Matching agents to fighting clients [*]

Kemal Kıvanç Aköz[†] Emre Doğan[‡] Danisz Okulicz[§]

December 30, 2021

Abstract

There are two sets of clients exogenously pre-matched in battles. Each client needs help from an agent to fight the battle. We study the stability of matchings of the form: agent-client × opponent client-opponent agent. Clients and agents are objectively ranked. The existence and characteristics of stable matchings depend on the structure of the prematching of clients. We propose a novel notion of comparing two-sided matchings in terms of assortativity, by representing them as partial orders. More extensive partial orders correspond to more positive assortative prematchings. If the prematching is close to negative assortative, a stable matching always exists. In any stable matching the induced matching between agents can be at most as positive assortative as the prematching. We provide two domains of preferences for which a stable matching always exists. We show that stability and core notions are independent, but stable matchings are always efficient.

JEL classification: C78, D62
Keywords: Matching, stability, assortative

[*] We would like to thank Steven Kivinen, Alexander Nestorov, David Pérez-Castrillo, the participants of HSE-NES Workshop in Microeconomic Theory for their useful comments, and Pavel Stupatsky for his excellent research assistance.

[†] HSE University (e-mail:kakoz@hse.ru).
[‡] HSE University (e-mail:edogan@hse.ru)
[§] HSE University (e-mail:dokulicz@hse.ru).

"Good lawyers worry about facts.
Great lawyers worry about their opponents."
Suits (Season 1; Ep. 7)

1 Introduction

Consider a set of legal disputes. Each dispute represents a conflict between two individuals – a plaintiff and a defendant, but both the plaintiff and the defendant need a lawyer to represent them. How are lawyers matched with the disputes? Alternatively, consider electoral races in multiple districts. Which politicians run where and against whom? There is a similarity between these two scenarios. In both cases, there is a set of predetermined battles between two sides: be it legal disputes between the plaintiffs and the attorneys, or political conflicts between the local left-wing and right-wing partisans. However, the battles are not resolved by the sides fighting directly, but rather through an agent representing a given side: lawyers in the case of legal disputes and political candidates in political races. In such environments which matchings are stable when both the sides of the battle and the agents have preferences over their entire match? Do stable matchings exist in general?

In this paper, we propose a four-sided matching model with conflict of interest and provide an analysis of stability in this setting. There is a finite number of battles and each battle consists of two clients that are in conflict with each other. We call the two-sided one-to-one matching of clients through battles a *prematching*. Each client needs the help of an agent in her battle. Agents are split into two sets according to the sides of the battles they can be matched with. We analyze the matchings between agents and clients, where each match is a quadruple of the form agent-client × opponent client- opponent agent. Clients have preferences over pairs of their agent and the opponent agent (following e.g., Becker 1973, Burdett and Coles 1997, Chade and Eeckhout 2020). Agents have preferences over pairs of their client and opponent agent. There is an objective ranking of both the agents and the clients. Clients prefer to be matched with better agents and their opponents to match with worse agents. Agents prefer to be matched with better clients and to face worse opponent agents. As the prematching is given, the objective ranking of clients captures the preferences of the agents over battles. Agents may have heterogeneous preferences over how they prioritize matching with better clients over worse agents. Moreover, there does not need to be an agreement on the ranking of battles across sides.

The focus of our analysis is the stability of matchings. We adopt the classical notion of pairwise stability [Gale and Shapley, 1962] to our setting. We call a matching (pairwise) stable when no agent-client pair on the same side forms a blocking pair, treating the matching of the other side as given. This notion is consistent with environments with conflicts where communication within each side is much easier than across sides. Although we do not provide a theory about how stable matchings arise, we believe that there are several environments in which stability is important. First, we can think of completely centralized markets in which some authority can decide on the allocation. An example of such an environment is the allocation of public attorneys and prosecutors to criminal cases with indigent defendants. There is evidence that market participants consider unstable allocations unjust. For example, Fleeta Drumgo sued the Superior Court of Marin Country for not granting his request to appoint Richard Hodge as his attorney, while Richard Hodge was ready and willing to do so (Drumgo vs Superior Court, 1973). Although Drumgo lost the case, the ruling was not unanimous with one judge dissenting, arguing that the criminal justice system may lose legitimacy if the indigent defendants are not allowed to choose their representation from a set of lawyers willing to represent them. Due to similar concerns, in 2015 Comel county (Texas) introduced a system in which indigent defendants could choose their own lawyers, successfully improving the perception about the legitimacy of the system in the eyes of the defendants [Nugent-Borakove and Cruz, 2017]. Second, our model is important for markets in which there are two separate clearinghouses, each responsible for one side of the battle. An example of such an environment is the allocation of candidates to political races by political parties. If a party consistently assigns politicians in an unstable way, popular politicians may decide to leave the party and run independently. Indeed, there is a large body of evidence in which political parties introduce primary elections in order to mitigate intra-party conflict [Ichino and Nathan, 2012, De Luca et al., 2002]. Primaries have the potential to mitigate the instability of assignments. If a Democratic candidate ran in California rather than in Texas, and they were believed to be more popular among voters in California than the current party nominee there, they could challenge the nominee in primaries and block the allocation. Finally, stable allocations are a natural first prediction for decentralized markets [Burdett and Coles, 1997, Echenique and Yariv, 2012]. The market for attorneys in legal cases fits our structure especially well. The fees for attorneys are heavily regulated and there is little variance in the contracts offered on the market with a standard contract charging 33% of the amount won (see Helland and Tabarrok 2003 for an overview of regulation in the US). As a result, the market for civil legal services essentially becomes a matching market without transfers (or contracts). If an allocation in such a market consistently deviates from a stable allocation, it means that some good lawyers consistently accept cases of little value, while they could be working on valuable cases. As such, it is natural to expect that over time the market allocation should become close to stable.

Our central contribution is to observe that stable matchings and their characteristics crucially depend on the structure of the prematching of clients. In particular, a stable matching is guaranteed to exist whenever battles which are attractive for one side tend to be less attractive for the other side. In other words, if a prematching is close to negative assortative then a stable matching always exists. The prematching determines whether two strong agents can oppose one another in a stable matching. We show that, in a sense, the matching of the agents has to be more negative assortative than the prematching.

We first study an important case where the prematching of clients is a negative assortative matching (NAM), capturing situations in which the battles are primarily differentiated by their difficulty. That is, if a battle is easy and attractive for agents on one side, it should be difficult and unattractive for the agents on the other side. There always exists a stable matching, in particular a positive assortative matching (PAM) matching of agents and clients is stable for all preferences. There can be multiple stable matchings, however, and in any such matching agents need to be matched negatively assortatively.

In Section 4, we allow for any type of prematching of clients so that we can analyze situations in which some battles are appealing for agents on both sides. In general, prematchings can be neither positive nor negative assortative. We develop a novel tool of describing assortativity in the intermediate cases. With that aim we propose a relation of Positive Assortative dominance (PA-dominance) over the set of battles. We say a battle PA-dominates another battle when the former is more attractive for agents on both sides. The PA-dominance relation allows us to think of prematchings as partially ordered sets. More extensive partially ordered sets correspond to more positive assortative prematchings. This notion can be used for any two-sided matching with objective rankings, including the matching of agents on opposing sides.

Our first observation is that the existence of a stable prematching is not guaranteed, as is the case in general for multi-sided matchings [Alkan, 1988]. We show that a stable matching exists for any preference profile if and only if the prematching is *bipartite*, i.e., it can be partitioned into two sets in such a way that all PA-dominance relations go from one set to another. We provide an algorithm based on serial dictatorship which finds a stable matching if the prematching is bipartite. Additionally, we provide two preference domains for which a stable matching exists irrespective of the prematching. Literature on multiple-sided matchings and matchings with externalities shows that stable matchings exist under various forms of lexicographic preferences[Danilov, 2003, Eriksson et al., 2006, Huang, 2010, Dutta and Massó, 1997]. In our setting we are able to generalize this result to a domain of threshold preferences. An agent with threshold preferences has an ordered partition of battles and always prefers battles from better groups but is concerned only with the opponent within each group. Furthermore, we show that the domain of threshold preferences is rich enough that for any matching that is stable under some preference profile it is also stable under some threshold preference profile. Finally, we consider a setting in which clients can be split into two types (similarly to e.g., Chade and Eeckhout 2020). In this environment, the agents are not interested in the exact identity of their client, but only in whether she is of a "good" or a "bad" type. Then, a stable matching always exists.

In Section 4.2, we move to the analysis of the characteristics of stable matchings. Namely, we describe a set of potentially stable matchings, that is, matchings which can be stable for some preference profile given a prematching. A matching is potentially stable if and only if the PA dominance relation is preserved from the pairs of agents to the pairs of clients that those agents match. Using this characterization, we provide a map from prematchings to the set of potentially stable agent matchings that can be observed at any given prematching. We prove that an agent matching is supported by a prematching if and only if it is at most as positive assortative as the prematching. That is, the partially ordered set describing the prematching needs to be isomorphic to some extension of the partially ordered set describing the matching of agents. This result suggests that the negative assortative matchings of agents should be commonly observed, whereas the positive assortative matchings of agents should be rare.

Finally, in Section 5 we explore the relation between our stability notion, efficiency and core. We first prove that stability implies efficiency. In our setting core significantly differs from pairwise stability because it requires cooperation from all four sides of the matching, which is hardly possible in conflict scenarios. Moreover, the core often has little predictive power, and it is possible that the entire set of matchings is in the core. Still, if the clients are more concerned with their opponents' agents than with their own agents, stable matchings may not belong to the core.

There are several articles studying matching problems in environments where the conflict is relevant. Kamali Shahdadi [2018] proposes a theoretical model of assigning public attorneys to indigent defendants, when the attorneys are subject to moral hazard. Agan et al. [2021] and Shem-Tov [2020] provide an empirical analysis of the assignment of public attorneys. However, they study the problem as a two-sided matching problem of defendants to attorneys, ignoring possible effects of the resulting matchings of prosecutors and attorneys. To our knowledge Iossa and Jullien [2012] is the only article considering the market for legal services as a four-sided market, in which both the attorney and the plaintiff hire lawyers and the outcome of the case depends on the entire match. However, Iossa and Jullien [2012] focus on the interplay of the career concerns of judges and the certification system of lawyers. As such, the lawyers can be of only two types: certified or not certified. Moreover, they sidestep the problems with defining and analyzing stability by assuming that the matching procedure is sequential and one side can commit to the choice of lawyer. Galasso and Nannicini [2011] consider the allocation of politicians to political races by two competing parties. The politicians can only be one of two types (expert or loyalist), and the parties allocate the politicians in order to maximize the likelihood of winning the election. Parties are assumed to have absolute power over the allocations, and the resulting assignments do not need to be stable. Our analysis complements Galasso and Nannicini [2011] as it describes allocations which are implementable when parties have no leverage over their members. In that sense, the results of Galasso and Nannicini [2011] are more applicable in countries with strong political parties while ours work better in countries with weak political parties.

Instead of four-sided matching, our setting can be thought of as matching with across-markets externalities. A matching of agents and clients on one side generates an externality on the other side and the prematching of the clients describes the externality structure. Sasaki and Toda [1996] and Hafalir [2008] propose stability notions in which the agents form a conjecture about the effect that their pair will have on the whole market before they decide to block. Our stability notion is more restrictive, as it allows the agents and the clients to ignore the reaction of the opponents while forming a blocking pair. It is similar to, for example, Pycia and Yenmez [2021] and Mumcu and Saglam [2010] in one-to-one matching environments. Similarly to our setting, the existence of stable matchings when externalities are present is not ensured. We provide the conditions for existence not only in terms of preference domains but also in terms of the externality structure. We follow recent literature in describing the assortativity properties of matchings with externalities [Chen, 2019, Chade and Eeckhout, 2020, Chen, 2021]. In particular, Chade and Eeckhout [2020] propose an example of firms competing in a Cournot duopoly in different markets and hiring workers which will influence the cost functions of the firms. As the competing firms are exogenously fixed into pairs, this example is conceptually similar to our framework. Leaving aside the technical differences (e.g., they study a transferable utility framework), while Chade and Eeckhout [2020] assume that the prematching of firms is PAM, one of the main goals of our article is to study the connection between the assortativity of the prematching and the assortativity of stable agent matchings. As such, we allow for arbitrary prematchings, and show that the client prematching always generates a tight upper bound on how positive assortative the matching of agents can be.

