Essays in Matching Theory

Thesis by
Peter Nathanael Doe

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2025
Defended May 2, 2025

Peter Nathanael Doe
ORCID: 0009-0002-6957-8911

ACKNOWLEDGEMENTS

This thesis would not have been possible without the generous support of numerous advisors, mentors, friends, and family.

My time at Caltech has been doubly enriched by the support of my advisors, Federico Echenique and Luciano Pomatto. Federico has shaped my intellectual journey more than anyone else. He taught me how to ask clear questions and find useful answers. The generosity with which he offered his time and support was fundamental to my success at Caltech. Luciano generously adopted me as his student partway through my journey, and his advising has profoundly impacted my understanding of research. His clarity of thought and style of presentation inspired my work. I have learned a tremendous amount from both of them, and I count myself especially fortunate to have worked with both.

I am also indebted to Omer Tamuz, who served as the chair of my committee. Through his guidance and good humor, Omer showed me how to articulate my ideas clearly and to enjoy my research. I am grateful as well to Charlie Sprenger, who served on my committee. Charlie's insightful feedback helped me to think critically about my research and its implications.

Beyond my committee, I have benefited immensely from numerous meetings with Pete Caradonna and Axel Niemeyer. Pete's mentorship extended beyond research; he taught me how to thrive in the role of an academic. In every meeting, Axel took time to fully understand my research, and he consistently pushed me to strengthen my arguments and reexamine my assumptions.

As a graduate student, I have had the privilege of forming many friendships with fellow students and postdoctoral scholars. Among these, several have been influential in the writing of this thesis. Jack Adeney, Fan Wu, Ke Shi, Nghia Tran Pham, Sumit Goel, Wade Hann-Caruthers, and Fedor Sandomirskiy have all offered insightful comments throughout this process, generously fielding my ideas and helping to refine them. Polina Detkova, Egor Stoyan, Camila Farrés, and Mitchell Linegar graciously volunteered their time to help me learn to present my work well and gave valuable comments.

Outside of Caltech, the support that the members of the Chapel at Pasadena have given to my family made this journey possible. They have been the welcoming arms of Christ to my family, opening their homes and building a community of friendship.

Special thanks are owed to my friend and mentor Glenn Biasi, who has consistently pointed my eyes to Christ.

My deepest thanks go to my wife, Vanessa. Her steadfast love and unending patience have made this work possible. She uprooted her life to follow me across the country and has tirelessly labored to raise our children Nathanael and Aviana.

Finally, the providence of God brought this work to completion. In the words of the hymnist, "*Here by Thy great help I've come, and I hope by Thy good pleasure safely to arrive at home.*"

ABSTRACT

This dissertation consists of three essays on matching theory. The first two essays examine provide new cooperative solutions for two problems arising within matching markets in practice. The third contributes a theoretical analysis of the causes and effects of a market failure within the medical residency match.

Chapter 1 analyzes a matching market in which some agents have made prior commitments to each other. Typically, matching market models ignore prior commitments. I analyze two-sided matching markets with pre-existing binding agreements between market participants. In this model, a pair of participants bound to each other by a pre-existing agreement must agree to any action they take. To analyze their behavior, I propose a new solution concept, the *agreeable core*, consisting of the matches which cannot be renegotiated without violating the binding agreements. My main contribution is an algorithm that constructs such a match by a novel combination of the Deferred Acceptance and Top Trading Cycles algorithms. The algorithm is robust to various manipulations and has applications to numerous markets including the resident-to-hospital match, college admissions, school choice, and labor markets.

In Chapter 2, I turn to the problem of increasing the efficiency of student assignments in school choice subject to constraints imposed by policymakers. In school choice, policymakers consolidate a district's objectives for a school into a priority ordering over students. They then face a trade-off between respecting these priorities and assigning students to more-preferred schools. However, because priorities are the amalgamation of multiple policy goals, some may be more flexible than others. This paper introduces a model that distinguishes between two types of priority: a between-group priority that ranks groups of students and must be respected, and a within-group priority for efficiently allocating seats within each group. The solution I introduce, the *unified core*, integrates both types. I provide a two-stage algorithm, the DA-TTC, that implements the unified core and generalizes both the Deferred Acceptance and Top Trading Cycles algorithms. This approach provides a method for improving efficiency in school choice while honoring policymakers' objectives.

Chapter 3 introduces a a behavioral model of early matching within the context of the National Resident Matching Program, the system by which graduating medical students are matched to hospital residency programs. In my model, two hospitals compete to match to a continuum of doctors. Each hospital can make early offers or wait until the match is produced through the matching program. Some doctors have a behavioral preference to match early while others do not. I show that the less-desirable hospital benefits from the option to make early offers. My results provide a theoretical foundation for behavior widely documented within the medical ethics and graduate medical education literature and confirm beliefs commonly held by residency program directors.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

Chapter 1

MATCHING WITH PRE-EXISTING BINDING AGREEMENTS: THE AGREEABLE CORE

1.1 Introduction

In matching markets, pre-existing agreements are common. For example, when a student is admitted to a college through an Early Decision program, she commits to attend the college; she is bound to the college, and it now controls her right to participate in the regular admission cycle. When a professional athlete signs on to a sports team, that team purchases her right to sign on to other teams. Both examples include market participants—whether students or athletes—who have bound themselves to others. They are denied the right to find a new partner unless they are released from their agreements.

The standard model of matching markets ignores these interdependencies. It gives participants unrestricted rights to form new agreements, regardless of their earlier agreements. That is clearly unrealistic.

I propose a new model to solve this problem that makes it possible to analyze such markets. At bottom, my model requires that any action taken by one person must receive the approval of the person to whom she is bound. For example, a professional athlete can only seek another position with the approval of her team. Without the approval, she faces penalties for breaching her agreement. To manage these constraints, I introduce the concept of an *agreeable* group of participants. A group is agreeable if no member of the group is bound to someone outside of the group by a pre-existing agreement. In my example, an agreeable group only contains the athlete if it also contains her team, and vice versa. Critically, neither the athlete nor the team needs to be released from an agreement by anyone outside of the group.

My solution, the *agreeable core*, consists of the outcomes that cannot be renegotiated by any agreeable group. For a candidate outcome μ, the agreeable core considers every agreeable group and checks whether the group can achieve a better outcome for its members. If no such agreeable group exists, then the agreeable core includes μ. In the professional sports example, an agreeable group may contain some athletes and their respective teams. The agreeable core allows those athletes and teams to renegotiate their new contracts *before they are signed* so long as every member of the group benefits. For example, a team may condition the release of a player from their pre-existing contract on whether it secures a more-preferred replacement to sign a new contract. No person outside of this group can impede the negotiation because the group is agreeable.

Notably, I show that there are only two ways that agreements are dissolved in the agreeable core (Proposition 3). First, some agreements are dissolved unconditionally by both parties, which corresponds to the legal concept of *mutual separation*. Both parties are able to find better alternative partners regardless of the action the other takes. The outcome would be the same with or without the agreement. For example, this occurs when a team and an athlete jointly agree to cancel their contract; whether or not the contract initially existed is irrelevant to their future decisions. Second, the remaining agreements are only dissolved through "trades," which correspond to *multilateral agreements*. In a trade, two or more participants exchange the partners to whom they bound. For example, this occurs when two teams trade players. I show that at every outcome in the agreeable core, every dissolved agreement is of one of these forms.

The main contribution of this chapter is a two-stage algorithm, the *Propose-Exchange* algorithm (PE), which always produces an outcome in the agreeable core. The novel feature of the PE is how it leverages Proposition 3 to partition the participants according to how they dissolve their agreements. The PE uses a cascading process to determine which agreements can be dissolved unconditionally. Among participants who unconditionally dissolve their agreements, the PE then uses Deferred Acceptance algorithm (DA) from two-sided matching theory to assign a match in the core (Gale and Shapley, 1962). For the participants who are bound by some agreements, the PE allows participants to trade their partners as in the Top Trading Cycles algorithm (TTC) from the object allocation literature (Shapley and Scarf, 1974).

The PE algorithm can replace existing algorithms in markets that suffer from a lack of participation. Two prominent applications of market design—"The Match" conducted by the National Resident Matching Program (NRMP) and open-enrollment programs—are only incomplete markets. In the NRMP, some residents are offered posts outside of The Match. Prospective residents are forced to decide between accepting an early offer and participating in The Match. In open-enrollment programs, students can simultaneously hold offers from both the school district and private schools, leading to market congestion. Both problems arise because some agents accept offers through a decentralized system. The PE algorithm resolves this problem by integrating the centralized market with the decentralized market. Both the NRMP and open-enrollment programs use a version of the DA or TTC, so the PE can implement either. Incorporating the decentralized market is also straightforward: simply take the outcome of the decentralized market as the set of binding agreements. Because the PE (and the agreeable core in general) leaves no agent worse off than they are with their pre-existing binding agreements, the PE encourages agents to participate who normally would not. Participating in the PE is a weakly dominant strategy for agents who have created binding agreements in the decentralized market.

Second, the agreeable core provides an explainable solution in matching with minimum constraints. In some applications agents have minimum quotas that the designer must respect. In the context of matching residents to hospitals in Japan, the Japanese government seeks to guarantee that some regions receive a minimum number of residents (Kojima, Tamura and Yokoo, 2018). In public-school open-enrollment, the designer may have a preference for maintaining socioeconomic diversity at the schools; these are frequently written as minimum constraints assigned to different socioeconomic tiers; see Fragiadakis and Troyan (2017) for a discussion of these examples. In the United States Military Academy, cadets are assigned to positions subject to minimum manning constraints (Fragiadakis and Troyan, 2017). To accommodate minimum constraints in the agreeable core, the designer only needs to create artificial binding agreements. For example, if the designer adds an agreement between a hospital and a resident, then the hospital is guaranteed to match to (at least one) resident. The agreeable core provides a robust justification for the outcome: no other outcome could be reached without violating either agents' preferences or the minimum constraints.

The results of this chapter are grounded in the formalization of binding agreements as an *initial match* denoted μ_0. In this formalization, each participant is initially matched to at most one other participant. For concreteness, I label one side *workers* and the other side *firms*, and I refer to groups of agents as *coalitions*. The initial match rules out any participant being "double-booked"; otherwise, one participant may be bound to two others, creating ambiguity as to which agreement has precedence. Similarly, the initial match only allows for binding agreements to be two-way. For example, this formulation requires that if a student is bound to a college, then that college is bound to this student. There are ways to allow for some types of one-way agreements, but these require modifying participants' preferences. In this formalization, agreeable coalitions of agents are those which only include one participant if and only if her initial match is also a member of the coalition.

The agreeable core is an entirely different approach compared to previous research on matching with an initial match. Previous research emphasizes the properties of specific algorithms, such as strategy-proofness or efficiency (Combe and Schlegel, 2024; Combe, Tercieux and Terrier, 2022; Guillen and Kesten, 2012; Hafalir, Kojima and Yenmez, 2023; Hamada et al., 2017). In contrast, this chapter first develops a solution concept and then constructs an algorithm. The advantage is that the outcome in the agreeable core can be justified without relying upon the properties of the particular algorithm used to select it. My approach complements existing research because outcome-oriented solutions (which only require static definitions) are easier to explain and justify to stakeholders than algorithmic properties.[1] The trade-off is that the PE algorithm does not have the same incentive properties that are often baked into existing algorithms; however, I show that the PE satisfies a weakened version of strategy-proofness.

The Propose-Exchange algorithm is novel in its combination of both the Deferred Acceptance and Top Trading Cycles algorithms and has no similar predecessors. To the best of my knowledge, the only other algorithm capable of implementing both the DA and the TTC is the Stable Improvement Cycles algorithm of Abdulkadiroğlu (2011), which operates in a very different fashion. My use of the DA to divide the matching problem into two is entirely new and has promising applications in other markets with an initial match.

The rest of the chapter proceeds as follows. In Section 1.2 I motivate the agreeable core through an illustrative example. Section 1.3 presents the model. In Section 1.4 I present the proof of my main result, the Propose-Exchange algorithm that always produces a match in the agreeable core. Section 1.5 contains several results related to the manipulability of the Propose-Exchange algorithm. I defer a discussion of the related literature until Section 1.6, where I discuss how the agreeable core presents an alternative understanding of several economic applications.

[1]For example, the statement *your child is at the highest ranked school you listed where she is above the school's cutoff* is easier for parents to understand than some axiomatizations of the Deferred Acceptance, such as *we used the only algorithm that satisfies non-wastefulness, population monotonicity, weak Maskin monotonicity, and mutual best*; see (Morrill, 2013a). I emphasize that axiomatic approaches have significant value both in research and practice, but in some applications other justifications are more helpful.

1.2 A Motivating Example

In this section I introduce an example to illustrate my main definitions. This example highlights the limitations of the standard solution concept—the core—in matching markets where an initial match exists (the pre-existing binding agreements). By way of reminder, a match is in the *core* if no group of agents, known as a *blocking coalition*, can strictly improve their outcomes by forming an alternative match solely among themselves. The core does not account for the binding agreements and fails to improve upon the initial match.

Example 1. There are four workers (w_1, w_2, w_3, and w_4) and four firms (f_A, f_B, f_C, and f_D). All workers prefer f_A to f_B to f_C to f_D, except worker w_1 who swaps the order of f_A and f_B. All firms prefer w_3 to w_1 to w_2 to w_4, except for firm f_A who swaps the order of w_1 and w_2. Worker w_1 and firm f_A have a contract, as do worker w_2 and firm f_B, and also w_4 and firm f_D. Worker w_3 and firm f_C do not have a contract. In the language of my model, these contracts are the initial match μ_0 to which any agent can appeal (the set of pre-existing binding agreements which cannot be dissolved without the agreement both parties). Any outcome must guarantee that all agents are weakly better off than under the initial match. The initial match is essential because it limits the participants' flexibility in forming new contracts. The preferences are summarized in Figure 1.1, with the initial match circled.

Consider the core of this market. At any core outcome, worker w_3 must be matched to firm f_A because they mutually rank each other as best; otherwise, the coalition of $\{w_3, f_A\}$ blocks the match. However, this implies that either w_1 or w_2 is *not* matched to f_A or f_B and thus is worse-off than under μ_0. This a violation of the initial match μ_0. Therefore there is no match in the core that improves upon the initial match.

The failure of the core to provide a match that improves upon the initial match arises from the blocking coalitions allowed. Allowing every subset of agents to block is too permissive and ignores the initial match μ_0. The core is usually justified by arguing that agents in a blocking coalition could form contracts among only themselves, which allows for coalitions such as $\{w_3, f_A\}$.

Although the core is unsatisfactory, there are two Pareto improvements of the initial match, indicated in Figure 1.2. In both, w_1 is matched to f_B and w_2 is matched to f_A. The first Pareto improvement, labeled $\bar{\mu}$, matches w_3 to f_C and w_4 to f_D. Every blocking coalition contains $\{w_3, f_A\}$ or $\{w_3, f_B\}$ because no firm wants w_4

w_1	w_2	w_3	w_4
f_B	f_A	f_A	f_A
f_A	f_B	f_B	f_B
f_C	f_C	f_C	f_C
f_D	f_D	f_D	f_D
$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$

f_A	f_B	f_C	f_D
w_3	w_3	w_3	w_3
w_2	w_1	w_1	w_1
w_1	w_2	w_2	w_2
w_4	w_4	w_4	w_4
$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$

= initial match μ_0.

Figure 1.1: Preferences in Example 1, listed from most to least preferred, with $\varnothing$ indicating a preference for remaining unmatched; for example, this first column reads w_1 strictly prefers f_B to f_A to f_C to f_D to being unmatched. The circles indicate the initial match μ_0; for example, w_1 is initially matched (that is, under contract) to f_A.

more than its partner in $\bar{\mu}$, and both w_1 and w_2 are matched to their most-preferred partners. Consider $\{w_3, f_A\}$ first. Both w_3 and f_A prefer each other to the proposed match $\bar{\mu}$. But would worker w_1 release f_A from her contract to go and match to w_3? Worker w_1's release of f_A is contingent upon w_1 signing a contract with f_B, but f_B has the same constraint: w_2 must be induced to release f_B, which cannot be done without guaranteeing that w_2 matches to f_A. But the premise of this blocking coalition is that f_A will match to w_3 instead of w_2, so w_2 would not consent to this plan. In the language of my model, the coalition $\{w_3, f_A\}$ is not agreeable and thus cannot renegotiate its contracts; a similar argument follows for the coalition $\{w_3, f_B\}$.

The story is different for the other Pareto improvement, labeled $\dot{\mu}$. In this match, w_3 is matched to f_D and w_4 to f_C. Here, the coalition $\{w_3, f_C\}$ blocks the match. Because neither w_3 nor f_C is under contract, no agent can prevent them from renegotiating a new match. This coalition qualifies as agreeable. The agreeable core intuitively selects the first match but not the second.

To illustrate the mechanics of the Propose-Exchange algorithm, the following steps outline how tentative matches are proposed and refined until no further improve-

w_1	w_2	w_3	w_4
f_B	f_A	f_A	f_A
f_A	f_B	f_B	f_B
f_C	f_C	f_C	f_C
f_D	f_D	f_D	f_D
$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$

f_A	f_B	f_C	f_D
w_3	w_3	w_3	w_3
w_2	w_1	w_1	w_1
w_1	w_2	w_2	w_2
w_4	w_4	w_4	w_4
$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$

(a) First Pareto Improvement, $\bar{\mu}$

w_1	w_2	w_3	w_4
f_B	f_A	f_A	f_A
f_A	f_B	f_B	f_B
f_C	f_C	f_C	f_C
f_D	f_D	f_D	f_D
$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$

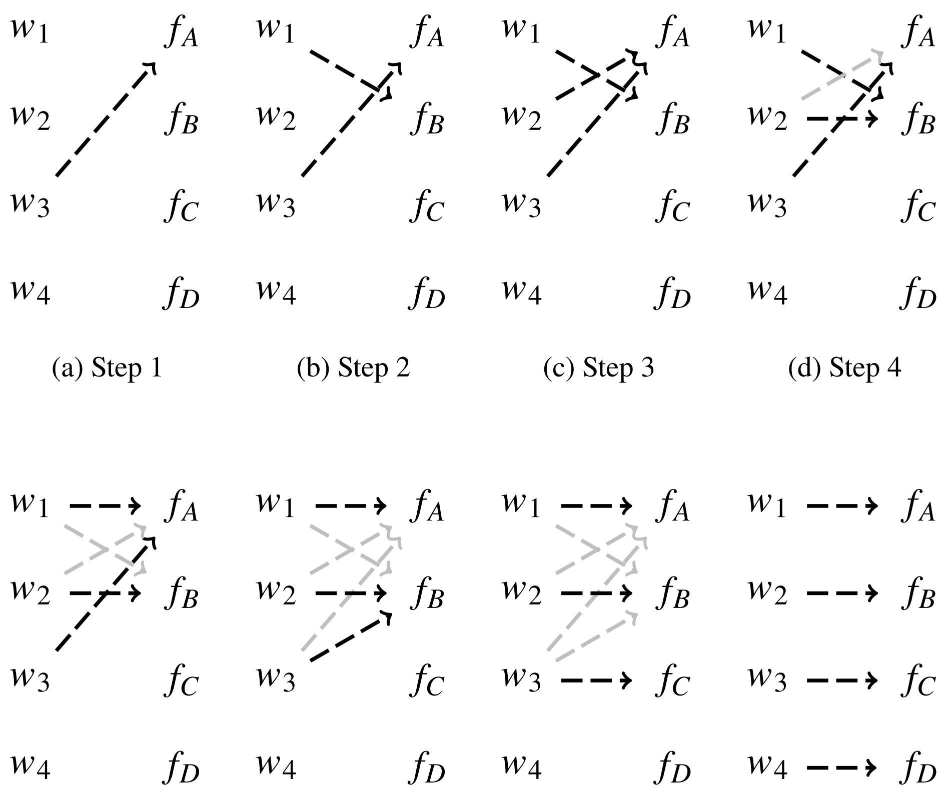
f_A	f_B	f_C	f_D
w_3	w_3	w_3	w_3
w_2	w_1	w_1	w_1
w_1	w_2	w_2	w_2
w_4	w_4	w_4	w_4
$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$

(b) Second Pareto improvement, $\dot{\mu}$

Figure 1.2: Pareto improvements of μ_0.

Note: throughout I use solid lines to denote the initial match and dashed lines to denote possible other matches

ments can be made. To compute the first Pareto improvement ($\bar{\mu}$), I leverage the Propose-Exchange algorithm. In this example the Propose stage takes worker w_3, who is initially unmatched, declares him "active." The Propose stage allows active workers to make proposals to their favorite firm which has not rejected them so far. In the first step, both w_3 proposes to f_A, who tentatively accepts him. Because f_A receives a proposal she prefers to her initial worker w_1, w_1 is now declared "active" as well. This guarantees that every firm weakly prefers the outcome of the Propose stage to the initial match μ_0 because she only releases her initial worker once she has a more-preferred tentative match. In the second step, w_1 proposes to f_B, who tentatively accepts w_1. Again, because f_B receives a proposal she prefers to her initial worker w_2, w_2 is now declared "active." In the third step, w_2 proposes to f_A, who rejects him. In the fourth step, w_2 proposes to his initial firm f_B; the Propose



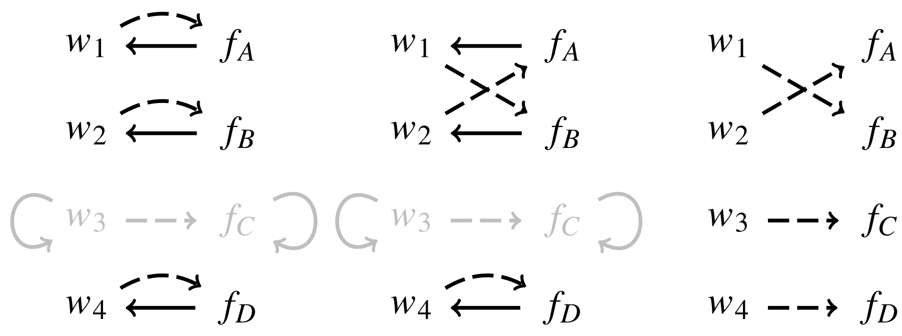
(a) Step 1 (b) Step 2 (c) Step 3 (d) Step 4

(e) Step 5 (f) Step 6 (g) Step 7 (h) Output μ_1

Figure 1.3: A visualization of the steps of the Propose stage. The black dashed lines indicate active proposals, and the light gray dashed lines indicate rejected proposals. Note that w_1 and w_2 only make proposals *after* f_A and f_B have each received a proposal, respectively. Worker w_4 never makes a proposal because f_D never receives a proposal.

stage requires that f_B accept w_2's proposal and reject w_1. This guarantees that every worker weakly prefers the outcome of the Propose stage to the initial match μ_0. Continuing in this fashion, w_1 proposes to his initial firm f_1, which causes w_3 to be rejected. Worker w_3 proposes to f_B and is rejected, and then to f_C and is tentatively accepted. These steps are visualized in Figure 1.3. The outcome of the Propose stage is denoted μ_1 and is depicted in panel (h). However, an agreeable blocking coalition still exists because workers w_1 and w_2 would prefer to exchange their initial firms f_A and f_B, and these firms also would prefer the exchange.

The Exchange stage modifies the outcome of the Propose stage to remove the agreeable blocking coalition $\{w_1, w_2, f_A, f_B\}$ of μ_1. In the Exchange stage, workers



(a) μ_0 (solid), μ_1 (dashed) (b) Step 1 (c) Output μ_2

Figure 1.4: A visualization of the steps of the Exchange stage. Agents w_3 and f_C are excluded because their μ_0- and μ_1-partners differ. There is only one step because each active agent is in a cycle in Step 1.

w_1, w_2, and w_4 and firms f_A, f_B, and f_D are active because they have not improved their initial match through the Propose stage, while w_3 and f_3 are inactive. In the first step, w_1 points to f_B because f_B is w_1's most-preferred firm. Again, w_2 points to f_A because f_A is w_2's most-preferred active firm. Worker w_4 would point to either f_A or f_B, but neither prefer him to their initial match, so w_4 is only allowed to point to his own firm f_D. This will guarantee that every firm weakly prefers the outcome of the Exchange stage to the initial match μ_0. Each active firm points to her initial worker. The cycle $w_1 \to f_B \to w_2 \to f_A \to w_1$ forms, and w_1 and w_2 are both permanently matched to the firms they point at. The cycle $w_4 \to f_D$ also forms, and w_4 is permanently matched to f_D. The output is μ_2, which is depicted in Figure 1.4. As expected, μ_2 is the first Pareto improvement that was discussed, the unique element of the agreeable core.

The Propose-Exchange algorithm involves both a "free market" phase in the Propose stage (but with participation restrictions on w_1, w_2, and w_4) as well as a "trading" phase in which could w_1, w_2, and w_4 exchanged their firm. The match at every stage of the Propose-Exchange algorithm is an improvement of the initial match.

1.3 Model

In this section I present a one-to-one matching model. Although many of my applications are many-to-one (e.g. many students match to one school), I defer a discussion of the nuances until Section 1.6. In most examples, the many-to-one case is a simple extension of the one-to-one model. Below I introduce the elements of a *matching problem*, which is a tuple $(W, F, \succ, \mu_0)$ consisting of workers, firms, a profile of preferences, and an initial match.

There is a set of workers W and the set of firms F, and the union of both is the set of agents $A \equiv W \cup F$. For clarity of exposition I use masculine pronouns for workers and feminine pronouns for firms. Every worker $w \in W$ has preference $\succsim_w$ over $F \cup \{w\}$ and every $f \in F$ has a preference $\succsim_f$ over $W \cup \{f\}$. A preference for oneself is a preference to be unmatched: if a prefers a to b this means that a prefers to remain unmatched than to match to b. Throughout I assume that $\succsim_a$ is complete, reflexive, transitive, and anti-symmetric, that is, that a can rank partners from most to least preferred with no ties.

A match is a function μ that takes in an agent a and returns the agent $\mu(a)$ that he or she is matched, where $a = \mu(a)$ means that a is unmatched. Formally, *match* is a function $\mu : A \longrightarrow A$ such that:

1. if $w \in W$ then $\mu(w) \in F \cup \{w\}$; and
2. if $f \in F$ then $\mu(f) \in W \cup \{f\}$; and
3. $\mu(\mu(a)) = a$.

The first two require that the match is two-sided: every worker matches to a firm (or is unmatched) and every firm matches to a worker (or is unmatched). The third requires that every agent is matched to the agent matched to him or her. If $\mu(a) = a$ then a is *μ-unmatched*; otherwise, $\mu(a)$ is the *μ-partner* of a (or possibly μ-firm or μ-worker). I write $\mu \succsim_X \mu'$ to mean $\mu(x) \succsim_x \mu'(x)$ for all $x \in X$.

There is an *initial* match μ_0. The initial match limits the set of matches I consider to the set of matches I consider to those satisfying the following:

Definition 1. Match μ is *individually rational* if $\mu \succsim_A \mu_0$.

The interpretation is that if an agent prefers their initial match μ_0 to the proposed match μ, they retain the right to demand μ_0, as it represents an enforceable agreement.

The Core

Here I formally introduce the core, which is the set of all individually rational matches not blocked by any coalition of agents. A coalition blocks a match if it can collectively form a match within the coalition that everyone weakly prefers to the current match. Formally, a *coalition* $C \subseteq A$ is a nonempty[2] subset of agents who may form a match among themselves. Let $\mu(C) \equiv \{\mu(a) : a \in C\}$. Note that if $\mu(C) \subseteq C$, then $\mu(C) = C$. If a coalition weakly prefers a match μ' to μ and μ' only matches agents in C to agents in C, then C may block μ; formally,

Definition 2. Coalition C *blocks* μ *through* ν if $\nu \succsim_C \mu$, $\nu(a) \succ_a \mu(a)$ for at least one $a \in C$, and $\nu(C) = C$.

The *core* is the set of all individually rational matches not blocked by any coalition through any match.[3]

The Agreeable Core

Example 1 demonstrates that the core may be empty. The nonexistence of a match that is both individually rational and unblocked by every coalition of agents motivates restricting either the matches a coalition can block through or the coalitions considered. The choice is nontrivial and hinges upon the interpretation of the initial match.

If the matches that a coalition can block through are restricted, then the natural requirement is that any coalition can block but only through an individually rational match μ. The interpretation is that the initial match is inviolable *ex post*. In order to block a match, a coalition needs only to suggest an individually rational match; as long as all agents are weakly better off than at μ_0, no agent can complain about his or her partner. However, it is easy to construct examples where this solution is empty.

The alternative is to restrict the set of coalitions but not the matches they can block through. The interpretation is that the initial match is not only inviolable *ex post* but also that any new contract formed by an agent requires the *ex ante* approval of his or her μ_0-partner. I consider only coalitions meeting the following criterion:

Definition 3. A coalition C is *agreeable* if $\mu_0(C) = C$.

A coalition C is agreeable if any contract in μ_0 does not contain both an agent in C and an agent not in C. By restricting my attention to agreeable coalitions, I

[2] Coalitions throughout the chapter are assumed to be nonempty. For ease of exposition this quantifier will not be listed.

[3] Formally, this is the *strong core* because I consider all *weak* blocks (allowing some coalition members to be indifferent between μ' and μ). In two-sided matching without indifferences all weak blocks are strong blocks. Because the coalitions I will consider later will usually contain agents who do not change partners, I use the strong core as it is smaller.

require that every agent in a blocking coalition of μ guarantees his or her μ_0-partner a weakly better partner at match μ' than at μ. To guarantee such an improvement, the μ_0-partner's partner at μ' must also be included in the coalition, which implies that the μ_0-partner's μ'-partner must also be included in the coalition, and so on. Definition 4 formalizes this idea.

Definition 4. The *agreeable core* is the set of individually rational matches not blocked by any agreeable coalition.

The agreeable core puts a strong requirement on blocking coalitions: every agent in the coalition *and their μ_0-partners* must be made weakly better off. My interpretation is that if some agent a is harmed by a block and his or her μ_0-partner is in the blocking coalition, then a can veto the block by refusing to dissolve the initial contract. The important nuance is that the harmed agent can veto μ' even if he or she prefers μ' to μ_0.

The veto power inherent in the agreeable core allows one member of a initial match to dictate the matches his or her partner can form. The picture to have in mind is both agents in a initial match simultaneously searching for better matches. They both agree to cancel their initial match *simultaneous to both confirming new partners.* Because the match of one partner influences who is willing to match with the other, both must agree not only to cancel their initial match but also approve of the other's new match. By only considering agreeable coalitions, I allow agents to veto a blocking coalition before the coalition acts.

I find the following justification for the agreeable core helpful in explaining the agreeable core and how I allow agents veto blocking coalitions *ex ante*. For a given initial match μ_0, agents are considering forming the individually rational match μ. Before μ is realized among the agents (say, before the agents cancel their initial agreements and form the μ agreements), a coalition considers enforcing some match μ' among themselves. If some agent a is in the coalition but $\mu_0(a)$ is not in the coalition, then $\mu_0(a)$ may refuse to permit a to form μ' unless $\mu_0(a)$ is certain he or she will prefer μ' to μ. Hence, $\mu_0(\mu)$ must also be in the coalition.

