Matching With Pre-Existing Binding Agreements: The Agreeable Core

Peter Doe

October 28, 2024

Abstract

I analyze two-sided matching markets with pre-existing binding agreements between market participants. In this model, a pair of participants who are bound to each other by a pre-existing agreement must mutually agree to any action taken by the pair. I propose a new solution concept, the *agreeable core*, consisting of the matches which cannot be renegotiated without violating the binding agreements. The main contribution of this paper is an algorithm that constructs such a match by a novel combination of the Deferred Acceptance and Top Trading Cycles algorithms. I show that my algorithm is robust to various manipulations and it has applications to numerous markets including the resident-to-hospital match, college admissions, school choice, and labor markets.

1 Introduction

I analyze two-sided matching markets with pre-existing binding agreements between market participants. These agreements impose interdependencies among participants' rights. 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.

I propose a new model to analyze these markets. This is necessary because the standard model endows participants with the unrestricted right to form new agreements and does not respect the pre-existing agreements. At a basic level, my model requires that any action taken by one person must receive the approval the person she is bound to. For example, the professional athlete can only seek other employment with the approval of her team. Without this 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. No person outside of this group can impede the negotiation because the group is agreeable.

In the agreeable core participants are granted maximum flexibility in dissolving agreements and forming new matches. Critically, when a group considers renegotiating a candidate outcome, it does so before the pre-existing agreements are dissolved; after the pre-existing agreements are dissolved, the candidate outcome becomes the set of binding agreements. In effect, each participant has the flexibility to secure a replacement before releasing her partner from their binding agreement. 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. Without this flexibility, market congestion—where participants delay decision-making due to the lack of guarantees on replacement agreements—can result when participants with binding agreements hesitate to dissolve them without certainty of better outcomes For instance, if a player is released from her pre-existing contract, the particular team she signs on to next directly affects which other players are available to her original team. The agreeable core resolves this by allowing the formation of a new agreement immediately after the old agreement is dissolved.

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. 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 a 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." 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 paper 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 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 residencies are offered 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 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 a decentralized market is also straightforward: take the outcome of the decentralized market as the set of binding agreements. Because the PE (and the agreeable core in general) guarantees that no agent is worse off then the binding agreements they enter the market with, the PE offers an incentive for participation to agents 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 (seeFragiadakis and Troyan (2017) for 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 cores 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 paper 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 from 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 paper 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. Arguably, outcomes are easier to explain than algorithms.[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 paper proceeds as follows. In Section 2 I motivate the agreeable core through an illustrative example. Section 3 presents the model. In Section 4 I present the proof of my main result, the Propose-Exchange algorithm that always produces a match in the agreeable core. Section 5 contains several results related to the manipulability of the Propose-Exchange algorithm. I defer a discussion of the related literature until Section 6, where I discuss how the agreeable core

[1] For example, the statement *your child is at highest ranked school you listed where she is above the school's cutoff* is easier for parents to understand compared to *we used the only algorithm that satisfies non-wastefulness, population monotonicity, weak Maskin monotonicity, and mutual best*; see Morrill (2013*a*).

presents an alternative understanding of several economic applications.

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, 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

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: 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.

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 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 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 the 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 improvements 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

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
$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$