In our framework, agent matchings are intermediated by the prematching of battles. In this sense, our paper is related to the models of two-sided matchings where each matched couple is also matched to an object so that all agents have preferences over both agents and objects. Examples of such models include matching with contracts [Hatfield and Milgrom, 2005], matching with projects [Nicolo et al., 2019], and matching with ownership [Combe, 2021]. In a related model of matching through intermediaries, Raghavan [2021] studies a three-sided matching where firms and workers match by mutually accepting the match offers generated by the intermediaries. Our model is different from these as the battles that intermediate the agent matchings consist of prematched clients who also have preferences over matches. The solution concepts that are relevant in our setting, in particular the (pairwise) stability and core, are defined by allowing clients to participate in blocking pairs and coalitions, respectively.

Our paper is organized as follows. Section 2 lays out our model. Section 3 studies the environment in which the prematching is negative assortative. Section 4 generalizes the results for arbitrary prematchings. Finally, Section 5 discusses efficiency and the core in our setting.

2 The Model

There are two equal-sized sets of clients $B = \{b_1, \ldots, b_k, \ldots, b_n\}$ and $D = \{d_1, \ldots, d_l, \ldots, d_n\}$ for some finite number n. Let $N = \{1, \ldots, n\}$ denote the set of indices. Each client in B is prematched with a unique client in D to compete for some prize. $C \subseteq B \times D$ denotes a one-to-one matching between clients, and we call such a client matching a *prematching* as we take it as given for each matching problem. We denote an element of C by $c_{kl} \equiv (b_k, d_l)$ and think of c_{kl} as a *battle* fought between the clients b_k and d_l. $\mathcal{C}$ denotes the set of all possible prematchings. The clients require the help of agents in their battles. Let $A = \{a_1, \ldots, a_i, \ldots, a_n\}$ and $E = \{e_1, \ldots, e_j, \ldots e_n\}$ be the sets of agents that the clients in B and D, respectively, can be matched with. In the Online Appendix we relax this assumption and allow the agents to change sides.[1]

Given $C \in \mathcal{C}$, we are interested in one-to-one matchings between agents and clients and the induced matching between the agents. A matching $\mu \subseteq A \times B \times D \times E$ is a set of quadruples such that $|\mu| = n$, for all $h \in N$, a_h, b_h, d_h, and e_h respectively appear at exactly one quadruple in μ, and finally $(a_i, b_k, d_l, e_j) \in \mu$ implies $c_{kl} \in C$. Let $\mathcal{M}^C$ denote the set of all feasible matchings under C. For a given $\mu \in \mathcal{M}^C$, as the prematching C is fixed, the *match* of agent a_i, $\mu(a_i)$, effectively consists of a b-client and an opponent agent. Therefore, $\mu(a_i) = (b_k, e_j)$ if $(a_i, b_k, d_l, e_j) \in \mu$, and in that case we say $\mu^e(a_i) = e_j$ and $\mu^b(a_i) = b_k$. Similarly, the match of b_k is $\mu(b_k)$ and $\mu(b_k) = (a_i, e_j)$ if $(a_i, b_k, d_l, e_j) \in \mu$. Analogous functions for e-agents and d-clients are defined in the same way. A matching μ can be split into several partial matchings that are two-sided one-to-one matchings. For example, a matching between a-agents and battles, or a-agents and e-agents. Given μ, $(a_i, b_k, d_l, e_j) \in \mu$ only if (a_i, c_{kl}) is an element of the partial matching μ^{ac}; μ^{ec} and μ^{ae} are similarly defined. To refer to an arbitrary matching between the agents we use $\eta \subseteq A \times E$ and call it an *agent matching*. η is a set of pairs such that $|\eta| = n$ and for all $h \in N$, a_h and e_h respectively appear at exactly one pair in η. H denotes the set of all agent matchings.

All agents and clients have strict and rational preferences over the set of all their potential matches. For all $a_i \in A$, $b_k \in B$, $d_l \in D$, and $e_j \in E$, let $\succ^a_i$, $\succ^b_k$, $\succ^d_l$,

[1] Online Appendix can be accessed from https://drive.google.com/file/d/1v93V2j32Vc8r7t4vxjL1iafdT9QxBWXY/view?usp=sharing

and $\succ_j^e$ denote the preference of a_i, b_k, d_l, and e_j, respectively. Preferences of agents and clients over matchings follow solely from their preferences over their matches and they are indifferent between any two matchings if their match is the same under these matchings. For any given $C \in \mathcal{C}$ and for any agent $a_i \in A$, the alternative set of a_i is $B \times E$. Any alternative (b_k, e_j) corresponds to the match (a_i, c_{kl}, e_j) where $c_{kl} \in C$. Similarly, for any $b_k \in B$, $d_l \in D$, and $e_j \in E$, the alternative sets are $A \times E$, $A \times E$, and $A \times D$, respectively. As the prematching is fixed, the preference relation over clients already captures the preferences of agents over the opposing client and the whole battle.

We assume that there is an objective ranking of both agents and clients. All agents prefer to match with a "better" client and a "worse" opponent agent, and similarly all clients prefer to match with a "better" agent and their opponent client to match with a "worse" agent. Assumption 1 below formalizes this restriction.

Assumption 1. *For all $a_i \in A$ and $e_j \in E$,*

$$\begin{aligned}
\forall k, k' \in N, \quad k < k' \quad &\Longrightarrow \quad (b_k, e_j) \succ_i^a (b_{k'}, e_j) \\
\forall l, l' \in N, \quad l < l' \quad &\Longrightarrow \quad (a_i, d_l) \succ_j^e (a_i, d_{l'}) \\
\forall j', k \in N, \quad j > j' \quad &\Longrightarrow \quad (b_k, e_j) \succ_i^a (b_k, e_{j'}) \\
\forall i', l \in N, \quad i > i' \quad &\Longrightarrow \quad (a_i, d_l) \succ_j^e (a_{i'}, d_l),
\end{aligned}$$

and for all $b_k \in B$ and $d_l \in D$

$$\begin{aligned}
\forall i, i', j \in N, \quad i < i' \quad &\Longrightarrow \quad (a_i, e_j) \succ_k^b (a_{i'}, e_j) \\
\forall i, j, j' \in N, \quad j < j' \quad &\Longrightarrow \quad (a_i, e_j) \succ_l^d (a_i, e_{j'}) \\
\forall i, j, j' \in N, \quad j > j' \quad &\Longrightarrow \quad (a_i, e_j) \succ_k^b (a_i, e_{j'}) \\
\forall i, i', j \in N, \quad i > i' \quad &\Longrightarrow \quad (a_i, e_j) \succ_l^d (a_{i'}, e_j)
\end{aligned}$$

As Assumption 1 introduces an objective ordering of both agents and clients, in which lower ranking corresponds to a "better" agent or client. In practical terms Assumption 1 requires that the degree of specialization of agents is limited. For example, politicians can be characterized by some overall popularity level which is not strongly dependent on the state they are running in; lawyers are characterized by some skill which is relatively easily transferable across cases, etc.

Fixing A, B, D, E, and a prematching C, a matching problem is defined solely by a preference profile $\succ = (\succ^a_h, \succ^b_h, \succ^d_h, \succ^e_h)_{h \in N} \in \mathcal{R}$, where $\mathcal{R}$ denotes the set of all preference profiles that satisfy Assumption 1.

Assumption 1 allows for heterogeneous preference profiles where different agents differently prioritize being matched to a better client or a worse opponent agent. Definition 1 below introduces lexicographic preferences which uniformly prioritize either the clients or opponent agents.

Definition 1. *Two types of lexicographic preferences for a-agents and b-clients are defined as follows (they are similarly defined for e-agents and d-clients):*
(i) $\succ^a_i$ *is client-lexicographic if for all* k, k', j, j' *with* $k < k'$, $(b_k, e_j) \succ^a_i (b_{k'}, e_{j'})$.
(ii) $\succ^a_i$ *is opponent-lexicographic if for all* k, k', j, j' *with* $j' < j$, $(b_k, e_j) \succ^a_i (b_{k'}, e_{j'})$.
(iii) $\succ^b_k$ *is agent-lexicographic if for all* i, i', j, j' *with* $i < i'$, $(a_i, e_j) \succ^b_k (a_{i'}, e_{j'})$.
(iv) $\succ^b_k$ *is opponent-lexicographic if for all* i, i', j, j' *with* $j' < j$, $(a_i, e_j) \succ^b_k (a_{i'}, e_{j'})$.

Threshold preferences, as defined below, substantially generalize lexicographic preferences.

Definition 2. *A threshold preference* $\succ^a_i$ *for agent* a_i *is defined solely by an ordered partition* $\{B^1_i, \ldots, B^{m_i}_i\}$ *of* B *in the following way:*

$$\begin{aligned} b_k \in B^r_i, b_{k'} \in B^{r'}_i, \text{ and } r < r' \quad &\Longrightarrow \quad (b_k, e_j) \succ^a_i (b_{k'}, e_{j'}) \text{ for all } j, j' \\ b_k, b_{k'} \in B^r_i \text{ and } j' < j \quad &\Longrightarrow \quad (b_k, e_j) \succ^a_i (b_{k'}, e_{j'}) \end{aligned}$$

$\succ^e_i$ *is defined similarly by an ordered partition* $\{D^1_j, \ldots, D^{m_j}_j\}$ *of* D.

Threshold preferences capture the main source of the heterogeneity of preferences in our model. Although rankings over both clients and opponents are fixed, different agents can care to a different degree about the strength of their opponent and the quality of their client. If all agents have threshold preferences, those whose preferences are characterized by a finer partition are primarily concerned with the quality of their client, whereas those whose preferences are characterized by a coarser partition are primarily concerned with the strength of their opponent. Indeed, a client-lexicographic preference and an opponent-lexicographic preference for an agent are special cases of a threshold preference where each one is defined by a partition of cardinality n and 1, respectively.

2.1 Solution concept

We define our stability notion in the following way. Fix a prematching $C \in \mathcal{C}$ and a matching $\mu \in \mathcal{M}^C$. We say that the agent-client pair (a_i, b_k) (or (e_j, d_l)) is a *blocking pair* at μ or (a_i, b_k) ((e_j, d_l)) blocks μ if

$$\begin{aligned} (b_k, \mu^e(b_k)) \succ_i^a \mu(a_i) \quad &\text{and} \quad (a_i, \mu^e(b_k)) \succ_k^b \mu(b_k) \\ (\text{or } (\mu^a(d_l), d_l) \succ_j^e \mu(e_j) \quad &\text{and} \quad (\mu^a(d_l), e_j) \succ_l^d \mu(d_l)) \end{aligned}$$

Definition 3. *Given a prematching C and a preference profile $\succ \in \mathcal{R}$, a matching $\mu \in \mathcal{M}^C$ is stable if there is no blocking pair at μ.*

Our stability notion is an adaptation of the pairwise stability for two-sided matchings. An agent and a client on one side treat the matching of the opposing side as given. Stability requires that no agent and client simultaneously prefer to be matched with each other to their current matches.

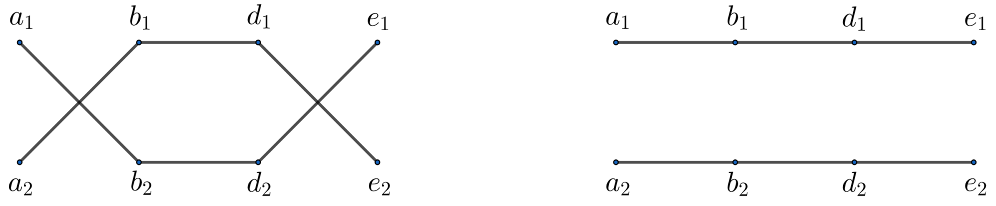

Figure 1: For any preference profile, the matching on the left is not stable; (a_1, b_1) is a blocking pair. The matching on the right is stable if a_1 and e_1 have client-lexicographic preferences. If a_1 has opponent-lexicographic preferences, (a_1, b_2) is a blocking pair.

To understand our stability notion in the context of classical pairwise stability [Gale and Shapley, 1962], fix not only the prematching of clients but also one side of the agent-client matching. We call a matching which does not have a blocking pair on one side a stable response for that side.

Definition 4. *A matching μ is called an ab (ed) stable response if there is no ab (ed) blocking pair.*

In a sense, the relation between our stability notion and a stable response corresponds to the relation between a Nash equilibrium and a best response.

Remark 1. *μ is stable if and only if it is simultaneously an ab and ed stable response.*

We are concerned not only with finding stable matchings given the preferences, but also with making testable predictions for situations in which only the rankings of agents and battles are known. For this purpose we propose a notion of a potentially stable matching.

Definition 5. *Given a prematching C, we say that a matching $\mu \in \mathcal{M}^C$ is potentially stable if there is a preference profile $\succ \in \mathcal{R}$ such that μ is stable at $\succ$.*

Note that the matching in Figure 1 on the left is not potentially stable as (a_1, b_1) is a blocking pair independent of the preference profile. On the other hand, the matching on the right is potentially stable. Understanding possible agent matchings is one of our primary concerns in this work. Therefore, we define the notion of agent matchings supported by a given C as follows.