Perhaps surprisingly, the set of matches not blocked by any agreeable coalition is not a subset of the individually rational matches. My definition of blocking coalition does not allow an agent to demand μ_0, and hence the restriction to individually rational matches is substantive. For a simple example, restrict Example 1 to just contractor 1 and city A. The match $\mu(1) = 1$ and $\mu(A) = A$ is not blocked by any coalition but does not Pareto improve μ_0.

I devote Section 1.4 to developing the machinery to prove my main result, namely, that the agreeable core is never empty. In the remainder of this section I briefly touch on several aspects of the agreeable core that do not require my more involved techniques. Section 1.3 shows that the agreeable core is always Pareto efficient, and conversely if μ_0 is Pareto efficient then $\{\mu_0\}$ is the agreeable core. As alluded to in the introduction, my model features several connections with both the classical model of stability (Gale and Shapley, 1962) and more recent models of reassignment (Combe, Tercieux and Terrier, 2022; Pereyra, 2013). In Section 1.3 and Section 1.3 I develop these connections; as an expository device and a prelude to my algorithm, I highlight the two leading algorithms in two-sided matching—the Deferred Acceptance and the Top Trading Cycles algorithms—and their adaptations used in the literature to guarantee individual rationality.

Efficiency

In this subsection I investigate the efficiency of the agreeable core. My first observation is that no match in the agreeable core is Pareto dominated:[4] if ν Pareto dominates μ, then the grand coalition A (which is always agreeable) blocks μ through ν. My second observations is a kind of a converse: if μ_0 is not Pareto dominated, then μ_0 is in the agreeable core. To see this, suppose (toward a contradiction) that some agreeable coalition C blocks μ_0 through μ. But then because $\mu_0(C) = \mu(C) = C$, I can define μ' that agrees with μ for agents in C and agrees with μ_0 everywhere else. But μ' then Pareto dominates μ_0, a contradiction to the supposition that μ_0 is Pareto efficient.

Remark 1. *Every μ in the agreeable core is Pareto efficient.*[5] *Moreover, μ_0 is Pareto efficient if and only if the μ_0 is the unique element of the agreeable core.*

Remark 1 assures us that the agreeable core satisfies the most common efficiency standard.

---

[4] I say that ν *Pareto dominates* μ if every agent weakly prefers ν to μ and at least one agent strictly prefers ν to μ.

[5] If μ is not Pareto dominated by any ν, then μ is *Pareto efficient*.

Connection to Stability

In this subsection I discuss the parallels between the agreeable core and the classic theory of stability introduced by Gale and Shapley (1962). The models are the same except that the classical model does not include an initial match in the primitives. This connection allows me to leverage a significant tool from two-sided stability, the Deferred Acceptance algorithm (DA), in my analysis

In the classic model, a *blocking pair* of a match is any worker and firm pair such that both prefer each other to their match. A match is *stable* if all agents prefer their match to being unmatched and there are no blocking pairs of the match. It is well-known (Roth and Sotomayor, 1990) that the set of stable matches is the core that I defined previously. My definition of the agreeable core guarantees that if $\mu_0(a) = a$ for all $a \in A$, then the agreeable core corresponds to the core because every coalition is agreeable. Therefore stability is the special case of the agreeable core when μ_0 leaves all agents unmatched.

Gale and Shapley (1962) gives an efficient algorithm for constructing a stable match: the Deferred Acceptance algorithm (Algorithm 1). Initially, the DA "activates" every worker and designates every agent as "currently unmatched." At every step of the DA, some active worker matched proposes to the firm he prefers the most among those he has not proposed to yet (if he would rather be unmatched, he is matched to himself and deactivated). Every firm then reviews the proposals she receives and her current match and rejects all but her most preferred proposal or match. The process continues until no more workers are matched active.

Algorithm 1 Deferred Acceptance (DA) algorithm

Notation: when I write $\mu^{\mathrm{DA}}(a) \leftarrow w$, I mean that a is matched to w and w is deactivated. If another worker w' was matched to a, then a rejects w', w' is matched to himself, and w' is activated.

set $\mu^{\mathrm{DA}}(f) \leftarrow f$ for all $f \in F$.
activate every worker.
while some worker w is activated **do**
    w proposes to his most-preferred firm f that he has not yet proposed to; if he would rather be unmatched, instead he proposes to himself and is deactivated, and we set $\mu^{\mathrm{DA}}(w) \leftarrow w$.
    if f prefers w to $\mu^{\mathrm{DA}}(f)$ then set $\mu^{\mathrm{DA}}(f) \leftarrow w$.
    else f rejects w.
end while
return μ^{DA}

Although guaranteed to produce a match unblocked by any coalition, the DA fails to satisfy individual rationality (see Pereyra (2013) and Combe, Tercieux and Terrier (2022)). There are two ways in which individual rationality can fail. First, a worker may strictly prefer his μ_0-partner to his match. Pereyra (2013) resolves this issue by requiring that each firm accepts her μ_0-partner if he proposes to her. This modification guarantees that workers find the outcome individually rational because no worker proposes to a less preferred firm without being rejected by his μ_0-partner.

In my setting firms also have individual rationality constraints. The DA fails to accommodate these because a worker makes proposals (and may be matched to another firm) even though his μ_0-firm has not received a proposal she prefers to the worker. I will see in Section 1.4 how to resolve this tension by limiting which workers can propose.

Connection to Reassignment

In this subsection I highlight the connection between the agreeable core and the standard model of reassignment. Recent research in reassignment seeks to find a match through a strategyproof mechanism that is both individually rational and maximizes some objective function (see (Combe, Tercieux and Terrier, 2022; Dur and Ünver, 2019) for two such examples). Because the agreeable core is motivated with first principles (the core) rather than with an objective in mind (obtaining a strategyproof mechanism), there are substantial differences in definitions and results. However, both approaches employ the same method: the Top Trading Cycles algorithm (TTC).

The TTC finds a match such that no coalition of workers can reallocate their μ_0-firms among themselves and improve their matches. The TTC starts with every worker and firm "active." At every step, every active firm points at the worker she is initially matched to, and every active worker points at his favorite active firm. At every step a cycle must form. The TTC assigns each worker in the cycle to the firm he points at, and then the agents in the cycle become inactive. The process terminates when no agents are active.

I define the TTC in Algorithm 2.

If some agents are matched by μ_0, then the TTC may not be individually rational. To accommodate this, Combe, Tercieux and Terrier (2022) and Combe (2023) make the following two modifications. First, a firm must point to her μ_0-worker so long as he is active. This guarantees that $\mu^{\text{TTC}} \gtrsim_W \mu_0$. Second, no worker may point

Algorithm 2 Top Trading Cycles (TTC) algorithm

set $\mu^{\text{TTC}}(a) = a$ for all a.
every agent is activated.
while at least one agent is active **do**
    every active worker points to his most-preferred of the active firms.
    every active firm points to her most preferred of the active workers.
    choose an arbitrary cycle $(w_1, f_2, \ldots w_{2k-1} \equiv w_1, f_{2k} \equiv f_2)$ such that every agent points to the next agent in the cycle.
    all agents in the cycle are deactivated.
    match every w_k to f_{k+1}.
end while
return μ^{TTC}

to a firm if that firm prefers her μ_0-partner to the worker. This guarantees that $\mu^{\text{TTC}} \gtrsim_W \mu_0$.

In my setting, however, these modifications are not enough. As I saw in Section 1.3, the agreeable core equals the set of stable matches when all agents are μ_0-unmatched. At least in this case firms must be given power to decide between the workers pointing to them, as in the DA. In Section 1.4 I incorporate this by limiting which workers and firms participate in the TTC.

1.4 A Proof of Existence: The Propose-Exchange Algorithm

In this section I present a computationally efficient and economically meaningful algorithm that always produces a match μ_2 (defined through this section) in the agreeable core. My algorithm is the *Propose-Exchange* algorithm (PE) and is composed of two stages. The Propose stage resembles the Deferred Acceptance algorithm and eliminates any block by a coalition that either includes an agent who is unmatched in the initial match or who becomes unmatched by the block. The Exchange stage resembles the Top Trading Cycles algorithm and eliminates all blocks that involve reshuffling initial partners among themselves. For readers unfamiliar with the Deferred Acceptance and the Top Trading Cycles algorithms, I refer the reader to Section 1.3 and Section 1.3, respectively.

The PE directly implies that the agreeable core exists and provides some insight into its structure. My main result is the following:

Theorem 1. *μ_2 is in the agreeable core.*

The proof (and definition of μ_2) occupies the remainder of this section. I first introduce a particular directed graph representation of the matching problem in Section 1.4, then introduce the Propose stage in Section 1.4, and finally the Exchange stage in Section 1.4. I conclude this section by noting how the introduction of initial matches creates additional complexity in analyzing the structure of the agreeable core. All omitted proofs are contained in Section 1.A.

A Graph-Theoretic Depiction

Despite my parsimonious definition of the agreeable core, so far testing whether μ is in the agreeable core requires checking whether any coalition can block μ through any μ', which is only feasible in small examples. My main result from this subsection is a characterization of blocking coalitions in terms of paths in a directed graph, which is computationally efficient. I use the language of graph theory to formalize my ideas.

A *digraph* G is a pair (V, E) where V is a set of *vertices* and E is a set of *ordered* pairs of vertices called (directed) *edges*, possibly including an edge from a vertex to itself, called a *loop*. The one nuance to my construction is that I allow for loops to be repeated once in E; formally, E is a *multiset*, but this will not cause any confusion.

I consider digraphs where the vertices are agents, and the edges represent matches. Edges going from F to W (and loops) are drawn from μ_0, while the edges going from W to F (and possibly repeated loops) are drawn from μ and any blocking pairs of μ. I abuse notation and write μ_0 for both the function and for the set of ordered pairs:

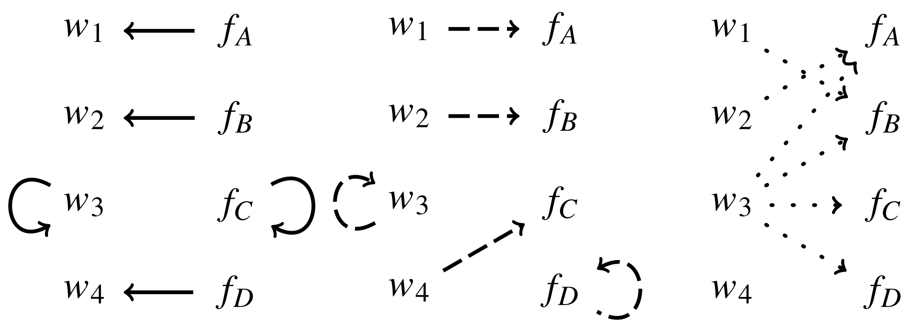
$$\mu_0 = \{(f, w) : \mu_0(f) = w\} \cup \{(a, a) : \mu_0(a) = a\}$$
$$\mu = \{(w, f) : \mu(w) = f\} \cup \{(a, a) : \mu(a) = a\}.$$

It is critical to understand that μ_0 and μ go in *opposite* directions (except for any loops). I always follow the convention that edges from the initial match travel from F to W, so although the matches μ_0 or μ may change, from context the direction of the edges is always clear. To include the blocking pairs, I define

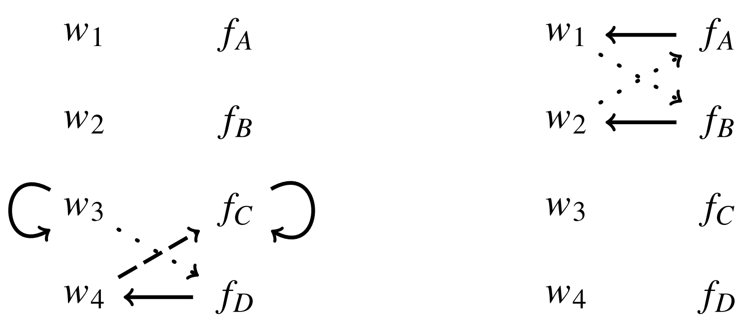
$$I(\mu) = \{(w, f) : f \succ_w \mu(w) \text{ and } \succ_f \mu(f)\} \cup \{(a, a) : a \succ_w \mu(a)\}.$$

My main digraph of interest is $(A, \mu_0 \cup \mu \cup I(\mu))$. That is, the vertices are agents, the first set of edges connects initial partners, and the second set of edges connects all pairs that weakly prefer each other over their μ-partners.

Figure 1.5 depicts the three types of edges using the set-up of Example 1 and a match μ that modifies the initial match μ_0 by leaving f_D unmatched and matching



(a) Initial Match μ_0 (b) Proposed Match μ (c) Blocking Pairs of μ



(d) Acyclic Blocking Path (e) Cyclic Blocking Path

Figure 1.5: The blocking digraph $(A, \mu_0 \cup \mu \cup I(\mu))$ is the union of the digraphs in subfigures (a), (b), and (c). Subfigures (d) and (e) depict the two kinds of blocking paths.

w_4 to f_C. Subfigure (a) includes the edges from μ_0, which are either loops (in the case of w_3 and f_C) or point from F to W. Subfigure (b) includes the edges from μ, which point from W to F. Subfigure (c) includes the blocking pairs of μ, which point from W to F.

A (simple) *path* in (V, E) is a vector of edges $P = (e_1, \ldots, e_n)$ such that the second coordinate of e_k equals the first coordinate of e_{k+1} for $1 \leq k < n$ and no vertex appears in more than two edges. Recall that a loop may appear twice (in both μ_0 and $\mu \cup I(\mu)$) so it is possible for path to consist of exactly two loops. I say a *vertex is in a path* if the path contains an edge that contains the vertex. I sometimes abuse notation and write P for the vertices in P.

A path P is *complete* if every vertex contained in the path is contained in exactly two edges of the path. A path is *alternating* if it no two consecutive edges (including loops) alternate between μ_0 and $\mu \cup I(\mu)$.[6] Two complete and alternating paths are depicted in subfigures (d) and (e) of Figure 1.5. For an arbitrary complete and alternating path P in $(A, \mu_0 \cup \mu \cup I(\mu))$, I define $\mu^P(a)$ as follows:

- if (w, f) is in P, then $\mu^P(w) = f$.
- if a is *not* in P but $\mu(a)$ is in P, then $\mu^P(a) = a$;
- if a is *not* in P and $\mu(a)$ is not in P, then $\mu^P(a) = \mu(a)$.

That is, μ^P matches $a \in P$ to the agent whom a shares an edge from $\mu \cup I(\mu)$ in P with; other matches are left unchanged where possible. By Lemma 1.A.1 in the appendix, every agent in P is contained in one edge from μ_0 and one edge is from $\mu \cup I(\mu)$, so μ^P is well defined and $\mu^P(P) = P$.

My main result of this subsection is that a path that is complete, alternating, and contains an edge from $I(\mu)$ corresponds to an agreeable blocking coalition in $(A, \mu_0 \cup \mu \cup I(\mu))$. I formalize this as follows:

Definition 5. Path P is a *blocking path of* μ if P is a complete and alternating path in $(A, \mu_0 \cup \mu \cup I(\mu))$ that contains at least one edge from $I(\mu)$.

A blocking path of μ is aptly named as it corresponds to a blocking coalition of μ.

Proposition 1. *An individually rational match μ is in the agreeable core if and only if μ admits no blocking paths. Moreover, if P is a blocking path of μ then P blocks μ through μ^P.*

Proposition 1 provides a test that is linear in the number of edges to see if μ is in the agreeable core.[7]

The Propose-Exchange algorithm is built on a partition of paths between those that form cycles and those that do not.

Definition 6. Let $P = (e_1, \ldots, e_n)$. If the first coordinate of e_1 is the second coordinate of e_n, then P is *cyclic*; otherwise, P is *acyclic*.

---

[6] Although the directed nature of the digraph makes most paths alternating, by formally requiring that a path is alternating I rule out the case that (w, f) and (f, f) may both be from $\mu \cup I(\mu)$.

[7] A depth first search initiated from every edge in $I(\mu)$ is sufficient.

As the name suggests, cyclic paths start with an agent and then return to that agent. In $(A, \mu_0 \cup \mu \cup I(\mu))$, a cyclic, complete, and alternating path corresponds to agents (who are μ_0-matched) trading their μ_0-partners among themselves. Acyclic paths that are also complete and alternating start with a loop and end with a loop, forming a line in the digraph. See subfigures (d) and (e) of Figure 1.5 for example cyclic and acyclic blocking paths. In $(A, \mu_0 \cup \mu \cup I(\mu))$, an acyclic, complete, and alternating path corresponds to agents trading their μ_0-firms among themselves, except that two agents are unmatched by one or both sets of edges. The Propose-Exchange algorithm works by first producing a match μ_1 that admits no acyclic blocking paths, then finding a series of Pareto improvements of μ_1 to produce a match μ_2 that has no cyclic blocking paths.

The Propose Stage

The first stage of my algorithm outputs a match μ_1 by systematically removing all acyclic blocking paths from $(A, \mu_0 \cup \mu \cup I(\mu))$. An acyclic blocking path P in $(A, \mu_0 \cup \mu \cup I(\mu))$ corresponds to a series of trades, but the agents at either end of the path are either μ_0-unmatched or μ^P-unmatched. These may be thought of as a cycle that includes the "unmatched" agent.

The Propose algorithm is a variation of the Deferred Acceptance algorithm. The DA is designed for markets where all agents are unmatched under μ_0 and is defined in Algorithm 1. I noted in Section 1.3 that the DA may fail individual rationality for both workers and firms. I provide guarantees to the agents in the Propose algorithm by only allowing a worker to make a proposal once his μ_0-firm has received a more preferred proposal and by requiring that a firm accept any proposal from her μ_0-worker. These adjustments, shown in italics, are essential to the success of the Propose stage. The Propose stage algorithm is defined in Algorithm 3. By construction, μ_1 is individually rational. If w strictly prefers μ_0 to μ_1, then w would have proposed to μ_0 (and not been rejected). Again, if $\mu_0(f)$ is matched by μ_1 to a firm other than f, then f received a proposal she prefers to $\mu_0(f)$ and hence she prefers μ_1 to μ_0.

I then show that at the end of the Propose algorithm, no blocking path of μ_1 is acyclic.

Lemma 1. *μ_1 admits no acyclic blocking paths.*

My proof leverages that an acyclic blocking path P in $(A, \mu_0 \cup \mu \cup I(\mu))$ always

Algorithm 3 Propose Stage algorithm

Notation: when I write $\mu_1(a) \leftarrow w$, I mean that a is matched to w and w is deactivated. If another worker w' was matched to a, then a rejects w', w' is matched to himself, and w' is activated.

set $\mu_1 \leftarrow \mu_0$
activate every worker.
if *w's μ_0-firm prefers w to being unmatched,* ***then deactivate*** w.
while some worker w is active **do**
w proposes to his most-preferred firm f that he has not yet proposed to; if he would rather be unmatched, instead he proposes to himself and is deactivated, and we set $\mu_1(w) \leftarrow w$.
if *f is w's μ_0-partner,* ***then*** *set $\mu_1(f) \leftarrow w$ and have f reject all future proposals.*
else if f prefers w to $\mu_1(f)$ *and to being unmatched*, then set $\mu_1(f) \leftarrow w$.
else f rejects w.
end while
return μ_1

begins with either a worker who is μ_0-unmatched and hence proposes or a firm who is μ^P-unmatched (and hence her μ_0-worker starts out active). Because the start and finish of the path are connected by workers who (weakly) prefer the firm they receive in the block, I can show that every worker in the path must have had the opportunity to propose. I then show that the path must terminate with either a worker who is μ_0-matched or a firm who is μ_0-unmatched, neither of which would reject the proposal made through the path. I conclude by showing that every firm accepts the proposal from her μ^P-partner, which contradicts that $\mu \neq \mu^P$.

The **while** step admits ambiguity because which worker is selected to propose is not specified. I show in Proposition 2 that the order in which workers are selected is irrelevant.

Proposition 2. *The output of the Propose stage is independent of the order the workers are called to propose in.*

The Exchange Stage

In the second stage of the algorithm, I eliminate all cyclic blocking paths. I do this by allowing agents to trade their initial agreements. Cyclic blocking paths in $(A, \mu_0 \cup \mu \cup I(\mu))$ correspond to workers and their μ_0-firms rearranging their initial matches among themselves. No agent in a cyclic path is unmatched by either μ or

μ_0. A cyclic blocking path represents an inefficient allocation for C: the coalition could have rearranged their initial matches among themselves and obtained a better match.

The Exchange algorithm is an adaptation of the Top Trading Cycles algorithm to find these cycles and remove them. The difficulty with using solely the TTC in my setting is that the TTC does not give firms the ability to select *between* workers. Although firm's preferences limit the set of acceptable workers, which worker is matched to the firm ultimately depends on the worker the firm is required to point at. If only some workers or firms are matched by μ_0, then the firm's lack of choice can lead to violations of the agreeable core. I resolve this by only applying the TTC to workers and firms who did not both find better partners through the Propose algorithm, with my addition indicated in italics. This modification guarantees that the Exchange stage is a Pareto improvement of μ_1; by selecting a Pareto improvement, I do not create any new acyclic blocking paths in the blocking digraph. The Exchange algorithm is defined in Algorithm 4.

Algorithm 4 Exchange Stage algorithm

set $\mu_2(a) = \mu_1(a)$ *for all* a.
every w *such that* $\mu_1(w) = \mu_0(w)$ *is activated with* $\mu_0(w)$.
while at least one worker is active **do**
    every active worker points to his most-preferred of the active firms *who prefer him to her* μ_0*-worker*.
    every active firm points *to her* μ_0*-worker.*
    choose an arbitrary cycle $(w_1, f_2, \ldots w_{2k-1} \equiv w_1, f_{2k} \equiv f_2)$ such that every agent points to the next agent in the cycle.
    all agents in the cycle sit down.
    match every w_k to f_{k+1}.
end while
return μ_2

My first observation is that the Exchange algorithm makes no agents worse off than under μ_1. Workers only point to firms they prefer to μ_0, and by my simplification of workers' preferences, firms can only be pointed at by workers they prefer to μ_0. The result is that at the end of the Exchange algorithm, μ_2 admits no cyclic blocking paths.

Lemma 2. *μ_2 admits no cyclic blocking paths.*

My proof leverages that if w strictly prefers f to $\mu_2(w)$, then f must sit down at least one step *before* w. A cyclic blocking path then implies that the firms in the path sit down on average strictly before the workers in the path sit down. However, because every worker's μ_0-firm is in the path and they sit down in the same step, it must be that the firms in the path sit down on average in the same step as the workers in the path sit down. This contradiction rules out cyclic blocking paths.

Existence

I am now ready to prove that μ_2 is in the agreeable core.

Proof of Theorem 1: Suppose (toward a contradiction) that μ_2 is not in the agreeable core. Then by Proposition 1 the digraph $(A, \mu_0 \cup \mu_2 \cup I(\mu_2))$ contains a blocking path P. By Lemma 2, P is acyclic. But P is also blocking path in $(A, \mu_0 \cup \mu_1 \cup I(\mu_1))$ because $\mu_2 \cup I(\mu_2) \subseteq \mu_1 \cup I(\mu_1)$ and $I(\mu_2) \subseteq I(\mu_1)$. By Lemma 1, P is not acyclic. This is a contradiction, which proves the claim. □

The importance of the Propose-Exchange algorithm in my proof cannot be understated. However, the algorithm has practical implications because it is also computationally efficient. The Propose stage runs in polynomial time because each worker can make at most $|F| + 1$ proposals. Similarly, one cycle is removed in every iteration of the Exchange stage, and at most $|F|$ cycles can be removed. An efficient algorithm is necessary for implementing the agreeable core in applications.

Structure

In this subsection, I highlight the difficulty in characterizing the underlying structure of the agreeable core and how it relates to other classes of algorithms commonly used to compute core outcomes. Although the set of stable matches has a well-understood structure which I summarize in the following paragraph, the agreeable core is not as tame. The hurdle in the analysis comes from the Exchange stage. To the best of my knowledge, there are no results from the literature that apply to the agreeable core when every agent is μ_0-matched.

I briefly summarize the main structural results on the set of stable matches. First, a *lattice* is a partially ordered set $(L, \geq)$ such that any two elements of L have a unique *least upper bound*, called the *join* of x and y, and a unique *greatest lower bound*, called the *meet* of x and y. That is, there is a unique $x \vee y$ such that if $z \geq x$ and $z \geq y$ then $z \geq x \vee y$, and there is a unique $x \wedge y$ such that if $x \geq z$ and $y \geq x$ then $x \wedge y \geq z$. A key result in two-sided matching is that the set of stable matches forms

w_1	w_2	w_3	f_A	f_B	f_C
f_C	f_A	f_A	w_2	w_1	w_1
f_B	f_B	f_C	w_3	w_2	w_3
f_A	$\varnothing$	$\varnothing$	w_1	$\varnothing$	$\varnothing$
$\varnothing$	f_C	f_C	$\varnothing$	w_3	w_2

(a) Initial match—μ_0

w_1	w_2	w_3	f_A	f_B	f_C
f_C	f_A	f_A	w_2	w_1	w_1
f_B	f_B	f_C	w_3	w_2	w_3
f_A	$\varnothing$	$\varnothing$	1	$\varnothing$	$\varnothing$
$\varnothing$	f_C	f_C	$\varnothing$	w_3	w_2

(b) w_2, f_A, and f_B's preferred match—$\bar{\mu}$

w_1	w_2	w_3	f_A	f_B	f_C
f_C	f_A	f_A	w_2	w_1	w_1
f_B	f_B	f_C	w_3	w_2	w_3
f_A	$\varnothing$	$\varnothing$	w_1	$\varnothing$	$\varnothing$
$\varnothing$	f_C	f_C	$\varnothing$	w_3	w_2

(c) w_1, w_3, and f_C's preferred match—$\dot{\mu}$

Figure 1.6: An example showing that the outcomes in the agreeable core cannot be ordered by $\succsim_W$.

a lattice with the partial order $\succsim_W$.[8] The join of two matches μ and ν is the match that gives every worker w his more preferred partner from $\{\mu(w), \nu(w)\}$ and every f her less preferred partner from $\{\mu(f), \nu(f)\}$; the meet is given symmetrically. This implies that there is a conflict of interest between the workers and the firms: if every worker weakly prefers a stable μ to a stable ν, then every firm weakly prefers ν to μ. Moreover, there is a *worker optimal* stable match and a *firm optimal* stable match.

To show that the agreeable core fails to be a lattice, consider the following example. Let $\mu_0(w_1) = f_A$, $\mu_0(w_2) = f_B$, and $\mu_0(w_3) = f_C$, and preferences are given as in Section 1.4. Both the pair w_2 and f_B and the pair w_3 and f_C prefer to participate in a cycle with the pair w_1 and f_A, but w_1 and f_A have opposing preferences over the two possible cycles. Worker w_1 prefers firm f_C and firm f_A prefers worker w_2, and so either cycle may be in the agreeable core. The agreeable core consists uniquely of the $\bar{\mu}$ match and the $\dot{\mu}$ match, a pair which is not ordered by $\succsim_W$. In this example there is no worker optimal match.

[8]Donald Knuth attributes this to John H. Conway.

Despite the impossibility of recovering a complete lattice over the agreeable core as in the classic model of stability, I show that a narrower result continues to hold. Given that the lattice structure failed in the example because two competing cycles exist in the agreeable core, an astute reader may conjecture that the lattice structure continues to hold for workers and firms who do not lie in such cycles. Suggestively, say that a is a *free agent* in μ if a lies on an acyclic, complete, and alternating path of (A, μ_0, μ). My first proposition justifies my terminology:

Proposition 3. *If μ is in the agreeable core, then there are no blocking pairs among free agents in μ. Moreover, every free agent a in μ weakly prefers $\mu(a)$ to being unmatched.*

The proof of Proposition 3 shows that these agents are "free" to form blocking pairs because each can satisfy a sequence formed by alternating edges from μ_0 and μ. Free agents resemble the agents in the classic model: their μ_0-partner (if any) is not concerned with the partner she finds.

However, an obstacle arises because the free agents depend on μ; that is, a may be a free agent in μ but not in ν. What I can show is that, if μ and ν share the same set of free agents and they agree on the agents who are not free, then $\mu \vee \nu$ is in the agreeable core. Toward that end, I say that μ and ν are *structurally similar* if they have the same set of free agents and $\mu(a) = \nu(a)$ for every agent which is not free. The following lemma shows that structurally similar matches in the agreeable core play nicely with the join and meet operators defined previously:

Lemma 3. *Let μ and ν be structurally similar matches in the agreeable core. Then $\mu \vee \nu$ is a match. The same holds for $\mu \wedge \nu$.*

Notably, $\mu \vee \nu$ may not be structurally similar to μ and ν.[9] The (possible) structural differences between $\mu \vee \nu$ and μ force us to discard any hope of obtaining a lattice-like result. However, the join and meet operators still produce matches in the agreeable core:

Theorem 2. *Let μ and ν be structurally similar matches in the agreeable core. Then $\mu \vee \nu$ and $\mu \wedge \nu$ are both in the agreeable core.*

[9] I have an example demonstrating this (available upon request), but it is too lengthy to include because it involves eight workers and eight firms.

The conflict of interest continues to hold for structurally similar matches. That is, if μ and ν are in the agreeable core *and are structurally similar*, then if every worker weakly prefers μ to ν, then every firm weakly prefers ν to μ. Conversely, in the classic matching framework, $\mu_0(a) = a$ for every agent and thus every agent is free. Every match is then structurally similar and hence my Theorem 2 generalizes standard results.

1.5 Incentives in the Propose-Exchange algorithm

This section addresses the incentive properties of the Propose-Exchange algorithm. The results provide insight into how robust the PE is to manipulation by participants. This is crucial for implementing the PE in practice because the output of the PE is only guaranteed to be in the agreeable core if the inputs are accurate. I find that while the PE is more susceptible to more kinds of manipulations than either the DA or the TTC, the new manipulations are difficult to execute.

I consider two kinds of manipulations in these subsections. In the first, I allow a worker to arbitrarily misreport his preference.[10] In the second, I allow a worker and a firm to create an artificial initial match, a misreport of μ_0.

For clarity through this section, I write $\succsim'_w$-Propose stage to indicate the operation of the Propose stage on the matching problem when w's preference $\succsim_w$ is replaced by $\succsim'_w$. A similar shorthand is used when μ_0 is replaced by μ'_0.

Preference Manipulation

In this subsection I discuss preference manipulations by workers. I allow a worker w to misreport his preference $\succsim_w$ by reporting $\succsim'_w$ instead. The intuition is that a worker may benefit from manipulating which agents (including himself) are active in the Exchange stage. I show that there may exist a worker who can profitably misreport his preference in the PE. However, this problem is not unique to the PE, but exists for every algorithm that produces a match in the agreeable core. These results are in contrast to their parallels in existing theory: no worker can profitably misreport his preferences in the DA or TTC (Dubins and Freedman, 1981; Dur and Ünver, 2019). I connect these results by showing that only workers who participate in both stages of the PE can profitably misreport their preferences.

Formally, *mechanism* ψ is a function of $(W, F, \succsim, \mu_0)$ that returns a match $\psi(W, F, \succsim, \mu_0)$.

