	1.00	1.00	1.00
St	0.85	0.88	0.89	0.89	0.89	0.89	0.89
Sz	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$r = 5$, direct marriages							
Dv	0.00	0.49	0.71	0.83	0.91	0.96	1.00
St	0.00	0.05	0.10	0.15	0.20	0.25	0.30
Sz	0.91	0.94	0.95	0.97	0.97	0.98	0.98
$r = 5$, long marriages							
Dv	0.00	0.96	1.00	1.00	1.00	1.00	1.00
St	0.85	0.92	0.92	0.92	0.92	0.92	0.92
Sz	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Panel B: Welfare on assortative societies							
$r = 2$, direct marriages							
Dv	0.00	0.28	0.49	0.66	0.80	0.91	1.00
St	0.78	0.79	0.80	0.81	0.82	0.83	0.83
Sz	0.92	0.93	0.93	0.94	0.95	0.95	0.96
$r = 2$, long marriages							
Dv	0.00	0.87	0.98	1.00	1.00	1.00	1.00
St	0.84	0.87	0.87	0.87	0.87	0.87	0.87
Sz	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$r = 5$, direct marriages							
Dv	0.00	0.49	0.71	0.83	0.91	0.96	1.00
St	0.78	0.82	0.84	0.86	0.87	0.88	0.88
Sz	0.92	0.94	0.96	0.97	0.98	0.98	0.98
$r = 5$, long marriages							
Dv	0.00	0.97	1.00	1.00	1.00	1.00	1.00
St	0.84	0.91	0.91	0.91	0.91	0.91	0.91
Sz	1.00	1.00	1.00	1.00	1.00	1.00	1.00

*Average over 10,000 random simulations, $n = 50$.
Sz equals the percentage of agents married.
Standard errors in the order of 1.0e-04.

Table 5: Welfare with $\frac{p}{q} = 10$

p	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Welfare on Euclidean societies										
$r = 2$, direct marriages										
0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00	
0.18	0.17	0.17	0.17	0.18	0.18	0.18	0.18	0.19	0.19	
0.75	0.79	0.81	0.83	0.84	0.85	0.86	0.86	0.87	0.87	
$r = 2$, long marriages										
0.75	0.87	0.91	0.94	0.95	0.96	0.97	0.97	0.98	0.98	
0.34	0.52	0.73	0.88	0.96	0.99	1.00	1.00	1.00	1.00	
0.84	0.88	0.88	0.88	0.89	0.89	0.89	0.89	0.89	0.89	
$r = 5$, direct marriages										
0.91	0.98	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	
0.35	0.35	0.35	0.35	0.35	0.35	0.35	0.36	0.36	0.36	
0.76	0.80	0.83	0.85	0.86	0.87	0.87	0.88	0.88	0.89	
$r = 5$, long marriages										
0.79	0.89	0.93	0.94	0.96	0.96	0.97	0.97	0.98	0.98	
0.60	0.76	0.90	0.96	0.99	1.00	1.00	1.00	1.00	1.00	
0.87	0.91	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	
0.94	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	

*Average over 10,000 random simulations, $n = 50$.
Sz equals the percentage of agents married.
Standard errors in the order of 1.0e-04.

Bibliography

Aleksandrov, M., H. Aziz, S. Gaspers, and T. Walsh (2015): "Online fair division: analysing a food bank problem," in *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015*, 2540–2546.

Amir, G., O. Angel, and A. Holroyd (2016): "Multidimensional Poisson matching," *unpublished.*

Arnsperger, C. (1994): "Envy-freeness and distributive justice," *Journal of Economic Surveys*, 8, 155–186.

Arrow, K. J. (1998): "What has economics to say about racial discrimination?" *Journal of Economic Perspectives*, 12, 91–100.

Ashlagi, I., Y. Kanoria, and J. D. Leshno (2017): "Unbalanced random matching markets: The stark effect of competition," *Journal of Political Economy*, 125, 69–98.

Ashlagi, I. and A. E. Roth (2014): "Free riding and participation in large scale, multi-hospital kidney exchange," *Theoretical Economics*, 9, 817–863.

Backstrom, L. and J. Kleinberg (2014): "Romantic partnerships and the dispersion of social ties: A network analysis of relationship status on Facebook," in *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, New York, NY, USA: ACM, CSCW '14, 831–841.