Definition 6. *Given a prematching C, we say that an agent matching η is supported by C if there is a preference profile $\succ \in \mathcal{R}$ and a matching $\mu \in \mathcal{M}^C$ such that μ is stable at $\succ$ and $\eta = \mu^{ae}$.*

For both matchings in Figure 1, the induced agent matching is $\mu^{ae} = \{(a_1, e_1), (a_2, e_2)\}$. As there is a profile for which one of these matchings is stable, $\eta = \{(a_1, e_1), (a_2, e_2)\}$ is supported by $C = \{c_{11}, c_{22}\}$.

3 The Negative Assortative Prematchings

As we discuss in Section 4, the set of stable matchings depends on the prematching C, which we can think of as the given environment. A natural environment to start with is one in which battles that are ranked highly by one side of the conflict have a low ranking for the other side. Civil litigation can be thought of as an example of such an environment. If there is strong evidence or a case law supporting plaintiff's position, the trial will be easy to win for the plaintiff, but difficult to win for the defendant. In a political race if a left-wing party tends to be popular in a given region, a left-wing candidate will have an easy time winning the race while a right-wing candidate will have a hard time. More generally, we are thinking of a situation in which the primary concern of both agents and clients is winning a battle, and the likelihood that the battle is won can be influenced by the talent of the agent and some innate characteristics of the battle itself. Only one side can win, and the larger the probability of winning is for one side, the lower it is for the other side. In this sense client b_1 fights in the battle with the most favorable characteristics for the ab side, and by extension the least favorable characteristics for the de side. Hence, client b_1 should be prematched with client d_n; similarly client b_2 should be prematched with client d_{n-1}, and so on. This is the negative assortative prematching C^{NAM} which we can formally define as follows. $C = C^{NAM}$ if for all k and l, $c_{kl} \in C$ implies $l = n - k + 1$. The negative assortative agent matching η^{NAM} is similarly defined, i.e., $\eta = \eta^{NAM}$ if for all i and j, $(a_i, e_j) \in \eta^{NAM}$ implies $j = n - i + 1$. Theorem 1 characterizes stable matchings in this environment.

Theorem 1. *If $C = C^{NAM}$, the following hold:*

i) *For any $\succ \in \mathcal{R}$, $\bar{\mu} = \{(a_i, b_i, d_{n-i+1}, e_{n-i+1})_{i \in N}\}$ is always a stable matching.*

ii) *μ is potentially stable if and only if $\eta^{NAM} = \mu^{ae}$.*

iii) *η^{NAM} is the only client matching supported by C^{NAM}.*

All proofs are in Appendix A unless otherwise stated.

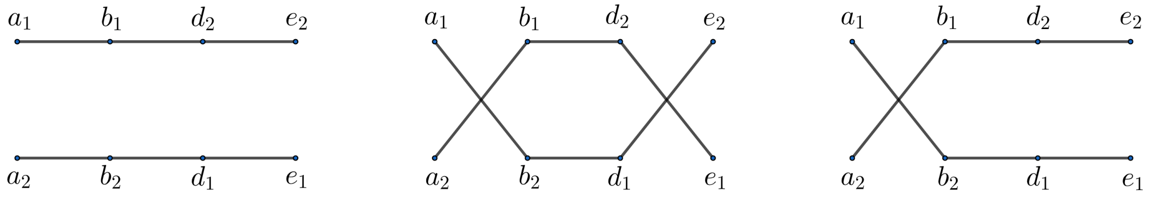

Figure 2: The leftmost matching is stable under any preferences, as both a_1 and e_1 are matched with the best client and the worst opponent agent. The matching in the middle is stable when a_1 and e_1 have lexicographic preferences for opponent agents, but not stable if one of them has client lexicographic preferences. The rightmost matching is never stable as (a_1, b_1) forms a blocking pair.

There are three important implications of Theorem 1. First, if the client matching is C^{NAM}, a stable matching always exists. Second, there is a large set of potentially stable matchings given C^{NAM}, consisting of all those where agents are matched negatively assortatively. Third, η^{NAM} is the only agent matching supported by C^{NAM}. These three observations will guide our analysis in the general setup.

4 All Prematchings

In this section, we allow for any prematching. For example, an extreme case opposite to C^{NAM} is a positive assortative prematching C^{PAM}, where a-agents and e-agents agree on the ranking of the battles. Formally, $C = C^{PAM}$ if for all k and l, $c_{kl} \in C$ implies $k = l$ (η^{PAM} can be similarly defined for the agent matchings). We can think of C^{PAM} representing an environment in which the battles are primarily differentiated by characteristics influencing the costs and benefits of fighting them. For example, legal cases can have different complexities, and a case which requires a lot of costly preparation for the plaintiff may also require a lot of preparation for the defendant. Political races can differ by prestige and importance, and a race important for the left-wing may also be important for the right-wing, and vice versa.

In many environments multiple factors can play a role, and some battles that the a-agents like may be disliked by the e-agents, whereas other battles can be favored by both sides. These environments are not characterized by C^{NAM} nor C^{PAM}, but rather by some C in-between those two extremes. To describe these intermediate scenarios we introduce the notion of Positive Assortative dominance (henceforth, PA-dominance) of battles. We say that a battle c_{kl} PA-dominates $c_{k'l'}$, if c_{kl} is better for both sides than $c_{k'l'}$. For that to happen it needs to be that the client on any given side of the battle c_{kl} is better than the client on the same side of the battle $c_{k'l'}$. Hence, in a sense, the two pairs of agents are positively assortatively prematched relative to each other. Formally, we define the PA-dominance relation as follows.

Definition 7. *Given $C \in \mathcal{C}$, we say that $(b_k, d_l) \in C$ PA-dominates $(b_{k'}, d_{l'}) \in C$ (or simply c_{kl} PA-dominates $c_{k'l'}$) whenever $k < k'$ and $l < l'$. $\tau(C)$ denotes the PA-dominance relation defined over the set C, $t(C) = |\tau(C)|$, and $\Gamma_\tau(C)$ is the directed graph representing $\tau(C)$.*

The notion of PA-dominance allows us to think of prematchings as partial orders. Roughly speaking more extensive partial orders represent more positive assortative matchings. Indeed, the positive assortative prematching is represented by a linear order in which any pair of clients can be compared in terms of PA-dominance, and negative assortative matching is represented by an empty partial order in which no two pairs can be compared. $t(C)$ corresponds to the number of pairs that are related to each other in terms of PA-dominance in the prematching C. We can also interpret it as a measure of distance between C and C^{NAM}. As $\tau(C)$ is a partial order, $\Gamma_\tau(C)$ is a directed acyclic graph. Naturally, the idea of PA-dominance applies not only to prematchings, but to all two-sided matchings with an objective ranking on each side, including an agent matching η.

In this section we generalize the findings of Section 3 to the general setup. In Section 4.1 we discuss the conditions under which a stable matching is guaranteed to exist. Section 4.2 describes the set of potentially stable matchings for a given C and the set of agents matchings supported by a given C.

4.1 Existence

In the standard two-sided matching context, the existence of a stable matching is established by the celebrated Gale-Shapley algorithm. However, existence is not guaranteed for a three-sided matching [Alkan, 1988]; that is, there are preference profiles at which there is no stable matching. As we discuss in Section 5, our stability notion and the notion of core, which requires no deviation from elements (four-tuples) of a matching, are independent. Therefore, literature on matching does not immediately provide arguments or intuition for the existence of stable matchings in our model. In the case of C^{NAM}, the existence of a stable matching is established by Theorem 1. Indeed, the result is stronger: there is a matching which is stable for any preference profile. This matching has the following special feature. The best agents on each side are matched with the best clients and the worst agents on the other side, in other words, they get the option which is universally best for any preference satisfying Assumption 1. Given that, the second best agents have their best options among those available, and so on. Whenever such a matching is feasible, it is stable under any preference satisfying Assumption 1. Therefore, existence follows from the negatively assortative structure of C^{NAM}. Consider the following example with C^{PAM}.

Example 1. *Let* $n = 3$*,* $C = C^{PAM} = \{c_{11}, c_{22}, c_{33}\}$*, and* $\succ$ *be defined as follows:* a_2 *and* e_2 *have client-lexicographic preferences. Moreover,* $\succ_1^a$*:* $\ldots \succ_1^a (b_1, e_2) \succ_1^a (b_3, e_3) \succ_1^a (b_2, e_2) \succ_1^a (b_1, e_1) \succ_1^a \ldots$*, and* $\succ_1^e$*:* $(a_3, d_1) \succ_1^e (a_3, d_2) \succ_1^e (a_2, d_1) \succ_1^e (a_1, d_1) \succ_1^e (a_2, d_2) \succ_1^e (a_1, d_2) \succ_1^e (a_3, d_3) \succ_1^e \ldots$*.*

Here, there is no stable matching. To see that, suppose μ is stable at $\succ$. Note that $\mu^d(e_1) \in \{d_1, d_2\}$ and $\mu^d(e_2) \in \{d_1, d_2\}$. Therefore, $\mu^d(e_3) = d_3$. If $\mu^b(a_1) \in \{b_1, b_2\}$, $\mu^d(e_1) = d_1$. Then, (a_1, b_3) blocks μ. If $\mu^b(a_1) = b_3$, $\mu^b(a_2) = b_1$ and $\mu^d(e_2) = d_1$. Then, (a_1, b_1) blocks μ.

What if C is neither C^{NAM} nor C^{PAM}? Indeed, for $n = 3$, it is possible to study every prematching and show that the existence of a stable matching is guaranteed for all other cases by brute force. However, we will show C^{PAM} is not the only prematching where we may not have stable matchings for some problems when $n > 3$. Indeed, even if a matching does not contain a positive assortative chunk of three consecutive battles, a stable matching may fail to exist. The common factor leading to nonexistence for these prematchings is that they all have three battles which can be linearly ordered in terms of PA-dominance. We call all the prematchings which do not include this structure bipartite.

Definition 8. *A prematching $C \in \mathcal{C}$ is bipartite (with respect to PA-dominance) if there is a partition $\{C^1, C^2\}$ such that $c \in C$ PA-dominates $c' \in C$ only if $c \in C^1$ and $c' \in C^2$.*

Intuitively, bipartiteness requires that no battle simultaneously PA-dominates and is PA-dominated. Note that C^{NAM} is bipartite as no battles are related in terms of PA-dominance at all, and the above condition trivially holds. Indeed, if three battles in C can be linearly ordered in terms of PA-dominance, as in Example 1, it is not bipartite since a dominated battle PA-dominates another battle. If there are no such three battles, C is obviously bipartite. Moreover, bipartite prematchings are represented by bipartite graphs.

It is easy to see that bipartidness cannot be violated if $t(C) < 3$, and for $n = 3$ C^{PAM} is the only prematching which has three such relations. Hence, five out of six prematchings are bipartite. While n grows, the ratio of bipartite matchings to those which are not bipartite decreases. However, it is still not very rare that a prematching is bipartite. Consider the following prematching: $C = \{c_{15}, c_{26}, c_{32}, c_{44}, c_{51}, c_{63}\}$. C is nontrivially far from C^{NAM}. If we use $t(.)$ as a measure, $t(C) = 4$ while $t(C^{NAM}) = 0$, and still C is bipartite. In contrast, $\hat{C} = \{c_{16}, c_{25}, c_{34}, c_{41}, c_{52}, c_{63}\}$ is not bipartite, even though, $t(\hat{C}) = 3$.

We show in Theorem 2 that it is possible to generalize the idea of the argument proving nonexistence for Example 1 if C is not bipartite. Moreover, we demonstrate that if C is bipartite a stable matching exists for any preference profile. The proof is constructive.

We propose an algorithm to search for a stable matching. The algorithm starts from a matching which is PAM of a-agents to b-clients and is an *ed* stable response. Then, we change the matching by eliminating any *ab* blocking pairs. However, the new matching we obtain can have *ed* blocking pairs. In the next step we eliminate them, and repeat the procedure until there are no blocking pairs left on any side. Eliminating blocking pairs on a given side can be achieved by a serial dictatorship.