[10] It is well-known that a firm can manipulate the DA by misreporting her preference, so I only consider the problem from the worker's perspective.

w_1	w_2	w_3
f_A	f_B	f_C
f_B	f_A	f_A
f_C	f_C	f_B

f_A	f_B	f_C
w_2	w_1	w_3
w_3	w_2	w_2
w_1	w_3	w_1

(a) Outcome of DA before preference swap

w_1	w_2	w_3
f_A	f_B	f_A
f_B	f_A	f_C
f_C	f_C	f_B

A	B	C
w_2	w_1	w_3
w_3	w_2	w_2
w_1	w_3	w_1

(b) Outcome of DA after preference swap

Figure 1.7: Worker w_3's exchange of the order of f_C and f_A in his preference leaves his partner unchanged, but causes workers w_1 and w_2 to receive new partners.

Definition 7. A mechanism is *preference manipulable* if there is at least one matching problem $(W, F, \succsim, \mu_0)$, worker w, and preference $\succsim'_w$ such that

$$\psi(W, F, \succsim_{-w}, \succsim'_w, \mu_0) \succsim_w \psi(W, F, \succsim, \mu_0).$$

In words, if for some example a worker w would rather report $\succsim'_w$ instead of $\succsim_w$, then ψ is preference manipulable.

A natural question arises as to whether a mechanism exists that is non-preference-manipulable and produces a match in the agreeable core. Proposition 4 provides a negative answer:

Proposition 4. *If $\psi(\succsim)$ is in the agreeable core for all $\succsim$, then ψ is preference manipulable.*

I prove Proposition 4 through a counterexample. The counterexample is driven my the possibility of bossiness within the DA. A mechanism is *bossy* if an agent can, by misreporting his preference, affect the matches of the other agents without changing his own. Consider the example in Figure 1.7. Worker w_3 can cause workers w_1 and w_2 to exchange partners by misreporting a preference for firm A. In the counterexample in the proof of Proposition 4, there is a worker w_1 who would like to exchange initial partners with w_2. Worker w_1 reduces w_2's ability to match to an initially unmatched firm by including that firm in his own preferences. Effectively, if w_2 is a free agent then w_1 will not be able to match to $\mu_0(w_1)$. Thus, w_1 manipulates w_2's options to keep w_2 matched to $\mu_0(w_2)$ to cause an exchange.

Theorem 3 formalizes this intuition. It shows that a worker only has two avenues through which to profit from a misreport. First, the worker may profit from finding a partner in the Exchange stage rather than the Propose stage. This is similar to truncating[11] his preferences. Second, the worker may find his partner in the Exchange stage but choose to manipulate which workers who participate in the Exchange stage, as in the counterexample previously discussed.

Theorem 3. *If worker w has a profitable misreport $\succsim'_w$, then w is active in both stages of the $\succsim'_w$-Propose-Exchange algorithm.*

Because whether a worker is active in the Propose stage is independent of his reported preferences, Theorem 3 further restricts the set of workers who can profitably misreport. A worker can only profitably misreport if he both has a μ_0-firm and is active in the Propose stage. For a market designer, these conditions are easy to verify and provide an upper bound on the number of workers who can profitably manipulate. Additionally, Theorem 3 highlights the informational requirements necessary to profitably misreport. A worker must be able to predict the outcome of the Exchange stage, which itself is a complicated object and depends on which agents participate in the Exchange stage.

Manipulating μ_0

In this subsection I complement the analysis of how preferences may be profitably misreported with an analysis of how the initial match may be profitably misreported. The concern is that because the initial match μ_0 affects the output of the PE, a pair of agents may find it in their interest to create a superfluous artificial agreement. I show that, while such a manipulation is possible, it usually requires an additional preference manipulation to be successful. I conclude that profitably misreporting the initial match requires a similar level of sophistication as a preference manipulation.

Formally, let μ_0 be given (and fixed throughout this subsection) with μ_2 the output of the μ_0-PE. Let worker w and firm f be both μ_0-unmatched, and let μ'_0 be formed from μ_0 by matching w and f. Let μ'_1 and μ'_2 be the respective outputs of the μ'_0-Propose and μ'_0-Exchange stages. If both w and f strictly prefer μ'_2 to μ'_1, then w and f can *profitably misreport* an initial match. Profitably misreporting an initial match requires that both w and f strictly gain from the deviation.

I show that, although it is possible for the PE to be manipulated in this way, its extent is quite limited and involves substantial risk for the worker. First, I show in Theorem 4 that any profitable misreport pushes w and f from the Propose stage

[11] moving his initial partner higher; see Roth and Rothblum (1999).

into the Exchange stage ($\mu'_1(w) = f$). The intuition is that if $\mu'_1(f) \neq w$, then f has received a better partner in the μ'_0-Propose stage and thus that all of the workers have received a worse partner. Second, Theorem 4 also shows that for any profitable misreport, w *cannot* be active in the μ'_0-Propose stage.

Theorem 4. *If w and f can profitably misreport an initial match, then $\mu'_1(w) = f$ and w is not active in the μ'_0-Propose stage.*

The interpretation of Theorem 4 is that f must prefer w to f's match when w is removed from the matching problem entirely. In effect, f faces little risk from the misreporting because w is as good as (if not better than) what f would receive if w were not present. For w however, an initial match with f could carry great risk if f is low on w's preferences relative to $\mu_1(w)$. This strategy may backfire because a mistake in w's calculations (or a misrepresentation by f) could render w assigned to f.

In summary, neither misreporting preferences or the initial match appears likely to succeed without detailed knowledge of the other participants' preferences. Misreports frequently expose misreporting agents to a large downside risk. These incentive findings inform the broader applicability of the PE, which I discuss in the following section.

1.6 Conclusion

This chapter has shown the strength of the agreeable core in providing a theory of equilibrium for a broad class of matching markets. The initial match organically models numerous real-world examples, and the Propose-Exchange algorithm is ready to be implemented in a variety of applications. In this closing section I discuss three topics. First, I review the connections between this chapter and existing research. Second, I provide guidance on applying the Propose-Exchange algorithm in several environments. Last, I close with a discussion of my modeling choices and possible extensions.

Connection to the Literature

This chapter develops a novel theory of matching under initial contracts that bridges object allocation and two-sided matching. It connects several literatures on two-sided matching. An exhaustive review of the literature is far beyond the scope of this chapter, so I list the only the most closely related work and its connections with this chapter.

I integrate the classic model of two-sided matching with recent advances in recontracting. In the classic model, a stable match always exists and can be found by the DA (Gale and Shapley, 1962). It is well known that the set of pairwise-stable matches corresponds to the core of a related cooperative game (Roth and Sotomayor, 1990). Later research largely discarded the connection with the core in favor of pairwise-stability notions. When considering matching with an initial match (in which the intersection of pairwise stable and individually rational outcomes may be empty), Pereyra (2013) and Guillen and Kesten (2012) generalize pairwise-stability by partitioning claims between valid and invalid claims and then removing all valid claims. This may be strongly inefficient (Combe and Schlegel, 2024; Combe, Tercieux and Terrier, 2022), and hence a mechanism with minimal envy is considered (Kwon and Shorrer, 2023). Although efficient, these minimal envy mechanisms are inscrutable to participants: the designer allows some claims but not others only because doing so minimizes some objective. This chapter advances this literature by reconnecting the initial back to the core, a more interpretable solution. I both minimize envy as in Kwon and Shorrer (2023) but also provide a clear definition of valid and invalid claims as in Pereyra (2013).

Research in school choice has made extensive use of both the DA and TTC. Abdulkadiroğlu and Sönmez (2003) suggests the Deferred Acceptance algorithm from Gale and Shapley (1962) or the Top Trading Cycles algorithm from Shapley and Scarf (1974) as desirable and implementable solutions. Both algorithms run in polynomial time, are relatively easy to describe, and are strategyproof. The DA is fair (no blocking pairs) while the TTC is efficient (Pareto efficient for the students). A handful of researchers seek to combine or modify the two algorithms to reconcile these properties, allowing certain priority violations (Abdulkadiroğlu, 2011; Dur, Gitmez and Yılmaz, 2019; Kesten, 2006; Kwon and Shorrer, 2023; Reny, 2022; Troyan, Delacrétaz and Kloosterman, 2020; Morrill, 2013b; Dur and Morrill, 2017). Papers in this vein typically define a set of properties of a mechanism (such as the allowable priority violations, efficiency, strategyproofness, etc.), and then present a satisfactory algorithm, typically a variation of the DA or TTC. My work complements this approach by an algorithm derived from first principles rather than with specific objectives in mind. My approach draws from cooperative game theory rather than emphasizing certain desirable properties of the final allocation.

A connected branch of matching theory develops methods for matching with minimum quotas. Schools are modeled as having both a maximum capacity for students but also a minimum required quota of students. One approach is to allow for wasted seats but not envy (Fragiadakis and Troyan, 2017). A separate approach uses an auxiliary "master list" (Ueda et al., 2012) or "precedence list" (Fragiadakis et al., 2016; Hamada et al., 2017) as a means to break ties: if two students wish to take an empty seat but the minimum quota requires that only one may do so, the list determines which worker can. The algorithms described in both approaches typically either sacrifice efficiency (based on the DA) or fairness (based on the TTC), and both require that all agents are mutually acceptable. I develop both approaches by endogenizing the master list into the initial match and not requiring any assumptions on preferences. Although a master list is natural in some applications, whether a master list or the initial match is more appropriate depends on the application.

Surprisingly, no authors have connected matching with minimum quotas and the matching with an initial match. I combine these subfields with the observation that, if the initial match provides a guarantee for both workers *and firms*, then minimum quotas are the special case when every firm is assigned workers equal to its minimum quota in the initial match. The initial match provides a different justification for why some blocking pairs are allowable but others are not, one which I think applies well to school choice.

Finally, the paper closest in spirit to ours is Abdulkadiroğlu and Sönmez (1999), "House Allocation with Existing Tenants." Their model is one-sided, and they show that a hybrid of the Serial Dictatorship algorithm and the TTC algorithm provides an efficient improvement over the initial match. I present a two-sided model with a hybrid algorithm between the DA and the TTC. Although my models are different, my approach is remarkably similar to theirs.

Applications

The PE can unify out-of-match residencies with the NRMP, creating a larger overarching match that nests both and guarantees Pareto efficiency while allowing for early matches. It is well-known that a fraction of medical residencies are offered independently of the centralized clearinghouse operated by the NRMP. These out-of-match residency programs entice prospective residents to sign binding contracts prior to the operation of the NRMP because these contracts provide guarantees to risk-averse residents. Because the rules of the NRMP forbid residents from participating if they have already accepted an out-of-match offer, these two markets operate independently.[12] The out-of-match offers introduce inefficiency by dividing the market temporally. Under the PE, the out-of-match market operates essentially unchanged: programs can entice residents with early offers. However, if the NRMP uses the Propose-Exchange algorithm, the residents and programs who have already formed contracts are allowed to participate as agents under an initial match. Remark 1 guarantees that the final match is Pareto efficient. A similar construction can be used to integrate Early Decision agreements into the regular college admission cycle.

The PE also allows for asymmetrical obligations, such as professional sports contracts or tenured positions, which bind participants unequally. For example, an athlete's contract with a team may allow the team to trade the athlete to another without the athlete's consent, but the athlete cannot "trade" his team without the team's consent. Similarly a tenured professor or teacher's contract allows her to leave her institution unilaterally, and restricts the institutions ability to remove her; see Combe, Tercieux and Terrier (2022) for an application to the French public school system. To incorporate this one-way obligation into the PE, I modify the participants' preferences. For the professor w tenured at (that is, initially matched to) institution f, I modify f's preference $\succsim_f$ by moving w to the bottom of $\succsim_f$. This guarantees that w is *never* required to remain at f, but always may choose to do so. Without an initial match, the standard model is instead forced to move w to the top of $\succsim_f$; this achieves the same result (w can always match to f), but suffers from inefficiency (Pereyra, 2013). The one-way contracts that allow for trades, as in professional sports, can similarly be included under additional assumptions.[13]

The initial match can also be leveraged to achieve minimum quotas that balance individual preferences and institutional needs. Examples of minimum quotas are minimum enrollment at a school or in a class, or guarantees that some "rural"

[12]Recently, the NRMP has implemented the "All-In" policy in an attempt to curtail residency programs from offering out-of-match residencies. The All-In policy requires that any residency program participating in the NRMP offer residencies exclusively through the NRMP.

[13]For instance, a "tradable" contract can be included through modifying the athlete's preference by putting the team and being unmatched at the bottom of the preference. Therefore, the athlete is always matched to a team, but the identity of the team can change. However, there is a tension: if the athlete can express a preference for being unmatched, then the team can terminate the athlete at will. Hence, in this model it is essential that the athlete can only be traded to a set of teams which he prefers to being unmatched.
Again, there is a limit to who can have tradable contracts. If a team is allowed to trade an athlete, then the PE algorithm must have the teams propose and point. This precludes any athlete from trading her team. In professional sports this is a reasonable assumption, but caution is needed in more general applications.

hospitals are matched to residents. For instance, a minimum quota of students may be required for a school to operate or for a class to be offered. The PE can incorporate these quotas by using the initial match to assign the minimum number of students to the school or class. By then modifying the school's or class's priority order (preferences) over students by moving the initially assigned students to the bottom, just above being unmatched, the designer guarantees that the school or class will enroll at least its minimum quota. The initial assignments are only binding if no other student desires the school or class. In this way, the initial match requires the minimal restriction on students' choices while meeting the institutional objective. The agreeable core provides a clear justification for why some students' choices are restricted. If a restricted student would like to attend another school, then at least one school would not meet its minimum quota or some student would be harmed.

Future Directions

The many-to-one setting introduces complex constraints because firms participating in an agreeable coalition must consider multiple binding agreements. The motivation behind my focus on one-to-one matching is driven by two competing models of a firm in many-to-one markets. In the first model, each firm is modeled as a collection of unit-demand sub-firms, each endowed with the master firm's preference over individual workers. This is the model used in most applications because eliciting a single ranking over workers from each firm is easier than a preference over sets of workers. The agreeable core then treats each sub-firm as an individual agent. A worker is initially matched to a single sub-firm, and he must include *that* sub-firm in any agreeable coalition. This model straightforwardly extends the one-to-one theory, and the same results hold.[14] In the second model, each firm is treated as an agent with a preference over *sets* of workers. Even in the classic model without an initial match, restrictions such as substitutability need to be placed on firm preferences to guarantee existence.[15] Beyond the question of existence, the requirement that $\nu(C) = C$ for a coalition C blocking with match ν implies that the size of an agreeable blocking coalition increases dramatically. For example, if a firm seeks to join a coalition, that coalition must include all of its initial workers (who themselves are possibly matched to other firms) and all of the workers it will

[14]There are some interesting additional questions in this environment, such as how a worker should construct his preference over two identical sub-firms which are initially matched to different workers, and whether a firm could rearrange the initial matches of its sub-firms to construct a new agreeable and blocking coalition.

[15]See Echenique and Oviedo (2004) for a unified treatment of the many-to-many case.

match to (who themselves may be initially matched to other firms). In a market with many workers initially matched, agreeable coalitions quickly must contain almost every agent in the model. The usefulness of the agreeable core in this context is unclear, and adapting it to these environments is a future avenue of research.

The agreeable core can provide insights into the formation of the initial match μ_0. The model is agnostic as to how μ_0 is determined. It could be interesting to use the agreeable core or the Propose-Exchange algorithm in combination with a model of the formation of μ_0 to understand pre-matching dynamics. Because the initial match is instrumental in the PE, developing a theory of pre-match formation could be insightful for other market-design applications. Theorem 4 addresses one such question, but more questions abound.

1.A Appendix to Chapter 1: Omitted Proofs

Throughout the appendix I abuse notation and write $a \in e$ to mean that either the first or second coordinate of e is a.

Lemma 1.A.1. *If P is a complete and alternating path in $(A, \mu_0 \cup \mu \cup I(\mu))$, then every agent contained in P is in exactly one edge from μ_0 and one edge from $\mu \cup I(\mu)$.*

Proof. Let $P = (e_1, \ldots, e_n)$ be a complete and alternating path in $(A, \mu_0 \cup \mu \cup I(\mu))$ and let a be contained in P. If $n = 2$, then the statement is trivial because completeness implies every $a \in P$ is in both e_1 and e_2 and P alternating implies that one of $\{e_1, e_2\}$ is in μ_0 and the other is in $\mu \cup I(\mu)$. Hence, let $n \geq 3$.

Again, if $a \in e_k \cap e_{k+1}$ for $k \geq 1$ then the statement is true because completeness implies e_k and e_{k+1} are the only edges in P containing a, both e_k and e_{k+1} cannot be from μ_0 by construction, and P alternating implies that both e_k, and e_{k+1} cannot be from $\mu \cup I(\mu)$. Therefore, one of $\{e_k, e_{k+1}\}$ is from μ_0 and the other from $\mu \cup I(\mu)$. Hence, let $a \in e_1 \cap e_n$ and thus P is cyclic. Let a be a worker; the argument is symmetric if a is a firm.

Because there is a bijection[16] between the workers and firms contained in P and every agent in P is contained in two edges of P, n is even. Therefore, if $e_1 \in \mu_0$ then $e_n \in \mu \cup I(\mu)$, and if $e_1 \in \mu \cup I(\mu)$ then $e_n \in \mu_0$. This proves the result. □

Proof of Proposition 1: Let μ be individually rational.

[16] namely, μ_0

For the ($\Rightarrow$) *direction:* I prove the contrapositive; that is, if μ admits a blocking path, then μ is not in the agreeable core. Let $P = (e_1, \ldots, e_n)$ be a blocking path in $(A, \mu_0 \cup \mu \cup I(\mu))$. Note that $\mu_0(P) = P$ and $\mu^P(P) = P$.

By the definition of $I(\mu)$, it follows that $\mu^P \succsim_P \mu$. Because P is blocking, there is an edge e in P that is also in $I(\mu)$. Hence, both agents in e strictly prefer μ^P to μ. Therefore, P is an agreeable blocking coalition and μ is not in the agreeable core.

For the ($\Leftarrow$) *direction:* I prove the contrapositive; that is, if μ is not in the agreeable core then μ admits a blocking path. Let μ be not in the agreeable core. Then there exists an agreeable blocking coalition C that blocks μ through ν.

Let a_1 be an agent in C such that $\nu(a_1) \succ_{a_1} \mu(a_1)$; such an agent exists by the definition of a blocking coalition. I will construct a path P from a_1 by iteratively adding alternating edges from μ_0 and ν to $\{a_1, \nu(a_1)\}$, first with increasing indices and then with decreasing indices. I assume that $a_1 \in W$; the other case follows from a symmetric argument.

Starting with $e_1 \equiv (a_1, \nu(a_1))$ and $P_1 \equiv (e_1)$, do the following iteratively. Choose an edge e_{k+1} from μ_0 or ν that is not already present in P_k such that the second coordinate of e_k is the first coordinate of e_{k+1}, then define P_{k+1} by appending e_{k+1} to P_k. Continue until no more edges may be added in this way. Finally, repeat the same process starting from e_1, but *prepending* edges $e_0, e_{-1}, \ldots$ to P_k.

Observe that P is a path in $(A, \mu_0 \cup \mu \cup I(\mu))$ because $\nu \succsim_C \mu$. Next, observe that because every agent in P is contained in at most two edges (one from μ_0 and the other form ν); every agent in P is contained in at least two edges because edges are added until no more can be added without including repeats and therefore P is complete. Also, P is alternating because $e_{2k} \in \mu_0$ and $e_{2k-1} \in \nu$. Finally, observe that $e_1 \in I(\mu)$. Therefore, P is a blocking path of μ. Therefore $(A, \mu_0 \cup \mu \cup I(\mu))$ contains a blocking path, completing the proof. $\square$

Introduction to the proofs of Lemma 1 and Proposition 2:

Before proving Lemma 1, I first introduce some notation and a short result:

Definition 1.A.1. I say that loop $e = (a, a)$ is a *proposal source* if either

1(a) : $(a, a) \in \mu_0$ and $a \in W$, or

1(b) : $(a, a) \notin \mu_0$ and $a \in F$.

I say that loop $e = (a, a)$ is a *proposal sink* if e in not a proposal source; that is, if either

2(a) : $(a, a) \notin \mu_0$ and $a \in W$ or

2(b) : $(a, a) \in \mu_0$ and $a \in F$.

A straightforward parity argument shows that if $P = (e_1, \ldots, e_n)$ is a complete, alternating, and acyclic path in $(A, \mu_0 \cup \mu \cup I(\mu))$, then e_1 is a proposal source and e_n is a proposal sink.

Lemma 1.A.2. *Let $P = (e_1, \ldots, e_n)$ be a complete, alternating, and acyclic path in $(A, \mu_0 \cup \mu \cup I(\mu))$ with $n \geq 3$. Then e_1 is a proposal source and e_n is a proposal sink.*

Proof. Because P is acyclic and complete, e_1 and e_n are both loops. Let $e_1 = (a_1, a_1)$ and $e_n = (a_{n-1}, a_{n-1})$. Similarly, let $e_2 = (a_1, a_2)$ and $e_{n-1} = (a_{n-2}, a_{n-1})$. Because $n \geq 3$, $a_1 \neq a_2$ and $a_{n-2} \neq a_{n-1}$.

Consider the following cases:

1. $a_1 \in W$: Then because there are no edges between two distinct workers, it follows that $a_2 \in F$. Therefore, $e_2 \in \mu \cup I(\mu)$. This implies that $e_1 \in \mu_0$. Therefore e_1 is a proposal source.
2. $a_1 \in F$: Then because there are no edges between two distinct workers, it follows that $a_2 \in W$. Therefore, $e_2 \in \mu_0$. This implies that $e_1 \in \mu \cup I(\mu)$. Therefore e_1 is a proposal source.

Symmetric arguments show that e_n is a proposal sink. □

Proof of Lemma 1:

Suppose (toward a contradiction) that $P = (e_1, \ldots, e_n)$ is an acyclic blocking path of μ_1. Because P is acyclic and complete, e_1 and e_n are both loops and $n \geq 3$. By

Lemma 1.A.2, e_1 is a proposal source and e_n is a proposal sink. Let

$$
\begin{aligned}
e_1 &= (a_1) \\
e_2 &= (a_1, a_2) \\
&\vdots \\
e_{n-1} &= (a_{n-2}, a_{n-1}) \\
e_n &= (a_{n-1}).
\end{aligned}
$$

I argue by induction that every worker $a_k \in P$ makes a proposal during the Propose algorithm. Because every agent contained in P weakly prefers μ^P to μ_1, it follows that every worker contained in P *who proposes* proposed to his μ^P-partner. In my base case I show that the worker with the lowest index contained in P proposes during the Propose algorithm. There are two possibilities:

1. a_1 *is a worker:* Because e_1 is a proposal source by definition $\mu_0(a_1) = a_1$. Hence a_1 begins the Propose algorithm activated. Therefore, a_1 proposes during the Propose algorithm.

2. a_1 *is a firm:* Because e_1 is a proposal source, by definition $\mu_0(a_1) \neq a_1$. Therefore $\mu_0(a_1) = a_2$. Because a_1 prefers μ^P to μ_0 and $\mu^P(a_1) = a_1$ because e_1 is loop, it follows that a_2 is activated at the start of the Propose algorithm. Therefore, a_2 proposes during the Propose algorithm.

For the inductive step, suppose $a_{k-1} \in W$ makes a proposal; I will show that the worker with the next highest index makes a proposal. If $k - 1 \geq n - 2$, then a_{k-1} is the worker with the highest index and the claim is vacuous; therefore, suppose $k - 1 < n - 2$. Because $\mu^P(a_{k-1}) = a_k$, it follows that a_{k-1} proposes at some point to a_k. Because μ_1 is individually rational and $\mu_0(a_k) = a_{k+1}$, it follows that a_k weakly prefers a_{k-1} to a_{k+1}. Therefore a_{k+1} is activated at some point and thus a_{k+1} makes at least one proposal during the Propose algorithm, concluding my inductive argument.

Next, I show that an agent contained in a proposal sink never rejects a proposal from their μ^P-partner. If a_{n-1} is a worker, then he never rejects a proposal from himself. If a_{n-1} is a firm, then $\mu_0(a_{n-1}) = a_{n-1}$ by definition. Because a_{n-1} prefers μ^P to both μ_0 and μ_1 and because a_{n-1} receives no proposals she prefers to $\mu_1(a_{n-1})$ (by construction of μ_1), it follows that a_{n-1} does not reject a proposal from $\mu^P(a_{n-1})$.

Finally, I show that no worker contained in P is rejected by his μ^P-partner. To see this, suppose (toward a contradiction) that $k-1$ is the largest index such that a_{k-1} is rejected by $\mu^P(a_{k-1})$. Because a proposal sink does not reject a proposal by his or her μ^P-partner, it follows that $k-1 < n-2$ (that is, a_{k-1} is not one of the last two agents in the path).

Because a_k prefers a_{k-1} to $\mu_1(a_k)$ and yet a_k rejects a_{k-1}, it must be that $\mu_0(a_k) = \mu_1(a_k)$ (by construction of μ_1). Therefore a_k is matched to a_{k+1} by both μ_0 *and* μ_1. Because matches are bijective, I have $\mu_1(a_{k+1}) = \mu_0(a_{k+1}) = a_k$. Consider that, because P is a complete and $n \geq 3$, it follows that $\mu^P(a_{k+1}) \neq \mu_1(a_{k+1})$. Therefore a_{k+1} must be rejected by $\mu^P(a_{k+1})$, a contradiction to my supposition that $k-1$ is the largest index for which a worker is rejected by his μ^P-match.

Therefore, because no worker in P is rejected by his μ^P-partner, it follows that μ^P agrees with μ_1 on P. Hence, every edge in P from $\mu_1 \cup I(\mu_1)$ is from μ_1. But because P is a blocking path, it must contain an edge from $I(\mu_1)$. Because $\mu \cap I(\mu_1) = \emptyset$, this is a contradiction. Therefore no blocking path of μ_1 is acyclic. □

Proof of Proposition 2: I say that a *proposal order* is a function that, at every step of the Propose phase, indicates which worker makes the next proposal. Let T and T' be two proposal orders, and let the output of the Propose stage using order T be μ and using T' be μ'. Suppose (toward a contradiction) that $\mu \neq \mu'$. Let

$$U = \{w \in W : \mu(w) \neq \mu'(w)\}$$
$$V = \{w \in W : w \text{ proposes under both } T \text{ and } T'\}.$$

There are two cases:

1. $U \cap V \neq \emptyset$: WLOG, there is some worker in $U \cap V$ who strictly prefers μ' to μ. Let w be the first such worker who is rejected by $f' \equiv \mu'(w)$ in the Propose stage under T. Because f' is w's μ'-partner, this implies that f' prefers w to being unmatched. Therefore, f' must reject w in favor of some w^*. Because f' is w's μ'-partner, this implies that w^* does not propose to f' under T'.

   Because w^* makes a proposal under T, it follows that there is some sequence of workers $w_1, \ldots, w_{n-1}, w^* \equiv w_n$ such that w_1 or $\mu_0(w_1)$ is a proposal source, and each w_k makes the first proposal to $\mu_0(w_{k+1})$ under T. Let k be the greatest index such that $w_k \in V$. Then it follows that w_k strictly prefers μ' to μ. Therefore, w_k must be rejected by $\mu'(w_k)$ earlier than w is rejected by f' under T, a contradiction.

2. $U \cap V = \emptyset$: Observe that U is nonempty by supposition. Let $w^* \in U$ and WLOG let w^* strictly prefer μ to μ'. Thus, w^* must make a proposal under T. It follows that there is some sequence of workers $w_1, \ldots, w_{n-1}, w^* \equiv w_n$ such that w_1 or $\mu_0(w_1)$ is a proposal source, and each w_k makes the first proposal to $\mu_0(w_{k+1})$ under T. Observe that $w_1 \in V$. Furthermore, if $w_k \in V$, then $\mu(w_k) = \mu'(w_k)$ by supposition. Hence, w_{k+1} makes a proposal under both T and T'. Therefore, $w_{k+1} \in V$. It follows that $w^* \in V$, a contradiction to the supposition that $U \cap V$ is empty.

Therefore, $\mu = \mu'$. □

Proof of Lemma 2: Suppose (toward a contradiction) $P = (e_1, \ldots e_n)$ is a cyclic blocking path in $(A, \mu_0 \cup \mu_2 \cup I(\mu_2))$. Because there is a bijection[17] between the workers and firms contained in P, n is even. Define $m \equiv \frac{n}{2}$.

From P (after a possible relabeling) define a vector of agents $(a_1, a_2, \ldots, a_n \equiv a_0)$ such that $(a_{k-1}, a_k) = e_{k-1}$, $a_1 \in W$, and $e_1 \in I(\mu_2)$. Because P is alternating, every odd agent is a worker and every even agent is a firm.

I first show that every agent in P is active in the Exchange stage. To see this, suppose (toward a contradiction) that some worker a_k in P is not active during the Exchange stage. Then a_k makes a proposal during the Propose stage to a_{k+1}. Therefore, a_{k+2} makes a proposal during the Propose stage. I can iterate this argument to show that every worker in P makes a proposal during the Propose stage. Because P is a blocking path, each firm in P prefers her respective proposal to her μ_1-partner. Because a_{k-2} is rejected by a_{k-1}, it necessarily follows that $\mu_1(a_k) = a_{k-1}$. Therefore, a_k is active in the Exchange stage, a contradiction. Therefore, every agent in P is active during the Exchange stage.

Let t_k be the iteration of the **while** . . . **do** loop of the Exchange algorithm that a_k sits down in.[18] During the Exchange algorithm every worker a_{2k-1} points to firm a_{2k}; hence, firm a_{2k} sits down weakly earlier than worker a_{2k-1}. In symbols, $t_{2k-1} \geq t_{2k}$ for all $1 \leq k \leq m$. Because $e_1 \in I(\mu_2)$, it follows that $t_1 > t_2$. Therefore,

$$\sum_{k=1}^{m} t_{2k-1} > \sum_{k=1}^{m} t_{2k}.$$

[17]namely, μ_0

[18]That is, if a_k sits down on the fourth iteration of the while loop, then $t_k = 4$.

However, every worker a_{2k+1} sits down at the same time firm a_{2k} sits down. In symbols, $t_{2k+1} = t_{2k}$ for all $1 \leq k \leq m$. Therefore,

$$\sum_{k=1}^{m} t_{2k+1} = \sum_{k=1}^{m} t_{2k}.$$

Because $\sum_{k=1}^{m} t_{2k+1} = \sum_{k=1}^{m} t_{2k-1}$, I reach a contradiction. □

Proof of Proposition 3: For the first claim, suppose (toward a contradiction) that w and f are both free agents in μ who also both prefer each other to $\mu(w)$ and $\mu(f)$, respectively. I construct a blocking path in $(A, \mu_0 \cup \mu \cup I(\mu))$, a contradiction to the supposition that μ is in the agreeable core.

Because w is a free agent in μ, w lies on an acyclic, complete, and alternating path P_w of (A, μ_0, μ). Rewrite P_w such that

$$P_w = \big(e_1, \ldots, e_{k-1}, \big(\mu_0(w), w\big), \big(w, \mu(w)\big), \ldots\big).$$

Similarly, there is a complete and alternating P_f such that

$$P_f = \big(\ldots, \big(\mu(f), f\big), \big(f, \mu_0(f)\big), e_{k+1}, \ldots, e_n\big).$$

There are two cases:

1. *P_w and P_f do not intersect:* Then

$$\big(e_1, \ldots, e_{k-1}, \big(w, f\big), e_{k+1}, e_n\big)$$

   is a blocking path of μ.