Banerjee, A., E. Duflo, M. Ghatak, and J. Lafortune (2013): "Marry for what? Caste and mate selection in modern India," *American Economic Journal: Microeconomics*, 5, 33–72.

Becker, G. (1981): *A Treatise on the family*, Harvard University Press.

Becker, G. S. (1973): "A Theory of Marriage: Part I," *Journal of Political Economy*, 81, 813–846.

——— (1974): "A Theory of Marriage: Part II," *Journal of Political Economy*, 82, S11–S26.

Bisin, A. and T. Verdier (2000): ""Beyond the melting pot": Cultural transmission, marriage, and the evolution of ethnic and religious traits," *Quarterly Journal of Economics*, 115, 955–988.

Bogomolnaia, A. and J.-F. Laslier (2007): "Euclidean preferences," *Journal of Mathematical Economics*, 43, 87 – 98.

Bogomolnaia, A. and H. Moulin (2004): "Random matching under dichotomous preferences," *Econometrica*, 72, 257–279.

Bogomolnaia, A., H. Moulin, F. Sandomirskiy, and E. Yanovskaya (2017): "Competitive division of a mixed manna," *ArXiv e-prints*.

Bogomolnaia, A., H. Moulin, and R. Stong (2005): "Collective choice under dichotomous preferences," *Journal of Economic Theory*, 122, 165 – 184.

Bollobás, B. (2001): *Random graphs*, Cambridge Studies in Advanced Mathematics, Cambridge University Press, 2 ed.

Bouveret, S. and J. Lang (2008): "Efficiency and envy-freeness in fair division of indivisible goods: logical representation and complexity," *Journal of Artificial Intelligence Research*, 32, 525–564.

Bouveret, S. and M. Lemaître (2016): "Characterizing conflicts in fair division of indivisible goods using a scale of criteria," *Autonomous Agents and Multi-Agent Systems*, 30, 259–290.

Brainard, W. and H. Scarf (2005): "How to compute equilibrium prices in 1891," *American Journal of Economics and Sociology*, 64, 57–83.

Brams, S. and M. Kilgour (2001): "Competitive fair division," *Journal of Political Economy*, 109, 418–443.

Bratter, J. and R. King (2008): "But will it last?: Marital instability among interracial and same-race couples," *Family Relations*, 57, 160–171.

Browning, M., P.-A. Chiappori, and Y. Weiss (2014): *Economics of the family*, Cambridge Surveys of Economic Literature, Cambridge University Press.

Budish, E. (2011): "The combinatorial assignment problem: approximate competitive equilibrium from equal incomes," *Journal of Political Economy*, 119, 1061–1103.

Budish, E., G. Cachon, J. Kessler, and A. Othman (2017): "Course match: A large-scale implementation of approximate competitive equilibrium from equal incomes for combinatorial allocation," *Operations Research*, 65, 314–336.

Budish, E. and E. Cantillon (2012): "The multi-unit assignment problem: theory and evidence from course allocation at Harvard," *American Economic Review*, 102, 2237–71.

Budish, E., Y.-K. Che, F. Kojima, and P. Milgrom (2013): "Designing random allocation mechanisms: theory and applications," *American Economic Review*, 103, 585–623.

Cacioppo, J. T., S. Cacioppo, G. C. Gonzaga, E. L. Ogburn, and T. J. VanderWeele (2013): "Marital satisfaction and break-ups differ across on-line and off-line meeting venues," *Proceedings of the National Academy of Sciences*, 110, 10135–10140.

Caestecker, F. and D. Fraser (2008): "The extraterritorial application of the Nuremberg laws. Rassenschande and mixed marriages in European liberal democracies," *Journal of the History of International Law*, 10, 35–81.

Chambers, C. and T. Hayashi (2017): "Can everyone benefit from economic integration?" *unpublished.*

Chambers, C. and M. B. Yenmez (2017): "Choice and matching," *American Economic Journal: Microeconomics*, forthcoming.

Che, Y.-K. and O. Tercieux (2015): "Efficiency and stability in large matching markets," *unpublished.*

Chiappori, P.-A., R. McCann, and B. Pass (2016): "Multidimensional Matching," Papers 1604.05771, arXiv.org.