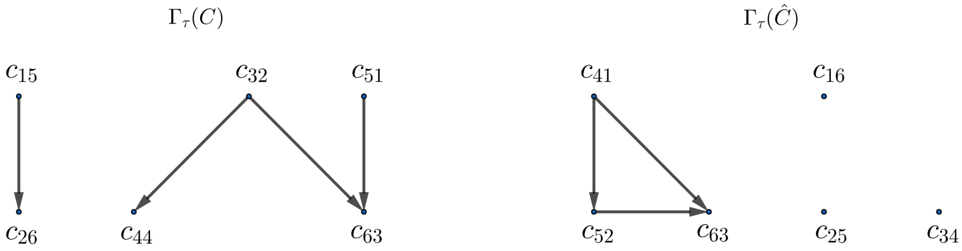

Figure 3: C can be partitioned in $C^1 = \{c_{15}, c_{32}, c_{51}\}$ and $C^2 = \{c_{26}, c_{44}, c_{63}\}$ in line with Definition 8. This is equivalent to the condition that $\Gamma_\tau(C)$ is a bipartite directed graph with all arrows routing from the nodes in the same set and pointing to the nodes in the complement of this set. $\Gamma_\tau(\hat{C})$ cannot be justified as a bipartite graph as an arrow routs from c_{52} while there is another arrow pointing to c_{52}.

Remark 2. *μ is an ab stable response if and only if it can be generated by the following procedure. Fix the matching of e-agents to battles μ^{ec}, and allow a_1 to choose her best alternative in μ^{ec}, a_2 to choose the best option among the remaining alternatives in μ^{ec}, and so on.*

Given Remark 2, we can implement an algorithm searching for a stable matching by sequentially generating *ab* and *ed* stable responses.

Algorithm 1 (The stable response algorithm)**.** *Fix C. At step 0, start with the matching μ_0 such that μ_0^{ac} is PAM ($\mu^b(a_i) = b_i$ for all $i \in N$) and μ_0 is an ed stable response. For all the odd steps $s \in \{1, 3, 5...\}$ of the algorithm, $\mu_s^{ec} = \mu_{s-1}^{ec}$ and μ_s is an ab stable response. For all the even steps $s \in \{2, 4, 6...\}$ of the algorithm, $\mu_s^{ac} = \mu_{s-1}^{ac}$ and μ_s is an ed stable response. The algorithm ends if $\mu_s = \mu_{s-1}$.*

Remark 3. *If the stable response algorithm ends at some finite step s, then the resulting matching μ_s is stable.*

Although Algorithm 1 in general may not terminate even if a stable matching exists. As long as the prematching is bipartite it is guaranteed to stop. To build an intuition for why it is the case we consider the following example.

Example 2. *Let $n = 4$, $C = \{c_{12}, c_{21}, c_{34}, c_{43}\}$, and $\succ$ be defined as follows: a_1 and e_1 have client-lexicographic preferences, a_2 and e_2 have opponent-lexicographic preferences. Moreover, $\succ_3^a$: $\ldots \succ_3^a (b_4, e_3) \succ_3^a (b_2, e_1) \succ_3^a (b_3, e_2) \succ_3^a \ldots$, and $\succ_3^e$: $\ldots \succ_3^e (a_3, d_4) \succ_3^e (a_1, d_2) \succ_3^e (a_2, d_3) \succ_3^e \ldots$*

In Example 2 the algorithm generates the following sequence of matchings: $\mu_0 = \{(a_1, c_{12}, e_4),\ (a_2, c_{21}, e_1),\ (a_3, c_{34}, e_3),\ (a_4, c_{43}, e_2)\}$, $\mu_1 = \{(a_1, c_{12}, e_4),\ (a_3, c_{21}, e_1),\ (a_2, c_{34}, e_3),\ (a_4, c_{43}, e_2)\}$, $\mu_2 = \{(a_1, c_{12}, e_3),\ (a_3, c_{21}, e_1),\ (a_2, c_{34}, e_4),\ (a_4, c_{43}, e_2)\}$. Then, $\mu_3 = \mu_2$ and the algorithm terminates. A closer look at this sequence reveals that at each step b_3 becomes better-off. At step $s = 1$ the position of b_3 improves because he is now matched with agent a_2 rather than a_3. As a result, d_4 (who is prematched with b_3) becomes less attractive for an e-agent at step 2 compared to step 0. At step $s = 2$ agent e_3 leaves d_4 and is replaced by agent e_4. The position of b_3 improves again.

This is no coincidence. In general, each step of the algorithm follows a similar pattern. Strong a-agents matched with b-clients in C^1 tend to leave to b-clients in C^2. As a response, strong e-agents matched with d-clients in C^2 leave to be matched with d-clients in C^1. This feeds back to the remaining strong a-agents matched with b-clients in C^1, etc. As a result, at every step of the algorithm at least some b-clients in C^2 become better-off while none become worse-off. Once this observation is made, it is enough to note that there is a limit to how well-off a client can be and the algorithm needs to cease. We summarize our results in Theorem 2.

Theorem 2. *Let $C \in \mathcal{C}$. A stable matching $\mu \in \mathcal{M}^C$ exists for all $\succ \in \mathcal{R}$ if and only if C is bipartite.*

Theorem 2 characterizes all the environments in which a stable matching is guaranteed to exist for all the preference profiles. In the remainder of this section we focus on the complementary question, and we propose two preference domains for which a stable matching always exists.

First, we consider threshold preferences. It is enough that agents on one side have threshold preferences, and Algorithm 1 (or its analogue starting from μ_0 being ab stable) is guaranteed to find a stable matching in a single step. Moreover, as it will become clearer in Section 4.2, any potentially stable matching can be rationalized using only threshold preferences.

Proposition 1. *Let C be any prematching. If $\succ \in \mathcal{R}$ is such that every e-agent has threshold preferences, then Algorithm 1 finds a stable matching at $\succ$.*[2]

[2] A mirror image of Algorithm 1, in which μ_0 is PAM of e-agents to d-clients and an ab stable response is guaranteed to find a stable matching if every a-agent has a threshold preference.

Second, we consider agent preferences with clients of binary type. In applied theory it is often assumed that market participants can be split into two groups or types: “high” and “low”, “good” and “bad”, etc. Members of each group are almost identical. For example, political races can be split into important and insignificant, defendants can be split into likely guilty and likely innocent (see Chade and Eeckhout 2020 for an application for technology development in a matching environment). In our setting, a situation in which clients have binary types can be defined as follows.

Definition 9. *A preference profile $\succ \in \mathcal{R}$ is induced by b-clients of binary type if there exists a partition of B into B^H and B^L such that for all $i, j, j', k, k' \in N$ if $(b_k, e_j) \succ_i^a (b_{k'}, e_{j'})$ and $b_k \in B^H$ ($b_k \in B^L$), then for all $b_{k''} \in B^H$ ($b_{k''} \in B^L$) $(b_{k''}, e_j) \succ_i^a (b_{k'}, e_{j'})$.*

An analogous definition can be used for d-clients. Indeed, Definition 9 says that the clients can be split into two sets, and as far as the agents are concerned, the characteristics of each client are fully summarized by the set they belong to. The ranking of the clients within a set is just an arbitrary rule of resolving ties. Binary type restriction is not related to threshold preferences. On one hand, it is more restrictive because it requires that the clients are partitioned into two groups and the partition is the same for every agent. On the other hand, it allows the agents to prefer being matched with a “low” client and a weak opponent to a “high” client and a strong opponent. Indeed, threshold preference domain is a Cartesian product preference domain while profiles induced by clients of binary type is not. Example 3 shows the distinction between the two cases.

Example 3. *Consider the following preference profile of a-agents.*
$\succ_1^a$: $(b_1, e_3) \succ_1^a (b_2, e_3) \succ_1^a (b_1, e_2) \succ_1^a (b_2, e_2) \succ_1^a (b_3, e_3) \succ_1^a (b_3, e_2) \succ_1^a (b_1, e_1) \succ_1^a (b_2, e_1) \succ_1^a (b_3, e_1)$.
$\succ_2^a$: $(b_1, e_3) \succ_2^a (b_2, e_3) \succ_2^a (b_3, e_3) \succ_2^a (b_1, e_2) \succ_2^a (b_2, e_2) \succ_2^a (b_1, e_1) \succ_2^a (b_2, e_1) \succ_2^a (b_3, e_2) \succ_2^a (b_3, e_1)$.
$\succ_3^a$ *is opponent lexicographic.*

The preference profile in Example 3 is induced by b-clients of binary type. To be precise $B^H = \{b_1, b_2\}$, $B^L = \{b_3\}$. Still, the preferences of all agents are different and only agent a_3 has threshold preferences.

Even though the assumption that types of clients are binary is a restriction on the preferences of agents there is a tight connection between it and bipartite prematchings.

As the “high” and “low” clients are almost identical, the exact way in which they are ranked is irrelevant for the stability of the matching. Hence, if some matching is stable for a bipartite prematching, it will remain stable for any prematching, as long as the set of “high” and “low” clients remains the same.

Proposition 2. *If a preference profile is induced by b-clients or d-clients who have binary types, then a stable matching exists.*

Note that for both Proposition 1 and Proposition 2 it is enough that agents on one side have restricted preferences. It allows for some flexibility of the model. Consider a situation in which the plaintiffs are split into two groups: “good” plaintiffs hold cases which are simple and easy to win, “bad” plaintiffs hold cases which are complex and hard to win. Such a division may not be achievable on the defendant side, for example, as the cases which are easy to win for the defendant are also complex and time-consuming. Still, Proposition 2 ensures that a stable matching exists.

4.2 Potentially stable matchings and supported agent matchings

In Theorem 1 we found an exact relation between the prematching C^{NAM} and the supported agent matching. In this section, we provide a characterization which shows that the set of supported agent matchings expands in a nontrivial fashion as the prematching becomes more positive assortative. Before we move to this result, we fully characterize the set of potentially stable matchings.

Theorem 3. *Fix $C \in \mathcal{C}$. $\mu \in \mathcal{M}^C$ is a potentially stable matching if and only if for all (a_i, b_k, d_l, e_j), $(a_{i'}, b_{k'}, d_{l'}, e_{j'}) \in \mu$, if (a_i, e_j) PA-dominates $(a_{i'}, e_{j'})$, then (b_k, d_l) PA-dominates $(b_{k'}, d_{l'})$.*

A consequence of Theorem 3 is that the PA-dominance relation between pairs of agents must be preserved by their clients in a matching for stability. The intuition behind this result is that if two strong agents are matched with each other in some battle that both of them consider unattractive, at least one strong agent would immediately block the matching with some client that holds a more attractive battle and has a weaker opponent.

Indeed, without precise information about the preferences of the agents, the extent to which our model can make a testable prediction is exactly described by preserving the PA-dominance from agent matchings to client matchings. We use threshold preferences to show that any matching preserving the PA-dominance relation is indeed stable for some preference profile. We characterize the preference of each agent by a bipartition of clients, the first element of the partition being the set of all clients that are at least as good as the one that the agent is matched with. This gives us the following remark.

Remark 4. *If matching μ is potentially stable, then it is stable at some preference profile $\succ \in \mathcal{R}$ at which all agents have threshold preferences.*

Theorem 3 builds intuition on which agent matchings can be supported by a given prematching. It suggests that the agent matching can be, in some way, only as positive assortative as the prematching. Otherwise, the PA-dominance relation could not be preserved from agents to clients. Indeed, that is the reason for which η^{NAM} is the only supported agent matching by C^{NAM}. However, the agent matching can always be less positive assortative than the prematching, as the PA-dominance relation needs to be preserved only from agent matchings to prematchings but not the other way around. Indeed, any agent matching which is negative assortative is potentially stable independent of the choice of C. To see how an agent matching cannot be supported if it is "too positively assortative" consider the following example.

Example 4. *$n = 3$, $C = \{(b_1, d_1), (b_2, d_3), (b_3, d_2)\}$, $\eta = \{(a_1, e_1), (a_2, e_2), (a_3, e_3)\}$.*

Note that $t(\eta) = 3$ while $t(C) = 2$. Therefore, it is immediately obvious that for any matching $\mu \in \mathcal{M}^C$ such that $\mu^{ae} = \eta$, PA-dominance relation over the agent matching cannot be preserved by the C. Then, η is not supported by C according to Theorem 2, or equivalently, regardless of the preferences of the agents and the clients, agents cannot match as in η given the prematching C under any stable matching. The cardinality of $\tau(\eta)$ should not be more than that of $\tau(C)$. This is a necessary condition for η to be supported by C. This fact will be a corollary of our next result. With the help of the following example we show that it is not a sufficient condition for η to be supported by C.

Example 5. *$n = 4$, $C = \{(b_1, d_3), (b_2, d_2), (b_3, d_1), (b_4, d_4)\}$, $\eta = \{(a_1, e_1), (a_2, e_4), (a_3, e_3), (a_4, e_2)\}$.*

Note that $t(C) = t(\eta) = 3$. However, to verify that η is not supported by C, it is enough to consider the graphs $\Gamma_\tau(C)$ and $\Gamma_\tau(\eta)$ which are depicted in Figure 4. η

is supported by C only if there is $\mu \in \mathcal{M}^C$ which can be thought of as a one-to-one matching between C and η that satisfies the condition in Theorem 3. Therefore, the pair (a_1, e_1) needs to be matched with one of the nodes in $\Gamma_\tau(C)$. As (a_1, e_1) PA-dominates three *ae* pairs that are part of such a matching μ, the *bd* pair with which (a_1, e_1) is matched under μ also needs to PA-dominate at least three *bd* pairs. This is obviously impossible given $\Gamma_\tau(C)$.