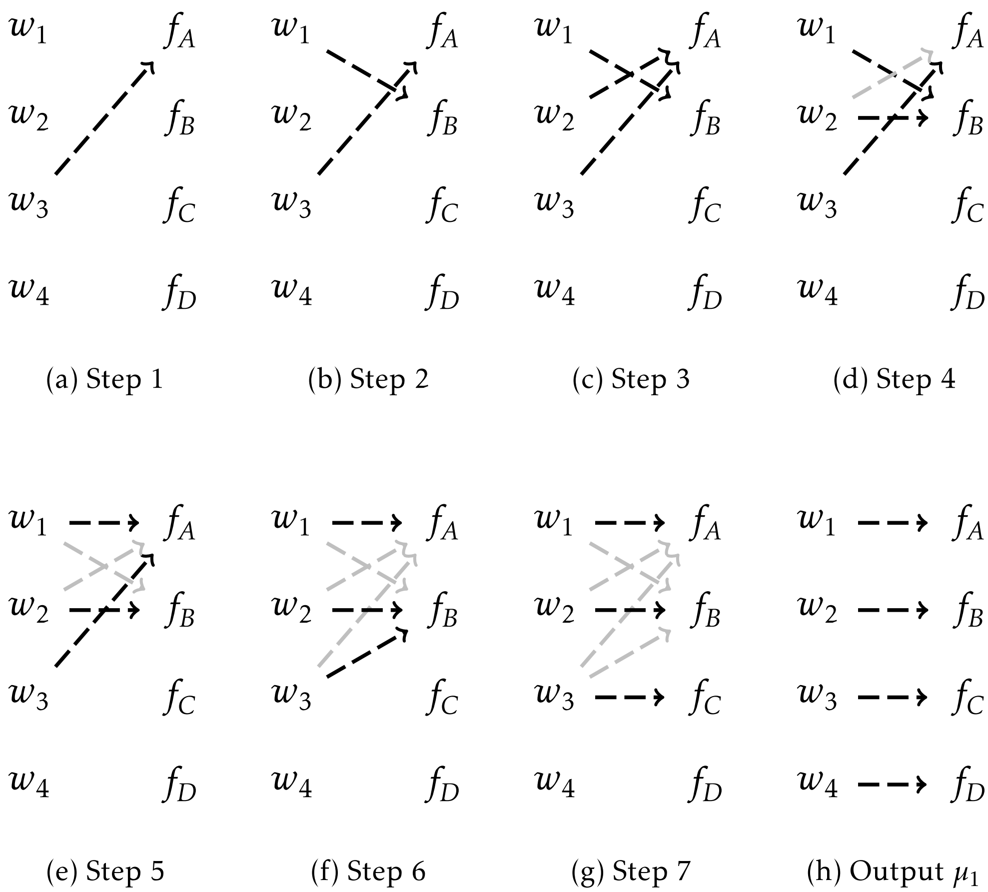
(b) Second Pareto improvement, $\dot{\mu}$

Figure 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 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 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 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 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 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



(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 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.

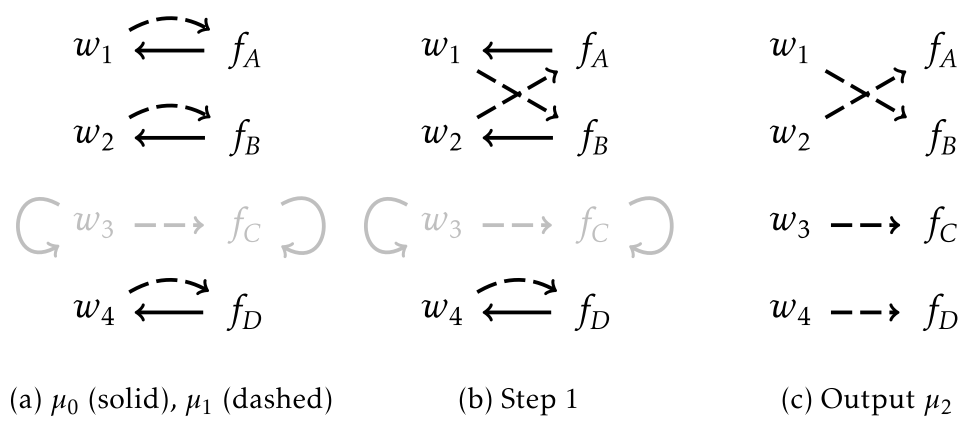


Figure 4: A visualization of the steps of the Exchange stage. Agents w_3 and f_C are excluded because their μ_0- and μ_1-partners differ. There is only one step because each active agent is in a cycle in Step 1.

match.

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 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 $a = \mu(a)$ mean that a is unmatched. Formally, *match* is a function $\mu : A \to A$ 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 $\mu \gtrsim_X \mu'$ to mean $\mu(x) \gtrsim_x \mu'(x)$ 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 $\mu \gtrsim_A \mu_0$.

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.

3.1 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

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

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]

3.2 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*

[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.

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 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 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 3.3 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 3.4 and Section 3.5 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.

3.3 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 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.

3.4 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

[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*.

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.

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 4.2 how to resolve this tension by limiting which workers can propose.

Algorithm 1 Deferred Acceptance (DA) algorithm

Notation: when I write $\mu^{\text{DA}}(a) \leftarrow w$, 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 $\mu^{\text{DA}}(f) \leftarrow f$ 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 $\mu^{\text{DA}}(w) \leftarrow w$.

if f prefers w to $\mu^{\text{DA}}(f)$ then set $\mu^{\text{DA}}(f) \leftarrow w$.

else f rejects w.

end while

return μ^{DA}

3.5 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 $\mu^{\text{TTC}} \succsim_W \mu_0$. Second, no worker may point to a firm if that firm prefers her μ_0-partner to the worker. This guarantees that $\mu^{\text{TTC}} \succsim_W \mu_0$.

In my setting, however, these modifications are not enough. As I saw in Section 3.4, 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 Section 4.3 I incorporate

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}

this by limiting which workers and firms participate in the TTC.