Chiappori, P.-A., S. Oreffice, and C. Quintana-Domeque (2012): "Fatter attraction: Anthropometric and socioeconomic matching on the marriage market," *Journal of Political Economy*, 120, 659–695.

Chipman, J. S. (1974): "Homothetic preferences and aggregation," *Journal of Economic Theory*, 8, 26 – 38.

Cohen, R., M. L. Barr, and V. Starnes (2014): "Living donor lung transplantation," in *Living donor advocacy: An evolving role within transplantation*, ed. by J. Steel, New York, NY: Springer, 75–89.

Crawford, V. P. (1991): "Comparative statics in matching markets," *Journal of Economic Theory*, 54, 389 – 400.

Cutler, D., E. L. Glaeser, and J. L. Vigdor (1999): "The Rise and Decline of the American Ghetto," *Journal of Political Economy*, 107, 455–506.

Debreu, G. (1982): "Existence of competitive equilibrium," in *Handbook of Mathematical Economics*, ed. by K. Arrow and M. Intrilligator, Elsevier, vol. 2, 697 – 743.

Duncan, B. and S. J. Trejo (2011): "Intermarriage and the intergenerational transmission of ethnic identity and human capital for Mexican Americans," *Journal of Labor Economics*, 29, 195–227.

Dutta, B. and D. Ray (1989): "A concept of egalitarianism under participation constraints," *Econometrica*, 57, 615–635.

Dworkin, R. (1981): "What is equality? Part 2: equality of resources," *Philosophy & Public Affairs*, 10, 283–345.

Echenique, F. and R. G. Fryer (2007): "A Measure of Segregation Based on Social Interactions," *Quarterly Journal of Economics*, 122, 441–485.

Eisenberg, E. (1961): "Aggregation of utility functions," *Management Science*, 7, 337–350.

Eisenberg, E. and D. Gale (1959): "Consensus of subjective probabilities: the Pari-Mutuel method," *Annals of Mathematical Statistics*, 30, 165–168.

Erdős, P. and A. Rényi (1959): "On random graphs I." *Publicationes Mathematicae (Debrecen)*, 6, 290–297.

Ergin, H., T. Sönmez, and U. Ünver (2017): "Dual-donor organ exchange," *Econometrica*, forthcoming.

Fisman, R., S. S. Iyengar, E. Kamenica, and I. Simonson (2008): "Racial preferences in dating," *Review of Economic Studies*, 75, 117–132.

Foster, J. E. and E. A. Ok (1999): "Lorenz dominance and the variance of logarithms," *Econometrica*, 67, 901–907.

Fryer, R. G. (2007): "Guess who's been coming to dinner? Trends in interracial marriage over the 20th century," *Journal of Economic Perspectives*, 21, 71–90.

Furtado, D. (2012): "Human capital and interethnic marriage decisions," *Economic Inquiry*, 50, 82–93.

——— (2015): "Ethnic intermarriage," in *International Encyclopedia of the Social & Behavioral Sciences*, ed. by J. D. Wright, Oxford: Elsevier, 118 – 122, second ed.

Furtado, D. and N. Theodoropoulos (2010): "Why does intermarriage increase immigrant employment? The role of networks," *The B.E. Journal of Economic Analysis & Policy*, 10, 1–33.

Gale, D. and L. Shapley (1962): "College admissions and the stability of marriage," *American Mathematical Monthly*, 69, 9–15.

Garcia, D. R. (2008): "Academic and racial segregation in charter schools," *Education and Urban Society*, 40, 590–612.

Gee, L. K., J. Jones, and M. Burke (2017): "Social networks and labor markets: How strong ties relate to job finding on Facebooks social network," *Journal of Labor Economics*, 35, 485–518.

Gilbert, E. (1959): "Random Graphs," *Annals of Mathematical Statistics*, 30, 1141–1144.

Goel, D. and K. Lang (2009): "Social ties and the job search of recent immigrants," NBER Working Papers 15186, National Bureau of Economic Research.

Goldman, J. and A. Procaccia (2015): "Spliddit: unleashing fair division algorithms," *SIGecom Exchanges*, 13, 41–46.

Granovetter, M. S. (1973): "The strength of weak ties," *American Journal of Sociology*, 78, 1360–1380.

Hitsch, G. J., A. Hortasu, and D. Ariely (2010): "Matching and sorting in online dating," *American Economic Review*, 100, 130–63.