2. *P_w and P_f do intersect:* Then let i be the greatest index less than k such that e_i is in P_f. Let e_j be the edge in P_f such that $e_i = e_j$. Therefore the path

$$\big(e_j, \ldots, e_{k-1}, \big(w, f\big), e_{k+1}, \ldots e_{j-1}\big)$$

   is a blocking path of μ.

In either case there is a blocking path of μ. But then μ is not in the agreeable core, a contradiction.

For the second claim I can repeat the argument from the first claim, substituting the edge (w, w) for $\{w, \mu(w)\}$ in path P_w and (f, f) for $\{\mu(f), f\}$ in path P_f. □

Lemma 1.A.3. *Let μ and ν be structurally similar matches in the agreeable core. Then $(\mu \vee \nu)(w) \in F$ if and only if $\mu(w) \in F$ <u>or</u> $\nu(w) \in F$. Similarly, $(\mu \vee \nu)(f) \in W$ if and only if $\mu(f) \in W$ <u>and</u> $\nu(f) \in W$. A symmetric result holds for $\wedge$.*

Proof. Both statements clearly hold for every agent that is not free in μ (and ν because μ and ν are structurally similar). Hence, I show that the statements hold for the free agents in μ.

For the first statement:

- *For the* $(\Rightarrow)$ *direction:* I show that if $\mu(w) \notin F$ and $\nu(w) \notin F$, then $(\mu \vee \nu)(w) \notin F$. Then $\mu(w) = \nu(w) = w$, which implies $(\mu \vee \nu)(w) = w$. Thus $(\mu \vee \nu)(w) \notin F$.
- *For the* $(\Leftarrow)$ *direction:* I show that if $\mu(w) \in F$ or $\nu(w) \in F$, then $(\mu \vee \nu)(w) \in F$. To see this, note that if $\mu(w) = f$ or $\nu(w) = f$, then w strictly prefers f to being unmatched (w) by Proposition 3. Therefore, $\mu \vee \nu$ cannot leave w unmatched and therefore $(\mu \vee \nu)(w) \in F$.

For the second statement:

- *For the* $(\Rightarrow)$ *direction:* I show that if either $\mu(f) \notin W$ or $\nu(f) \notin W$, then $(\mu \vee \nu)(f) \notin W$. Then $\mu(f) = f$ or $\nu(f) = f$. By Proposition 3, f weakly prefers both $\mu(f)$ and $\nu(f)$ being unmatched. By the definition of $\vee$, $(\mu \vee \nu)(f) = f$. Therefore, $(\mu \vee \nu)(f) \notin W$.
- *For the* $(\Leftarrow)$ *direction:* I show that if $\mu(f) \in W$ and $\nu(f) \in W$, then $(\mu \vee \nu)(f) \in W$. Then $\{\mu(f), \nu(f)\} \subseteq W$. Therefore $(\mu \vee \nu)(f) \in W$.

This completes the proof. □

Proof of Lemma 3: I draw my proof from the proof of Theorem 2.16 in Roth and Sotomayor (1990). I show that $\mu \vee \nu$ is a match; the argument for $\mu \wedge \nu$ is symmetric.

Because the free agents are the same in μ and ν, I need only to show that $\mu \vee \nu$ is a match on the free agents of μ and ν; all other matches are left unchanged because μ and ν are structurally similar. It is immediate from the definition of $\vee$ that items 1 and 2 from the definition of a match hold. That is, I only need that

$(\mu \vee \nu)(a) = b \iff (\mu \vee \nu)(b) = a$. Of course, if $a = b$ then the statement is tautological; hence, I prove for $w \in W$ and $f \in F$:

$$(\mu \vee \nu)(w) = f \iff (\mu \vee \nu)(f) = w.$$

For the ($\Rightarrow$) *direction:* I show that $(\mu \vee \nu)(w) = f$ implies $(\mu \vee \nu)(f) = w$. I consider the case when $\mu(w) = f$; the other case is symmetric. Suppose (toward a contradiction) that $(\mu \vee \nu)(f) \neq w$. Then $(\mu \vee \nu)(f) = \nu(f)$. Then f strictly prefers w to $\nu(f)$ and w strictly prefers f to $\nu(w)$, so w and f is a blocking pair of ν, a contradiction by Proposition 3. This completes this direction.

For the ($\Leftarrow$) *direction:* I show that $(\mu \vee \nu)(f) = w$ implies $(\mu \vee \nu)(w) = f$. I define a sequence of sets, then study their cardinality. Let

$$\begin{aligned} W' &\equiv \{w \in W : (\mu \vee \nu)(w) \in F\} \\ &= \{w \in W : \mu(w) \in F \text{ or } \nu(w) \in F\} && \because \textit{Lemma } 1.A.3. \end{aligned}$$

and

$$\begin{aligned} F' &\equiv \{f \in F : (\mu \vee \nu)(f) \in W\} \\ &= \{f \in F : \mu(f) \in W \text{ and } \nu(f) \in W\} && \because \textit{Lemma } 1.A.3. \end{aligned}$$

Observe the following relations:

$$\begin{aligned} |F'| &= |\mu(F')| && \because \mu \text{ is a match} \\ \mu(F') &\subseteq W' && \because \text{Definition of } F' \text{ and } W'. \end{aligned}$$

Therefore $|F'| \leq |W'|$. Similarly,

$$\begin{aligned} |W'| &= |(\mu \vee \nu)(W')| && \because (\Rightarrow) \text{ implication} \\ (\mu \vee \nu)(W') &\subseteq F' && \because (\Rightarrow) \text{ implication.} \end{aligned}$$

Therefore $|W'| \leq |F'|$ and thus $|W'| = |F'|$. Therefore $|(\mu \vee \nu)(W')| = |F'|$ and thus $(\mu \vee \nu)(W') = F'$.

The final string of implications is as follows: If $(\mu \vee \nu)(f) \in W$, then $f \in F'$. If $f \in F'$, then there exists w in $w \in W'$ such that $(\mu \vee \nu)(w) = f$. This completes this direction.

Therefore, $\mu \vee \nu$ satisfies item 3 from the definition of a match and thus $\mu \vee \nu$ is a match. $\square$

Lemma 1.A.4. *Let μ and ν be structurally similar matches in the agreeable core. Then $\mu \vee \nu \subseteq \mu \cup \nu$ and $I(\mu \vee \nu) \subseteq I(\mu) \cup I(\nu)$. The same holds for $\mu \wedge \nu$.*

Proof. By construction, $\mu \vee \nu$ only contains matches from μ and ν and thus $\mu \vee \nu \subseteq \mu \cup \nu$.

Let $\{w, f\} \in I(\mu \vee \nu)$ and let A^F be the free agents in μ (and ν because μ and ν are structurally similar). There are three cases:

1. $|\{w, f\} \cap A^F| = 0$: Then $(\mu \vee \nu)(w) = \mu(w)$ and $(\mu \vee \nu)(f) = \mu(f)$ by construction, so $\{w, f\} \in I(\mu)$.
2. $|\{w, f\} \cap A^F| = 1$: Suppose that $w \in A^F$; the other case is symmetric. Then either $(\mu \vee \nu)(w) = \mu(w)$ or $(\mu \vee \nu)(w) = \nu(f)$; again, let $(\mu \vee \nu)(w) = \mu(w)$ and the other case is symmetric. Then $(\mu \vee \nu)(f) = \mu(f)$ by construction, so $\{w, f\} \in I(\mu)$.
3. $|\{w, f\} \cap A^F| = 2$: This contradicts Proposition 3 and thus cannot happen.

In the cases that do not lead to a contradiction I see that $\{w, f\} \in I(\mu) \cup I(\nu)$, which completes the proof. □

Definition 1.A.2. A *crossing edge* at μ contains both a free agent and an agent who is not free at μ.

Lemma 1.A.5. *Let μ and ν be structurally similar matches in the agreeable core. Then any blocking path of $\mu \vee \nu$ must contain two crossing edges at μ. All crossing edges at μ of any blocking path of $\mu \vee \nu$ are contained in either $I(\mu)$ or $I(\nu)$.*

A symmetric result holds for $\mu \wedge \nu$.

Proof. Let A^F denote the free agents in μ (and ν because μ and ν are structurally similar), and let P be a blocking path of $\mu \vee \nu$.

I first prove that all crossing edges at μ of any blocking path of $\mu \vee \nu$ are contained in either $I(\mu)$ or $I(\nu)$. To see this, let $(a, b) \in P$ be a crossing edge with $a \in A^F$ and $b \in A \backslash A^F$. Because $\mu_0(A^F) = A^F$, it follows that $(a, b) \in \mu \vee \nu \cup I(\mu \vee \nu)$. Because $\mu(A^F) = A^F$ and $\nu(A^F) = A^F$ by construction it follows that $\{a, b\} \notin \mu \vee \nu$. Therefore $(a, b) \in I(\mu \vee \nu)$.

Next, I show that $P \not\subseteq A^F$ and $P \not\subseteq A \backslash A^F$. To see this, consider both cases (toward a contradiction in each case):

1. *Suppose* $P \subseteq A^F$: Then exists an edge e in P such that $e \in I(\mu \vee \nu)$. By Lemma 1.A.4, $e \in I(\mu)$ (the other case is symmetric). If $e = (w, f)$, then e constitutes a blocking pair and contradicts Proposition 3. If $e = (a, a)$, then a strictly prefers being unmatched to μ and contradicts Proposition 3. Therefore, $P \nsubseteq A^F$.

2. *Suppose* $P \subseteq A \backslash A^F$: Note that $\mu \vee \nu$ agrees with μ on $A \backslash A^F$. If P blocks $\mu \vee \nu$ then P blocks μ, a contradiction to the supposition that μ is in the agreeable core. Therefore, $P \nsubseteq A \backslash A^F$.

Therefore, P intersects both A and $A \backslash A^F$. By the definition of a path, there exists some crossing edge at μ in P.

Third, to see that two crossing edges at μ exist, suppose not. Let (a, b) be the crossing edge at μ in P such that $a \in A^F$ and $b \notin A^F$. As observed earlier, $(a, b) \in I(\mu \vee \nu)$. By Lemma 1.A.4, it follows that $(a, b) \in I(\mu)$ (the other case is symmetric). Suppose that $a \in W$; the other case is symmetric. Then P may be written

$$P = (\overbrace{e_1, \dots, e_{k-1}, (\mu_0(a), a),}^{\text{contained in } A^F} \quad \underbrace{(a, b),}_{\text{contained in } I(\mu)} \quad \overbrace{e_k \dots, e_K}^{\text{contained in } A \backslash A^F} \;).$$

Note that every edge from e_k to e_K exists in $(A, \mu_0 \cup \mu \cup I(\mu))$ because $\mu \vee \nu$ agrees with μ for these agents. Because $a \in A^F$, there is an alternating, complete, and acyclic path P^a in $(A, \mu_0 \cup \mu \cup I(\mu))$ such that

$$P^a = (e_1^a, \dots, e_{l-1}^a, (\mu_0(a), a), (a, \mu(a)), e_l^a \dots, e_L^a).$$

Because $\mu_0(A^F) = A^F$ and $\mu(A^F) = A^F$ by construction, it follows that every agent in P^a is in A^F. Observe that the path

$$P^* = (e_1^a, \dots, e_{l-1}^a, (\mu_0(a), a), (a, b), e_k \dots, e_K)$$

is a blocking path of μ, a contradiction. Hence, there are at least two edges that intersect both A^F and $A \backslash A^F$. $\square$

Proof of Theorem 2: I show that $\mu \vee \nu$ is in the agreeable core; the argument for $\mu \wedge \nu$ is symmetric. By Lemma 3, $\mu \vee \nu$ is a match. Because μ and ν are both individually rational, $\mu \vee \nu$ is individually rational. The remaining step is to show that there are no blocking paths of $\mu \vee \nu$.

Suppose (toward a contradiction) that $\mu \vee \nu$ is blocked by an agreeable coalition. By Proposition 1, there is a blocking path P of $\mu \vee \nu$. Let A^F denote the free agents in μ (and ν because μ and ν are structurally similar).

By Lemma 1.A.5, there are two crossing edges at μ in P, and both of these is in $I(\mu \vee \nu)$. There are two cases:

1. *There exists two crossing edges e_k and e_K at μ in path P such that the edges $e_{k+1}, \ldots, e_{K-1}$ (if any) are contained within $A \backslash A^F$.* Let $\{w, f\} = e_k$ and $(w', f') = e_K$ with $a_k, a_K \in A^F$ and $b_k, b_K \in A \backslash A^F$. Because $\mu \vee \nu \succsim_W \mu$ and $\mu \vee \nu \succsim_W \nu$, it follows that one of $(w, f) \in I(\mu)$ and $(w, f) \in I(\nu)$. By Lemma 1.A.4, let $(w', f') \in I(\mu)$ (the other case is symmetric).

   Because $w \in A^F$, there exists an acyclic, complete, and alternating path P^w of $(A, \mu_0 \cup \mu \cup I(\mu))$:

   $$P^w = (e_1^w, \ldots, e_{i-1}^w, \{\mu_0(w), w\}, \{w, \mu(w)\}, \ldots).$$

   Similarly because $f' \in A^F$:

   $$P^{f'} = (\ldots, (\mu(f'), f'), (f', \mu_0(f')), e_{j-1}^{f'}, \ldots, e_1^{f'}).$$

   Then the path

   $$P^* = (\underbrace{e_1^w, \ldots, e_{i-1}^w, (\mu_0(w), w),}_{P^w} \overbrace{(w, f), e_{k+1}, \ldots, e_{K-1}, (w', f'),}^{P}$$
   $$\underbrace{(f', \mu_0(f')), e_{j-1}^{f'}, \ldots, e_1^{f'})}_{P^{f'}}$$

   is a blocking path of μ, a contradiction to the supposition that μ is in the agreeable core.

2. *There does not exist two crossing edges e_k and e_K at μ in path P such that the edges $e_{k+1}, \ldots, e_{K-1}$ (if any) are contained within $A \backslash A^F$.* Let (a, b) be a crossing edge of μ of P with $a \in A^F$. Let $b \in W$; the other case is symmetric. The supposition implies that P must be acyclic and hence can be written

   $$P = (\underbrace{e_1, \ldots, e_{k-1},}_{\text{contained in } A \backslash A^F} (b, a), (a, \mu_0(a)), \ldots).$$

Because $a \in A^F$, there exists an acyclic, complete, and alternating path P^a of $(A, \mu_0, \mu \cup I(\mu))$:

$$P^a = (\ldots, (\mu(a), a), (a, \mu_0(a)), e^a_{i-1}, \ldots, e^a_1).$$

Then the path

$$P^* = (\overbrace{e_1, \ldots, e_{k-1},}^{P} \underbrace{(a, \mu_0(a)), e^a_{i-1}, \ldots, e^a_1}_{P^a})$$

is a blocking path of μ because μ and $\mu \vee \nu$ agree on the agents in $A \backslash A^F$. This is a contradiction to the supposition that μ is in the agreeable core.

Therefore, there are no blocking paths of $\mu \vee \nu$, which implies that $\mu \vee \nu$ is in the agreeable core. □

Proof of Proposition 4: Consider the following counterexample. There are three workers denoted by the numbers 1, 2, and 9, and three firms denoted by the letters A, B, and Z. Workers 1 and 2 are reference matched to A and B, respectively, while worker 9 and firm Z are each reference matched to him or herself. Formally,

$$\begin{array}{lll} \mu_0(1) = A & \mu_0(2) = B & \mu_0(9) = 9 \\ \mu_0(A) = 1 & \mu_0(B) = 2 & \mu_0(Z) = Z. \end{array}$$

A profile of preferences $\succ$ and an alternate profile of worker preferences are given in Figure 1.8. I use the circles to indicate match μ°, the squares to indicate match $\mu^\square$, and $\tilde{}$ to indicate $\tilde{\mu}$.

$$\begin{array}{lll} \mu^\circ(1) = B & \mu^\circ(2) = A & \mu^\circ(9) = Z \\ \mu^\circ(A) = 2 & \mu^\circ(B) = 1 & \mu^\circ(Z) = 9 \end{array}$$

$$\begin{array}{lll} \mu^\square(1) = A & \mu^\square(2) = Z & \mu^\square(9) = B \\ \mu^\square(A) = 1 & \mu^\square(B) = 9 & \mu^\square(Z) = 2 \end{array}$$

$$\begin{array}{lll} \tilde{\mu}(1) = A & \tilde{\mu}(2) = B & \tilde{\mu}(9) = Z \\ \tilde{\mu}(A) = 1 & \tilde{\mu}(B) = 2 & \tilde{\mu}(Z) = 9. \end{array}$$

P_1

$\succ_1$	$\succ_2$	$\succ_9$
B°	Z	B
Z	A°	Z°
A	B	

P_2

$\succ'_1$	$\succ_2$	$\succ_9$
B°	$Z^{\square}$	$B^{\square}$
Z	A°	Z°
$A^{\square}$	B	

P_3

$\succ'_1$	$\succ'_2$	$\succ_9$
B	$Z^{\square}$	$B^{\square}$
Z	A	$\tilde{Z}$
$\tilde{A}^{\square}$	$\tilde{B}$	

P_4

$\succ'_1$	$\succ'_2$	$\succ'_9$
B	$Z^{\square}$	$B^{\square}$
Z	A	Z
$A^{\square}$	B	

$\succ_A$	$\succ_B$	$\succ_Z$
A	$9^{\square}$	$\tilde{9}^{\circ}$
2°	1°	1
$\tilde{1}^{\square}$	$\tilde{2}$	$2^{\square}$

Figure 1.8: Tables provide preferences $\succ$ and alternate worker preferences $\succ'$. A grayed-out firm in $\succ'$ indicates that the worker matching to himself more than to that firm. If the table does not specify a preference over an alternative, then they are worse than every alternative listed.

I keep the firm preference profile fixed at $\succ_A$, $\succ_B$, and $\succ_Z$ for the firms and only specify preferences for the workers.

To prove the result, suppose that ψ is not preference manipulable. I consider the sequence of preference profiles P_1, P_2, P_3, and P_4 formed by swapping $\succ'_1$ for $\succ_1$, then $\succ'_2$ for $\succ_2$, and then $\succ'_9$ for $\succ_9$. I use the non-manipulability of ψ to restrict ψ to a unique match in each case. I then show that at P_3 worker 9 can profitably deviate to $\succ'_9$, a contradiction to the non-manipulability of ψ.

First, I limit the scope of matches I consider. Consider any μ and any P_j.

- If $A \succ_1 \mu(1)$ then 1 strictly prefers $\mu_0(1)$ to $\mu(1)$, hence μ is not in the agreeable core; the same holds for $B \succ_2 \mu(2)$, $2 \succ_B \mu(B)$, and $1 \succ_A \mu(A)$.
- If $j \neq 4$ and $Z \succ_9 \mu(9)$, then $\{9, Z\}$ is an agreeable coalition that blocks μ.
- If $j = 4$ and $Z \succ_9 \mu(9)$, then μ in the agreeable core implies that $\mu(1) \neq Z$ and hence $B \succ_9 \mu(9)$ implies that $\{2, 9, B, Z\}$ is an agreeable coalition that blocks μ; hence, if μ is in the agreeable core then $\mu(9) = B$.

- If $\mu(1) = Z$ and $\mu(2) = A$, then for P_1 $\{1, A, Z\}$ is an agreeable blocking coalition and for P_2, P_3, and P_4 $A \succ'_1 Z$. Hence for all P_j $\mu(1) = Z$ and $\mu(2) = A$ imply that μ is not in the agreeable core.

It follows that every worker is matched to a firm, and thus every firm is matched to a worker. Therefore, any match in the agreeable core only occurs between agents who are listed on each other's preferences in Figure 1.8. An exhaustive search reveals that μ°, $\mu^\square$, and $\tilde{\mu}$ are the only matches that meet these criteria.

For P_1, the agreeable core is $\{\mu^\circ\}$ because:

- ✓ μ° is the output of the PE algorithm and hence is in the agreeable core.
- ✗ $\mu^\square$ is blocked by the agreeable coalition $\{1, A, Z\}$ with any deviation μ' such that $\mu'(1) = Z$ and $\mu'(A) = A$.
- ✗ $\tilde{\mu}$ is blocked by the agreeable coalition $\{1, 2, A, B\}$ with any deviation μ' such that $\mu'(1) = B$ and $\mu'(2) = A$.

Hence, $\psi(P_1) = \mu^\circ$.

For preferences P_2, the agreeable core is $\{\mu^\circ, \mu^\square\}$ because:

- ✓ μ° does not match any worker to a firm he dropped from his preference, so every blocking coalition under these preferences forms under the prior preferences.
- ✓ $\mu^\square$ is the output of the PE algorithm and hence is in the agreeable core.
- ✗ $\tilde{\mu}$ is blocked by the agreeable coalition $\{1, 2, A, B\}$ with any deviation μ' such that $\mu'(1) = B$ and $\mu'(2) = A$.

If $\psi(P_2) = \mu^\square$, then consider the deviation by worker 1 of misreporting $\succ_1$ at P_2. Because $\mu^\circ(1) \succ'_1 \mu^\square(1)$, this is a profitable deviation. Therefore, because ψ is not preference manipulable, $\psi(P_2) = \mu^\circ$.

For preferences P_3, the agreeable core is $\{\mu^\square, \tilde{\mu}\}$ because:

- ✗ μ° matches worker 2 to firm A, which violates the requirement that $\mu(2) \succsim_2 B$.
- ✓ $\mu^\square$ is the output of the PE algorithm and hence is in the agreeable core.

✓ $\tilde{\mu}$: Observe that Z cannot be strictly better off in any blocking coalition, and thus 2 cannot be strictly better any blocking coalition. Furthermore, any agreeable coalition that makes 1 strictly better off must include B and hence, because the coalition is agreeable, 2. Therefore, any agreeable blocking coalition cannot make any worker strictly better off. Hence, $\tilde{\mu}$ is also in the agreeable core.

If $\psi(P_3) = \mu^{\square}$, then consider the deviation by worker 2 of reporting $\succ_2'$ at P_2. Because $\mu^{\square}(2) \succ_2 \mu^{\circ}(2)$, this is a profitable deviation. Therefore, because ψ is not preference manipulable, $\psi(P_3) = \tilde{\mu}$.

In this final step, I note that the core under P_4 is the singleton $\mu^{\square}$. To see this, observe that μ° and $\tilde{\mu}$ each match a worker to a firm he lists below his reference match, and therefore none of these three matches is in the agreeable core. $\mu^{\square}$ is the output of the PE algorithm and hence is in the agreeable core. However, consider the deviation by worker 9 of reporting $\succ_9'$ at P_3. Because $\mu^{\square}(9) \succ_9 \tilde{\mu}(9)$, this is a profitable deviation. Therefore, ψ is preference manipulable, a contradiction. □

Introduction to the proofs of Theorem 3:

Lemma 1.A.6. *For any μ_1, there is no w and f such that all three conditions are true:*

1. *w is active in the Propose stage; and*
2. *$\mu_0(f) \neq \mu_1(f)$; and*
3. *(w, f) is a blocking pair of μ_1.*

Proof. Toward a contradiction, suppose (w, f) is such a pair. Because w is active and w strictly prefers f to $\mu_1(w)$, w makes a proposal to f. Because $\mu_0(f) \neq \mu_1(f)$ and f strictly prefers w to $\mu_1(f)$, f does not reject the proposal from w. This is a contradiction to the supposition that (w, f) is a blocking pair. Therefore, no such pair exists. □

Proof of Theorem 3: Suppose (toward a contradiction) that w can profitably misreport $\succsim_w'$ but that w is not active in both the $\succsim_w'$-Propose and $\succsim_w'$-Exchange stages. First I consider the case when w is not active in the $\succsim_w'$-Propose stage, and then the case when w is not active in the $\succsim_w'$-Exchange stage. Before continuing, I note that

w's preferences do not affect whether w is active in the $\succsim_w$-Propose or $\succsim'_w$-Propose stages.

Suppose w is not active in the $\succsim'_w$-Propose stage. The rest of the proof follows directly from the non-manipulability of the Top Trading Cycles algorithm. This is well-known in the literature; see Ma (1994) for one such proof, and footnote 4 of Dur and Ünver (2019) for a list of references to other proofs. This is a contradiction to the supposition that w can profitably misreport $\succsim'_w$.

The remainder of the proof is built on the proof of the blocking lemma of Roth and Sotomayor (1990).

For the remainder of the proof, suppose that w is active in the $\succsim'_w$-Propose stage but not in the $\succsim'_w$-Exchange stage. Therefore, w is active in the $\succsim_w$-Propose stage as well. Let μ'_1 be the output of the $\succsim'_w$-Propose stage. Let W' be the set of workers who strictly prefer μ'_1 to μ_1 and are active in the $\succsim_w$-Propose stage. By supposition, $w \in W'$, so W' is nonempty. Because μ_1 is individually rational, every worker in W' is active in the $\succsim'_w$-Propose stage but not active in the $\succsim'_w$-Exchange stage.

Next, I show that there always exists a worker w^* and firm f^* such that the following four conditions hold:

1. w^* is active in the $\succsim'_w$-Propose stage; and
2. $\mu_0(f^*) \neq \mu'_1(f^*)$; and
3. (w^*, f^*) is a blocking pair of μ'_1; and
4. $w^* \neq w$.

There are two cases:

1. $\mu'_1(W') = \mu_1(W')$: First, I show that every w' who is active in the $\succsim_w$-Propose stage is also active in the $\succsim'_w$-Propose stage. To see this, note that there is a sequence of workers $w_1, \ldots, w_n \equiv w'$ such that w_k is acceptable[19] to $\mu_0(w_{k+1})$ and w_k is the first worker to propose to $\mu_0(w_{k+1})$ in the $\succsim_w$-Propose stage. Toward a contradiction, suppose that some workers in the sequence are not active in the $\succsim'_w$-Propose stage, and let w_k be the one with the lowest index. Obviously, $k \neq 1$. By construction, w_{k-1} is active in the $\succsim'_w$-Propose stage and

[19]That is, f^* prefers $\tilde{w}$ to $\mu_0(f^*)$.

prefers μ_1' to μ_1 because w_{k-1} does not propose to $\mu_0(w_k)$. By supposition, $\mu_1'(W') = \mu_1(W')$. Therefore, there is some acceptable $\tilde{w} \in W'$ who proposes to $\mu_0(w_k)$ in the $\succsim_w'$-Propose stage. Hence w_k is active in the $\succsim_w'$-Propose stage, a contradiction. Therefore w' is active in the $\succsim_w'$-Propose stage.

Let $F' \equiv \mu_1'(W')$. Fix an arbitrary order of proposals and let f^* be the last firm in F' to receive a proposal from an acceptable worker in W' in the $\succsim_w$-Propose stage. Because μ_1' is individually rational, each worker in W' is acceptable to her μ_1'-partner. Because W' is nonempty and every worker in W' makes a proposal in the $\succsim_w$-Propose stage, such a firm exists.

Because every worker in W' strictly prefers μ_1' to μ_1 and is active in the $\succsim_w$-Propose stage, every firm in F' must have rejected at least one proposal from an acceptable worker in W' in the $\succsim_w$-Propose stage (namely, the firm's μ_1'-partner). Thus f^* was matched to some $w^* \in W$ when she received this last proposal and f^* rejects w^*. Note that w^* cannot be in W'; otherwise, after being rejected by f^*, w^* would have proposed to another firm in F' because $\mu_1(W') = F'$. Hence, $w^* \neq w$. Note that w^* is active in the $\succsim_w$-Propose stage, so he is also active in the $\succsim_w'$-Propose stage. ***This satisfies conditions 1 and 4.***

Next, note that $\mu_0(f^*) \neq \mu_1'(f^*)$ because $\mu_1(f^*) \in W'$ and no worker in W' is active in the $\succsim_w'$-Exchange stage (see earlier comment). ***This satisfies condition 2.***

Finally, note that f^* strictly prefers w^* to $\mu_1'(f^*)$ because f^* must have rejected $\mu_1'(f^*)$ but w^* was tentatively accepted immediately prior to f^* accepting $\mu_1(f^*)$ in the $\succsim_w$-Propose stage. Because w^* is active in both the $\succsim_w$- and $\succsim_w'$-Propose stages and $w \notin W'$, it follows that w weakly prefers μ_1 to μ_1' Because w^* strictly prefers f to $\mu_1(w)$ and w weakly prefers μ_1 to μ_1', it follows that w^* strictly prefers f to $\mu_1'(w^*)$. Therefore, (w^*, f^*) is a blocking pair of μ_1'. ***This satisfies condition 3.***

This completes this case.

2. $\mu_1'(W') \neq \mu_1(W')$: Fix an arbitrary order of proposals and let f^* be the first firm in $\mu_1'(W') \backslash \mu_1(W')$ to receive a proposal from $\mu_1'(f^*)$ in the $\succsim_w'$-Propose stage. Note that $\mu_0(f^*) \neq \mu_1'(f^*)$ because $\mu_1'(f^*) \in W'$ and no worker in W' is active in the $\succsim_w'$-Exchange stage (see earlier comment). ***This satisfies condition 2.***

Let $w^* \equiv \mu_1(f^*)$. Note that $w^* \notin W'$ and thus $w^* \neq w$. ***This satisfies condition 4.***

Let $w' \equiv \mu_1'(f^*)$. Note that w' proposes to f^* in the $\succsim_w$-Propose stage because $w' \in W'$. Therefore, w^* is active in the $\succsim_w$-Propose stage.

Next, I show that w^* is active in the $\succsim_w'$-Propose stage. To see this, note that there is a sequence of workers $w_1, \ldots, w_n \equiv w^*$ such that in the $\succsim_w$-Propose stage, w_k is acceptable to $\mu_0(w_{k+1})$ and w_k is the first worker to propose to $\mu_0(w_{k+1})$. Toward a contradiction, suppose that some workers in the sequence are not active in the $\succsim_w'$-Propose stage, and let w_k be the one with the lowest index. Obviously, $k \neq 1$. By construction, w_{k-1} is active in the $\succsim_w'$-Propose stage and prefers μ_1' to μ_1 because w_{k-1} does not propose to $\mu_0(w_k)$. Therefore, w_{k-1} must propose to $\mu_1'(w_{k-1})$ at an earlier step of the $\succsim_w'$-Propose stage than w' proposes to f^*, a contradiction to the supposition that w' is the first such worker to do so. Hence w_k is active in the $\succsim_w'$-Propose stage, a contradiction. Therefore w^* is active in the $\succsim_w'$-Propose stage. ***This satisfies condition 1.***

Note that w^* strictly prefers f^* to $\mu_1'(w^*)$ because $w^* \notin W'$, w^* is active in both Propose stages, and $f^* = \mu_1(w^*) \neq \mu_1'(w^*)$. Similarly, $w^* \neq \mu_0(f^*)$ because μ_1' is individually rational. Because w' is rejected by f^* in favor of w^* in the $\succsim_w$-Propose stage, it follows that f^* strictly prefers w^* to w'. Therefore, (w^*, f^*) is a blocking pair of μ_1'. ***This satisfies condition 3.***

This completes this case.