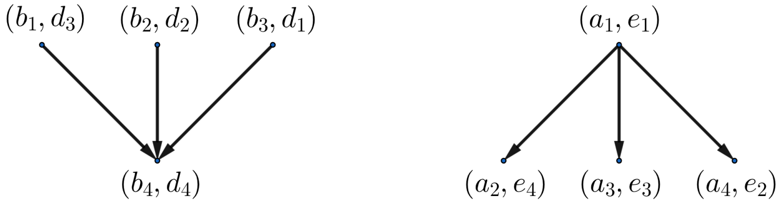

Figure 4: The graph on the left corresponds to $\Gamma_\tau(C)$ for C in Example 5 and the graph on the right corresponds to $\Gamma_\tau(\eta)$ for C in Example 5.

The exercise above gives us a clear hint about the relevant structure we need in order to characterize the supported agent matchings by a given C. Disregarding the labels in the graphs of η and C, the graph of $\tau(\eta)$ should be a subgraph $\tau(C)$. The following notion of subgraph isomorphism summarizes this condition.

Definition 10. *For any $C \in \mathcal{C}$ and $\eta \in H$, $\Gamma_\tau(C)$ is subgraph isomorphic to $\Gamma_\tau(\eta)$ if $\Gamma_\tau(\eta)$ can be obtained from $\Gamma_\tau(C)$ by using only two types of changes: relabeling vertices and deleting edges.*

In our setting $\Gamma_\tau(C)$ ($\Gamma_\tau(\eta)$) represents the partial order induced by the PA-dominance relation on a prematching (an agent matching). If $\Gamma_\tau(C)$ is subgraph isomorphic to $\Gamma_\tau(\eta)$, then $\tau(C)$ is isomorphic to some extension of $\tau(\eta)$. In that sense if $\Gamma_\tau(C)$ is subgraph isomorphic to $\Gamma_\tau(\eta)$, we can say that C is more positive assortative than η. Indeed, $\Gamma_\tau(C^{PAM)}$ is a complete graph and is subgraph isomorphic to $\Gamma_\tau(\eta)$ for any η. On the contrary $\Gamma_\tau(C^{NAM})$ is an empty graph and is subgraph isomorphic only to η^{NAM}.

Theorem 4. *A matching of agents η can be supported by a prematching of clients C if and only if $\Gamma_\tau(C)$ is subgraph isomorphic to $\Gamma_\tau(\eta)$.*

The main message of Theorem 4 is that agent matchings close to η^{NAM} should be more common than agent matchings close to η^{PAM}. Those closer to η^{NAM} are supported by almost any prematching while those that are closer to η^{PAM} are only supported by very positive assortative prematchings. In particular, η^{NAM} is supported by any C, and η^{PAM} only by C^{PAM}.

Additionally, Theorem 4 identifies settings in which our model gives sharp testable predictions even without knowledge of individual preferences. If $t(C)$ is small, the majority of agent matchings can be eliminated. For example, no agent matching with $t(\eta) > t(C)$ is supported by C. More generally, if C is subgraph isomorphic to C' then the set of agent matchings supported by C' is a subset of agent matchings supported by C. Hence, as we move toward C^{NAM} the predictions of our model become sharper.

Corollary 1. *For any C, C', and η, if $t(C) < t(\eta)$, η is not supported by C. Moreover, if C is subgraph isomorphic to C', any agent matching supported by C' is also supported by C.*

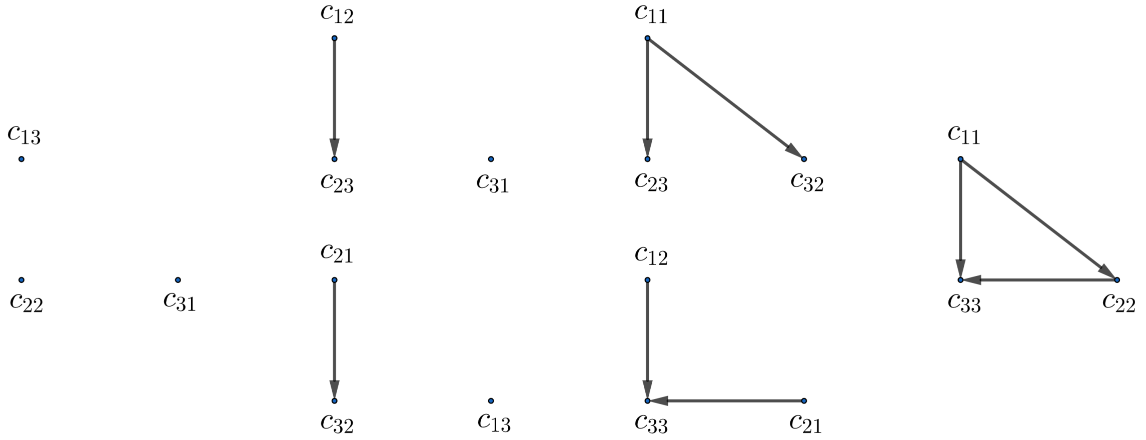

Figure 5: Graphs of all six prematchings when $n = 3$. Each graph is subgraph isomorphic to any graph in the column to the left of it. Graphs of prematchings $\{c_{12}, c_{23}, c_{31}\}$ and $\{c_{13}, c_{21}, c_{32}\}$ are simultaneusly subgraph isomorphic to each other (as they differ only in the labels). Graphs of prematchings $\{c_{11}, c_{23}, c_{32}\}$ and $\{c_{12}, c_{21}, c_{33}\}$ are not comparable.

Finally, Theorem 4 provides the structure determining which agent matchings can be supported by which prematchings: partial order induced by PA-dominance relation. It allows us to identify structurally identical prematchings which will generate exactly the same testable predictions in terms of supported agent matchings. Consider Example 6 to see two distinct prematchings with the same underlying structure.

Example 6. $C = \{c_{15}, c_{23}, c_{34}, c_{41}, c_{52}\}$, $C' = \{c_{14}, c_{25}, c_{33}, c_{41}, c_{52}\}$.

5 Efficiency and Core

The standard core stability notion coincides with pairwise stability for two-sided one-to-one matchings without externalities. An immediate consequence of this is that pairwise stable matchings are also efficient. In our model, the core boils down to requiring no deviations by quadruples for the same reason the core is equivalent to pairwise stability in two-sided one-to-one matching environments. Our stability notion, however, is defined relative to deviations by agent-client couples from the same side but not by coalitions across the sides. Therefore, there is no prior reason to expect that stable matchings in our environment will coincide with the ones in the core or the efficient ones. In this section, we show that every stable matching is efficient but core stability and our pairwise stability notion are logically independent. Furthermore, we provide a domain of preferences within those satisfying Assumption 1, which we call "non-spiteful" client preferences, on which our stability notion is a refinement of the core.

We start our analysis with the welfare analysis based on Pareto efficiency.

Definition 11. *Given any $C \in \mathcal{C}$ and $\succ \in \mathcal{R}$, a matching μ is efficient if no matching μ' Pareto dominates μ, i.e., if for any other matching μ' such that an agent or a client in $A \cup B \cup D \cup E$ prefers μ' to μ, there is an agent or a client in $A \cup B \cup D \cup E$ who prefers μ to μ' and.*

Below, we provide examples of inefficient matchings for two entirely opposite pre-matchings.

Example 7. *Let $|N| = 3$ and $C = C^{NAM}$ and consider the following two matchings:*

$\mu = \{(a_2, b_1, d_3, e_1), (a_3, b_2, d_2, e_3), (a_1, b_3, d_1, e_2)\}$,

$\mu' = \{(a_3, b_1, d_3, e_2), (a_1, b_2, d_2, e_1), (a_2, b_3, d_1, e_3)\}$.

Suppose that the preferences for the agents and the clients are as follows. a_1, a_3, e_1, and e_3 have client-lexicographic preferences. a_2 and e_2 have opponent-lexicographic preferences. b_1, b_3, d_1, and d_3 have opponent-lexicographic preferences. b_2 and d_2 have agent-lexicographic preferences. Then, μ' Pareto dominates μ.

Example 8. *Let $|N| = 4$ and $C = C^{PAM}$ and consider the following two matchings:*

$\mu = \{(a_2, b_1, d_1, e_2), (a_1, b_2, d_2, e_3), (a_3, b_3, d_3, e_1), (a_4, b_4, d_4, e_4)\}$,

$\mu' = \{(a_1, b_1, d_1, e_1), (a_2, b_2, d_2, e_4), (a_4, b_3, d_3, e_2), (a_3, b_4, d_4, e_3)\}$.

Suppose that the preferences for the agents and the clients are as follows. a_1, a_4, e_1, and e_4 have client-lexicographic preferences. a_2, e_2, a_3, and e_3 have opponent-lexicographic preferences. b_2, b_3, d_2, and d_3 have opponent-lexicographic preferences. b_1, b_4, d_1, and d_4 have agent-lexicographic preferences. Then, μ' Pareto dominates μ.

It is easy to see that the matchings μ in Example 7 and Example 8 are not stable for the preferences at which they are Pareto dominated. In both Example 7 and Example 8, the pair (a_1, b_1) blocks μ.

Proposition 3. *Let $C \in \mathcal{C}$ and $\succ \in \mathcal{R}$. If $\mu \in \mathcal{M}^C$ is stable at $\succ$, then μ is efficient at $\succ$.*

A matching in the core is stable against all deviations by any coalition K such that $K \in A \cup B \cup D \cup E$. Note that whenever there is a coalition that blocks a matching, there should also be a quadruple of the form (a_i, b_k, d_l, e_j) that blocks the matching. Therefore, it is sufficient to define the core according to deviations by such quadruples.

In the literature on multi-sided matchings, the stability notion is often defined in a consistent way with the core [See Alkan, 1988, for example] provided here. Our motivation for adopting the stability notion, which does not require deviations across the sides, is that for most applications it is easier to imagine situations where agents and clients on the same side can make agreements while the communication between competing agents is much more limited. Still, stable matchings which belong to the core can be thought of as especially robust. They survive even the possibility of reaching an "across the aisle agreement" of two opposing clients. As we show, under some restrictions on the preferences of the clients, any stable matching is in the core.

Definition 12. *For any $\succ \in \mathcal{R}$ and $C \in \mathcal{C}$, a quadruple $(a_i, b_k, d_l, e_j) \in A \times B \times D \times E$ blocks $\mu \in \mathcal{M}^C$ if all agents and clients in the quadruple prefer to match with each other over their current matches in μ. A subset of matchings $CO(\succ) \subseteq \mathcal{M}^{\mathcal{C}}$ is called the core if for any matching $\mu \in CO(\succ)$ there does not exist a blocking quadruple.*

A matching is in the core, if there does not exist a quadruple (a_i, b_k, d_l, e_j) that prefers to coordinate and match each other instead of their current matches. Such a coordination requires an alignment of preferences of competing clients b_k and d_l as well as the agents a_i and e_j. However, it is easy to come up with a preference profile at which every matching is in the core. Consider the $\succ$ such that all b-clients have agent-lexicographic preferences and all d-clients have opponent-lexicographic preferences. Then, b-clients and d-clients have completely opposite preferences over the set of their potential matches $A \times E$. Therefore, any matching is in $CO(\succ)$ independent of the prematching.

We show below that unless clients are willing to de-escalate the intensity of their battle by simultaneously decreasing the quality of their agents, stability is a refinement of the core.

Proposition 4. *Fix $C \in \mathcal{C}$ and $\succ \in \mathcal{R}$, and let $\mu \in \mathcal{M}^C$ be stable at $\succ$. A quadruple (a_i, b_k, d_l, e_j) blocks μ with $\mu^a(b_k) = a_{i'}$ and $\mu^e(d_l) = e_{j'}$ only if $i' < i$ and $j' < j$.*

In many applications, it is easy to imagine situations where clients are more concerned with their own agents rather than their opponents'. Then they do not prefer to have a worse agent even if that means the opponent client is getting a worse agent as well. In such cases, whenever there is a quadruple that blocks a matching, it could only be because both agents are better than those in their current matchings. However, this is not possible under Proposition 4. Therefore, any stable matching would belong to the core. "Non-spiteful" preferences defined below formalize this intuition.