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 3.4 and Section 3.5, 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 4.1, then introduce the Propose stage in Section 4.2, and finally the Exchange stage in Section 4.3. 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 Appendix A.

4.1 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.

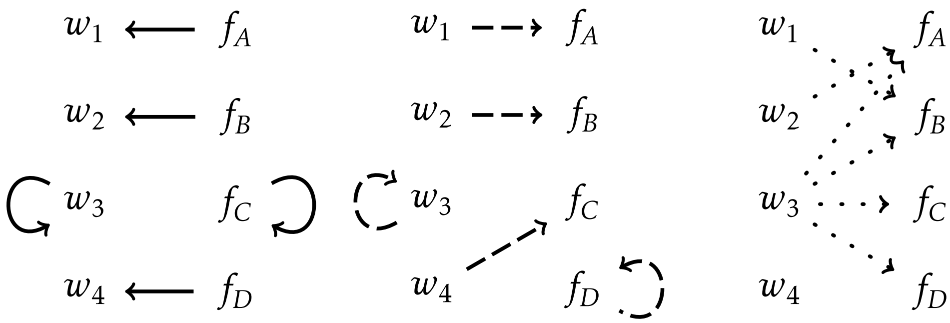
Figure 5 depicts the three types of edges using the set-up of Example 1 and a match μ that modifies the initial match μ_0 by leaving f_D unmatched and matching w_4 to f_C. Subfigure (a) includes the edges from μ_0, which are either loops (in the case of w_3 and f_C) or point from F to W. Subfigure (b) 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 $P = (e_1, \dots, e_n)$ 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.

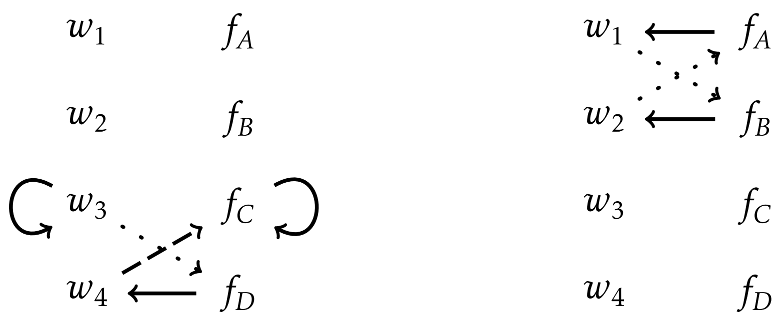
A path P is *complete* if every vertex contained in the path is contained in exactly two edges of the path. A path is *alternating* if it no two consecutive edges (including loops) alternate between μ_0 and $\mu \cup I(\mu)$.[6] Two complete and alternating paths are depicted in subfigures (d) and (e) of Figure 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 $\mu^P(w) = f$.

[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 $\mu \cup I(\mu)$.



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



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

Figure 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.

- 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 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*.

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 corre-

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

sponds 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 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.

4.2 The Propose Stage

The first stage of my algorithm outputs a match μ_1 by systematically removing all acyclic blocking paths from $(A, \mu_0 \cup \mu \cup I(\mu))$. A 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 3.4 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.

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

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 a acyclic blocking path P in $(A, \mu_0 \cup \mu \cup I(\mu))$ always 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.*

4.3 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.

4.4 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. $\square$

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.

4.5 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 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 4.5. 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

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

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.

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 a 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

[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.

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

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

(a) Initial match—μ_0

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

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

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

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

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

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

Figure 6: An example showing that the outcomes in the agreeable core cannot be ordered by $\succsim_W$.

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.*

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.

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.

[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.

5.1 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 $\psi(W, F, \succsim, \mu_0)$.

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

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 before preference swap

Figure 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 $w1$ and w_2 to receive new partners.

can, by misreporting his preference, affect the matches of the other agents without changing his own. Consider the example in Figure 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.

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

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.

5.2 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 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 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.

6 Conclusion

This paper 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 close with a discussion of my modeling choices and possible extensions. Second, I provide guidance on applying the Propose-Exchange algorithm in several environments. Third, I review the connections between this paper and existing research.

6.1 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.[12] 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.[13] 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 match to (who themselves may be ini-

[12]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?

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

tially 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.

6.2 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.[14] 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 resi-

[14] 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.

dents 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 tenured professor or teacher's contract allows her leave her institution unilaterally, 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.[15]

[15]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.

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" 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.

6.3 Connection to the Literature

This paper 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 paper, so I list the only the most closely related work and its connections with this paper.

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. My paper 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 plethora 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, 2013*b*; 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.

References

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.

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.

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

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*.

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.

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

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].

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

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.

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

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*.