Because only w misreports, w^* in each case has the same preferences. Therefore, the conditions of lemma 1.A.6 are met, a contradiction to the supposition that μ_1' is the output of the $\succsim_w'$-Propose stage. This completes the proof.

□

Proof of Theorem 4: This proof has two parts. In the first, I show that $\mu_1'(w) = f$. In the second, I show that w is not active the μ_0'-Propose stage.

Suppose (toward a contradiction) that $\mu_1'(w) \neq f$. I show that every worker who proposes in the μ_0-Propose stage weakly prefers μ_1 to μ_1'. This contradicts the supposition that w strictly prefers μ_1' to μ_1.

First, choose an arbitrary proposal order for the μ_0-Propose stage such that w only makes his first proposal if he is the only active worker. Use the notation $(\tilde{w}, \tilde{f})$ to indicate that $\tilde{w}$ proposes to $\tilde{f}$, and let $(w_1, f_1), (w_2, f_2), \ldots, (w_n, f_n)$ be the order of proposals. By Proposition 2 the output of the Propose stage is independent of the proposal order.

Second, I argue by induction that there is a proposal order for the μ_0'-Propose stage such that the first n proposals are $(w_1, f_1), (w_2, f_2), \ldots, (w_n, f_n)$. In the base case, consider (w_1, f_1). There are two cases:

1. $w_1 \neq w$: Then w_1 or $\mu_0(w_1)$ is a proposal source in μ_0. Thus w_1 or $\mu_0(w_1)$ is a proposal source in μ_0'. Therefore w_1 is active at the start of the μ_0'-Propose stage.

2. $w_1 = w$: Then w is the only active worker at the start of the μ_0-Propose stage. Because $\mu_1'(w) \neq f$, this implies that w is active at some point in the μ_0'-Propose stage. Therefore, w is active at the start of the μ_0'-Propose stage.

Therefore there is a proposal order such that (w_1, f_1) is the first proposal in the μ_0'-Propose stage.

For the inductive step, suppose that there is a proposal order such that $(w_1, f_1), (w_2, f_2), \ldots, (w_{k-1}, f_{k-1})$ are the first $k-1$ proposals in the μ_0'-Propose stage. There are two cases:

1. $w_j \neq w$ *for any* $j < k$: Observe that there are weakly more rejections in the μ_0'-Propose stage. Therefore, the set of active agents is weakly larger in the μ_0'-Propose stage, with the possible exception of w. If $w_k = w$, then w is the only active worker in the μ_0-Propose stage. Because $\mu_1'(w) \neq f$, this implies that w is active at some point in the μ_0'-Propose stage. Therefore w must be active at the k^{th} step of the μ_0'-Propose stage. Therefore w_k must be active at the k^{th} step of the μ_0'-Propose stage

2. $w_j = w$ *for some* $j < k$: Observe that there are weakly more rejections in the μ_0'-Propose stage. Therefore, the set of active agents is weakly larger in the μ_0'-Propose stage because w has been active at least once. Therefore w_k must be active at the k^{th} step of the μ_0'-Propose stage.

Therefore, w makes weakly more proposals in the μ_0'-Propose stage, a contradiction to the supposition that w and f profitably misreport the initial match. Therefore, $\mu_1'(w) = f$.

Suppose (toward a contradiction) that w is active in the Propose phase μ_0'-Propose stage. Let

$$w_1 \equiv w, f_1 \equiv \mu_2'(w_1), w_2 \equiv \mu_0'(f_1), \ldots, f_n \equiv f$$

be the cycle in which w and f sit down in in the μ_0'-Exchange stage.

Consider any w_k in this cycle. If w_k is active in the μ_0'-Propose stage, then w_k proposes to f_k in the μ_0'-Propose stage because $\mu_1'(w_k) = \mu_0'(w_k)$. Because $\mu_2'(f_k) = w_k$, it follows that f_k weakly prefers w_k to $\mu_0'(f_k)$. Because f_k rejects w_k at some point of the μ_0'-Propose stage, it then follows that w_{k+1} is active in the μ_0'-Propose stage. By supposition, w is active in the μ_0'-Propose stage.

Therefore, w_n is active in the μ_0'-Propose stage. Therefore, w_n proposes to f in the μ_0'-Propose stage but f rejects w_n. Because f strictly prefers $\mu_2'(f)$ to $\mu_2(f)$, and weakly prefers $\mu_2(f)$ to being unmatched, it follows that f does not reject a proposal from w_n, a contradiction. Therefore, w is not active in the μ_0'-Propose stage. $\square$

Chapter 2

A UNIFIED THEORY OF SCHOOL CHOICE

2.1 Introduction

School choice programs[1] allocate school seats using *priorities* (such as sibling status or proximity): when a school is oversubscribed, these priorities determine which students are admitted. The interpretation of these priorities is fundamental to the program's design, and there are two leading interpretations. The stronger interpretation posits that priorities are the right to attend the school ahead of those with lower priorities. In the weaker interpretation, priorities reflect "better opportunities" to attend a school, all else equal (Abdulkadiroğlu and Sönmez, 2003).

These interpretations underpin the central trade-off in school choice between fairness and efficiency. *Fairness* means no student is denied admission to a school in favor of a lower-priority student; *efficiency* means students are assigned to schools in such a way that no student can be improved without harming another student. By requiring the stronger interpretation, the designer arrives at a fair match because no student prefers a school that a lower-priority student is assigned to. However, allowing for the weaker interpretation shifts the set of permissible assignments to include an efficient match. The main dilemma is that no algorithm delivers a match that is both fair and efficient.

Despite the appeal of efficiency, policymakers consistently choose fairness.[2] Empirically, the cost is substantial: for instance, Abdulkadiroğlu, Pathak and Roth (2009) finds that the assignments of 4300 eighth-grade students in New York City could have been improved without harming any other students. In response to this, recent research focuses on weaker versions of fairness to allow for efficiency gains.

However, an examination of why policymakers side with fairness reveals that they are frequently concerned with the violations of some priorities but not with others. For instance, a priority derived from a high test score may have a stronger interpretation

[1] What economists refer to as *school choice*—the ability to choose between public schools—is frequently called "open enrollment" in the media. Although, in common parlance, "school choice" refers to voucher or tax credit programs for defraying the costs of private schooling or homeschooling, I will use the standard language in economics.

[2] New Orleans Recovery School District and the Common App is the only example I know of that sought an efficient outcome in some rounds of the assignment process, but this was abandoned after the first year.

than a priority assigned based on proximity to the school. Because economists have not considered how policymakers may treat some priorities differently than others, this omission has precluded a promising approach to reconciling fairness and efficiency. The key to understanding how policymakers view priorities lies in the two-step process that is typically used to construct priorities.

In the first step, the policymaker identifies a set of student characteristics to prioritize and uses these characteristics to partition students into priority groups.[3] Then, the policymaker assigns a priority over the priority groups that reflects the policymaker's objectives for the school; this is the *between-group* priority. In the second step, within each group, another *within-group* priority is provided to break ties, which could depend upon other student characteristics or a random lottery. A student's priority at a school is lexicographically determined first by her between-group priority and then by her within-group priority.

Policymakers understand that because priorities arise from different sources, they require different interpretations. Allocating a seat to a student in a lower priority group ahead of a student in a higher priority group may be unacceptable, but assignments within groups may be more subtle. The prominent case of Boston Public Schools (BPS)—the first district to carefully consider its assignment mechanism—illustrates this difference.

BPS created two priority groups: students with a sibling attending the school and all other students, with the former given higher between-group priority than the latter. Within each priority group, BPS gave higher within-group priority to students residing in the school's walk zone, and broke any remaining ties with a random lottery. When comparing the efficient algorithm with the fair algorithm, Superintendent Payzant implies that the between-priority derived from having a sibling present should be treated differently from the within-priorities:

> [The efficient system] presents the opportunity for the priority of one student at a given school to be "traded" for the priority of a student at another school, [. . .] There may be advantages to this approach, [. . .] It may be argued, however, that certain priorities – e.g., sibling priority – apply only to students for particular schools and should not be traded away. (Abdulkadiroğlu et al., 2006)

[3]For example, these characteristics frequently include whether the student has a sibling attending the school, whether the student is within a particular geographic region, the student's test scores, etc., and a priority group consists of the students with the same set of characteristics.

The implication from his silence concerning walk-zone priority is that, in some cases, walk zone priorities *could* be traded away. Sibling priority has the stronger interpretation while walk zone priority has the weaker interpretation of a better opportunity to attend the school.[4]

Despite the different interpretation of priorities within these two steps, economists usually remain agnostic about the source of the priority. The nuances between priority groups are dismissed, and all priorities are treated equally. The existing options afforded to policymakers do not allow them to express these nuances. This leaves policymakers in a quandary because a more efficient match is better than not, but violating the stronger priorities may be impermissible. When BPS faced this decision, they erred on the side of respecting priorities and implemented the fair mechanism even though only a subset of their priorities required the stronger interpretation.

In this paper, I introduce a model that explicitly distinguishes between two types of priority—between-group and within-group—and I propose the unified core to connect the two. In my model, each school's priority consists of two layers. The first layer is the between-group priority, a weak preference that specifies which students belong to which priority group and how those groups are ordered. The second layer is the within-group priority, a strict preference that refines the between-group priority.[5] Accordingly, only the between-group priority is interpreted as a right to attend the school ahead of lower-priority students, and it cannot be violated. On the other hand, the within-group priority has the milder interpretation of an opportunity to attend the school; it is a means of allocating the school seat, but it is not inviolable. The interpretation of the within-group priority is weak enough to guarantee efficiency within a group while still using the priority to allocate the seat.

The unified core connects the two cores that underlie the fair and efficient algorithms, and is implemented by carefully combining the algorithms that implement the fair

[4]Readers familiar with this story will note that my exposition of the priority groups is slightly different than that of Abdulkadiroğlu et al. (2006). Formally, BPS designated five priority groups: continuing students, sibling-walk, sibling, walk, and all others. However, as Payzant's quote illustrates, sibling priority and walk zone priority are normatively different. Conveniently, the order on the priority groups allows for walk zone priority to be viewed as a tiebreaker within the sibling priority group. If BPS ranked these groups differently, my model would not apply directly. For ease of exposition, I exclude the continuing students from the model; continuing students can be easily added to the model by including an additional priority group.

[5]In most applications, the within-group priority consists of a random lottery. However, I refrain from referring to it as a *tiebreaker* because the within-group priority may include substantive components, such as geographic location (as in the case of BPS). The emphasis, however, is that these priorities have a weaker interpretation.

and efficient cores. The core underlying the fair algorithm is the *fair* core of Gale and Shapley (1962) (henceforth GS), which allows a student to claim a seat at a school if she has a higher priority than a student currently there. This stems from the strong interpretation of priorities and treats the school as a strategic player and the priority as the school's preference. The result is a fair match because any unfair match can be blocked by the student whose priority is violated.

The core underlying the efficient algorithm is from Rong, Tang and Zhang (2022) (henceforth RTZ),[6] who develop their *efficient* core to provide a foundation for the use of several efficient algorithms in school choice. RTZ's efficient core allows a student to take possession of a school provided no higher-priority students can veto that action. In effect, the efficient core allows for a form of tradable priorities, where a low-priority student can receive a priority from another student so long as no higher-priority students are harmed. As is paralleled in many economic models, the power to trade freely (in this case, trade school seats) guarantees that an efficient allocation is reached.

Whether policymakers prefer the outcomes of the fair core or the efficient core ultimately depends on the source of the students' priorities. If the priority reflects different priority groups, then the fair core is appealing; if it reflects within-group tiebreakers such as a random lottery, the efficient core allows more students to attend the schools that they prefer. However, neither the fair core nor the efficient core allows for combinations of the two priorities.

The unified core resolves this tension and connects the two cores by allowing the policymaker to explicitly designate two kinds of priorities. By treating the between-group priorities as in the fair core while the within-group priorities as in the efficient core, the unified core generalizes both cores. Students can always claim a seat at a school if they have a higher between-group priority than a currently assigned student, as in GS. Within-group priorities, however, may be traded amongst students within the same priority class, as in RTZ. This guarantees a form of "within-group" efficiency. The way I implement this mirrors RTZ: a student can take a seat provided no within-group interrupters in her priority group can block her.

The main challenge in crafting a solution for this model is integrating the differing normative implications of the two priorities. Unlike the model of Erdil and

[6] It is worth mentioning here that RTZ's core identifies precisely the *just* assignments of Morrill (2015).

Ergin (2008) (henceforth EE),[7] the within-group priorities are not merely random tiebreakers. Within-group priority may encode substantive differences (as in the case of BPS) that must be respected, just not in the same fashion as the between-group priority. The difficulty lies in creating a unified model that allows for both interpretations, one between groups and the other within groups.

The power of the unified core lies in its ability to interpolate between the frameworks of GS and RTZ. When the between-group priority is a strict ranking where each student forms a singleton priority group, the unified core corresponds to the fair core of GS. On the other hand, when the between-group priority is indifferent over all students (there is a single priority group), then the unified core corresponds to the efficient core of RTZ. At intermediate stages, the unified core blends both RTZ and GS in a principled manner.

To make the unified core practical, I introduce a novel two stage algorithm—the DA-TTC algorithm—that always produces a match in the unified core. In the first stage, I allow students to freely apply to their most-preferred schools as in the Deferred Acceptance algorithm of GS. Schools tentatively accept the highest-priority applicants and immediately reject the rest. The process continues until no more students are rejected, at which point the first stage terminates and returns the match μ_1.

In the second stage, I adapt Gale's Top Trading Cycles algorithm but with a restriction on the allowable trades (Shapley and Scarf, 1974). Essentially, students may only trade their seats (that they have from μ_1) to other students in the same priority group. I iteratively find groups of students who can, by trading their assigned schools among themselves, be matched to their most-preferred school out of the ones currently available. This algorithm is the crux of this paper, and is presented in Section 2.3. I label the resulting match μ_2. Once all students have been removed, the second stage terminates and returns the match μ_2.

My main result establishes that μ_2 belongs to the unified core. The advantage of my approach is a practical algorithm to finding an element of the unified core. The DA-TTC is also computationally feasible.[8]

---

[7] I discuss the differences in models in Section 2.1, and the differences in our techniques in Section 2.4. In short, EE uses only between-group priorities; however, I must use more care because of the within-group priorities. This prevents a direct application of the Stable Improvement Cycles algorithm.

[8] It runs in polynomial time.

A key advantage of the unified core in applications is the simplicity of its definition.[9] First, there are no between-group priority violations, so the fairness property of the fair core is preserved. However, when there is a within-group priority violation, the explanation is straightforward: rectifying the violation would harm a student with higher within-group priority. By focusing on a reasonable outcome, the definition of the unified core lends itself to applications in school choice.

The rest of this paper is structured as follows. First, I close out this section with a discussion of the relevant papers, emphasizing the differences between the unified core, related concepts, and the methods used. Section 2.2 introduces the model, the bulk of which is devoted to the distinction between within-group and between-group priorities. In Section 2.3, I develop my main result (Theorem 2.1) which shows that a match in the unified core can always be found using an efficient algorithm. I conclude in Section 2.4 with a discussion of the result and further avenues of research.

Related Literature

This paper connects several strands of research.

First, as mentioned previously, the model in this paper has substantial overlap with Erdil and Ergin (2008) (henceforth, EE). The key difference between my model and EE's model is that I include the within-group priorities in the model primitives, while they take these priorities as randomly generated from the between-group priorities. Although mathematically similar, because the within-group priorities may carry normative implications, the outcomes in the unified core are restricted. In EE, any student within the same priority group can be assigned to the school so long as the match is Pareto optimal among the matches that are stable with respect to the between-group priorities. In contrast, in the unified core, a student of higher within-group priority can claim a seat so long as no higher priority students in the same group are harmed. The within-group priority disciplines how seats are allocated within priority groups.

Second, ever since Abdulkadiroğlu and Sönmez (2003), researchers have been trying to modify the Deferred Acceptance algorithm to accommodate efficiency gains (Ehlers and Morrill, 2020; Kesten, 2010; Troyan, Delacrétaz and Kloosterman, 2020; Reny, 2022). These approaches, however, take the priorities as given and weaken the definition of the core. The unified core complements this approach by

[9]Indeed, Payzant expressed reservation concerning the TTC because of the complexity of the algorithm. I return to this topic in the discussion.

weakening the priorities and then adapting the core to this framework. The upshot of the unified core is that the definition of the core is simpler, but the downside is that the preferences must have a two-stage structure to make any efficiency gains.

Third, a separate approach has been to provide foundations for the use of the Top Trading Cycles algorithm in school choice (Abdulkadiroğlu and Che, 2010; Morrill, 2013a; Abdulkadiroğlu et al., 2017; Chen, Chen and Hsu, 2021; Rong, Tang and Zhang, 2022; Dur and Paiement, 2024). These axiomatic definitions ground this work, and I extend our understanding of these mechanisms by integrating them into the fairness framework of GS.

The paper closest in spirit to mine is Abdulkadiroğlu (2011). The main difference between his paper and mine is the restriction they place on between-priorities: in their model, between-group priorities either express indifference or a strict preference over all students. Additionally, his focus is on student optimality rather than on the core. He uses the within-priorities in a similar way, but only in his algorithm; the unified core provides the underpinnings for his solution.

2.2 Model

In this section I present the formal model of a *school choice problem* and my key definitions. In the first subsection, I introduce the standard primitives, except that I replace a school's one priority with two priorities: the between-group priority and within-group priority. In the second subsection, I develop the theory of the unified core. In the third subsection, I turn to the solutions of GS, RTZ, and EE, and highlight the differences and advantages of the unified core.

The Primitives

There is a finite set of *students* I and a finite set of *schools* S, collectively called *agents*. Each school has a *capacity* $q_s \geq 1$ of seats. A *match* is a function $\mu : I \to I \cup S$ with the following two properties:

1. if $i \in I$, then $\mu(i) \in S \cup \{i\}$; and
2. if $s \in S$, then $|\mu^{-1}(s)| \leq q_s$.

The first requirement guarantees that the market is two-sided: students are matched to schools or unmatched (this is denoted as being matched to oneself). The second requirement guarantees that capacities are not exceeded.

Every student has a strict preference $\succ_i$ over $S \cup \{i\}$; the associated weak preference is $\succsim_i$. A match μ is *individually rational* if for every $i \in I$, $\mu(i) \succsim_i i$. When considering a set of students $C \subseteq I$, I say that C *prefers* ν to μ if for every $i \in C$, $\nu(i) \succsim_i \mu(i)$ and the comparison is strict for some student in C.

Schools, however, come equipped with two orders over I, the between-group priority and the within-group priority.[10] The *between-group priority* order is a weak preference $\succsim_s$.[11] It represents the priorities that require the stronger interpretation as a right to attend the school ahead of others. When $\succsim_s$ does not rank two students strictly, it means that they are in the same priority group. Formally, denote the *priority group* of student i at s as

$$[a]_s \equiv \{j \in I : j \sim_s i\}.$$

The *within-group priority* order is a strict preference $\succ_s^*$. It represents the priorities that are allowed the weaker interpretation of a better opportunity to attend the school. In practice, the priority that a policymaker uses is the lexicographic combination of $\succsim_s$ and $\succ_s^*$: first, $\succsim_s$ is used to partition students into groups, and then $\succ_s^*$ is used to break ties within groups. Rather than define a combined priority, I assume that $\succ_s^*$ is a refinement of $\succsim_s$; that is, if $a \succ_s b$, then $a \succ_s^* b$. That is, $\succ_s^*$ is what most researchers refer to as school priority, whereas for me it is the result of the policymaker breaking ties within priority groups.

Because this model is many-to-one, it is necessary to construct the school's between-group priority over groups of students.[12] Toward this end, I make the standard assumption that the school's between-group priority over groups of students is *responsive* (Roth, 1985). In words, this means that for equally-sized groups, the first has greater between-group priority than the second if every student in the first can be paired with a student in the second such that the first has greater between-group priority. For unevenly sized groups, the smaller group never has greater between-group priority than the larger group, but the larger group has greater between-group priority if there is a subset of the larger group that is the same size as the smaller group and has greater between-group priority.

[10] Implicit here is the assumption that every student is acceptable to every school. Extending this model to incorporate unacceptable students is straightforward, but adds unnecessary notation.

[11] The irreflexive portion is $\succ_s$ (the strict preference relation) and the reflexive portion is $\sim_s$ (the indifference relation).

[12] Extending the within-group priority in a similar manner is unnecessary in my analysis.

Formally, I construct $\succsim_s$ for groups of students in the following way. For $C, C' \subseteq I$ such that $|C| = |C'|$, I write $C \succsim_s C'$ if the students in C and C' can be indexed such that $c_1 \succsim_s c'_1, c_2 \succsim_s c'_2, \ldots, c_{|C|} \succsim_s c'_{|C|}$. When $C \succsim_s C'$ and $C' \succsim_s C$, I write $C \sim_s C'$.[13] If $C \succsim_s C'$ but not $C' \succsim_s C$, then I write $C \succ_s C'$. When $|C| \neq |C'|$, I write $C \succsim_s C'$ if $|C| > |C'|$ and there is some $\tilde{C} \subseteq C'$ such that $\tilde{C} \succsim_s C'$.

Unified Core

With the primitives in hand, I turn to a discussion of the core of the model. As introduced earlier, policymakers view between-group and within-group priorities differently. To accommodate both types of priorities, I introduce two separate types of blocks. In both cases the conditions on student preferences are the same: the coalition of students must prefer the new match to the old one. What changes is the definition of what the coalition can enforce.

The between-group priorities resemble the preferences in the GS two-sided model. However, within my framework I focus on students as the active players. To accommodate this, Definition 2.1 allows a student to block a match if her between-group priority has been violated. This definition mirrors the standard blocking definition of GS. It rules out any violations of the between-group priority order.

Definition 2.1. The coalition C can *between-group enforce* match ν over match μ if for every $i \in C$, either there is a student $j \in \mu^{-1}(\nu(i))$ such that $i \succ_s j$ or $|\mu^{-1}(\nu(i))| < q_{\nu(i)}$.[14]

The definition of blocking is standard: a coalition can *between-group block* match μ if it can between-group enforce some match ν that the coalition prefers to μ. Note that if a match is not individually rational, then it is between-group blocked by some student who wishes to be unmatched.

Between-group blocks are useful in school choice models, but do not fully model the priorities given by schools. As seen in EE, using between-group blocks alone can provide substantial welfare improvements for students. However, the within-priorities may also carry normative value as Payzant implied. To address this, I turn to a treatment of within-group priorities.

[13] Note that $\succsim_s$ may not be a complete relation over groups of students: there are some coalitions C and C' such that neither $C \succsim_s C'$ nor $C' \succsim_s C$.

[14] Traditionally in GS-style models, the schools are included in the coalition; however, in RTZ-style models (see next paragraph), this is not the case. I present this student-only version for use here. The downside is that the definition of enforcement relies on μ; the upside is fewer cases in the definition of the unified core.

The within-group priorities are modeled as in the RTZ framework. In their model, a coalition of students can enforce a match to some schools only if no group of students *outside* the coalition could veto such a match by asserting their stronger claims. I import this definition into my model by restricting attention only to students within the same priority group. To formalize this, I first define the *within-group upper contour set*:

$$U_s(i) = \{j \in [i]_s : j \succsim_s^* i\}.$$

In words, j is in $U_s(i)$ if j has a higher within-group priority at s. When considering whether i can claim a seat at s, $U_s(i)$ are the students who could prevent this match because they have a stronger within-group claim to s. For example, suppose $q_s = 1$; if i wishes to match to s, then i must guarantee that students in $U_s(i)$ are not harmed by this action. I call students in $U_s(i)$ but not in C the *within-group interrupters* of i. A sufficient condition is that there are no within-group interrupters. When $q_s > 1$, it is not necessary to eliminate every within-group interrupter: if n students from C match to s, then $q_s - n$ within-group interrupters do not need to be included because the within-group interrupters (even when acting together) are unable to prevent the coalition members from claiming their seats at the school.[15] Definition 2.2 formalizes this intuition.

Definition 2.2. A coalition $C \subseteq I$ can *within-group enforce* match ν over match μ if for every school $s \in \nu(C)$, the following two conditions hold:

1. $\nu^{-1}(s) \sim_s \mu^{-1}(s)$, and
2. the sum of the number of within-group interrupters across all students in C that are matched to s is less than or equal to $q_s - |\nu^{-1}(s)|$.

Again, the definition of blocking is symmetric: a coalition can *within-group block* match μ if it can within-group enforce some match ν that the coalition prefers to μ.

The main difference between within-group blocks and between-group blocks is their treatment of schools. Schools must strictly benefit (according to the between-priority) from between-group blocks, but may be indifferent in a within-group

[15] The reader may wonder why I (and RTZ) include this slackness in the definition of enforceability. Without it, a student with top priority might trade *multiple* seats away. The intuition is that, if every within-group interrupter is included, then including the highest-priority student might necessitate (through a chain of enforceability claims) including two or more lower-priority students who are matched to the school. An example demonstrating this is available upon request.

block. The within-group block compensates by placing a stronger condition upon the student coalition.

With both types of blocks in hand, I turn to defining the unified core in definition 2.3.

Definition 2.3. μ is in the *unified core* if it is neither between-group blocked nor within-group blocked.

In Section 2.3, I provide a two-stage algorithm that always finds a match in the unified core. Before introducing the algorithm, I first discuss the two most-closely related solutions.

Comparison with Related Models

In this subsection, I briefly introduce the models of GS and RTZ.

In the fairness framework of GS, a match μ is *fair* if there is no student i and s such that $s \succ_i \mu(i)$ and either there is a $j \in \mu^{-1}(s)$ such that $i \succ_s^* j$ or $|\mu^{-1}(s)| < q_s$. Note that when the between-group priority orders all students strictly, then the set of fair matches and the unified core are the same.

In contrast, the efficiency framework of RTZ uses priorities to allocate school seats as if they were objects. The within-group priority represents a form of ownership of the school and allows for a form of trading. The key difference between the efficient core and the unified core is that the efficient core assumes that there is a single priority group: any group of students can use their priorities to trade their schools. This is made by a slight change in definitions: the *efficient upper contour set* is

$$U_s^*(i) = \{j \in I : j \succ_s^* i\}$$

and for a coalition C and match ν, the *efficient interrupters* of $i \in C$ are the students in $U_{\nu(i)}^*(i)$ but not C. The definition of *efficient enforcement* is then the same as within-group enforcement except that the condition $\nu^{-1}(s) \sim_s \mu^{-1}(s)$ is dropped and the efficient interrupters are counted instead of within-group interrupters. The definition of efficient blocking and the efficient core is similar. When the between-group priority places all students within the same priority group, the efficient core and the unified core are the same.

The power of the unified core is its ability to interpolate between the efficient and fair cores. By allowing for the between-group priority to range between a strict ranking of all students and complete indifference, the unified core captures the nuances of policymakers' objectives and the two-stage nature of priorities. In example 21, I show how the unified core models what the fair and efficient cores cannot.

Example 21. Consider a school choice problem with six students and six schools, each school having one seat. The students' preferences and the schools' between- and within-priority orders are as follows (unlisted preferences/priorities are irrelevant):

$\succ_1$	$\succ_2$	$\succ_3$	$\succ_{1'}$	$\succ_{2'}$	$\succ_{3'}$
a	b	a	a'	b'	a'
b	a	c	b'	a'	c'
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

$\succ_a$	$\succ_b$	$\succ_c$	$\succ_{a'}$	$\succ_{b'}$	$\succ_{c'}$
$1, 2, 3$	$1, 2$	3	$2'$	$1'$	$3'$
$\vdots$	$\vdots$	$\vdots$	$3'$	$2'$	$\vdots$
			$1'$	$\vdots$	
			$\vdots$		

$\succ^*_a$	$\succ^*_b$	$\succ^*_c$	$\succ^*_{a'}$	$\succ^*_{b'}$	$\succ^*_{c'}$
2	1	3	$2'$	$1'$	$3'$
3	2	$\vdots$	$3'$	$2'$	$\vdots$
1	$\vdots$		$1'$	$\vdots$	
$\vdots$			$\vdots$		

Notice that this example consists of two parallel problems side-by-side. In the first, the students 1 and 2 are the highest-priority students at the school the other most-prefers (schools a and b, respectively), and in the second, students $1'$ and $2'$ are the highest-priority at the school the other most-prefers (schools a' and b', respectively). The fair core is the match:

$$\mu^{\text{GS}} = \begin{pmatrix} 1 & 2 & 3 & 1' & 2' & 3' \\ b & a & c & b' & a' & c' \end{pmatrix}.$$

The fair core does not incorporate the difference between the between-group and within-group priorities; hence it is less efficient than allowed. However, the efficient core is also unsatisfactory:

$$\mu^{\text{RTZ}} = \begin{pmatrix} 1 & 2 & 3 & 1' & 2' & 3' \\ a & b & c & a' & b' & c' \end{pmatrix}.$$

The difficulty with RTZ is that it does not recognize that student $3'$ can claim a seat at school a' ahead of student $1'$.

The unified core, however, correctly identifies the difference between these priority groups:

$$\mu^{\mathrm{UC}} = \begin{pmatrix} 1 & 2 & 3 & 1' & 2' & 3' \\ a & b & c & b' & a' & c' \end{pmatrix}.$$

2.3 Analysis

In this section, I present my main result (Theorem 2.1). It states that the unified core is non-empty for every school choice problem and provides an algorithm for finding one such match.

I use a two-stage algorithm to prove the result. In the first stage, I leverage the Deferred Acceptance algorithm (DA) to produce a match that is not between-group blocked (Gale and Shapley, 1962). In the second stage, I build a variation of the Top Trading Cycles algorithm (TTC) to remove any within-group blocks from the previous match while not adding any between-group blocks (Shapley and Scarf, 1974). The match produced is in the unified core.

Removing between-group blocks: The Deferred Acceptance Algorithm

In this subsection, I present the first stage of the DA-TTC: the Deferred Acceptance algorithm from GS. In the DA, students initially start unmatched. In the first round, students "propose" to their favorite school. Every school tentatively accepts students up to its capacity (picking the highest according to its within-group priority) and immediately rejects the rest. In subsequent rounds, students who are not tentatively accepted propose to their favorite school among those that have not rejected them. Every school considers the new proposals simultaneously with the students it has tentatively accepted, and again rejects all but the top students up to its capacity. This process continues until every student is matched or has been rejected by every acceptable school. I formally write this in Algorithm 5.

The DA was originally designed by GS to construct a fair match. Fairness is a stronger requirement than not being between-group blocked, this result applies to my model. Lemma 2.1 translates the standard result into my model. The proof is standard and is relegated to Section 2.A.

Lemma 2.1. *μ_1 is not between-group blocked by any coalition.*

Algorithm 5 Deferred Acceptance Algorithm

every student i points to the $\succsim_i$-highest agent
while more than q_s students point to some school s **do**
  s rejects all but the $\succ_s^*$-top q_s students who proposed to s
  every student i points to the $\succsim_i$-highest agent *who has not rejected him yet*
end while
μ_1 assigns each student to the agent he last proposed to

Removing within-group blocks: The Top Trading Cycles Algorithm

In this subsection, I present the second stage of the DA-TTC: the Top Trading Cycles algorithm of Shapley and Scarf (1974). In the TTC, each student "owns" a seat at the school she is matched to in μ_1. Students are allowed to trade the seats that they own with others, but only within priority groups.

