Centralized Matching with Incomplete Information

Marcelo Ariel Fernandez[*], Kirill Rudov[†], and Leeat Yariv[‡§]

July 6, 2021

Appeared in American Economic Review: Insights, 2022, 4(1), 18-33

Abstract

We study the impacts of incomplete information on centralized one-to-one matching markets. We focus on the commonly used Deferred Acceptance mechanism (Gale and Shapley, 1962). We show that many complete-information results are fragile to a small infusion of uncertainty about others' preferences.

[*]Department of Economics, Johns Hopkins University, fernandez@jhu.edu

[†]Department of Economics, Princeton University, krudov@princeton.edu

[‡]Department of Economics, Princeton University, CEPR, and NBER, lyariv@princeton.edu

[§]We thank Muriel Niederle for many conversations that inspired this project as well as the Editor Dirk Bergemann, four anonymous reviewers, Yeon-Koo Che, Laura Doval, Federico Echenique, Nicole Immorlica, and Fuhito Kojima for very helpful comments and suggestions. We gratefully acknowledge financial support from the National Science Foundation through grant SES-1629613 and from the William S. Dietrich II Economic Theory Center at Princeton University.

1. Introduction

1.1. Overview. This paper demonstrates the impacts of incomplete information on matching clearinghouses' outcomes. The large literature on two-sided matching has been successful in its application to various markets, ranging from the matching of doctors to residency positions to the matching of kids to schools. Most of this literature assumes complete information: participants are perfectly informed of all other participants' preferences, in addition to their own. We introduce uncertainty about others' preferences in a centralized matching setting employing the celebrated Deferred Acceptance algorithm (DA), introduced by Gale and Shapley (1962). We show that several canonical results break down even when minimal such uncertainty is added.

Stability has been of central importance in the design of matching clearinghouses. In a variety of markets, studies have documented the persistence of stable centralized clearinghouses, and the collapse of others (Roth and Xing, 1994; Roth, 2002; McKinney, Niederle, and Roth, 2005). DA clearinghouses are prevalent in applications. They implement stable outcomes for reported preferences and are incentive compatible when cores are small.[1] Even with incomplete information, there is arguably a hope that generated outcomes would be ex-post (or complete-information) stable, at least with singleton cores. Otherwise, interactions following the clearinghouse's operations would potentially undo matches. For example, in the medical residency labor market, a newly-minted doctor matched to her, say, third choice could call her top two choices and form a blocking pair, even absent complete information, see Roth and Sotomayor (1992). Indeed, concern about such incidents was part of the impetus for introducing centralized clearinghouses to begin with. We therefore analyze conditions under which ex-post stable outcomes are to be expected when cores are small.

There are two main messages. First, equilibria implementing the ex-post stable outcomes always exist, and in some classes of incomplete-information economies—e.g., ones with as-

[1]With complete information, DA mechanisms exhibit many appealing features relative to other stable mechanisms. In particular, DA mechanisms cannot be improved upon in terms of manipulability, see Van der Linden (2019).

sortative preferences on either side—they are unique (Propositions 1, 2). Second, however, even complete-information versions of such economies are fragile to the addition of minimal uncertainty about others' preferences. Maintaining singleton cores, a vanishing fraction of market participants can be added so that many features of stable outcomes break down in equilibrium and a large fraction of participants is affected (Propositions 3, 4).

Our constructions illustrate that many classical results are overturned in the presence of incomplete information. Equilibrium outcomes corresponding to unstable matchings may be desirable for the *receiving* side in a DA clearinghouse, despite the traditional view of the proposing side as advantaged. Furthermore, the set of matched individuals may differ across equilibrium outcomes. Unlike in the complete-information benchmark, equilibrium selection can offer a useful instrument for influencing who gets matched. We also show that several frequently-used technical simplifications are invalid when information is incomplete. In particular, best-response strategy sets do not necessarily include truncation strategies.

Taken together, these results suggest the importance of accounting for details pertaining to the information market participants have in centralized matching clearinghouses.

1.2. Related Literature. Roth (1989) offers a particular example of a market with incomplete information on *both* sides in which there exists no stable mechanism implementing the (complete-information) stable matching for each preference realization. Importantly, in his example, preference profiles that can conceivably be realized are associated with *multiple* stable matchings.[2]

With incomplete information, Ehlers and Massó (2007) link singleton cores and truthful preference revelation as an ordinal Bayesian Nash equilibrium. We assume a singleton core state-by-state. Consequently, core outcomes can be implemented in equilibrium. We show cases that yield *other* equilibrium outcomes.[3]

[2]Roth and Rothblum (1999) consider a class of economies in which there are always optimal truncation strategies in the firm-proposing DA. Coles and Shorrer (2014) identify optimal truncation strategies in a large class of markets. We illustrate cases in which truncation strategies are sub-optimal.