Holroyd, A. E. (2011): "Geometric Properties of Poisson Matchings," *Probability Theory and Related Fields*, 150, 511–527.

Holroyd, A. E., R. Pemantle, Y. Peres, and O. Schramm (2009): "Poisson matching," *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 45, 266–287.

Hylland, A. and R. Zeckhauser (1979): "The efficient allocation of individuals to positions," *Journal of Political Economy*, 87, 293–314.

Hyslop, J. (1995): "White working-class women and the invention of apartheld: Purified Afrikaner nationalist agitation for legislation against mixed marriages, 1934-9," *Journal of African History*, 36, 57–81.

Immorlica, N. and M. Mahdian (2005): "Marriage, honesty, and stability," in *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, SODA '05, 53–62.

Jain, K. and V. Vazirani (2010): "Eisenberg – Gale markets: algorithms and game-theoretic properties," *Games and Economic Behavior*, 70, 84 – 106.

Kalmijn, M. (1998): "Intermarriage and homogamy: Causes, patterns, trends," *Annual Review of Sociology*, 24, 395–421.

Kantarevic, J. (2004): "Interethnic marriages and economic assimilation of immigrants," IZA Discussion Papers 1142, Institute for the Study of Labor (IZA).

Katta, A. and J. Sethuraman (2006): "A solution to the random assignment problem on the full preference domain," *Journal of Economic Theory*, 131, 231 – 250.

Kelso, A. S. and V. P. Crawford (1982): "Job matching, coalition formation, and gross substitutes," *Econometrica*, 50, 1483–1504.

Kojima, F. and P. A. Pathak (2009): "Incentives and stability in large two-sided matching markets," *American Economic Review*, 99, 608–27.

Kominers, S., M. Ruberry, and J. Ullman (2010): "Course allocation by proxy auction," in *Proceedings of the Sixth International Workshop on Web and Internet Economics, WINE 2010*, 551–558.

Koutsoupias, E. (2014): "Scheduling without payments," *Theory of Computing Systems*, 54, 375–387.

Kramarz, F. and O. N. Skans (2014): "When strong ties are strong: Networks and youth labour market entry," *Review of Economic Studies*, 81, 1164–1200.

Krishna, A. and U. Ünver (2008): "Improving the efficiency of course bidding at business schools: field and laboratory studies," *Marketing Science*, 27, 262–282.

Lee, S. (2017): "Incentive compatibility of large centralized matching markets," *Review of Economic Studies*, 84, 444.

Lewis, K. (2013): "The limits of racial prejudice," *Proceedings of the National Academy of Sciences*, 110, 18814–18819.

Livingston, G. and A. Brown (2017): "Intermarriage in the U.S. 50 years after Loving v. Virginia," *Pew Research Center Report*.

Maher, M. (2016): "A tennis assignment algorithm," *Mathematics Today*, 52, 130–131.

McGrath, A., G. Tsunokai, M. Schultz, J. Kavanagh, and J. Tarrence (2016): "Differing shades of colour: Online dating preferences of biracial individuals," *Ethnic and Racial Studies*, 39, 1920–1942.

Meng, X. and R. Gregory (2005): "Intermarriage and the economic assimilation of immigrants," *Journal of Labor Economics*, 23, 135–174.

Moulin, H. (1990): "Fair division under joint ownership: Recent results and open problems," *Social Choice and Welfare*, 7, 149–170.

——— (1992): "An application of the Shapley value to fair division with money," *Econometrica*, 60, 1331–1349.

——— (2003): *Fair division and collective welfare*, Cambridge, MA, USA: MIT Press.

Moulin, H. and W. Thomson (1988): "Can everyone benefit from growth?" *Journal of Mathematical Economics*, 17, 339 – 345.

Ortega, J. (2016): "Playing tennis without envy," *Mathematics Today*, 52, 268–269.

——— (2017): "Can everyone benefit from social integration?" Papers 1705.08033, arXiv.org.

Othman, A., C. Papadimitriou, and A. Rubinstein (2014): "The Complexity of Fairness Through Equilibrium," in *ACM Conference on Economics and Computation*, New York, NY, USA: ACM, EC '14, 209–226.

Othman, A., T. Sandholm, and E. Budish (2010): "Finding approximate competitive equilibria: Efficient and fair course allocation," in *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems*, Richland, SC, AAMAS '10, 873–880.