The way that I implement this is by first restricting which students and schools are active, and then restrict which schools a student can "point" to. At the start of the TTC, each student i is *active* if she is in the lowest priority group of $\mu_1(i)$ among the students matched to $\mu_1(i)$ (i.e. for every $j \in \mu_1^{-1}(\mu_1(i))$, $j \succsim_{\mu_1(i)} i$). Throughout the TTC, each school s is active if at least one student in $\mu_1^{-1}(s)$ is active (at the start of the TTC, every school such that $|\mu_1^{-1}(s)| \geq 1$ is active by construction).

In every round of the TTC, for each student, construct the list of *admissible* schools. A school s is admissible to student i if there is some active $j \in \mu_1^{-1}(s)$ such that $i \sim_s j$. Each student points to her most-preferred admissible school, and each active school points to the highest-priority active student in $\mu_1^{-1}(s)$ (by definition, a school is only active if there is an active student in $\mu_1^{-1}(s)$, so this is well-defined). Because every active student points to an active school, and every active school points to an active student, a cycle must form. Every student on the cycle is assigned a seat at the school she points to and becomes inactive. The process is repeated until no students (and hence no schools) are active. Algorithm 6 provides the formal definition.

Theorem 2.1. *μ_2 is in the unified core.*

The full proof of Theorem 2.1 is available in Section 2.A. Here, I outline the key points in the argument. A preliminary step is to show that μ_2 is not between-group blocked; given that μ_2 is a Pareto improvement of μ_1 for the students but is equivalent under the between-priorities, this is not difficult to show.[16] The rest of the proof

[16]But the importance of this cannot be overstated; as I discuss in Section 2.4, having to find a Pareto improvement of μ_1 is has restricted the kinds of algorithms available in school choice.

Algorithm 6 Top Trading Cycles Algorithm

initialize $\mu_2 \leftarrow \mu_1$
for all $i \in I$ **do** ▹ activates students in the lowest-priority group at their assigned school
  initialize i as *active* if $j \succsim_{\mu_1(i)} i$ for every $j \in \mu_1^{-1}(\mu_1(i))$
end for
for $s \in S$ **do**
  initialize s as *active* if there is an active student in $\mu_1^{-1}(s)$
end for
while there is an active student **do**
  for every active student i **do** ▹ students point to most-preferred admissible school
    i points to her $\succ_i$-most preferred active school in $\{s :$ exists active $j \in \mu_1^{-1}(s)$ s.t. $i \sim_s j\}$
  end for
  each active school points to the highest-priority active student in $\mu_1^{-1}(s)$
  a cycle $i_1, s_1, \ldots, i_n, s_n$ exists
  for all $1 \leq k \leq n$ **do** ▹ removes students in cycle from problem, then restarts
    set $\mu_2(i_k)$ to s_k
    deactivate i_k
    if there are no active students in $\mu_1^{-1}(s_k)$, then deactivate s
  end for
end while

supposes that C within-group blocks μ_2 with ν, and then finds a contradiction.

The major difficulty with applying a standard proof that the TTC is in the core is that, for a school s in $\nu(C)$, not every student in $\nu^{-1}(s)$ is necessarily in C. Put differently, using the TTC usually requires a well-defined set of owners, which the unified core lacks. The crux of the proof is a construction of a bijection T_s between students in $\nu^{-1}(s) \cap C$ and students in $\mu_2^{-1}(s) \cap C$, which identifies an owner for each relevant seat. I construct T_s by addressing students in both $\nu^{-1}(s) \cap C$ and $\mu_2^{-1}(s) \cap C$ separately from those in only $\nu^{-1}(s) \cap C$. For those in both, I make T_s the identity function. For those in only $\nu^{-1}(s) \cap C$, I leverage the $\nu^{-1}(s) \sim_s \mu_2^{-1}(s)$ of within-group enforcement requirement to find an equally-sized set of students in $(\mu_2^{-1}(s) \cap C) \backslash \nu^{-1}(s)$. I further show that if i is active in the TTC, then $T_{\nu(i)}(i)$ is also active in the TTC and $T_{\nu(i)}(i) \in U_{\nu(i)}(i)$. Hence, for active students, the T_s essentially identifies which student owns which seat at s.

With the bijection T_s in hand, I then can use standard methods. I construct a cycle of students $i_1, \ldots i_n$ in C where $\nu(i_k) = \mu_2(i_{k+1})$ as in most proofs involving the

TTC. I construct it such that least one student in this cycle must strictly prefer ν to μ_2. But I also show that every i_k is active in the Top Trading Cycles algorithm, so each must be deactivated only after the next student. This then leads to a situation in which i_k must be deactivated strictly before i_k is, a contradiction.

2.4 Discussion

In this section, I place Theorem 2.1 and the DA-TTC in conversation with existing results and point toward several avenues for future research. I first compare this result with that of EE, highlighting the complexities arising from the within-group blocks. I then turn to the question of finding a constrained efficient match. Finally, I turn to a more general discussion of the DA and TTC in school choice, and highlight the difficulties present in bridging these two algorithms.

With the proof of the non-emptiness of the unified core in hand, I turn to a discussion of the differences between the DA-TTC and the Stable Improvement Cycles algorithm (DA-SIC) of EE. For context, EE considers the matches that are not between-group blocked. The DA-SIC starts by running the DA using the within-group priorities to establish a baseline match that is not between-group blocked. The SIC then checks for a cycle of students who each prefer the next student's match such that, if the students exchange seats, no between-group blocks are created; this is the eponymous *stable* improvement cycle. The output of the SIC stage is a student-optimal match among those that are not between-group blocked. The critical difference between the SIC and the TTC is that the SIC allows for trades *across* priority groups while the TTC does not. The following example with three students and two schools (each with one seat) illustrates this:

$\succ_1$	$\succ_2$	$\succ_3$
a	b	a
b	a	b

$\succ_a$	$\succ_b$
2	1
$1,3$	$2,3$

$\succ_a^*$	$\succ_b^*$
2	1
3	3
1	2

$$\mu^{\text{DA}} = \begin{pmatrix} 1 & 2 & 3 \\ b & a & 3 \end{pmatrix}.$$

At the match μ^{DA}, EE identifies a stable improvement cycle of $1 \rightarrow a$ and $2 \rightarrow b$. However, allowing this pair to trade schools violates student 3's within-priority at both schools: after the trade, the coalition of $\{3\}$ can within enforce a match to either a or b. In effect, the stable improvement cycle is stable with respect to $\succsim$, but disregards $>^*$.[17] This allows for these trades between priority groups. This difference is not a mere technical nuance; in the case of BPS, it is precisely trades like this that concern the policymaker.

Thus far, I have dealt only with existence rather than constrained-efficiency. The major drive of EE is to find a student-optimal match subject to the stability constraints, which they show can be found by iteratively eliminating stable improvement cycles. Given the similarity between the DA-TTC and the DA-SIC, the natural question arises as to whether an analogous method can be used to eliminate "unified" stable improvement cycles. Unfortunately, this approach seems doomed to failure. Although for the between-priority the stable improvement cycle can be imported as-is, checking whether a cycle respects within-priority is significantly more complex. The crux is that a student's within-priority can be violated when a higher-priority student trades her seat to a lower-priority agent. However, in a cycle, whether such a "trade" can be found depends upon the matches of students outside the cycle. Put differently, a stable improvement cycle may involve trades across priority groups, but then a trade within the priority group must be found to protect the former trade. The problem is compounded by the possibility that the latter trade may itself be the result of another simultaneous trade. Whether a constrained efficient match can be found using this technique or others is an open question.

There are several open questions about algorithms similar to the DA-TTC. The DA-TTC is a part of a growing set of algorithms that can be roughly described as "DA baseline, then trade to improve." Starting with Abdulkadiroğlu and Sönmez (2003), several authors have used this formula, such as EE, Kesten (2010), and Doe (2024). The difficulty with combining these two families of algorithms is maintaining monotonicity on a set of matches. Echenique and Oviedo (2004) shows how the DA can be viewed as a monotonic function on matches; however, the TTC does not share this property (Echenique, Goel and Lee, 2024). When combining these two algorithms, care must be taken to guarantee that the trades made in the TTC do not upset the match found in the DA, which is why the previous papers allow students only to trade the school they received from the DA rather than any school they have the highest-priority for. Going the other way around is tougher: if the

[17] There is a second difference between the DA-TTC and the DA-SIC. Consider a school choice problem with one school (with one seat) and two students, both of whom desire the seat at the school. If the students are in the same priority group, then the DA-SIC could assign either student to the school. The DA-TTC, however, is constrained to only assign the higher-priority student. Put differently, the DA-TTC can emulate the TTC because the DA-TTC makes use of the within-group priority; the DA-SIC cannot.

designer first allows students to trade schools and then attempts to remove fairness violations, a student may trade for a school that is taken away; with one student-school pair removed from the group of trading students, the outcome of the TTC shifts unpredictably. The only paper to tackle both simultaneously is Abdulkadiroğlu (2011). Is this approach to connecting the DA and TTC the only viable option? Can conditions be placed upon priorities to ameliorate the non-monotonicity of the TTC in the presence of the DA?

Another obstacle to implementing the TTC is the complexity of the algorithm. Payzant alludes to this; after his previous quote, he continued: “Moreover, Top Trading Cycles is less transparent – and therefore more difficult to explain to parents – because of the trading feature executed by the algorithm, which may perpetuate the need or perceived need to ‘game the system.’” Concerns like Payzant’s have spurred a growing body of literature to understand the complexity of the TTC (Gonczarowski and Thomas, 2024; Leshno and Lo, 2017). In this paper, I instead focused on properties of the match rather than the process used to reach the match. The properties of a match may prove more transparent to stakeholders than the properties of an algorithm.

I close with a brief note about the DA and TTC. Within the literature on school choice, the vast majority of algorithms stem from the DA and the TTC. The monotonicity of the DA is well-suited for the lattice structure of the stable matches, and the TTC flexibly handles ownership economies (Gusfield and Irving, 1989; Papai, 2000). However, as I note in the previous paragraphs, these algorithms are difficult to combine. Additionally, given the complexity of the TTC and how few districts have attempted to implement it, relying on TTC-style algorithms to improve the efficiency of the DA may not be successful in applications. Developing general algorithms—as in Abdulkadiroğlu (2011)—to probe the efficiency-fairness frontier is a research direction of primary importance.

2.A Appendix to Chapter 2: Omitted Proofs

Lemma 2.A.1. *Every student weakly prefers μ_1 to being unmatched.*

Proof. Observe that no student rejects a proposal from herself (because students do not make rejections). This implies that no student prefers being unmatched more than μ_1. □

Proof of Lemma 2.1:

Suppose (toward a contradiction) that μ_1 is between-group blocked by a coalition C between-enforcing a match ν which it prefers.

First, by lemma 2.A.1, there is some school in $\mu_1(C)$.

Second, observe that a school only rejects students if it receives more proposals (cumulatively) than its capacity q_s. Therefore, if i strictly prefers a school to $\mu_1(i)$, then that school is filled to capacity at μ_1.

The previous two points imply the existence of a school s and two students i and j with the following properties:

1. $s \in \nu(C)$; and
2. $i \in C \cap \nu^{-1}(s)$ but $i \notin \mu_1^{-1}(s)$; and
3. $j \in \mu_1^{-1}(s)$ but $j \notin \nu^{-1}(s)$; and
4. $i \succ_s j$.

But then notice that i must have proposed to s but have been rejected. However, j must have proposed to s but *not* have been rejected. If j proposed to s before s rejects i, then s rejects j before rejecting i, a contradiction. If j proposed to s after s rejects i, then s must have rejected i in favor of a higher within-group priority student i^*; hence, j is also rejected, a contradiction.

Therefore, there is no coalition which between-group blocks μ_1. □

Lemma 2.A.2. *If $C \sim_s C'$, then for every $i \in I$: $|[i]_s \cap C| = |[i]_s \cap C'|$*

Proof. Suppose (toward a contradiction) that for some $i \in I$: $|[i]_s \cap C| \neq |[i]_s \cap C'|$. Without loss of generality, let i be in the $\succsim$-highest priority group such that $|[i]_s \cap C| \neq |[i]_s \cap C'|$. Again, without loss of generality, let $|[i]_s \cap C| > |[i]_s \cap C'|$. Consider that $C \sim_s C'$ implies $C' \succsim_s C$. But then C' and C cannot be indexed such that $c'_k \succsim_s c_k$ because every student in a $\succsim$-more preferred priority group must be paired with a student in that same priority group and the students in the $[i]_s$ priority group are imbalanced. This is a contradiction. □

Lemma 2.A.3. *The following statements are true:*

1. *If C within-group blocks μ_2 by within-enforcing ν, then C within-group blocks μ_1 by within-enforcing ν as well.*

2. *If C between-group blocks μ_2 by between-enforcing ν, then C within-group blocks μ_1 by within-enforcing ν as well.*

Proof. I show that C can within-group enforce or between-group enforce ν over μ_1 (respectively). Because every student weakly prefers μ_2 to μ_1, it follows that C prefers ν to μ_1 and hence within-group blocks or between-group blocks μ_1 with ν.

For the first statement, suppose that C within-group blocks μ_2 by within-enforcing ν. I show that C can within-group enforce ν over μ_1. To see this, notice that $U_s(i)$ is independent of μ_2. Hence, the sum of within-group interrupters across all students in $\nu^{-1}(s) \cap C$ is the same for both μ_2 and μ_1. Additionally, by definition, for every $s \in \nu(C)$: $\mu_2^{-1}(s) \sim_s \nu^{-1}(s)$. By the construction of μ_2: $\mu_2^{-1}(s) \sim_s \mu_1^{-1}(s)$. Hence, $\nu(s) \sim_s \mu_1^{-1}(s)$. Therefore, C can within-group enforce ν over μ_1.

For the second statement, suppose that C between-group blocks μ_2 by within-enforcing ν. I show that C can between-group enforce ν over μ_1. To see this, notice that for every school s, $\mu_2 \sim_s \mu_1$. Hence, if $\nu^{-1}(s) \succ_s \mu_2$, then $\nu^{-1}(s) \succ_s \mu_1$. Therefore, C can between-group enforce ν over μ_1.

This completes the proof. □

Lemma 2.A.4. *If C within-group blocks μ_2 with ν, then $i \in C$ strictly prefers $\nu(i)$ to $\mu_2(s)$ only if s is filled to capacity at μ_2.*

Proof. Suppose (toward a contradiction) that $i \in C$ strictly prefers $\nu(i) \equiv i$ to $\mu_2(s)$ but $|\mu_2^{-1}(s)| < q_s$.

By lemma 2.A.3, C within-group blocks μ_1 with ν. By the construction of μ_2, $|\mu_2^{-1}(s)| = |\mu_1^{-1}(s)|$. By the construction of μ_1, s only rejects i if more than $|q_s|$ students point to s. But less than q_s other students point at s, so s does not reject i. Therefore, i weakly prefers μ_1 to i. This contradicts that i strictly prefers s to μ_2. Therefore, $|\mu_2^{-1}(s)| = q_s$. □

Proof of Theorem 2.1: By lemma 2.A.3, μ_2 is not between-group blocked; otherwise, μ_1 would be between blocked, a contradiction to lemma 2.1.

Now consider within-group blocks. Suppose (toward a contradiction) that μ_2 is within-group blocked by a coalition C within-enforcing a match ν.

First, for every school $s \in \nu(C)$, I define an injection $T_s : \nu^{-1}(s) \cap C \to \mu_1^{-1}(s) \cap C$ in the following manner. Let $\nu^{-1}(s) \cap C = \{i_1, i_2, \ldots, i_n\}$ be indexed such that

$i_{k+1} >_s^* i_k$ for every k. Notice that by the construction of μ_1, if $i_k \in \mu_1^{-1}(s)$, then for $k' > k$, $i_{k'} \in \mu_1^{-1}(s)$. Therefore, there is a unique m (possibly taking the value of 0 or n) such that $i_m \notin \mu_1^{-1}(s)$ but $i_{m+1} \in \mu_1^{-1}(s)$. I define T_s piecewise based on the index k:

- For $m + 1 \leq k \leq n$, let $T_s(i_k) \equiv i_k$.

  Observe that $T_s(i_k) \in \mu_1^{-1}(s)$ by the definition of m. Notice that $T_s(i_k) \in C$ because $i_k \in \nu^{-1}(i) \cap C$. This piece of T_s is clearly injective.

- For $1 \leq k \leq m$, consider the following argument:

  Because C within-group enforces ν at μ_2, it follows that $\nu^{-1}(s) \sim_s \mu_2^{-1}(s)$. By construction, $\mu_1^{-1}(s) \sim_s \mu_2^{-1}(s)$. By construction of μ_1, every student in $\mu_1^{-1}(s)$ is strictly $>_s^*$-preferred to every student in $i_1, \ldots, i_m$. Hence by lemma 2.A.2, $i_1, \ldots, i_m \in [i_m]_s$.

  Again, because C within-group enforces ν over μ_2, there are at most $q_s - |\nu^{-1}(s)|$ within-group interrupters for student i_m. Because $|\nu^{-1}(s)| = q_s$ by lemma 2.A.4, it follows that there are no within-group interrupters of i_m. Hence, $U_s(i_m) \subseteq C$. But every student in $\mu_1^{-1}(s)$ is $>_s^*$-preferred to i_m. Thus, $\mu_1^{-1}(s) \cap [i_m]_s \subseteq U_s(i_m)$. Therefore, $\mu_1^{-1}(s) \cap [i_m]_s \subseteq C$.

  Finally, note that $|\mu_1^{-1}(s) \cap [i_m]_s| = |\nu^{-1}(s) \cap [i_m]_s|$ by lemma 2.A.2. Let M be the greatest index such that $i_M \in [i_m]_s$, and note that $\{i_1, \ldots i_M\} \subseteq \nu^{-1}(s) \cap [i_m]_s$. But then I have

  $$\left|\mu_1^{-1}(s) \cap [i_m]_s\right| \geq \left|\{i_1, \ldots, i_m, \ldots i_M\}\right|.$$

  However, because $\{i_{m+1}, \ldots, i_M\} \subseteq \mu_1^{-1}(s)$, this can be rewritten:

  $$\left|\left(\mu_1^{-1}(s) \cap [i_m]_s\right) \backslash \{i_{m+1}, \ldots, i_M\}\right| + \left|\{i_{m+1}, \ldots, i_M\}\right| \geq \left|\{i_1, \ldots, i_m, \ldots i_M\}\right|.$$

  This further implies

  $$\left|\left(\mu_1^{-1}(s) \cap [i_m]_s\right) \backslash \{i_{m+1}, \ldots, i_M\}\right| \geq \left|\{i_1, \ldots, i_m\}\right|. \tag{2.1}$$

  Define $T_s(i_k)$ as the k^{th} $>_s^*$ least-preferred student in $\left(\mu_1^{-1}(s) \cap [i_m]_s\right) \backslash \{i_{m+1}, \ldots, i_M\}$. By equation (2.1), $T_s(i_k)$ is well-defined and is an injection on this piece. Because $T_s(i_k) \notin \{i_{m+1}, \ldots, i_M\}$, this piece of T_s has no overlap with the first piece. Finally, by the above argument, $T_s(i_k) \in C$.

Hence, there is a well-defined injection $T_s : \nu^{-1}(s) \cap C \to \mu_1^{-1}(s) \cap C$.

Second, iteratively construct the following sequence of students. Let i_1 be some student in C such that $\nu(i_1) \neq \mu_2(i_1)$. Because C prefers ν to μ_2, such a student exists. Let $i_{k+1} \equiv T_{\nu(i_k)}(i_k)$.

Third, I claim that i_k is well-defined, $i_k \in C$, and $\mu_1(i_k) \neq \nu(i_k)$. I show this by induction. For the base case when $k = 1$, I make three observations:

1. That i_1 is well-defined is noted previously.
2. $i_1 \in C$ by definition.
3. Because i_1 strictly prefers $\nu(i_1)$ to $\mu_2(i_1)$ and i_1 weakly prefers μ_2 to μ_1 by construction, it follows that $\nu(i_1) \neq \mu_1(i_1)$.

For the inductive step, suppose that every student with index less than k is well-defined, is in C, and is not matched to the same school in ν and μ_1. I make three observations:

1. Note that $\nu(i_{k-1}) \in \nu(C)$. Similarly, note that $\nu(i_{k-1}) \in S$ because μ_2 is individually rational. Hence, $T_{\nu(i_{k-1})}$ is well-defined. Therefore i_k is well-defined.
2. $i_k \in C$ because the domain of $T_{\nu(i_{k-1})}$ is a subset of C.
3. Because $\nu(i_{k-1}) \neq \mu_2(i_{k-1})$ and i_{k-1} weakly prefers μ_2 to μ_1, it follows that $\nu(i_{k-1}) \neq \mu_1(i_{k-1})$. Thus, $i_{k-1} \notin \mu_1^{-1}(\nu(i_{k-1}))$. By the construction of $T_{\nu(i_{k-1})}$, it follows that $\mu_1(i_k) = \nu(i_{k-1})$ and $\nu(i_k) \neq \nu(i_{k-1})$. Therefore, $\mu_1(i_k) \neq \nu(i_k)$.

Fourth, because each $T_{\nu(i_k)}$ is an injection and each student i_k is matched to only one school in μ_1, it follows that if $i_k \neq i_l$, then $T_{\nu(i_k)} \neq T_{\nu(i_l)}$. Because there are a finite number of students, it follows that there is a (minimal) index n such that $i_1 = i_{n+1}$.

Fifth, consider the cycle $i_1, \ldots, i_n$. Because $T_{\nu(i_{k-1})}(i_{k-1}) \neq i_k$, it follows that $j \succsim_{\mu_1(i_k)} i_k$ for every $j \in \mu_1^{-1}(\mu_1(i_k))$. Hence, every student in the cycle is active the Top Trading Cycles stage. Notice that each i_k must be deactivated weakly later than i_{k+1} because i_k points to $\mu_1(i_{k+1})$. However, i_1 must be deactivated *strictly* later than i_2 because $\nu(i_1) \neq \mu_1(i_2)$. Therefore, i_1 must be deactivated strictly before i_1, a contradiction.

Therefore, there is no coalition that within-group blocks ν. □

Chapter 3

RANKED-TO-MATCH: THE EFFECTS OF EARLY MATCHING IN THE NRMP

3.1 Introduction

The National Resident Matching Program (NRMP) is the leading example of success in market design. Since 1952 the NRMP has brokered matches between medical school graduates and residency programs. Its persistence is viewed as a consequence of the stability of the implemented match (Roth, 1991) and the timing of the availability of information in the market. Recent research in medical ethics and medical education identifies coercive *post-interview communication* (PIC) as a potential source of market failure; residency programs are successfully arranging matches through coercive PIC with doctors prior to the operation of the NRMP match, but these matches may be inefficient. The unraveling of the medical resident market to a date before the NRMP operates has welfare implications that I examine in this paper.

I focus on communication by hospital residency programs that either states how a hospital ranks a doctor, asks a doctor how she ranks the hospital, or implies that a positive rank by the hospital of the doctor depends upon a commitment from the doctor to rank the hospital first on her list.[1] Such PIC is a violation of the Match Agreement.[2] I refer to PIC involving these questions or statements as *coercive*, and connect unraveling with coercive PIC. Early matches (which are a form of unraveling) are either formal offers for outside-of-match positions or informal agreements to mutually top-rank each other. Numerous free-form responses of surveyed doctors support a direct connection between coercive PIC and early matches. For instance, one doctor reported:

> "Many program directors explicitly stated that my position on their rank list depended on postinterview communication. That a commitment to rank them first would increase my chances of matching at their program." (Williams et al., 2019)

[1]For conciseness, I refer to hospital residency programs as "hospitals" and medical students as "doctors". Additionally, I use masculine pronouns for hospitals and feminine pronouns for doctors.

[2]See NRMP Match Code of Conduct, Section 6.5.

As discussed in section 1.2, PIC is the object of numerous studies in the medical ethics and medical graduate education literatures. In these studies, doctors consistently report changes to their submitted Rank-Order List (ROL) as a result of PIC. That coercive PIC results in early matches (and hence unraveling) is well-documented by this literature.

In my model two hospitals hold a common preference over doctors, but the doctors' preferences are uncertain. A hospital that makes many early offers risks that too many doctors may accept, and hence not enough positions remain for doctors who only participate in the NRMP match. Conversely, a hospital that makes too few early offers allows desirable doctors to be poached by the other hospital in the early match process. When choosing to offer early matches to doctors, hospitals are forced to weigh the uncertain response of the doctors against the risk of desirable doctors accepting early match offers from another hospital. This forces hospitals to hedge by making early offers to doctors that the hospital may not want (depending on what the doctors' preferences are) to prevent these doctors from accepting early offers from the other hospital. Because the less-desirable hospital matches to lower-ranked doctors, the less-desirable hospital is able to poach lower-ranked doctors that the more-desirable program has not made early offers to but may desire when the doctors' preferences are revealed.

My main result is that the program that is less-desirable in expectation benefits from early matching at the expense of the program that is more desirable. The utility transfer occurs because the less-desirable program successfully poaches doctors near the bottom of the more-desirable program's accepted doctors. As the effectiveness of coercive PIC increases, so does the transfer from the more-desirable program to the less-desirable program.

This result theoretically supports widely held beliefs in medical education. For instance, in a survey of program directors in internal medicine about the impact of a policy banning formal early matches (the All-In policy), Adams et al. (2012) states

> "The most commonly cited concern [about the All-In policy] was that smaller, nonuniversity programs and those in geographically less-desirable areas would suffer in recruitment."

My interpretation is that banning formal early matches makes coercive PIC less effective. Program directors predicted that reducing the effectiveness of coercive

PIC would transfer qualified doctors from lower-ranked to higher-ranked hospitals. I establish that coercive PIC is detrimental to more-desirable hospitals. I also find that the responsiveness of the doctors to coercive PIC has an effect on the distortion caused.

Institutional Background

Medical students spend four years in medical school before beginning residency. Residency programs provide medical graduates with hands-on experience in particular fields of medicine. A residency program lasts between three and seven years, and completion of the first year is required to become a licensed physician within the United States. In the 2022-2023 admissions cycle there were 48,156 applicants for 40,375 residency positions (Program, 2023).

The application cycle begins in September of the fourth year when the Electronic Residency Application System (ERAS) opens. Applicants have already completed the Step One and Step Two exams, and the Step Three exam is not completed until after matches are announced in March. In late September the applications are simultaneously released to residency programs. In early October, Medical Student Performance Evaluations are released by the medical schools to residency programs. Shortly after this, residency programs contact applicants to arrange interviews, with the majority of interviews concluded by the end of the year.

Each student and residency program submits a Rank-Order List (ROL) to the NRMP, with the deadline for both sides set in February. The NRMP computes a slightly modified version of the doctor-proposing deferred acceptance algorithm.[3] On Match Day in March, the NRMP announces to doctors and residency programs whom they have been matched with. The match outputted by the NRMP is a legally binding contract between doctors and residency programs.

For doctors who did not match and for residency programs with vacant seats remaining, the NRMP facilitates the Supplemental Offer and Acceptance Program (SOAP). SOAP is held for about a week shortly after the matches have been announced, and is colloquially referred to as the "scramble."

Notably, there is little to no new information revealed after the interview stage. After a doctor completes an interview at a residency program, typically there is no more

[3]The modifications have been shown by Roth and Peranson (1999) to have a minimal impact on the match achieved.

information about the program that the doctor learns.[4] Similarly, after a residency program interviews a doctor, the only institutional feature by which the program may (potentially) gain information would be through the release of the Fall semester grades, but I have come across no discussion of these as factors in the formulation of the ROL. In essence, all relevant information to make a ROL based on a doctor's or program's merits is available after interviews are conducted.

Literature Review

This paper bridges the medical ethics literature on inappropriate PIC with the market design literature on unraveling. Below, I summarize key contributions from each literature.

Studies on inappropriate PIC have been conducted within numerous medical specialties. The vast majority of studies are unincentivized surveys administered either to residents (prospective or current) or to program directors of residency programs, with a response rate of roughly 50% common among these surveys. Many address both coercive PIC and other kinds of inappropriate PIC, such as questions by hospitals concerning marital status, family plans, religion, and other interviews.

Surveys of (prospective) residents (either administered by a medical school to its students or by a residency to applicants) generally find substantial levels of inappropriate PIC. Fields surveyed include Pediatrics (Opel et al., 2007), Emergency Medicine (Thurman et al., 2009), Dermatology (Sbicca et al., 2010, 2012), Radiation Oncology (Holliday, Thomas Jr and Kusano, 2015; Berriochoa et al., 2016), Orthopedic Surgery (Camp et al., 2016), Internal Medicine (Cornett et al., 2017; Williams et al., 2019; Swan and Baudendistel, 2014), and Integrated Vascular Surgery (Fereydooni et al., 2022). Additionally, numerous surveys are not specific to one field: Anderson, Jacobs and Blue (1999); Pearson and Innes (1999); Miller et al. (2003); Jena et al. (2012); Monir et al. (2021). Several other studies find similar patterns within other matches: in the Urology match (Teichman et al., 2000; Sebesta et al., 2018; Handa et al., 2021) and in the military match (Ratcliffe et al., 2012).

There is not a standard template for the questions in these surveys. However, a few themes are apparent in the results. Doctors commonly report ($\sim$ 15% of respondents) being told they are "ranked-to-match," which is interpreted as meaning the doctor is ranked at the top of the program's ROL. Less frequently ($\sim$ 5% of respondents),

[4] Some doctors are invited back to the residency program for a "Second Look," but the surveys discussed in Section 3.1 indicate that most doctors do not find these opportunities to be informative.

doctors report being offered incentives to match early. Quite commonly ($\sim 25\%$ of respondents), doctors report changing their rank as a result of inappropriate (although not always coercive) PIC. Although the data quality is low due to the nature of the studies, the overall picture from these surveys is that coercive PIC occurs in some but not all cases, and that it has a meaningful impact on the final match.

Surveys to program directors are less frequent but still cover a wide spectrum of medical fields. Fields surveyed include General Surgery (Anderson and Jacobs, 2000), Family Practice (Carek et al., 2000), Dermatology (Sbicca et al., 2010), Obstetrics and Gynecology (Curran et al., 2012; Frishman et al., 2014), Internal Medicine (Chacko et al., 2018), Urology (Farber et al., 2019), and Otolaryngology (Harvey et al., 2019). Additionally, Grimm, Avery and Maxfield (2016) surveys program directors without reference to a specific field. Program directors commonly report ($\sim 50\%$ of respondents) that they feel doctors have made an informal commitment to rank the hospital first. Understandably, surveys of program directors imply that substantially less inappropriate PIC is initiated by programs than implied by surveys of doctors.