Definition 13. *Fix any prematching $C \in \mathcal{C}$. A preference profile $\succ$ is called non-spiteful between clients if for any $(b_k, d_l) \in c$, $(a_i, e_j) \succ^b_k (a_{i'}, e_{j'})$ and $(a_i, e_j) \succ^d_l (a_{i'}, e_{j'})$ implies $i < i'$ or $j < j'$.*

As a direct corollary of Proposition 4, stability is a refinement of the core on the domain of non-spiteful preferences between clients.

Corollary 2. *For any $C \in \mathcal{C}$ and $\succ \in \mathcal{R}$ which is non-spiteful between clients, every stable matching belongs to the core.*

Even if a preference profile is not non-spiteful between clients, the set of stable matchings is occasionally a subset of the core as a blocking quadruple also requires the coordination of agents and hence the preferences of agents matter. The following example illustrates a situation where this is not the case.

Example 9. $n = 3$, $C = C^{NAM}$. *All the agents have opponent-lexicographic preferences. $b_1, b_3, d_1,$ and d_3 have agent-lexicographic preferences. b_2 and d_2 are opponent lexicographic.*

Note that every matching μ such that $\mu^{ae} = \eta^{NAM}$ is stable. Hence, there are six stable matchings. However, the only potential blocking quadruple (a_3, b_2, d_2, e_3) at a stable μ, blocks μ if and only if $(a_2, b_2, d_2, e_2) \notin \mu$ as otherwise either b_2 or d_2 is matched with the best agent on her side. Therefore, only four out of six stable matchings are in the core. Then, as we have situations where every matching is in the core, stability and the core are logically independent.

Remark 5. *For any $n > 2$, the solution concepts core and stability are logically independent.*

In our framework, stability and the core are logically independent because the type of coalitions that can block a matching under these two concepts are different. For stability, an agent-client pair on the same side may block a matching; while for the core, a coalition of two agent-client pairs, one from each side can block. For each side of the agent-client matching, the matchings on the other side create externalities. Our stability notion does not internalize these externalities while the core does. A similar conclusion, that the core and stability are different, is proved by Sasaki and Toda [1996] in an environment of two-sided matchings with externalities.

6 Conclusion

We study a problem of matching agents to fighting clients. To do so, we develop a notion of stability applicable in this setting and characterize conditions under which stable matchings exist. Additionally, we characterize the matchings which can be stable for some preference profile. The central testable prediction of the model is that negative assortative agent matchings are more common than positive assortative agent matchings. Our analysis leaves a large space for extensions.

In the Online Appendix we relax the assumption that the agents form two distinct sets, and allow agents to switch sides. Several results, including Theorem 1 and Proposition 1, survive the extension. However, if the agents can switch sides of the battle, there is no unique way of expressing the prematching as a two-sided matching. Hence, all the results related to the notion of PA-dominance, including theorems 2–4, become more nuanced. As there are more potential blocking pairs, existence is harder to sustain. Although there is no longer a simple map from a prematching to a set of supportable matchings, the central message of Theorem 4 stands and negative assortative matchings of agents are easier to support than positive assortative matchings.

Although we characterize the stable matchings, we do not propose a theory on how they can be reached. Whether stable matchings can be achieved in a decentralized market or through some centralized allocation mechanism is left for further research. Moreover, as a benchmark model, we disregard the possibility of transferable utility and contracts while the matching is one-to-one and the structure of fought battles is exogenous. In several applications, some or all of these assumptions are violated. A particular example is the formation of research teams that compete in patent races.

We propose two applications of our setting: the allocation of candidates to political races and the allocation of lawyers to cases. Although our model provides clear insights about potential outcomes, a richer environment should be considered in order to have sharper predictions.

First, in political races incumbent advantage is known to play a significant role and modeling it requires at least partially relaxing the objective ranking assumption. Moreover, the interests of individual politicians and the interests of political parties may not be the same. In fact, our model suggests that they can be in conflict: strong political candidates may sometimes prefer to run in safer races. Whereas political parties may prefer to allocate candidates to tight races. How far the allocation from the stable allocation is will depend on the leverage that competing political parties have over their members.

Second, in the problem of the allocation of lawyers to legal cases, it is not necessarily true that stable allocations are desirable from the perspective of society in general. On one hand, stable allocations are guaranteed to be efficient and should be considered fair by the participants. On the other hand, the social welfare should also take into account how the allocation of lawyers influences the likelihood of making a correct judgement. However, this exercise requires an additional theory on how the talent of opposing lawyers influences the decisions of judges and juries.

A Proofs

In the proofs we write, with a slight abuse of the notation, that $d_l = C(b_k)$ if $c_{kl} \in C$, and for any $B' \subseteq B$, $C(B') = \{d \in D : d = C(b) \text{ for some } b \in B'\}$ ($C(D')$ is similarly defined). Moreover, we adopt the notation that $a_i < a_{i'}$ whenever $i < i'$ (and similarly for b-clients, d-clients, and e-agents).

Proof of Theorem 1

i) $\bar{\mu} \in \mathcal{M}^C$. Note that (a_i, b_k) is a blocking pair only if $a_i < \bar{\mu}^a(b_k)$ and either $b_k < \bar{\mu}^b(a_i)$ or $\bar{\mu}^e(a_i) < \bar{\mu}^e(b_k)$. Let $a_i < \bar{\mu}^a(b_k) = a_k$. As $i < k$, $b_i = \bar{\mu}^b(a_i) < b_k$ and $\bar{\mu}^e(b_k) = e_{n-k+1} < e_{n-i+1} = \bar{\mu}^e(a_i)$. Analogous reasoning holds for any e_j and d_l.

ii) *Sufficiency.* Let $\succ \in \mathcal{R}$ be such that all the agents in $A \cup E$ have opponent lexicographic preferences. It is straightforward to check μ such that $\eta^{NAM} = \mu^{ae}$ is stable at $\succ$.

Necessity. Suppose for a contradiction that μ is stable at $\succ$ and agents are not negatively assortatively matched in μ. Define $i_0 = \min\{i : \mu^e(a_i) \neq e_{n-i+1}\}$. Note that $\mu^e(a_{i_0}) < e_{n-i_0+1}$ since for all $j > n - i_0 + 1$ there is some $i' < i_0$ such that $\mu^e(a_{i'}) = e_j$ by the definition of i_0. Similarly, $a_{i_0} < \mu^a(e_{n-j+1})$. As μ is stable, a_{i_0} does not form a blocking pair with $\mu^b(e_{n-i_0+1})$. Hence, $\mu^b(a_{i_0}) < \mu^b(e_{n-i_0+1})$, and therefore, we have $\mu^d(e_{n-i_0+1}) < \mu^d(a_{i_0})$ as $C = C^{NAM}$. Then $(\mu^e(a_{i_0}), \mu^d(e_{n-i_0+1}))$ is a blocking pair. Contradiction.

iii) immediately follows from ii). ■

Proof of Theorem 2

(Necessity). Let $C \in \mathcal{C}$ be not bipartite, i.e., we have k,k',k'',l,l',l'' with $k < k' < k''$ and $l < l' < l''$ such that $c_{kl}, c_{k'l'}, c_{k''l''} \in C$. It suffices to construct $\succ \in \mathcal{R}$ such that no matching is stable at $\succ$. Let $k_0 = k$ and $l_0 = l$, and among all battles in $C \setminus \{c_{k_0 l_0}\}$, choose the battle with the best b-client that is PA-dominated by (b_{k_0}, d_{l_0}) and call this pair (b_{k_1}, d_{l_1}). Note that such a battle exists as $(b_{k'}, d_{l'})$ is a potential option. Choose the battle in C with the best b-client that is PA-dominated by (b_{k_1}, d_{l_1}) and call this pair (b_{k_2}, d_{l_2}). Note that $k_0 < k_1 < k_2$ and $l_0 < l_1 < l_2$ and by construction the following holds.

(I) For all k with $k_0 < k < k_1$, $c_{kl} \in C$ implies $l < l_0$.

(II) For all k with $k_1 < k < k_2$, $c_{kl} \in C$ implies $l < l_1$.

Define $C_1 = \{c_{1.}, ..., c_{k_0-1.}\}$, $C_2 = \{c_{k_0.}, ..., c_{k_2.}\}$, $C_3 = \{c_{k_2+1.}, ..., c_{n.}\}$, $A_1 = \{a_1, ..., a_{k_0-1}\}$, $A_2 = \{a_{k_0}, ..., a_{k_2}\}$, $A_3 = \{a_{k_2+1}, ..., a_n\}$, $B_2 = \{b_{k_0}, ..., b_{k_2}\}$, and $D_2 = C(B_2)$, where $c_{k.}$ represents the battle in which b_k is prematched with some d_l. Let $r = n - k_0 + 1$ and define $E_1 = \{e_1, ..., e_{n-k_2}\}$, $E_2 = \{e_{n-k_2+1}, ..., e_r\}$, and $E_3 = \{e_{r+1}, ..., e_n\}$. Let $\hat{l}$ be the smallest index such that $d_{\hat{l}} \in D_2$ and $l_0 < \hat{l} < l_2$, and let $\hat{k}$ be such that $c_{\hat{k}\hat{l}} \in C$. $\hat{l}$ exists as $l_1 \in D_2$. $D' = \{d_l \in D_2 : l < l_0\}$, and $m = n - k_2 + |D'| + 1$. Note that D' is possibly empty and $m < r$. Define $E' = \{e_{n-k_2+1}, ..., e_{m-1}\}$ and note that $|E'| = |D'|$. Consider the preference $\succ \in \mathcal{R}$ satisfying the following:

(i) All a-agents except a_{k_0} are client-lexicographic.

(ii) For all j, j' and for all k, k' with $k \leq k_2 < k'$, $(b_k, e_j) \succ^a_{k_0} (b_{k'}, e_{j'})$.

(iii) For all $j < r$ and for all $b, b' \in B_2 \setminus \{b_{k_0}\}$ with $b \neq b'$, $(b, e_r) \succ^a_{k_0} (b', e_j)$.

(iv) $(b_{k_0}, e_{m+1}) \succ^a_{k_0} (b_{k_2}, e_r) \succ^a_{k_0} (b_{k_0}, e_m)$

(v) For all j, l, and l', $(a_i, d_l) \succ^e_j (a_{i'}, d_{l'})$ if $i' \leq k_2 < i$ or $i' < k_0 \leq i$.

(vi) For all j, all $a, a' \in A_2$ with $a \neq a'$ and for all $l < l_2$, $(a, d_l) \succ^e_j (a', d_{l_2})$.

(vii) For all $e_j \in E'$, all $a, a' \in A_2$ with $a \neq a'$, and for all $l < l'$, $(a, d_l) \succ^e_j (a', d_{l'})$.

(viii) For all $a, a' \in A_2$ with $a \neq a'$, and for all $l > \hat{l}$, $(a, d_{\hat{l}}) \succ^e_m (a', d_l)$

(ix) $(a_{\hat{k}+1}, d_{\hat{l}}) \succ^e_m (a_{k_0+1}, d_{l_0}) \succ^e_m (a_{k_0}, d_{l_0}) \succ^e_m (a_{\hat{k}}, d_{\hat{l}})$

Now suppose for a contradiction that μ is stable at $\succ$. Using (i) and (ii) we have $\mu^c(A_1) = C_1$ and $\mu^c(A_3) = C_3$. Therefore, $\mu^c(A_2) = C_2$ and $\mu^b(A_2) = B_2$. Then using (v), we have $\mu^c(E_1) = C_3$ and $\mu^c(E_3) = C_1$. Therefore, $\mu^c(E_2) = C_2$ and $\mu^d(E_2) = D_2$.

Since $|E'| = |D'|$, we have $\mu^d(E') = D'$ by (vii). Then, using definition of D' and (viii), we have $\mu^d(e_m) \in \{d_{l_0}, d_{\hat{l}}\}$, otherwise e_m would block μ together with one of the two clients. I and II implies that for all $d_l \in D_2 \setminus d_{l_2}$, $l < l_2$. Then using (vi), $\mu^d(e_r) = d_{l_2}$ ($\mu^b(e_r) = b_{k_2}$).

If $\mu^b(a_{k_0}) \in \{b_{k_0+1}, ..., b_{k_2-1}\}$, (a_{k_0}, b_{k_2}) blocks μ by (iii) as $\mu^b(e_r) = b_{k_2}$. Then $\mu^b(a_{k_0}) \in \{b_{k_0}, b_{k_2}\}$. Suppose $\mu^b(a_{k_0}) = b_{k_0}$. Then, $\mu^b(a_i) = b_i$ for all $a_i \in A_2$ by (i). Hence, $\mu^d(a_{\hat{k}}) = d_{\hat{l}}$ and by (ix), $\mu^d(e_m) = d_{l_0}$. Then (a_{k_0}, b_{k_2}) blocks μ by (iv).