[3]Immorlica et al. (2020) examine how the design of school-choice systems affects information acquisition. Fernandez (2020) shows that DA only supports stable outcomes when agents avoid regret.

In the decentralized, cooperative setting, there has been a long-standing quest for a natural stability notion allowing for incomplete information. Liu et al. (2014) and Liu (2020) offer such a notion for matching markets with transfers and one-sided incomplete information, similar to ours. Bikhchandani (2017) suggests such a notion for similar markets that do not allow transfers.[4]

A growing empirical literature estimates preferences in centralized matching markets using constraints implied by stability (see review in Chiappori and Salanié, 2016 as well as Agarwal, 2015 and Hsieh, 2012). Our results provide caution for this approach when preference information may be incomplete. Market features guaranteeing small cores when information is complete (see, e.g., Ashlagi, Kanoria, and Leshno, 2017 and references therein), may be insufficient for ensuring unique stable equilibrium outcomes. Furthermore, observed small cores for *reported* preferences do not guarantee their truthfulness.

2. The Model

2.1. The Matching Economy. A *matching market* is a triplet $\mathcal{M} = (F, W, U)$ composed of a finite set of firms $F = \{f_i\}_{i\in[m]}$, where $[m] = \{1, 2, \ldots, m\}$; a finite set of workers $W = \{w_j\}_{j\in[n]}$; and match utilities $U = \{u^f_{ij}, u^w_{ij}\}_{(i,j)\in[m]\times[n]}$. For each pair (i, j), u^f_{ij} is firm f_i's utility from matching with worker w_j and u^w_{ij} is worker w_j's utility from matching with firm f_i. Let $u^f_{i\varnothing}$ and $u^w_{\varnothing j}$ denote the utilities of firm f_i and worker w_j from remaining unmatched, respectively. Without loss of generality, all utilities from being unmatched are normalized to zero, i.e. $u^f_{i\varnothing} = u^w_{\varnothing j} = 0$ for all i, j. We assume all preferences are strict. Throughout, we focus on markets where all worker-firm pairs are mutually acceptable; i.e., $u^f_{ij} > u^f_{i\varnothing}$ and $u^w_{ij} > u^w_{\varnothing j}$.[5]

An *economy* is a quintuple $\mathcal{E} = (F, W, \{U(\theta)\}_{\theta\in\Theta}, \Theta, \Psi)$, where Θ is a finite set of states

[4] Chakraborty, Citanna, and Ostrovsky (2010) study a school-choice setting with one-sided incomplete information. They suggest defining stability together with the underlying clearinghouse. Several papers allow for incomplete information when modelling decentralized interactions as a non-cooperative dynamic game (see Ferdowsian, Niederle, and Yariv, 2021 and references therein).

[5] Our negative results do not rely on all agents being acceptable.

with each $\theta \in \Theta$ corresponding to a different market $\mathcal{M}(\theta) = (F, W, U(\theta))$ with the same set of firms F and workers W, and Ψ is a probability distribution over states. Without loss of generality, we assume Ψ has full support on Θ.

2.2. The Game and Equilibrium Concept. We consider a *centralized matching economy game* in which, at the outset, a state $\theta \in \Theta$ is selected according to Ψ. All firms are informed of the realized state θ and their match utilities in that state—their "types" are fully revealing. Each worker w_j is privately informed only of his preferences, $u_{ij}^w(\theta)$ for $i \in [m] \cup \{\varnothing\}$, which constitute his "type." We assume workers' private information is not revealing of the state; i.e. for any $\theta, \theta' \in \Theta$, $u_{ij} \equiv u_{ij}^w(\theta) = u_{ij}^w(\theta')$. These match utilities induce an ordinal preferences profile $\succ = \{\succ_{f_i}, \succ_{w_j}\}_{(i,j)\in[m]\times[n]}$. In the Online Appendix, we consider more general environments, allowing for two-sided incomplete information or richer private types.

Participants simultaneously submit rank-ordered lists over acceptable partners to a centralized clearinghouse, which generates a matching using the firm-proposing DA algorithm.

Since truthful reporting is weakly dominant for firms (Roth, 1989), we focus on Bayesian Nash equilibria (hereafter, BNE) with truthful firms.[6]

In this paper, we illustrate how "small" deviations from the complete-information benchmark can alter outcomes dramatically in a heavily utilized clearinghouse. In particular, we introduce uncertainty about others' preferences only on one market side. Why is the uninformed side of the market (the workers) on the algorithm's receiving side? Otherwise, since firms are informed of the state, and truth-telling is weakly dominant for the proposing side, the analysis would reduce to that of complete information. Naturally, in general, information may be incomplete on both market sides and a designer may not know which side is better informed. As we show in the Online Appendix, our insights on the impacts of incomplete information hold for a general class of stable mechanisms, not just the firm-proposing DA.