Economists have focused more on the causes of unraveling than on its consequences. The three primary causes of unraveling identified in the literature are the following. First, several authors identify the instability of the mechanism as a cause (Roth, 1991; Sönmez, 1999). Second, others identify uncertainty over the preferences of one or both sides as a cause (Roth and Xing, 1994; Li and Rosen, 1998; Hałaburda, 2010; Niederle, Roth and Ünver, 2013; Ambuehl and Groves, 2020). The uncertainty that these models impose is on a common quality that is uncertain for one or both sides of the market but will be revealed. Third, a few authors identify costs as a source of unraveling; Damiano, Li and Suen (2005) examines search frictions as a source of unraveling, and Echenique and Pereyra (2016) identifies discounting combined with strategic complementarities as a source.

This paper is most similar to the papers which identify search costs as a cause of unraveling. From the institutional background provided in Section 3.1, the mechanism is stable and there is no uncertainty over the quality of either side of the market that will be revealed before the match is arranged. I depart from the prior literature by assuming that unraveling occurs and focusing on its welfare effects. The only other authors to address early matching in the context of the NRMP are Ashlagi et al. (2023). They use data from OB/GYN programs to study the welfare

	ER	LA
H	$H \succ L \succ \emptyset$ prefers to match early	$H \succ L \succ \emptyset$ prefers to match late
L	$L \succ H \succ \emptyset$ prefers to match early	$L \succ H \succ \emptyset$ prefers to match late

Figure 3.1: Doctor preferences conditioned on type

effects of institutionalizing an early match, and find that an institutionalized early match would have negative welfare consequences.

3.2 Model

Basics

There are two hospitals H and L (High and Low), each with capacity Cap. There is a continuum of doctors with scores distributed on the interval $[0, 1]$. Doctor types are further differentiated along two dimensions. First, each doctor has a *hospital preference* of H or L corresponding to preferences $H \succ L \succ \emptyset$ or $L \succ H \succ \emptyset$, respectively. Second, each doctor has a *match preference* of early (ER) or late (LA). I define $\Theta = [0, 1]\times\{H, L\}\times\{\text{ER}, \text{LA}\}$ as the doctor space, and let $f_{\text{Full}} : \Theta\times\Omega \to \mathbb{R}_+$ be the density of doctors for the state ω. For simplicity, I assume that the doctor's score, hospital preference, and match timing preference are mutually independent, and that the state ω only enters f_{Full} through the hospital preference term. Hence, I have

$$f_{\text{Full}}(x, I, P \mid \omega) = f(x) \cdot h(I \mid \omega) \cdot p(\text{P})$$

with the normalization

$$\begin{aligned} h(H \mid \omega) + h(L \mid \omega) &= 1 \\ p(\text{ER}) + p(\text{LA}) &= 1. \end{aligned}$$

I assume that ER doctors exhibit a behavioral bias toward a guaranteed partner: given the opportunity to match to her second-favorite hospital with certainty or to partake in the NRMP match, the ER doctor prefers the former. In contrast, LA doctors *never* reveal their preferences if asked and always reject early offers. I interpret this as LA doctors complying fully with the Match Agreement statements about persuasion.[5]

I assume that $p(\text{ER}) = r$ for $0 < r < 1$. I call r the *responsiveness* of the market because it reflects how likely a doctor is to respond to an offer to match before the stable match is arranged.

Hospital preferences over doctors are given by the following utility function. If μ is a matching, then the utility of hospital $I \in \{H, L\}$ is

$$u_I(\mu \mid \omega) = \int_\Theta \mu_\theta(I)\theta_1 f_{\text{Full}}(\theta \mid \omega) d\theta.$$

I denote the expected utility:

$$u_I(\mu) = \mathbb{E}_\omega \left[\int_\Theta \mu_\theta(I)\theta_1 f_{\text{Full}}(\theta \mid \omega) d\theta \right].$$

A *matching* is a function $\mu : \{H, L\} \times \Theta \to \{0, 1\}$ such that $\mu_\theta(H) + \mu_\theta(L) \leq 1$. A matching μ is *stable* if there no $I \in \{H, L\}$ and $\theta, \theta' \in \Theta$ such that $\theta_1 > \theta'_1$, $\mu_\theta(I) = 0$, $\mu_{\theta'}(I) = 1$, and $I \succ_{\theta_1} J$ for $\mu_\theta(J) = 1$.

Uncertainty

I assume that there is uncertainty over the aggregate hospital preferences of the doctors. I model this by making the state $\omega \in \{\omega_H, \omega_L\}$ a random variable and assume that

$$h(H \mid \omega_H) = 1$$
$$h(H \mid \omega_L) = \frac{1}{2}.$$

If $\omega = \omega_H$, then I say that H is *popular* and L is *unpopular*. Otherwise, I say that L is popular and H is unpopular. Which hospital is popular is unknown to both hospitals, but both hospitals share the same belief.

I choose 1 and $\frac{1}{2}$ here to simplify the analysis. When $h(H \mid \omega_L) \neq \frac{1}{2}$, finding the equilibrium of the model becomes substantially more difficult and counterintuitive. I consider that the clarity of exposition is worth the simplification. I conjecture that

[5] An alternate interpretation is that LA doctors are unsure of their own preferences, or have yet to conclude interviews, etc.

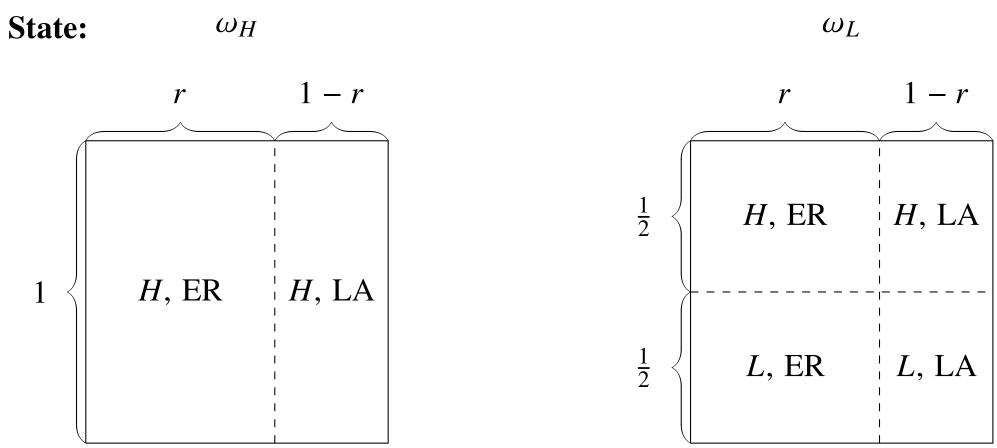


Figure 3.2: Distribution of doctor types conditioned on the state ω

my results still hold on a portion of the parameter space when this requirement is relaxed.

$$\Pr(H \text{ is popular}) = \Pr(\omega = \omega_H) = \frac{1}{2}.$$

I emphasize here that the only uncertainty facing a doctor is the aggregate preferences of the *other* doctors; every doctor knows her own preferences (whether H or L is preferred and whether the doctor is ER or LA) and $f(\cdot)$ and $p(\cdot)$.

To summarize, hospitals are uncertain over doctors' preferences. This uncertainty is both individual (hospitals do not know θ's preferences) and aggregate (hospitals do not know which hospital is popular). Each hospital and each doctor, however, is certain of his or her own preferences.

Structure

There are two phases. First, in the *early-matching* phase, each hospital I chooses a Lebesgue-measurable set Π_I to make private offers to, and let $\Pi = (\Pi_H, \Pi_L)$. I use λ to denote the Lebesgue measure. Each doctor type θ observes the set of offers made to her. Doctors of type θ then respond by choosing a function $\rho_\theta(I \mid \Pi)$ such that

$$\rho_\theta(I \mid \Pi) \in \{0, 1\} \qquad \forall I \in \{H, L\} \tag{3.1}$$

$$\rho_\theta(H \mid \Pi) + \rho_\theta(L \mid \Pi) \leq 1. \tag{3.2}$$

Line 1 requires that θ gives a binary response, and line 2 requires that θ accept at most one early offer. I note that ρ is a matching.

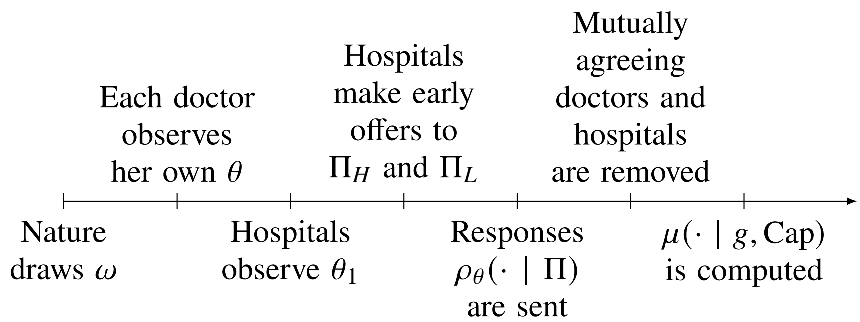


Figure 3.3: Timeline of events

If a doctor accepts an offer, it is a binding commitment between the hospital and the doctor; the doctor is matched with the hospital, and the hospital sets a new capacity $\text{Cap}_I(\rho, \omega) = \text{Cap}_I - \int_\Theta \rho_\theta(I \mid \Pi) f_{\text{Full}}(\theta \mid \omega) d\theta$. The density of remaining doctors is $g(\theta \mid \rho, \omega) = f_{\text{Full}}(\theta \mid \omega) \cdot (1 - \rho_\theta(H \mid \Pi) - \rho_\theta(L \mid \Pi))$.

Second, in the *late-matching* phase, the doctor-optimal stable match $\mu(\cdot \mid g, \text{Cap})$ is implemented among the remaining doctors and the remaining hospital capacities based on agents' true preferences *regardless of any early offers made*.[6] In this case (See Azevedo and Leshno (2016) for more details), each hospital sets a cutoff (H sets q_H and L sets q_L) and the match correspondence is

$$\mu_{(x,H,\text{P})}(H \mid g, \text{Cap}) = \mathbb{1}\{x \geq q_H\}$$
$$\mu_{(x,H,\text{P})}(L \mid g, \text{Cap}) = \mathbb{1}\{x \geq q_L \text{ and } x < q_H\}$$

and similarly for (x, L, P).

If a hospital s its capacity with early matches, it keeps only the highest-scoring doctors and discards the rest. That is, if for some I, $\int_\Theta \rho_\theta(I \mid \Pi) f_{\text{Full}}(\theta \mid \omega) d\theta >$ Cap, then I truncate $\rho_\theta(I \mid \Pi)$ to $\tilde{\rho}_\theta(I \mid \Pi)$ as follows. Define x such that $\int_{\{\theta \mid \theta_1 \geq x\}} \rho_\theta(I \mid \Pi) f_{\text{Full}} d\theta = \text{Cap}$. Then

$$\tilde{\rho}_\theta(I \mid \Pi) = \begin{cases} \rho_\theta(I \mid \Pi) & \theta_1 \geq x \\ 0 & \text{otherwise.} \end{cases}$$

However, to incorporate the reputation costs borne by hospitals that renege their promises, I include that if the hospital I that exceeded capacity in the early match, then it is penalized by having 3 deducted from his utility, where 3 is chosen so that neither hospital would choose to exceed capacity.

[6]The condition that hospitals only need rank accepting doctors first is crucial. Otherwise, the offers in the late match phase would be more difficult to analyze.

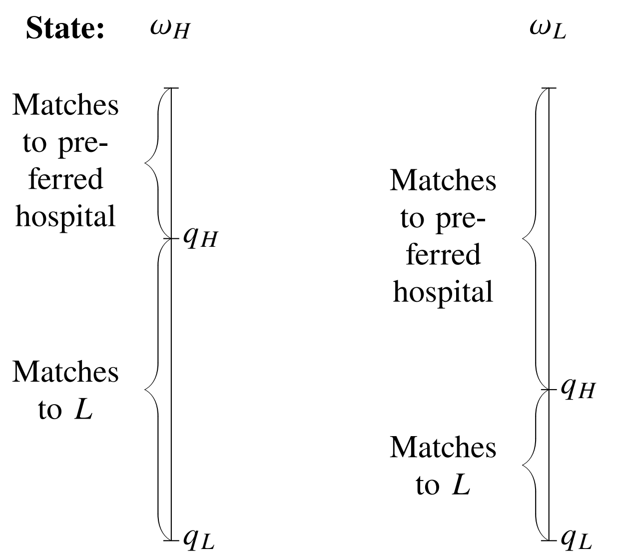


Figure 3.4: Cut-off structure

Equilibria

Each doctor responds deterministically based on her type θ to offers made. If $\theta_3 = \mathrm{LA}$ (the doctor has a preference for late matching), then $\rho_\theta(\cdot \mid \Pi) = 0$ (the doctor rejects all offers). If $\theta_3 = \mathrm{ER}$ (the doctor has a preference for early matching), then the doctor accepts the offer from her most-preferred hospital (if she receives an offer). Formally, $\rho_\theta(\theta_2 \mid \Pi) = \Pi_{\theta_2}(\theta_1)$ and $\rho_\theta(I \mid \Pi) = 1$ $\Pi_{\theta_2}(\theta_1) = 0$ and $\Pi_I(\theta_1) = 1$ for $I \neq \theta_2$, and $\Pi_I(\theta_1) = 0$ otherwise.

Let

$$\mathrm{LATE}_\theta(\cdot \mid \Pi, \omega) = \mu\Big(\cdot \mid g\big(\cdot \mid \rho_\theta(\cdot \mid \Pi), \omega\big), \mathrm{Cap}\big(\rho_\theta(\cdot \mid \Pi), \omega\big)\Big).$$

A strategy for hospital I is a Lebesgue-measurable set $\Pi_I \subseteq [0,1]$. To clarify the exposition, I provide the following definition to reduce the number of equilibria. Informally, two strategies are outcome-equivalent if they yield the same matchings (up to a set of doctors of measure zero) in both the early and late phases.

Definition 3.1. Two strategies Π and Π' are said to be *outcome-equivalent* if $\rho_\theta(I \mid \Pi) + \mathrm{LATE}(I \mid \Pi) = \rho_\theta(I \mid \Pi') + \mathrm{LATE}(I \mid \Pi')$ for almost every θ for both $I \in \{H, L\}$ and both $\omega \in \{\omega_H, \omega_L\}$.

My equilibrium concept is a pure-strategy Nash equilibrium. An *equilibrium* is a vector Π such that each hospital maximizes its expected utility given the strategy of the other. Formally,

$$\Pi_I \in \arg\max_{\Pi_I'} \mathbb{E}_\omega u\Big(\rho(I \mid \Pi) + \text{LATE}(I \mid \Pi, \omega)\Big).$$

Although pure-strategy equilibria are not guaranteed to exist, in my analysis I prove its existence in this model.

3.3 Analysis

My analysis has two main results. Theorem 3.1 establishes the existence and uniqueness of equilibrium. It further shows that, in equilibrium, hospitals use cutoff strategies. Theorem 3.2 conducts comparative statics with respect to r, and shows that hospital L's utility is increasing in r while H's utility decreases in r.

Formally, my results are the following. First, if Π_I is outcome-equivalent to $\tilde{\Pi}_I = [\pi_I, 1]$, then I say that Π_I is a *cutoff strategy*. That is, each hospital chooses a threshold score and makes early offers to all doctor types who score above that threshold. For simplicity, I refer cutoff strategies by the cutoff π.

Let

$$\pi_H^*(r) = 1 - \frac{4\text{Cap}}{3+r}$$
$$\pi_L^*(r) = 1 - \frac{6\text{Cap}}{3+r}.$$

I depict $\pi^*(r)$ in Figure 4. Since the cutoffs in the match phase are deterministic given Π and the state ω, I write $q_H(\Pi \mid \omega)$ and $q_L(\Pi \mid \omega)$.

My first result is that $\pi^*(r)$ is an equilibrium for every $r \in [0, 1]$, and is unique.

Theorem 3.1. $\pi^*(r)$ *is the unique equilibrium.*

Despite the flexibility in choosing Π, each hospital only uses cutoff strategies. Perhaps counterintuitively, in Lemma 3.1 I show that *every* best-response is a cutoff strategy.

To prove Theorem 3.1, I use backward induction. The critical step is the proof that if Π_I is a best-response, then Π_I is a cutoff strategy. I state this in Lemma 3.1:

Lemma 3.1. *If* Π_I *is a best response to* Π_J*, then there exists some* Π_I^* *that is outcome-equivalent to* Π_I *such that*

$$\Pi_I^* = [\pi_I, 1],$$

where $q_I(\Pi \mid \omega_J) \le \pi_I \le q_I(\Pi \mid \omega_I)$.

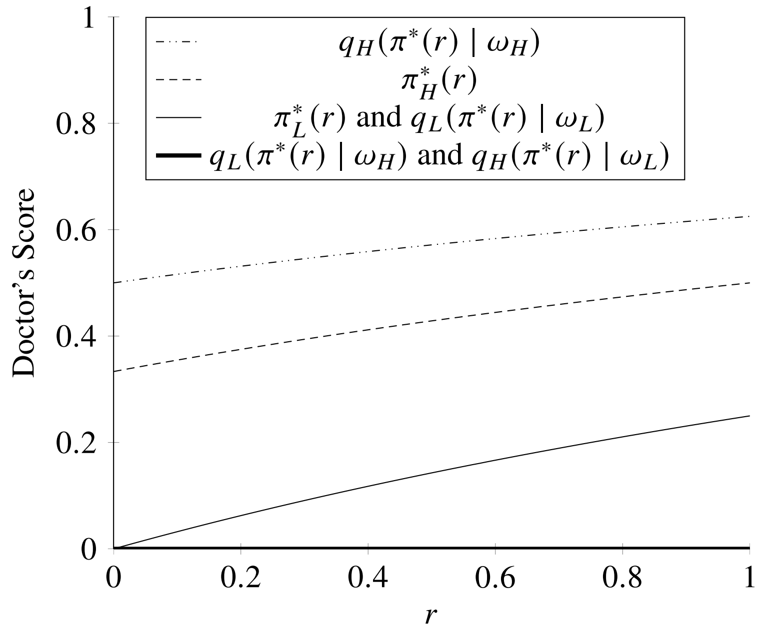


Figure 3.5: Illustration of the equilibrium when Cap = 1/2. Notice that hospital H over admits in the early stage when H is popular: $q_H(\pi^*(r) \mid \omega_H) \neq \pi_H^*(r)$. In contrast, hospital L never over admits: $q_L(\pi^*(r) \mid \omega_L) = \pi_L^*(r)$. Because Cap $\geq$ 1/2, every doctor is admitted by at least one hospital in the the stable match $(q_H(\pi^*(r) \mid \omega_L) = q_L(\pi^*(r) \mid \omega_H) = 0)$.

If there are multiple π that satisfy Lemma 3.1, I take π to be the smallest one.

Lemma 3.1 narrows the scope of possible deviations Π_I' tremendously. Because q is a deterministic function π, finding equilibria becomes equivalent to checking possible $\pi_I \in [0, 1]$.

The main difficulty in proving Lemma 3.1 is that early offers made by I to doctors in Π_J have different yields based on the state ω, whereas early offers made to doctors *not* in Π_J always result in perfect yield. A generic strategy Π_I could produce a variety of combinations of yield based on the state ω that cutoff strategies cannot replicate.

The insight that resolves this difficulty is that when hospital I drops doctors from Π_I, these doctors either accept early offers from hospital J (giving hospital I access to doctors on the margin of q_J) or are still available (so hospital I is just as well off). The proof of Lemma 3.1 leverages this in case 3.

Here I provide a sketch of the proof of Lemma 3.1. Toward a contradiction, I suppose that Lemma 3.1 does not hold. Then there are two sets of doctors A and B with positive measure such that B is above A with regards to θ_1, A is in Π_I but B is not in Π_I. The three cases are:

1. If B is above $q_I(\pi^* \mid \omega_I)$, then hospital I should always make early offers to doctors in B.
2. If B is below $q_I(\pi^* \mid \omega_I)$ and below $q_J(\pi^* \mid \omega_J)$, then hospital I should drop doctors in A. This follows because if $\omega = \omega_I$, hospital I does not want these doctors, and if $\omega = \omega_J$, then hospital J will not make regular offers to A (and any doctor hospital J gains from hospital I dropping A lets hospital I gain a doctor on the margin of $q_J(\pi^* \mid \omega_J)$).
3. If B is below $q_I(\pi^* \mid \omega_I)$ and above $q_J(\pi^* \mid \omega_J)$, then hospital I should drop some doctors in A and admit some doctors in B such that $q_I(\pi^* \mid \omega_I)$ remains the same. This works because, in state ω_I, hospital I shifts toward higher scoring doctors without affecting the cutoffs. If $\omega = \omega_J$, then hospital I is better off because any doctor in A that also received an early offer from hospital J now accepts the early offer from hospital J, making hospital J worse off. Because the set of matched doctors is the same (all doctors with $\theta_1 \in [1 - 2\text{Cap}, 1]$), I see that any loss to hospital J is a gain for hospital I.

These cases complete the proof.

Using Lemma 3.1, proving Theorem 3.1 involves tedious algebra and case work. The proof mostly consists of hypothesizing the order of the cutoffs q and π and verifying that these are correct. The proof is relegated to Section 3.A.

In my second theorem, I analyze the comparative statics of varying r. I see that hospital H's utility decreases in r.

Theorem 3.2. *$u_H(\pi^*(r))$ is decreasing in r.*

The proof of Theorem 3.2 relies only on calculating a derivative, so I relegate it to the supplementary *Mathematica* notebook. Because the same set of doctors is accepted for every r and no hospital exceeds capacity, it follows that hospital L's utility is increasing when hospital H's utility is decreasing. Hence, hospital L prefers larger r.

3.4 Discussion

Theorem 3.2 provides a theoretical foundation for the commonly held belief that the All-In policy would harm less competitive programs. In my model, r reflects

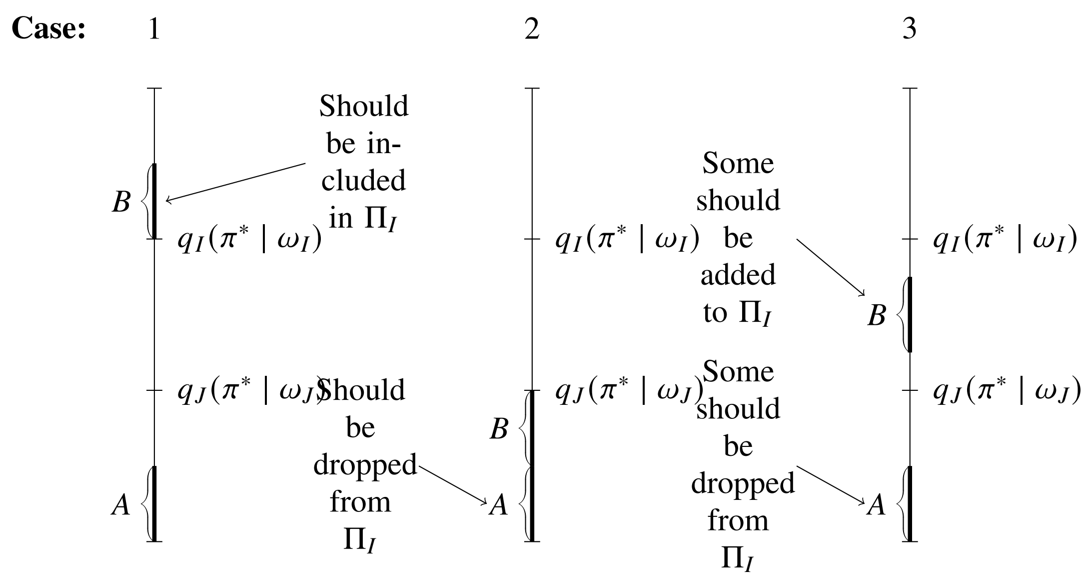


Figure 3.6: Illustration of the proof of Lemma 3.1. Hospital I makes offers to doctors in A but not in B. In Case 1, hospital I should always make offers to doctors in B because regardless of the state ω hospital I *always* wants these doctors. In Case 2, hospital I should never make offers to doctors in A because regardless of the state ω hospital I *never* wants these doctors. In Case 3, hospital I should replace offers made to doctors in A with offers made to doctors in B because hospital I prefers doctors in B and if the state is ω_J then every doctor in A is still available (or is early matched to hospital J, which hurts hospital J and thus helps hospital I).

how attractive early offers are to doctors; a higher r translates to more doctors being willing to accept early offers. If hospitals are unable to make contractual agreements with doctors prior to the NRMP, then some doctors who could have been persuaded to accept an early offer will not accept it. Similarly, stronger restrictions on PIC may prevent hospitals from as effectively persuading doctors to accept an early match. I interpret these as lowering the responsiveness of the market, r. Theorem 3.2 predicts that the less competitive hospitals would be harmed by these policies.

My model does not allow me to test different proposals to reduce early matching, but it does provide predictions about the number of early offers made and the hospitals that benefit from early matches. Counterintuitively, as doctors become less responsive to early offers (lower r), hospitals make *more* early offers. Policies intended to reduce the rate of early matches are likely to increase the number of early offers made even though the number of early matches decreases.

Beyond the direct policy implications, Theorem 3.1 is a technical advancement that could be useful for other applications in matching theory. The main difficulty with establishing that best-responses are only in cutoff strategies is that non-cutoff strategies are able to flexibly combine doctors included in other hospitals' early offer sets to produce complex lotteries over the early match ρ. Theorem 3.1 demonstrates that when there are just two hospitals, this flexibility is unnecessary. The two-hospital assumption makes the analysis tractable.

Two directions for future work are apparent. The first is to increase the number of hospitals. Extending Lemma 3.1 to this case may prove impossible, and I conjecture that hospitals will have a cutoff for every combination of other hospitals. The key insight from the proof of Lemma 3.1 is that when hospital I makes early offers to doctors below $q_J(\pi \mid \omega_J)$ (hospital J's cutoff when J is popular), these early offers are only beneficial to hospital I in the state that J is popular; hence the early offers to these doctors are unnecessary. It is not apparent how viable this lemma is with more than two hospitals.

In my model, hospital L gains utility from early offers because hospital L has lower cutoffs in the late match than hospital H. Hospital L does not need to hedge his bets concerning q_L because q_L is always below π_H. For a multi-hospital extension, it is unclear how this property would translate. It would be fruitful to further understand the properties of a hospital that cause it to gain from early offers.

A second extension is to consider the welfare of the doctors. It would be interesting to examine which types of doctors are impacted most by early offers, and whether some doctors are likely to gain from the early offer system.

3.A Appendix to Chapter 3: Omitted Proofs

Proof of Lemma 3.1: Toward a contradiction, suppose that Π_I is a best response to Π_J and that Π_I is not outcome-equivalent to Π_I^* for any π_I. Then there are sets A and B such that $\lambda(A) > 0$, $\lambda(B) > 0$, $\inf B \geq \sup A$, A is in Π_I, and B is not in Π_I. WLOG there are three cases:

1. $\inf B > q_I(\Pi \mid \omega_I)$: In this case, consider the deviation
$$\Pi_I^\beta = \Pi_I \cup \{b \in B \text{ s.t. } b \geq \beta\}.$$
Because $r > 0$ there exists some β^* such that $\beta^* \geq q_I(\Pi_I^{\beta^*}, \Pi_J \mid \omega_I)$ and $\lambda([\beta^*, 1] \cap B) > 0$.

   If $q_I(\Pi_I^{\beta^*}, \Pi_J \mid \omega_I) = q_I(\Pi \mid \omega_I)$ for every such B and β, then the supposition that Π_I is not outcome-equivalent to some Π_I^* is violated, a contradiction. Hence, take B and β^* such that $q_I(\Pi_I^{\beta^*}, \Pi_J \mid \omega_I) \neq q_I(\Pi \mid \omega_I)$.

Because $\Pi_I^* \supset \Pi_I$, it follows that

$$\int_\Theta \rho_\theta(I \mid \Pi_I^{\beta^*}, \Pi_J) f_{\text{Full}}(\theta \mid \omega) d\theta > \int_\Theta \rho_\theta(I \mid \Pi) f_{\text{Full}}(\theta \mid \omega) d\theta.$$

Hence, $q_I(\Pi_I^{\beta^*}, \Pi_J \mid \omega_I) > q_I(\Pi \mid \omega_I)$.

- If $\omega = \omega_I$, then I has traded doctors in the interval $[q_I(\Pi \mid \omega_I), q_I(\Pi_I^{\beta^*}, \Pi_J \mid \omega_I)]$ for doctors above β^*.
- If $\omega = \omega_J$, then I has traded doctors in the interval $[q_I(\Pi \mid \omega_J), q_I(\Pi_I^{\beta^*}, \Pi_J \mid \omega_J)]$ for doctors above β^*.

Hence, I is strictly better off in both states, a contradiction. This also establishes that $\pi_I \leq q_I(\Pi \mid \omega_I)$.

2. $\sup B < q_I(\Pi \mid \omega_I)$ and $\sup B < q_J(\Pi \mid \omega_J)$: In this case, consider the deviation

$$\Pi_I^\alpha = \Pi_I \backslash \{a \in A \text{ s.t. } a < \alpha\}.$$

Because $r > 0$ there exists some α^* such that $\lambda([0, \alpha^*] \cap A) > 0$ and $q_I(\Pi_I^{\alpha^*}, \Pi_J \mid \omega_I) \neq q_I(\Pi \mid \omega_I)$. and $\sup A < q_J(\Pi_I^{\alpha^*}, \Pi_J \mid \omega_I)$.

If $q_I(\Pi_I^{\alpha^*}, \Pi_J \mid \omega_I) = q_I(\Pi \mid \omega_I)$ for every such A and α, then the supposition that Π_I is not outcome-equivalent to some Π_I^* is violated, a contradiction. Hence, take A and α^* such that $q_I(\Pi_I^{\alpha^*}, \Pi_J \mid \omega_I) \neq q_I(\Pi \mid \omega_I)$.

Because $\Pi_I^* \subset \Pi_I$, it follows that

$$\int_\Theta \rho_\theta(I \mid \Pi_I^{\alpha^*}, \Pi_J) f_{\text{Full}}(\theta \mid \omega) d\theta < \int_\Theta \rho_\theta(I \mid \Pi) f_{\text{Full}}(\theta \mid \omega) d\theta.$$

Hence, $q_I(\Pi_I^{\alpha^*}, \Pi_J \mid \omega_I) < q_I(\Pi \mid \omega_I)$.

- If $\omega = \omega_I$, then I has traded doctors below α^* for doctors in the interval $[q_I(\Pi_I^{\alpha^*}, \Pi_J \mid \omega_I), q_I(\Pi \mid \omega_I)]$. This is a strict improvement.
- If $\omega = \omega_J$, then every doctor that I forgoes (doctors (θ_1, I, ER) such that $\theta_1 < \alpha^*$, $\theta_1 \in A$, and $\theta_1 \in \Pi_J$) is either still available (if J overfills his capacity in the early match) or implies that I can acquire a doctor in $[q_J(\Pi_I^{\alpha^*}, \Pi_J \mid \omega_I), q_J(\Pi \mid \omega_I)]$. This is a weak improvement.

Hence, I is strictly better off, a contradiction.