Now let $\mu^b(a_{k_0}) = b_{k_2}$. Therefore, $\mu^b(a_i) = b_{i-1}$ for all i with $k_0 < i \leq k_2$ by (i), and hence, $\mu^d(a_{\hat{k}+1}) = d_{\hat{l}}$. By (ix), we have $\mu^d(e_m) = d_{\hat{l}}$ as otherwise (in case $\mu^d(e_m) = d_{l_0}$) $(e_m, d_{\hat{l}})$ would block. Therefore, $\mu^e(b_{k_0}) \in \{e_{m+1}, ..., e_{r-1}\}$, but then (a_{k_0}, b_{k_0}) blocks μ by (iv).

Sufficiency. Let $C \in \mathcal{C}$ be as in the statement of the theorem and $\succ \in \mathcal{R}$. It suffices to show that the stable response algorithm terminates at a finite step according to Remark 2. Note that we can partition the battles into two groups $\{\{C^1\}, \{C^2\}\}$ such that C^1 is the set of undominated battles in terms of PA-dominance, C^2 is the set of dominated battles in terms of PA-dominance, and for all $q \in \{1, 2\}$ and for all $c, c' \in C^q$, c does not PA-dominate c'. Let $B^1 = \{b_k : c_{k.} \in C^1\}$, $B^2 = \{b_k : c_{k.} \in C^2\}$, and D^1, D^2 be defined similarly. For all $q \in \{1, 2\}$, let A^q_s (E^q_s) be the set of a-agents

(e-agents) that are matched with a b-client in B^q (d-client in D^q) at matching μ_s. For all $q \in \{1,2\}$, we call an odd (even) step s B^q (D^q) *improving* if for all $b_k \in B^q$ ($d_l \in D^q$), we have $\mu_s^a(b_k) \leq \mu_{s-1}^a(b_k)$ ($\mu_s^e(d_l) \leq \mu_{s-1}^e(d_l)$). We prove sufficiency using the following claims.

Claim 1. For all s and for all $q \in \{1,2\}$, A_s^q are positively assortatively matched with the clients in B^q, i.e., for all $b_k, b_{k'} \in B^q$, we have $\mu_s^a(b_k) < \mu_s^a(b_{k'})$. A similar argument holds for E_s^q and D^q.

Proof of Claim 1. Let $s = 0$. Then the claim holds by construction for both a-agents. To see that it also holds for e-agents, fix $c_{kl}, c_{k'l'} \in C^q$, without loss of generality assume $l < l'$, and let $\mu_0^e(d_l) = e_j$ and $\mu_0^e(d_{l'}) = e_{j'}$. As neither of the two battles in $\{c_{kl}, c_{k'l'}\}$ PA-dominates the other, we have $k' < k$. Since $\mu_0^a(b_{k'}) < \mu_0^a(b_k)$, we have $j < j'$ as otherwise (e_j, d_l) would be a blocking pair at μ_0 contradicting Remark 1. Now let the claim hold for the first $s-1$ steps for both a-agents and e-agents where $s-1$ is even. Let $q \in \{1,2\}$, $b_k, b_{k'} \in B^q$ with $k < k'$, $C(b_k) = d_l$, $C(b_{k'}) = d_{l'}$, $\mu_{s-1}^e(d_l) = e_j$ and $\mu_{s-1}^e(d_{l'}) = e_{j'}$. As c_{kl} and $c_{k'l'}$ are not related in terms of PA-dominance, we have $l' < l$ and by our induction assumption $j' < j$. Suppose for a contradiction that $\mu_s^a(b_{k'}) = a_{i'} < \mu_s^a(b_k) = a_i$. Then, $(a_{i'}, b_k)$ blocks μ_s, contradicting Remark 2 which states that there is no ab blocking pair. Proof for the even step s is similar. □

Claim 2. Step 1 is B^2 improving.

Proof of Claim 2. Suppose for a contradiction that step 1 is not B^2 improving. Then, there exist $b_k \in B^2$ such that $a_i = \mu_0^a(b_k) < \mu_1^a(b_k)$. Take the minimum k such that the aforementioned condition holds. Recall that μ_0 is a stable response of e-agents to $(\mu)_{ac}^{PAM}$ between a-agents and battles, and μ_1 is the outcome of a serial dictatorship of a-agents choosing partial matchings from $(\mu_0)_{ce}$. Then, in round i of this serial dictatorship procedure at step 1, no a-agents have chosen the partial matching in $(\mu_0)_{ce}$ associated with b_k as $a_i < \mu_1^a(b_k)$. Therefore, b_k is available for a_i at step 1.

Case 1: $a_i \in A_1^2$. Using Claim 1, we have $\mu_1^b(a_i) < b_k$. For all $k' < k$ such that $b_{k'} \in B^2$, $\mu_0^a(b_{k'}) < a_i$ since $\mu_0(a_{i'}) = b_{i'}$ for all $i' \in N$. This contradicts that k is the smallest index such that a b-client in B^2 receives a worse a-agent at step 1 compared to step 0.

Case 2: $a_i \in A_1^1$. Let $\mu_1^b(a_i) = b_{k'}$ and $\mu_0^a(b_{k'}) = a_{i'}$. Consider first the case that $i < i'$. Then $k < k'$ since $\mu_0(a_{i''}) = b_{i''}$ for all $i'' \in N$. Let $c_{kl}, c_{k'l'} \in C$, $\mu_0^e(d_l) = e_j$

and $\mu_0^e(d_{l'}) = e_{j'}$. Then, $l' < l$ as otherwise c_{kl} PA-dominates $c_{k'l'}$ which contradicts $b_{k'} \in B^1$. This implies that $j' < j$ as $(a_{i'}, d_{l'})$ is objectively better than (a_i, d_l) for all e-agents. Hence, $\mu_1^e(b_k) = e_j$ and $\mu_1^e(b_{k'}) = e_{j'}$. Then, a_i would have chosen (b_k, e_j) over $(b_{k'}, e_{j'})$ at step 1.

Now consider the case $i' < i$. Define $(a_i)_1^1 = |i'' < i : a_{i''} \in A_0^2 \text{ and } a_{i''} \in A_1^1|$ and $(a_i)_1^2 = |i'' < i : a_{i''} \in A_0^1 \text{ and } a_{i''} \in A_1^2|$. Then using Claim 1, if $(a_i)_1^2 < (a_i)_1^1$, $b_{k'}$ would not be available for a_i at round i of the serial dictatorship procedure at step 1, and if $(a_i)_1^1 < (a_i)_1^2$, b_k is not available for a_i. □

Claim 3. If an odd step s is B^2 improving, then for all $b_k \in B^1$, $\mu_{s-1}^a(b_k) \leq \mu_s^a(b_k)$.

Proof of Claim 3. Let $X = \{x_{h_1}, x_{h_2}, ...\}$ and $X' = \{x_{p_1}, x_{p_2}, ...\}$ with $|X| = |X'| = M$ be two ordered sets such that $h_1 < h_2 < \ldots$ and $p_1 < p_2 < \ldots$. We say $X \leq X'$, or equivalently X is component-wise not worse than X', if $h_m \leq p_m$ for all $m \in \{1, ..., M\}$. Let $\bar{A}_s^q$ be the ordered set derived from A_s^q where the first element in $\bar{A}_s^q$ is the a-agent with the lowest index, the second element is the a-agent with the second lowest index, and so on. Define the ordered sets $\delta_1(A_s^q) = \bar{A}_s^q \setminus \bar{A}_{s-1}^q$ and $\delta_2(A_s^q) = \bar{A}_{s-1}^q \setminus \bar{A}_s^q$ for $q \in \{1, 2\}$ and note that $|\delta_1(A_s^q)| = |\delta_2(A_s^q)|$ as $|\bar{A}_s^q| = |\bar{A}_{s-1}^q| = |D^q|$. Moreover, as both $\{A_{s-1}^1, A_{s-1}^2\}$ and $\{A_s^1, A_s^2\}$ are partitions of A, we have $\delta_1(A_s^1) = \delta_2(A_s^2)$ and $\delta_1(A_s^2) = \delta_2(A_s^1)$. Now, let s be B^2 improving, and hence, we have $\bar{A}_s^2 \leq \bar{A}_{s-1}^2$. Each new element in $\bar{A}_s^2$ is replaced by a unique element in $\bar{A}_{s-1}^2$ that has a higher index. Therefore, $\delta_1(A_s^2) \leq \delta_2(A_s^2)$, and hence, $\delta_2(A_s^1) \leq \delta_1(A_s^1)$. This implies that $\bar{A}_{s-1}^1 \leq \bar{A}_s^1$, and using Claim 1 we have the desired result. □

Claim 4. If an odd step s is B^2 improving, then step $s+1$ is D^1 improving.

Proof of Claim 4. Let s be B^2 improving and suppose for a contradiction that $s+1$ is not D^1 improving. Then, there exist $d_l \in D^1$ such that $e_j = \mu_s^e(d_l) < \mu_{s+1}^e(d_l)$. Take the minimum l such that the aforementioned condition holds. Recall that μ_{s+1} is a stable response of e-agents to $(\mu_s)_{ac}$. Therefore, μ_{s+1} is the outcome of a serial dictatorship of e-agents choosing partial matchings from $(\mu_s)_{ac}$. Then, in round j of this serial dictatorship procedure at step $s+1$, no e-agent has chosen the partial matching in $(\mu_s)_{ac}$ associated with d_l as $e_j < \mu_{s+1}^e(d_l)$. Therefore, d_l is available for e_j.

Case 1: $e_j \in E_{s+1}^1$. Using Claim 1, we have $\mu_{s+1}^d(e_j) < d_l$. For all $l' < l$ such that $d_{l'} \in D^1$, $\mu_s^e(d_{l'}) < e_j$ using Claim 1. This contradicts that l is the smallest index such that a d-client in D^1 receives a worse e-agent at step $s+1$ compared to step s.

Case 2: $e_j \in E^2_{s+1}$. Let $\mu^d_{s+1}(e_j) = d_{l'}$ and $\mu^e_s(d_{l'}) = \mu^e_{s-1}(d_{l'}) = e_{j'}$. Consider first the case that $j < j'$. Since step s is B^2 improving, the a-agent that $d_{l'}$ is matched with at step s (and hence at step $s+1$) is not worse than that at step $s-1$. Moreover, the a-agent that d_l is matched with at step s (and hence at step $s+1$) is not better than that at step $s-1$ using Claim 3. This contradicts that e_j chose d_l over $d_{l'}$ at step $s-1$ in round j of the serial dictatorship of e-agents since $j < j'$ implies both partial matches that are associated with d_l and $d_{l'}$ are available for e_j. Now, consider the case that $j' < j$. Define $(e_j)^1_{s+1} = |j'' < j : e_{j''} \in E^2_s$ and $e_{j''} \in E^1_{s+1}|$ and $(e_j)^2_{s+1} = |j'' < j : e_{j''} \in E^1_s$ and $e_{j''} \in E^2_{s+1}|$. Then, by Claim 1, if $(e_j)^1_{s+1} < (e_j)^2_{s+1}$, $d_{l'}$ is not available for e_j at round j of the serial dictatorship at step $s+1$; and if $(e_j)^2_{s+1} < (e_j)^1_{s+1}$, d_l is not available for e_j. □

We skip the proof for the following two claims which mimic the proof of Claims 3 and 4.

Claim 5. If an odd step s is D^1 improving, then for all $d_l \in D^2$, $\mu^e_{s-1}(d_l) \leq \mu^e_s(d_l)$.

Claim 6. If an even step s is D^1 improving, then step $s+1$ is B^2 improving.

We know from Claims 2 and 6 that at each odd step including step 1, the matchings of b-clients in B^2 are weakly improving in terms of the A agents they match. The same reasoning holds for d-clients in D^1. Then, as n is finite, the algorithm terminates at a finite step. ■

Proof of Proposition 1 Run Algorithm 1 and suppose that at step $s = 1$ there is some blocking pair. Using Remark 2 the blocking pair is of ed form. Take any two $(a_i, b_k, d_l, e_j), (a_{i'}, b_{k'}, d_{l'}, e_{j'}) \in \mu_1$ such that e_j and $d_{l'}$ form a blocking pair. Observe that then $j < j'$. Moreover, it cannot be that $d_l \in D^m_j$ and $d_{l'} \in D^{m'}_j$ for $m \neq m'$. In other words, d_l and d'_l are in the same element of e_j's partition of clients. If $m' > m$ then e_j always prefers to be matched with d_l rather than with d'_l and $(e_j, d_{l'})$ would not be a blocking pair. Similarly, if $m' < m$ then e_j always prefers to be matched with $d_{m'}$ rather than with d_m, but then μ_0 is not a stable response to any μ^{ac}.