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

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

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

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

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

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.

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

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.*

A 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 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 $P = (e_1, \dots, e_n)$ 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. □

Introduction to the proofs of Lemma 1 and Proposition 2:

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

Definition 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 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 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,

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

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. □

[17] namely, μ_0

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

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 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

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

and

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

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. □

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

Lemma 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 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, $P \not\subseteq A^F$.

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, $P \not\subseteq A \backslash A^F$.

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 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

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

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

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

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

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

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

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 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 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 $f' \in A^F$:

$$P^{f'} = (\ldots, (\mu(f'), f'), (f', \mu_0(f')), e^{f'}_{j-1}, \ldots, e^{f'}_1).$$

Then the path

$$P^* = (\underbrace{e^w_1, \ldots, e^w_{i-1}, (\mu_0(w), w),}_{P^w} \overbrace{(w, f), e_{k+1}, \ldots, e_{K-1}, (w', f'),}^{P} \underbrace{(f', \mu_0(f')), e^{f'}_{j-1}, \ldots, e^{f'}_1}_{P^{f'}}).$$

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. $\square$

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:

$$\mu_0(1) = A \qquad \mu_0(2) = B \qquad \mu_0(9) = 9$$
$$\mu_0(A) = 1 \qquad \mu_0(B) = 2 \qquad \mu_0(Z) = Z$$

A profile of preferences $\succ$ and an alternate profile of worker preferences are given in Figure 8. I use the circles to indicate match μ°, the squares to indicate match $\mu^{\blacktriangle}$, and $\tilde{}$ to indicate $\tilde{\mu}$.

$$\mu^{\circ}(1) = B \qquad \mu^{\circ}(2) = A \qquad \mu^{\circ}(9) = Z$$
$$\mu^{\circ}(A) = 2 \qquad \mu^{\circ}(B) = 1 \qquad \mu^{\circ}(Z) = 9$$

$$\mu^{\blacktriangle}(1) = A \qquad \mu^{\blacktriangle}(2) = Z \qquad \mu^{\blacktriangle}(9) = B$$
$$\mu^{\blacktriangle}(A) = 1 \qquad \mu^{\blacktriangle}(B) = 9 \qquad \mu^{\blacktriangle}(Z) = 2$$

$$\tilde{\mu}(1) = A \qquad \tilde{\mu}(2) = B \qquad \tilde{\mu}(9) = Z$$
$$\tilde{\mu}(A) = 1 \qquad \tilde{\mu}(B) = 2 \qquad \tilde{\mu}(Z) = 9$$

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.

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°	$Z^\blacktriangle$	$B^\blacktriangle$
Z	A°	Z°
$A^\blacktriangle$	B	

P_3

$\succ_1'$	$\succ_2'$	$\succ_9$
B	$Z^\blacktriangle$	$B^\blacktriangle$
Z	A	$\tilde{Z}$
$\tilde{A}^\blacktriangle$	$\tilde{B}$	

P_4

$\succ_1'$	$\succ_2'$	$\succ_9'$
B	$Z^\blacktriangle$	$B^\blacktriangle$
Z	A	Z
$A^\blacktriangle$	B	

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

Figure 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.

- 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 8. An exhaustive search reveals that μ°, $\mu^\blacktriangle$, 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.

✗ $\mu^\blacktriangle$ 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 $\{\mu^\circ, \mu^\blacktriangle\}$ 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.

- ✓ $\mu^{\blacktriangle}$ 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 $\psi(P_2) = \mu^{\blacktriangle}$, then consider the deviation by worker 1 of misreporting $\succ_1$ at P_2. Because $\mu^{\circ}(1) \succ_1' \mu^{\blacktriangle}(1)$, 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 $\{\mu^{\blacktriangle}, \tilde{\mu}\}$ because:

- ✗ μ° matches worker 2 to firm A, which violates the requirement that $\mu(2) \succsim_2 B$.

- ✓ $\mu^{\blacktriangle}$ 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 $\psi(P_3) = \mu^{\blacktriangle}$, then consider the deviation by worker 2 of reporting $\succ_2'$ at P_2. Because $\mu^{\blacktriangle}(2) \succ_2 \mu^{\circ}(2)$, 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 $\mu^{\blacktriangle}$. 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. $\mu^{\blacktriangle}$ 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 $\mu^{\blacktriangle}(9) \succ_9 \tilde{\mu}(9)$, this is a profitable deviation. Therefore, ψ is preference manipulable, a contradiction. □

Introduction to the proofs of Theorem 3:

Lemma 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. $\square$

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 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

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

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 $\mu_1'(W') \setminus \mu_1(W')$ 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 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. $\square$