3. $\sup B < q_I(\Pi \mid \omega_I)$ and $\inf B > q_J(\Pi \mid \omega_J)$: In this case, consider the deviation

$$\Pi_I^{\alpha,\beta} = \Big(\Pi_I \cup \{b \in B \text{ s.t. } b \geq \beta\}\Big) \backslash \{a \in A \text{ s.t. } a < \alpha\}.$$

There exists some α^* and β^* such that $\lambda([0,\alpha^*] \cap A) > 0$, $\lambda([\beta^*, 1] \cap B) > 0$, and $q_I(\Pi_I^{\alpha^*,\beta^*}, \Pi_J \mid \omega_I) = q_I(\Pi \mid \omega_I)$.

It follows that

$$\int_\Theta \rho_\theta(I \mid \Pi_I^{\alpha^*,\beta^*}, \Pi_J) f_{\text{Full}}(\theta \mid \omega) d\theta = \int_\Theta \rho_\theta(I \mid \Pi) f_{\text{Full}}(\theta \mid \omega) d\theta.$$

Observe also that

$$\int_\Theta \rho_\theta(I \mid \Pi_I^{\alpha^*,\beta^*}, \Pi_J) f_{\text{Full}}(\theta \mid \omega_J) d\theta \leq \int_\Theta \rho_\theta(I \mid \Pi) f_{\text{Full}}(\theta \mid \omega_J) d\theta$$

with strict inequality if $\lambda(A \cap [0, q_J(\Pi \mid \omega_J)]) > 0$. Hence $q_J(\Pi_I^{\alpha^*,\beta^*}, \Pi_J \mid \omega_J) \leq q_J(\Pi \mid \omega_J)$.

- If $\omega = \omega_I$, then I has traded doctors below α^* for doctors above β^*. This is a strict improvement.
- If $\omega = \omega_J$, observe that $q_J(\Pi_I^{\alpha^*,\beta^*}, \Pi_J \mid \omega_J) \leq q_J(\Pi \mid \omega_J)$. Hence, J acquires weakly worse doctors under $\Pi_I^{\alpha^*,\beta^*}$ than Π_I. Because the every doctor who was matched under Π is matched under $\Pi_I^{\alpha^*,\beta^*}, \Pi_J$, it follows that I is weakly better off.

Hence, I is strictly better off, a contradiction.

In all three cases, a contradiction was reached. Hence, Π_I is outcome-equivalent to $[\pi_I, 1]$ for some π_I.

To establish the final claim that $\pi_I \geq q_I(\Pi \mid \omega_J)$, suppose toward a contradiction that $\pi_I < q_I(\Pi \mid \omega_J)$. Then hospital I could instead respond with $\pi_I + \epsilon$ for $\epsilon > 0$. For ϵ small enough, $\pi_I + \epsilon < q_I(\pi_I + \epsilon, \Pi_J \mid \omega_J)$. Hospital I prefers to not be matched with doctors with scores in $[\pi_I, \pi_I + \epsilon]$ and instead be matched with doctors in the range $[q_I(\pi_I, \Pi_J \mid \omega_J), q_I(\pi_I + \epsilon, \Pi_J \mid \omega_J)]$ or $[q_I(\pi_I, \Pi_J \mid \omega_I), q_I(\pi_I + \epsilon, \Pi_J \mid \omega_I)]$. Hence, $\pi_I + \epsilon$ is a better response, a contradiction. □

Lemma 3.A.1. *If π_I is a best-response to π_J such that $\pi_I < \pi_J \leq q_J(\pi \mid \omega_J)$, then $\pi_I = q_I(\pi \mid \omega_I)$.*

Proof. Suppose toward a contradiction that $\pi_I < \pi_J \leq q_J(\pi \mid \omega_J)$ and $\pi_I \neq q_I(\pi \mid \omega_I)$. If $\pi_I > q_I(\pi \mid \omega_I)$ a contradiction is immediate because hospital I can profitably deviate to $\pi_I - \epsilon$ for some $\epsilon > 0$ that is small such that $\pi_I > q_I(\pi_I - \epsilon, \pi_J \mid \omega_I)$. If $\pi_I < q_I(\pi \mid \omega_I)$ then observe that $\pi_J < q_J(\pi \mid \omega_J)$ by lemma 1. Consider the deviation by hospital I to $\pi_I + \epsilon$ for $\epsilon > 0$ small such that $\pi_I + \epsilon \leq q_I(\pi_I + \epsilon, \pi_J \mid \omega_I)$.

- If $\omega = \omega_I$, then hospital I is strictly better off as he has exchanged doctors in $[\pi, \pi_I{+}\epsilon]$ for doctors in $[q_I(\pi_I{+}\epsilon, \pi_J \mid \omega_I), q_I(\pi \mid \omega_I)]$, a strict improvement.
- If $\omega = \omega_J$, then hospital I is weakly better off because every doctor in $[\pi, \pi_I{+}\epsilon]$ is still available to hospital I.

Hence, the deviation by hospital I to $\pi_I + \epsilon$ is a strict improvement. □

Corollary 1. *If π is a equilibrium such that $\pi_L < \pi_H$, then $\pi_L = q_L(\pi \mid \omega_L)$.*

Corollary 2. *There is no equilibrium π such that $\pi_L > \pi_H$.*

Proof. Suppose toward a contradiction that $\pi_L > \pi_H$. Then $q_L(\pi \mid \omega_L) > \pi_L$ by Lemma 1. Lemma 2 implies that $\pi_H = q_H(\pi \mid \omega_H)$. But by definition, $q_L(\pi \mid \omega_L) < q_H(\pi \mid \omega_H)$, a contradiction. □

Proof of Theorem 3.1: The proof is as follows. First, I conjecture that for some π the following holds:

$$q_H(\pi \mid \omega_L) = q_L(\pi \mid \omega_H) \leq \pi_L \leq q_L(\pi \mid \omega_L) \leq \pi_H < q_H(\pi \mid \omega_H) \qquad (*).$$

Under $(*)$ I can calculate q as a function of π. I then solve for the first-order conditions to find π^* and see that $(*)$ holds under π^*. I then check if either hospital has a profitable deviation $\tilde{\pi}$. When checking for profitable deviations, I use Lemmas 1 and 2 to make the search tractable.

To show uniqueness, I again use Lemmas 1 and 2, and also Corollaries 1 and 2, to rule out many possible alternative equilibria π'. I then check any remaining π' by hand.

I use the computer software *Mathematica* to assist with algebraic manipulation. The commands executed can be found in the Supplementary Materials.

First, observe that $(*)$ implies

$$\begin{aligned}
q_H(\pi \mid \omega_L) &= 1 - 2\text{Cap} \\
q_L(\pi \mid \omega_H) &= 1 - 2\text{Cap} \\
q_L(\pi \mid \omega_L) &= \frac{1 - 2\text{Cap} + \pi_H r - 2\pi_L r}{1-r} \\
q_H(\pi \mid \omega_H) &= \frac{1 - \text{Cap} - \pi_H r}{1-r}.
\end{aligned}$$

I then calculate

$$\begin{aligned}
u_H(\pi) =& \Big(\text{Cap}^2(6-8r) + r(1 + 3(-2+\pi_H)\pi_H - 2(-2+\pi_L)\pi_L - 2(\pi_H - \pi_L)^2 r) \\
&+ 8\text{Cap}(-1 + (1 + \pi_H - \pi_L)r)\Big)\Big(8(-1+r)\Big)^{-1} \\
u_L(\pi) =& \Big(\text{Cap}^2(10-8r) + r(-1 - 3(-2+\pi_H)\pi_H + 2(-2+\pi_L)\pi_L + 2(\pi_H - \pi_L)^2 r) \\
&+ 8\text{Cap}(-1 + (1 - \pi_H + \pi_L)r)\Big)\Big(8(-1+r)\Big)^{-1}.
\end{aligned}$$

I see that $u_H(\pi)$ is concave in π_H, and $u_L(\pi)$ is concave in π_L. Hence, I need only solve the first-order conditions, which yields

$$\begin{aligned}
\pi_H^*(r) &= 1 - \frac{4\text{Cap}}{3+r} \\
\pi_L^*(r) &= 1 - \frac{6\text{Cap}}{3+r}.
\end{aligned}$$

Under $\pi^*(r)$ I see that $(*)$ holds.

I now check for profitable deviations.

Suppose (toward a contradiction) that $\tilde{\pi}_L$ is a profitable deviation for hospital L that is a best-response to π_H^*. Consider the following cases:

- If $\tilde{\pi}_L > \pi_L^*$, then I observe $q_L(\tilde{\pi}_L, \pi_H^* \mid \omega_L) < q_L(\pi^* \mid \omega_L) < \tilde{\pi}_L$, a contradiction to Lemma 1.
- If $\tilde{\pi}_L < \pi_L^*$, then I observe $q_L(\tilde{\pi}_L, \pi_H^* \mid \omega_L) > q_L(\pi^* \mid \omega_L) > \tilde{\pi}_L$, a contradiction to Lemma 2.

Hence, hospital L has no profitable deviations.

Suppose (toward a contradiction) that $\tilde{\pi}_H$ is a profitable deviation for hospital H that is a best-response to π_l^*. Consider the following cases:

- If $\tilde{\pi}_H > \pi_H^*$ such that $\tilde{\pi}_H > q_H(\tilde{\pi}_H, \pi_L^* \mid \omega_H)$, then this is a contradiction to lemma 1.

- If $\tilde{\pi}_H > \pi_H^*$ such that $\tilde{\pi}_H \leq q_H(\tilde{\pi}_H, \pi_L^* \mid \omega_H)$, then observe that $q_L(\tilde{\pi}_H, \pi_L^* \mid \omega_L) \leq \tilde{\pi}_H$. Thus $(*)$ continues to hold. This is a contradiction to π_H^* maximizing $u_H(x, \pi_L^*)$ under $(*)$.

- If $\tilde{\pi}_H < \pi_H^*$ such that $\tilde{\pi}_H > \pi_L^*$, then observe that $q_L(\tilde{\pi}_H, \pi_L^* \mid \omega_L) < \pi_L^*$. This case is considered in the cell "DEVIATION 1" in the supplementary *Mathematica* notebook. There, I derive q given π_L^* and $\tilde{\pi}_H$. I then derive hospital H's utility, show that it is concave in $\tilde{\pi}_H$, and that the first order conditions are satisfied for $\tilde{\pi}_H > \pi_H^*$. Hence, the best-response must be $\tilde{\pi}_H = \pi_H^*$, a contradiction.

- If $\tilde{\pi}_H < \pi_H^*$ such that $\tilde{\pi}_H \leq \pi_L^*$ and $\tilde{\pi}_H \geq q_H(\tilde{\pi}_H, \pi_L^* \mid \omega_L)$, then observe that $q_L(\tilde{\pi}_H, \pi_L^* \mid \omega_L) = q_L(\tilde{\pi}_H, \pi_L^* \mid \omega_H)$ and $q_H(\tilde{\pi}_H, \pi_L^* \mid \omega_L) \geq q_L(\tilde{\pi}_H, \pi_L^* \mid \omega_L)$. I note that the reversal in the order of $q_L(\tilde{\pi}_H, \pi_L^* \mid \omega_L)$ and $q_H(\tilde{\pi}_H, \pi_L^* \mid \omega_L)$ occurs precisely when $\tilde{\pi}_H$ and π_L^* switch order. This case is considered in the cell "DEVIATION 2" in the supplementary *Mathematica* notebook. There, I derive q given π_L^* and $\tilde{\pi}_H$. I then derive hospital H's utility, show that it is concave in $\tilde{\pi}_H$, and that the first order conditions are satisfied for $\tilde{\pi}_H > \pi_H^*$. Hence, the best response must be $\tilde{\pi}_H = \pi_L^*$, a contradiction.

- If $\tilde{\pi}_H < \pi_H^*$ such that $\tilde{\pi}_H \leq \pi_L^*$ and $\tilde{\pi}_H < q_H(\tilde{\pi}_H, \pi_L^* \mid \omega_L)$, then observe by Lemma 1 this cannot be a best response, a contradiction.

Hence, π^* is an equilibrium.

To show uniqueness, I note that by corollaries 1 and 2 (and lemma 1), I need only consider other equilibria π' of the form $\pi'_L = \pi'_H$. Suppose (toward a contradiction) that such a π' is an equilibrium. Then by Lemma 1 $q_H(\pi' \mid \omega_H) \geq \pi'_H \geq q_H(\pi' \mid \omega_L)$ and $q_L(\pi' \mid \omega_L) \geq \pi'_L \geq q_L(\pi' \mid \omega_H)$. Observe that $q_H(\pi' \mid \omega_L) = q_L(\pi' \mid \omega_L)$. This implies that either $q_H(\pi' \mid \omega_H) = 1 - 2\text{Cap}$ or $q_L(\pi' \mid \omega_L) = 1 - 2\text{Cap}$. I then observe that $q_L(\pi' \mid \omega_L) \leq q_H(\pi' \mid \omega_H)$. Thus, $q_L(\pi' \mid \omega_L) = 1 - 2\text{Cap}$. Hence, $\pi'_H = \pi'_L = 1 - 2\text{Cap}$.

In the cell "ALTERNATE EQUILIBRIUM BEST RESPONSE" in the attached *Mathematica* notebook, I show that the deviation

$$\pi_H^*(r) = 1 + 2\text{Cap}\left(-1 + \frac{1}{3 - 2r}\right)$$

is a profitable deviation, a contradiction. Hence, π^* is the unique equilibrium, and the theorem is proved.

□

Bibliography

Abdulkadiroğlu, Atila. 2011. "Generalized Matching for School Choice." working paper.

Abdulkadiroğlu, Atila, and Tayfun Sönmez. 1999. "House Allocation with Existing Tenants." Journal of Economic Theory, 88(2): 233–260.

Abdulkadiroğlu, Atila, and Tayfun Sönmez. 2003. "School Choice: A Mechanism Design Approach." American Economic Review, 93(3): 729–747.

Abdulkadiroğlu, Atila, and Yeon-Koo Che. 2010. "The Role of Priorities in Assigning Indivisible Objects: A Characterization of Top Trading Cycles." Working Paper.

Abdulkadiroğlu, Atila, Parag A. Pathak, and Alvin E. Roth. 2009. "Strategy-proofness versus Efficiency in Matching with Indifferences: Redesigning the NYC High School Match." American Economic Review, 99(5): 1954–1978.

Abdulkadiroğlu, Atila, Parag Pathak, Alvin Roth, and Tayfun Sonmez. 2006. "Changing the Boston School Choice Mechanism." National Bureau of Economic Research w11965, Cambridge, MA.

Abdulkadiroğlu, Atila, Yeon-Koo Che, Parag Pathak, Alvin Roth, and Olivier Tercieux. 2017. "Minimizing Justified Envy in School Choice: The Design of New Orleans' OneApp." National Bureau of Economic Research w23265, Cambridge, MA.

Adams, Michael, Thomas B Morrison, Stephanie Call, Andrew J Halvorsen, Jared Moore, Maria Lucarelli, Steven Angus, and Furman S McDonald. 2012. "Changes in the national residency matching program policy: are internal medicine program directors "all-in"?" Journal of Graduate Medical Education, 4(2): 148–153.

Ambuehl, Sandro, and Vivienne Groves. 2020. "Unraveling over time." Games and Economic Behavior, 121: 252–264.

Anderson, Kimberly D, and Donald M Jacobs. 2000. "General surgery program directors' perceptions of the match." Current surgery, 57(5): 460–465.

Anderson, Kimberly D, Donald M Jacobs, and Amy V Blue. 1999. "Is match ethics an oxymoron?" The American journal of surgery, 177(3): 237–239.

Ashlagi, Itai, Ephy Love, Jason I Reminick, and Alvin E Roth. 2023. “Early vs Single Match in the Transition to Residency: Analysis Using NRMP Data From 2014 to 2021.” Journal of Graduate Medical Education, 15(2): 219–227.

Azevedo, Eduardo M, and Jacob D Leshno. 2016. “A supply and demand framework for two-sided matching markets.” Journal of Political Economy, 124(5): 1235–1268.

Berriochoa, Camille, Matthew C Ward, Michael A Weller, Emma Holliday, Aaron Kusano, Charles R Thomas Jr, and Rahul D Tendulkar. 2016. “Applicant interview experiences and postinterview communication of the 2016 radiation oncology match cycle.” International Journal of Radiation Oncology* Biology* Physics, 96(3): 514–520.

Camp, Christopher L, Paul L Sousa, Arlen D Hanssen, Matthew D Karam, George J Haidukewych, Daniel A Oakes, and Norman S Turner. 2016. “Orthopedic Surgery Applicants: What They Want in an Interview and How They Are Influenced by Post–Interview Contact.” Journal of surgical education, 73(4): 709–714.

Carek, Peter J, Kimberly D Anderson, Amy V Blue, and Brian E Mavis. 2000. “Recruitment Behavior and Program Directors: How Ethical Are Their Perspective About the Match Process?” FAMILY MEDICINE-KANSAS CITY-, 32(4): 258–260.

Chacko, Karen M, Shalini Reddy, Michael Kisielewski, Stephanie Call, Lisa L Willett, and Saima Chaudhry. 2018. “Postinterview communications: two surveys of internal medicine residency program directors before and after guideline implementation.” Academic Medicine, 93(9): 1367–1373.

Chen, Siwei, Yajing Chen, and Chia-Ling Hsu. 2021. “New axioms for top trading cycles.” arXiv:2104.09157 [econ].

Combe, Julien. 2023. “Reallocation with priorities and minimal envy mechanisms.” Economic Theory, 76(2): 551–584.

Combe, Julien, and Jan Christoph Schlegel. 2024. “Reallocation with priorities.” Games and Economic Behavior, 143: 287–299.

Combe, Julien, Olivier Tercieux, and Camille Terrier. 2022. “The Design of Teacher Assignment: Theory and Evidence.” The Review of Economic Studies, 89(6): 3154–3222.

Cornett, Patricia A, Chris Williams, Richard L Alweis, John McConville, Michael Frank, Bhavin Dalal, Richard I Kopelman, Vera P Luther, Alec B O’connor, and Elaine A Muchmore. 2017. “Problematic communications during 2016 fellowship recruitment in internal medicine.” Journal of community hospital internal medicine perspectives, 7(5): 277–281.

Curran, Diana S, Pamela B Andreatta, Xiao Xu, Clark E Nugent, Samantha R Dewald, and Timothy RB Johnson. 2012. "Postinterview communication between obstetrics and gynecology residency programs and candidates." Journal of Graduate Medical Education, 4(2): 165–169.

Damiano, Ettore, Hao Li, and Wing Suen. 2005. "Unraveling of Dynamic Sorting." Review of Economic Studies, 72: 1057–1076.

Doe, Peter. 2024. "Matching With Pre-Existing Binding Agreements: The Agreeable Core." Working Paper.

Dubins, Lester, and David Freedman. 1981. "Machiavelli and the Gale-Shapley Algorithm." The American Mathematical Monthly, 88(7): 485–494.

Dur, Umut, and Scott Paiement. 2024. "A characterization of the top trading cycles mechanism for the school choice problem." Mathematical Social Sciences, 129: 93–100.

Dur, Umut Mert, A. Arda Gitmez, and Özgür Yılmaz. 2019. "School choice under partial fairness." Theoretical Economics, 14(4): 1309–1346.

Dur, Umut Mert, and M. Utku Ünver. 2019. "Two-Sided Matching via Balanced Exchange." Journal of Political Economy, 127(3): 1156–1177.

Dur, Umut Mert, and Thayer Morrill. 2017. "The Impossibility of Restricting Tradeable Priorities in School Assignment." working paper.

Echenique, Federico, and Jorge Oviedo. 2004. "A Theory of Stability in Many-to-many Matching Markets." SSRN Electronic Journal.

Echenique, Federico, and Juan Sebastián Pereyra. 2016. "Strategic complementarities and unraveling in matching markets." Theoretical Economics, 11(1): 1–39.

Echenique, Federico, Sumit Goel, and SangMok Lee. 2024. "Stable allocations in discrete exchange economies." Journal of Economic Theory, 222: 105921.

Ehlers, Lars, and Thayer Morrill. 2020. "(Il)legal Assignments in School Choice." The Review of Economic Studies, 87(4): 1837–1875.

Erdil, Aytek, and Haluk Ergin. 2008. "What's the Matter with Tie-Breaking? Improving Efficiency in School Choice." The American Economic Review, 98(3): 669–689. Publisher: American Economic Association.

Farber, Nicholas J, Christopher J Neylan, Arnav Srivastava, Shashank S Pandya, Amy Kaplan, Kushan D Radadia, Eric A Singer, and Sammy E Elsamra. 2019. "The urology match process and the limited value of post-interview communication for program directors." Urology, 128: 23–30.

Fereydooni, Arash, Joel L Ramirez, Katherine L Morrow, Sabina Sorondo, Jason T Lee, Dawn M Coleman, and Venita Chandra. 2022. "Interview experience, post-interview communication, and gender-based differences in the integrated vascular surgery residency match." Journal of Vascular Surgery, 75(1): 316–322.

Fragiadakis, Daniel, and Peter Troyan. 2017. "Improving matching under hard distributional constraints: Improving matching under constraints." Theoretical Economics, 12(2): 863–908.

Fragiadakis, Daniel, Atsushi Iwasaki, Peter Troyan, Suguru Ueda, and Makoto Yokoo. 2016. "Strategyproof Matching with Minimum Quotas." ACM Transactions on Economics and Computation, 4(1): 1–40.

Frishman, Gary N, Kristen A Matteson, Jessica L Bienstock, Karen E George, Tony Ogburn, Phillip N Rauk, Peter F Schnatz, and Lee A Learman. 2014. "Postinterview communication with residency applicants: a call for clarity!" American journal of obstetrics and gynecology, 211(4): 344–350.

Gale, D., and L. S. Shapley. 1962. "College Admissions and the Stability of Marriage." The American Mathematical Monthly, 69(1): 9–15. Publisher: Mathematical Association of America.

Gonczarowski, Yannai A., and Clayton Thomas. 2024. "Structural Complexities of Matching Mechanisms." 455–466. Vancouver BC Canada:ACM.

Grimm, Lars J, Carolyn S Avery, and Charles M Maxfield. 2016. "Residency postinterview communications: more harm than good?" Journal of Graduate Medical Education, 8(1): 7–9.

Guillen, Pablo, and Onur Kesten. 2012. "Matching Markets with Mixed Ownership: The Case for a Real-Life Assignment Mechanism." International Economic Review, 53(3): 1027–1046. Publisher: [Economics Department of the University of Pennsylvania, Wiley, Institute of Social and Economic Research, Osaka University].

Gusfield, Dan, and Robert W Irving. 1989. The stable marriage problem: structure and algorithms. MIT press.

Hafalir, Isa, Fuhito Kojima, and M. Bumin Yenmez. 2023. "Efficient Market Design with Distributional Objectives." 849–849. London United Kingdom:ACM.

Hałaburda, Hanna. 2010. "Unravelling in two-sided matching markets and similarity of preferences." Games and Economic Behavior, 69(2): 365–393.

Hamada, Naoto, Chia-Ling Hsu, Ryoji Kurata, Takamasa Suzuki, Suguru Ueda, and Makoto Yokoo. 2017. "Strategy-proof school choice mechanisms with minimum quotas and initial endowments." Artificial Intelligence, 249: 47–71.

Handa, Nicole, Amanda X Vo, Jenny Guo, Mary Kate Keeter, and Stephanie J Kielb. 2021. "The Impact of Stricter Guidelines on Post-Interview Contact, Restricted Interview Topics and Differences by Gender in the Urology Match." Urology Practice, 8(1): 143–148.

Harvey, Rebecca S, Kevin J Kovatch, Mark EP Prince, Kelly M Malloy, and Marc C Thorne. 2019. "Trends in Post-Interview Communication Practices." The Laryngoscope, 129(3): 607–612.

Holliday, Emma B, Charles R Thomas Jr, and Aaron S Kusano. 2015. "Integrity of the national resident matching program for radiation oncology: national survey of applicant experiences." International Journal of Radiation Oncology* Biology* Physics, 92(3): 525–531.

Jena, Anupam B, Vineet M Arora, Karen E Hauer, Steven Durning, Nicole Borges, Nancy Oriol, D Michael Elnicki, Mark J Fagan, Heather E Harrell, Dario Torre, et al. 2012. "The prevalence and nature of postinterview communications between residency programs and applicants during the match." Academic Medicine, 87(10): 1434–1442.

Kesten, Onur. 2006. "On two competing mechanisms for priority-based allocation problems." Journal of Economic Theory, 127(1): 155–171.

Kesten, Onur. 2010. "School Choice with Consent *." Quarterly Journal of Economics, 125(3): 1297–1348.

Kojima, Fuhito, Akihisa Tamura, and Makoto Yokoo. 2018. "Designing matching mechanisms under constraints: An approach from discrete convex analysis." Journal of Economic Theory, 176: 803–833.

Kwon, Hyukjun, and Ran I Shorrer. 2023. "Justified-Envy-Minimal Efficient Mechanisms for Priority-Based Matching." working paper.

Leshno, Jacob D, and Irene Lo. 2017. "The Simple Structure of Top Trading Cycles in School Choice: A Continuum Model." Working Paper.

Li, Hao, and Sherwin Rosen. 1998. "Unraveling in matching markets." American Economic Review, 371–387.

Ma, Jinpeng. 1994. "Strategy-proofness and the strict core in a market with indivisibilities." International Journal of Game Theory, 23(1): 75–83.

Miller, John Bernard, Douglas C Schaad, Robert A Crittenden, Nancy E Oriol, and Carol MacLaren. 2003. "Communication between programs and applicants during residency selection: effects of the match on medical students' professional development." Academic Medicine, 78(4): 403–411.

Monir, Reesa L, Kristina Michaudet, Joseph G Monir, Kiarash P Rahmanian, Charlie Michaudet, Lou Ann Cooper, and Heather Harrell. 2021. "Impact of Match Violations on Applicants' Perceptions and Rankings of Residency Programs." Cureus, 13(1).

Morrill, Thayer. 2013a. "An alternative characterization of the deferred acceptance algorithm." International Journal of Game Theory, 42(1): 19–28.

Morrill, Thayer. 2013b. "Making Efficient School Assignment Fairer." working paper.

Morrill, Thayer. 2015. "Making just school assignments." Games and Economic Behavior, 92: 18–27.

Niederle, Muriel, Alvin E Roth, and M Utku Ünver. 2013. "Unraveling results from comparable demand and supply: An experimental investigation." Games, 4(2): 243–282.

Opel, Douglas, Richard Shugerman, Heather McPhillips, Wendy Sue Swanson, Sarah Archibald, and Douglas Diekema. 2007. "Professionalism and the match: a pediatric residency program's postinterview no-call policy and its impact on applicants." Pediatrics, 120(4): e826–e831.

Papai, Szilvia. 2000. "Strategyproof Assignment by Hierarchical Exchange." Econometrica, 68(6): 1403–1433.

Pearson, RD, and AH Innes. 1999. "Ensuring compliance with NRMP policy." Academic Medicine, 74(7): 747–8.

Pereyra, Juan Sebastián. 2013. "A dynamic school choice model." Games and Economic Behavior, 80: 100–114.

Program, National Resident Matching. 2023. "Results and Data 2023 Main Residency Match®." `https://www.nrmp.org/wp-content/uploads/2023/05/2023-Main-Match-Results-and-Data-Book-FINAL.pdf`.

Ratcliffe, Temple A, Steven J Durning, Anupam B Jena, Thomas Grau, Anthony R Artino Jr, Vineet M Arora, Karen E Hauer, Nicole Borges, Nancy Oriol, D Michael Elnicki, et al. 2012. "Postinterview communication between military residency applicants and training programs." Military medicine, 177(suppl_9): 54–60.

Reny, Philip J. 2022. "Efficient Matching in the School Choice Problem." American Economic Review, 112(6): 2025–2043.

Rong, Kang, Qianfeng Tang, and Yongchao Zhang. 2022. "The Core of School Choice Problems." Working Paper.

Roth, Alvin E. 1985. "The college admissions problem is not equivalent to the marriage problem." Journal of Economic Theory, 36(2): 277–288.

Roth, Alvin E. 1991. "A natural experiment in the organization of entry-level labor markets: Regional markets for new physicians and surgeons in the United Kingdom." The American economic review, 415–440.

Roth, Alvin E, and Elliott Peranson. 1999. "The redesign of the matching market for American physicians: Some engineering aspects of economic design." American economic review, 89(4): 748–780.

Roth, Alvin E., and Marilda A. Oliveira Sotomayor. 1990. Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis. Econometric Society Monographs, Cambridge University Press.

Roth, Alvin E., and Uriel G. Rothblum. 1999. "Truncation Strategies in Matching Markets—in Search of Advice for Participants." Econometrica, 67(1): 21–43.

Roth, Alvin E, and Xiaolin Xing. 1994. "Jumping the gun: Imperfections and institutions related to the timing of market transactions." The American Economic Review, 992–1044.

Sbicca, Jennifer A, Emily S Gorell, David H Peng, and Alfred T Lane. 2012. "A follow-up survey of the integrity of the dermatology National Resident Matching Program." Journal of the American Academy of Dermatology, 67(3): 429–435.

Sbicca, Jennifer A, Emily S Gorell, Matthew H Kanzler, and Alfred T Lane. 2010. "The integrity of the dermatology National Resident Matching Program: results of a national study." Journal of the American Academy of Dermatology, 63(4): 594–601.

Sebesta, Elisabeth M, Michael J Lipsky, Michele Nunez, Kimberly L Cooper, and Gina M Badalato. 2018. "The national resident matching program code of conduct: what is the perceived degree of compliance during the urology match process?" Urology, 122: 37–43.

Shapley, Lloyd, and Herbert Scarf. 1974. "On cores and indivisibility." Journal of Mathematical Economics, 1(1): 23–37.

Sönmez, Tayfun. 1999. "Can pre-arranged matches be avoided in two-sided matching markets?" Journal of Economic theory, 86(1): 148–156.

Swan, Emma C, and Thomas E Baudendistel. 2014. "Relationship between postinterview correspondence from residency program applicants and subsequent applicant match outcomes." Journal of Graduate Medical Education, 6(3): 478–483.

Teichman, Kimberly D Anderson, Marsha M Dorough, Catherine R Stein, Scott A Optenberg, and Ian M Thompson. 2000. "The urology residency matching program in practice." The Journal of urology, 163(6): 1878–1887.

Thurman, R Jason, Eric Katz, Wallace Carter, Jin Han, Eve Kayala, Nicole McCoin, and Alan B Storrow. 2009. "Emergency medicine residency applicant perceptions of unethical recruiting practices and illegal questioning in the match." Academic Emergency Medicine, 16(6): 550–557.

Troyan, Peter, David Delacrétaz, and Andrew Kloosterman. 2020. "Essentially stable matchings." Games and Economic Behavior, 120: 370–390.

Ueda, Suguru, Daniel Fragiadakis, Atsushi Iwasaki, Peter Troyan, and Makoto Yokoo. 2012. "Strategy-proof mechanisms for two-sided matching with minimum and maximum quotas (Extended Abstract)." Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems.

Williams, Christopher M, Richard L Alweis, Alec B O'Connor, Bhavin Dalal, Devesh Rai, Abdullah Abdullah, Richard Kopelman, Patricia Cornett, Michael O Frank, Vera P Luther, et al. 2019. "Inappropriate communication during internal medicine fellowship recruitment: a mixed-methods analysis." The American journal of medicine, 132(6): 770–775.