As $d_l, d_{l'} \in D^k_j$ and $(e_j, d_{l'})$ form a blocking pair, it needs to be that $i < i'$. For μ_1 to be a stable response to μ^{ec}_0 it must be that $k < k'$. Otherwise, using Assumption 1, a_i and $b_{k'}$ block. However, if $k < k'$ and d_l, $d'_l \in D^k_j$ then μ_0 is not a stable response to PAM of a-agents to b-clients, as e_j and $d_{l'}$ would block it. Contradiction. ■

Proof of Proposition 2 Without loss of generality we consider a case in which b-clients have a binary type. First, consider a case in which C is such that for all c_{lk},$c_{l'k'} \in C$ if $b_k, b'_k \in B^H$ with $k < k$ then $l > l'$. Then, it is a corollary of Theorem 2 that the stable matching exists.

Second, consider a general case. Construct a fictional prematching $\hat{C}$ such that (i) if $b_k \in B^H$ and $\hat{C}(b_k) = d_l$, then $C(d_l) \in B^H$, (ii) if $k \in B^L$ and $\hat{C}(b_k) = d_l$, then $C(d_l) \in B^L$, (iii) c_{kl},$c_{k'l'} \in C$ if $k < k'$ and $b_k, b_{k'} \in B^H$ then $l > l'$. Construction of the fictional prematching amounts to a permutation of ds within groups of "good" and "bad" bs in a way that within each group bs and ds are negatively assortatively matched. From Theorem 2 it follows that a stable matching exists in this setting. Take any matching which is stable at $\hat{C}$ call it $\hat{\mu}$, preserve the matching of a-agents to d-clients to e-agents, and rearrange b-clients so that the prematching is given by C and call it μ.

Suppose that the resulting matching is not stable and take some (a_i, b_k, d_l, e_j), $(a_{i'}, b_{k'}, d_{l'}, e_{j'}) \in \mu$ $(a_i, b_{\hat{k}}, d_l, e_j)$, $(a_{i'}, b_{\hat{k'}}, d_{l'}, e_{j'}) \in \hat{\mu}$ such that $(a_i, b_{k'})$ is a blocking pair. It cannot be that both $b_k, b_{k'} \in B^H$. Otherwise, both $b_{\hat{k}}, b_{\hat{k'}} \in B^H$ and $(a_i, b_{\hat{k'}})$ would block $\hat{\mu}$. Analogous argument shows that both k and k' cannot be simultaneously in B^L.

Suppose $b_k \in B^L$ and $b_{k'} \in B^H$, then $b_{\hat{k}} \in B^L$ and $b_{\hat{k'}} \in B^H$. Moreover, since $\hat{\mu}$ is stable, $(b_{\hat{k}}, e_j) \succ_i^a (b_{\hat{k'}}, e_j)$. But then, using the fact that b-clients have a binary type $(b_k, e_j) \succ_i^a (b_{k'}, e_j)$. Hence, $(a_i, b_{k'})$ is not a blocking pair. Analogous reasoning applies to $b_k \in B^H$ and $b_{k'} \in B^L$. Contradiction. ■

Proof of Theorem 3

Necessity. Let $\succ \in \mathcal{R}$ and $\mu \in \mathcal{M}^C$ be stable at $\succ$, moreover, let (a_i, b_k, d_l, e_j), $(a_{i'}, b_{k'}, d_{l'}, e_{j'}) \in \mu$ and (a_i, e_j) PA-dominate $(a_{i'}, e_{j'})$. If $k' < k$, $(a_i, b_{k'})$ is a blocking pair at μ. If $l' < l$, $(e_j, d_{l'})$ is a blocking pair at μ. Then, as μ is stable, we have $k < k'$ and $l < l'$.

Sufficiency. Take any matching $\mu \in \mathcal{M}^C$ and let $\succ \in \mathcal{R}$ be such that for all $i \in N$, if $\mu^b(a_i) = b_k$, $\succ_i^a$ is a threshold preference defined by the ordered partition $\{\{b_1, ..., b_k\}, \{b_{k+1}, ..., b_n\}\}$, and e-agents also have threshold preferences defined similarly.

Suppose that μ is not stable at $\succ$, and without loss of generality say that there is an ab blocking pair. Let (a_i, b_k, d_l, e_j), $(a_{i'}, b_{k'}, d_{l'}, e_{j'}) \in \mu$ and $(a_i, b_{k'})$ be a blocking pair at μ. Since $(a_i, e_{j'}) \succ_{k'}^b (a_{i'}, e_{j'})$, $i < i'$. As $(b_{k'}, e_{j'}) \succ_i^a (b_k, e_j)$, $b_{k'} \in \{b_1, ...b_k\}$;

and hence, $k' < k$ and $j < j'$ by definition of threshold preferences. Then, (a_i, e_j) PA-dominates $(a_{i'}, e_{j'})$ but (b_k, d_l) does not PA-dominate $(b_{k'}, d_{l'})$. ■

Proof of Theorem 4

Sufficiency. Take any C and η such that $\Gamma_\tau(C)$ is subgraph isomorphic to $\Gamma_\tau(\eta)$. As a result there exists a bijection $f(.)$ from η to C such that for any $(a_i, e_j), (a_{i'}, e_{j'}) \in \eta$, if there exists an edge from (a_i, e_j) to $(a_{i'}, e_{j'})$, then there is an edge from $f((a_i, e_j))$ to $f((a_{i'}, e_{j'}))$. Use this bijection to build the matching: $\mu = \{(a_i, f((a_i, e_j)), e_j)_{(a_i, e_j)\in\eta}\}$. Observe that by construction the PA-dominance relation is preserved from the agent matching to the client matching. Hence, using Theorem 3, η can be supported by C.
Necessity. Take any C and η such that η is supported by C. Hence, using Theorem 3, there exists a bijection $f(.)$ from η to C such that for any $(a_i, e_j), (a_{i'}, e_{j'}) \in \eta$ for which (a_i, e_j) PA-dominates $(a_{i'}, e_{j'})$ it is the case that $f((a_i, e_j))$ PA-dominates $f((a_{i'}, e_{j'}))$. Take this bijection and apply it for matching vertices of $\Gamma_\tau(\eta)$ and $\Gamma_\tau(C)$. Observe that, by construction, if there exists an edge from (a_i, e_j) to $(a_{i'}, e_{j'})$ then there exists an edge from $f((a_i, e_j))$ to $f((a_{i'}, e_{j'}))$. Hence, $\Gamma_\tau(C)$ is subgraph isomorphic to $\Gamma_\tau(\eta)$. ■

Proof of Proposition 3

Fix $C \in \mathcal{C}$ and $\succ \in \mathcal{R}$. Let $\mu \in \mathcal{M}^C$ be stable at $\succ$ and $(a_1, b_k, d_l, e_j) \in \mu$ for some $k, l, j \in N$. Suppose for a contradiction that there exists $\mu' \in \mathcal{M}^C$ that Pareto dominates μ. We first show that $(a_1, b_k, d_l, e_j) \in \mu'$. Suppose not and consider first the case $(a_1, b_{k'}, d_{l'}, e_{j'}) \in \mu'$ for $k \neq k'$. As $\mu' \succ^d_{l'} \mu$, $e_{j'}$ is a better e-agent than $\mu^e(d_{l'})$. Then, since $(b_{k'}, \mu^e(d_{l'})) \succ^a_1 (b_{k'}, e_{j'}) \succ^a_1 (b_k, e_j)$, $(a_1, b_{k'})$ would have blocked μ. Now suppose $(a_1, b_k, d_l, e_{j'}) \in \mu'$ for $j \neq j'$. Since $\mu' \succ^d_l \mu$, $e_{j'}$ is a better e-agent than e_j. This contradicts that $\mu' \succ^a_1 \mu$. Given that a_1's match did not change, we can repeat the same argument one-by-one for a_2, a_3, and so on. ■

Proof of Proposition 4 Let $C \in \mathcal{C}$ and $\succ \in \mathcal{R}$, $\mu \in \mathcal{M}^C$ be stable at $\succ$, and (a_i, b_k, d_l, e_j) block μ. Suppose without loss of generality for a contradiction that $a_i < \mu^a(b_k)$. As d_l is in the blocking quadruple, we have $e_j < \mu^e(d_l)$. Then, $(b_k, \mu^e(d_l)) \succ^a_i (b_k, e_j) \succ^a_i (\mu^b(a_i), \mu^e(a_i))$. Then (a_i, b_k) is a blocking pair at μ. ■

References

Amanda Agan, Matthew Freedman, and Emily Owens. Is your lawyer a lemon? incentives and selection in the public provision of criminal defense. *Review of Economics and Statistics*, 103(2):294–309, 2021.

Ahmet Alkan. Nonexistence of stable threesome matchings. *Mathematical Social Sciences*, 16(2):207–209, 1988.

Gary S Becker. A theory of marriage: Part i. *Journal of Political Economy*, 81(4): 813–846, 1973.

Ken Burdett and Melvyn G Coles. Marriage and class. *The Quarterly Journal of Economics*, 112(1):141–168, 1997.

Hector Chade and Jan Eeckhout. Competing teams. *The Review of Economic Studies*, 87(3):1134–1173, 2020.

Bo Chen. Downstream competition and upstream labor market matching. *International Journal of Game Theory*, 48(4):1055–1085, 2019.

Bo Chen. Labor market matching with ensuing competitive externalities in large economies. *Mathematical Social Sciences*, 109:12–17, 2021.

Julien Combe. Matching with ownership. *Journal of Mathematical Economics*, page 102563, 2021.

Vladimir I Danilov. Existence of stable matchings in some three-sided systems. *Mathematical Social Sciences*, 46(2):145–148, 2003.

Miguel De Luca, Mark P Jones, and María Inés Tula. Back rooms or ballot boxes? candidate nomination in argentina. *Comparative Political Studies*, 35(4):413–436, 2002.

Bhaskar Dutta and Jordi Massó. Stability of matchings when individuals have preferences over colleagues. *Journal of Economic Theory*, 75(2):464–475, 1997.

Federico Echenique and Leeat Yariv. An experimental study of decentralized matching. *Working Paper*, 2012.

Kimmo Eriksson, Jonas Sjöstrand, and Pontus Strimling. Three-dimensional stable matching with cyclic preferences. *Mathematical Social Sciences*, 52(1):77–87, 2006.

Vincenzo Galasso and Tommaso Nannicini. Competing on good politicians. *American Political Science Review*, 105(1):79–99, 2011.

David Gale and Lloyd S Shapley. College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15, 1962.

Isa E Hafalir. Stability of marriage with externalities. *International Journal of Game Theory*, 37(3):353–369, 2008.

John William Hatfield and Paul R Milgrom. Matching with contracts. *American Economic Review*, 95(4):913–935, 2005.

Eric Helland and Alexander Tabarrok. Contingency fees, settlement delay, and low-quality litigation: Empirical evidence from two datasets. *Journal of Law, Economics, and Organization*, 19(2):517–542, 2003.

Chien-Chung Huang. Circular stable matching and 3-way kidney transplant. *Algorithmica*, 58(1):137–150, 2010.

Nahomi Ichino and Noah L Nathan. Primaries on demand? intra-party politics and nominations in ghana. *British Journal of Political Science*, 42(4):769–791, 2012.

Elisabetta Iossa and Bruno Jullien. The market for lawyers and quality layers in legal services. *The RAND Journal of Economics*, 43(4):677–704, 2012.

Behrang Kamali Shahdadi. Matching with moral hazard: Assigning attorneys to poor defendants. *American Economic Journal: Microeconomics*, 10(3):1–33, 2018.

Ayşe Mumcu and Ismail Saglam. Stable one-to-one matchings with externalities. *Mathematical Social Sciences*, 60(2):154–159, 2010.

Antonio Nicolo, Arunava Sen, and Sonal Yadav. Matching with partners and projects. *Journal of Economic Theory*, 184:104942, 2019.

M. Elain Nugent-Borakove and Franklin Cruz. The power of choice. the implication of a system where indigent defendants choose their own counsel. *The Justice Management Institute Report*, 2017.

Marek Pycia and M Bumin Yenmez. Matching with externalities. *University of Zurich, Department of Economics, Working Paper*, (392), 2021.

Madhav Raghavan. Matching firms and workers through intermediaries. *Available at SSRN 3882447*, 2021.

Hiroo Sasaki and Manabu Toda. Two-sided matching problems with externalities. *Journal of Economic Theory*, 70(1):93–108, 1996.

Yotam Shem-Tov. Make-or-buy? the provision of indigent defense services in the us. *The Review of Economics and Statistics*, pages 1–27, 2020.