Pittel, B. (1989): "The average number of stable matchings," *SIAM Journal on Discrete Mathematics*, 2, 530–549.

Potarca, G. and M. Mills (2015): "Racial preferences in online dating across European countries," *European Sociological Review*, 31, 326–341.

Prendergast, C. (2017): "The allocation of food to food banks," *unpublished*.

Qian, Z. (1997): "Breaking the racial barriers: Variations in interracial marriage between 1980 and 1990," *Demography*, 34, 263–276.

Rosenfeld, M. J. and R. J. Thomas (2012): "Searching for a mate: The rise of the Internet as a social intermediary," *American Sociological Review*, 77, 523–547.

Roth, A., T. Sönmez, and U. Ünver (2005): "Pairwise kidney exchange," *Journal of Economic Theory*, 125, 151 – 188.

Roth, A. and M. Sotomayor (1992): *Two-sided matching: A study in game-theoretic modeling and analysis*, Econometric Society Monographs, Cambridge University Press.

Roth, A. E. (2002): "The economist as engineer: Game theory, experimentation, and computation as tools for design economics," *Econometrica*, 70, 1341–1378.

Rudder, C. (2009): "How your race affects the messages you get," OK Cupid Blog.

——— (2014): *Dataclysm: Who we are (When we think no one's looking)*, New York, NY, USA: Crown Publishing Group.

Schelling, T. (1972): “A process of residential segregation: Neighborhood tipping,” in *Racial Discrimination in Economic Life*, ed. by A. Pascal, Lexington, Massachusetts: Lexington Books - RAND, chap. 5, 157–184.

Schelling, T. C. (1969): “Models of segregation,” *The American Economic Review*, 59, 488–493.

——— (1971): “Dynamic models of segregation,” *The Journal of Mathematical Sociology*, 1, 143–186.

Shapley, L. and H. Scarf (1974): “On cores and indivisibility,” *Journal of Mathematical Economics*, 1, 23 – 37.

Shubik, M. (1971): “The “bridge game” economy: an example of indivisibilities,” *Journal of Political Economy*, 79, 909–912.

Sobel, J. (2009): “ReGale: Some memorable results,” *Games and Economic Behavior*, 66, 632 – 642, special Section In Honor of David Gale.

Sönmez, T. and U. Ünver (2010): “Course bidding at business schools,” *International Economic Review*, 51, 99–123.

Sprumont, Y. (1990): “Population monotonic allocation schemes for cooperative games with transferable utility,” *Games and Economic Behavior*, 2, 378 – 394.

——— (2008): “Monotonicity and solidarity axioms in economics and game theory,” in *Rational Choice and Social Welfare: Essays in Honor of Kotaro Suzumura*, ed. by P. Pattanaik, T. Koichi, X. Yongsheng, and Y. Naoki, Berlin, Heidelberg: Springer, 71–94.

Stevenson, B. and J. Wolfers (2007): “Marriage and divorce: Changes and their driving forces,” *Journal of Economic Perspectives*, 21, 27–52.

Thomson, W. (1983): “The fair division of a fixed supply among a growing population,” *Mathematics of Operations Research*, 8, 319–326.

U.S. Commission of Civil Rights (1977): "School desegregation in Little Rock, Arkansas," Tech. rep., Staff Report of the U.S. Commission on Civil Rights.

Uzawa, H. (1962): "Walras'existence theorem and Brouwer's fixed-point theorem," *Economic Studies Quarterly*, 13, 59–62.

Varian, H. (1974): "Equity, envy, and efficiency," *Journal of Economic Theory*, 9, 63 – 91.

Vazirani, V. (2007): "Combinatorial algorithms for market equilibria," in *Algorithmic Game Theory:*, ed. by N. Nisan, T. Roughgarden, E. Tardos, and V. Vazirani, Cambridge University Press, 103–134.

Wong, L. Y. (2003): "Why do only 5.5% of black men marry white women?" *International Economic Review*, 44, 803–826.

Yilmaz, Ö. (2011): "Kidney exchange: An egalitarian mechanism," *Journal of Economic Theory*, 146, 592 – 618.

Zhang, Y. and J. Van Hook (2009): "Marital dissolution among interracial couples," *Journal of Marriage and Family*, 71, 95–107.