[6]Any worker strategy that lists the most preferred firm first is weakly undominated, see Roth and Sotomayor (1992). In our analysis, the distinction between ex-ante and interim weakly undominated strategies has no bite, see details in the Online Appendix.

We return to this point when discussing our results.

In the complete-information benchmark, incentive compatibility issues are associated with multiplicity of stable matchings. In order to isolate the impacts of incomplete information on strategic behavior, we assume that each market in the support of Ψ has a unique stable matching.[7] All participants reporting their preferences truthfully constitutes a BNE in weakly undominated strategies. Therefore,

Proposition 0. *Any economy admits a BNE in weakly undominated strategies implementing the unique (complete-information) stable outcome in each state.*

Throughout, we refer to an outcome as "unstable" if it is unstable for the *true* preferences.

2.3. Preference Assumptions. The literature has not yet identified general necessary and sufficient conditions for a market to exhibit a unique stable matching. Throughout our analysis, we focus on the *Sequential Preference Condition* (hereafter, SPC), first introduced by Eeckhout (2000).

Formally, a market $\mathcal{M}' = (F', W', U')$ is a *sub-market* of the original market $\mathcal{M} = (F, W, U)$ if $F' \subseteq F$, $W' \subseteq W$, and U' is induced by U when restricted to $F' \times W'$. We say that (f, w) is a *top-top match* for sub-market $\mathcal{M}'$ if, for firm f, worker w is the favorite worker in $\mathcal{M}'$, and vice versa. A market satisfies the *SPC* if, up to relabeling, there exists an ordering of the firms $f_1, f_2, \ldots, f_m \in F$ and an ordering of the workers $w_1, w_2, \ldots, w_n \in W$ such that for any $i \leq \min(m, n)$, (f_i, w_i) is a top-top match for the sub-market induced by $\{f_j, w_j\}_{j \geq i}$. For these orderings, we say a pair (f_i, w_i), $i \leq \min(m, n)$, and its respective agents have *order* i.

Any market that satisfies the SPC has a unique stable matching that can be derived by partnering top-top pairs in sequence. The condition generalizes numerous others.[8] In

[7] Recent literature has identified various conditions under which large markets entail small cores (see, e.g., Immorlica and Mahdian, 2005 and Ashlagi et al., 2017). Our analysis speaks to such settings.

[8] These include α-reducibility (Clark, 2006), the co-ranking condition (Legros and Newman, 2010), the universality condition (Holzman and Samet, 2014), the aligned preferences condition (Ferdowsian, Niederle, and Yariv, 2021), and oriented preferences (Reny, 2021).

particular, the SPC is satisfied for markets in which firms (workers) share the same ranking of workers (firms), corresponding to firms (workers) having *assortative preferences*.

A given economy $\mathcal{E} = (\{\mathcal{M}(\theta)\}_{\theta\in\Theta}, \Theta, \Psi)$ satisfies the *SPC* if each market $\mathcal{M}(\theta)$ in its support satisfies the SPC for possibly state-specific orderings of firms $f_{1|\theta}, f_{2|\theta}, \ldots, f_{m|\theta}$ and workers $w_{1|\theta}, w_{2|\theta}, \ldots, w_{n|\theta}$.

As it turns out, the SPC alone does not guarantee a unique equilibrium outcome (see the Online Appendix). The following restriction, however, will be useful for identifying economies that do.

Definition 1. *An economy* $\mathcal{E} = (\{\mathcal{M}(\theta)\}_{\theta\in\Theta}, \Theta, \Psi)$ *satisfies the SPC* if it satisfies the SPC and, for any state* $\theta \in \Theta$*, and any order* $i \leq \min(m, n)$*,*

if $f \succ_{w_{i|\theta}} f_{i|\theta}$*, then for any* θ'*, there exists* $i' < i$ *(that may depend on* θ'*) with* $f = f_{i'|\theta'}$*.*

The SPC* implies that if, in some state, a worker of a given order i prefers some firm over his stable partner in this state, then in *any* state, this firm must prefer her stable (state-specific) partner over the corresponding (state-specific) worker of the same order i. In other words, if some firm is "unreachable" for a worker, it is "unreachable" for other workers of the same order in any of the possible states.

Our economy game, and the SPC*, do not treat firms and workers symmetrically. In particular, an economy with (possibly state-dependent) assortative preferences for firms may violate the SPC* (see the Online Appendix).[9] In general, the SPC* holds whenever the order of firms induced by the SPC is state-independent.

3. Motivating Example

We start by illustrating a simple incomplete-information economy in which unstable outcomes result from equilibria in weakly undominated strategies.

[9] Any economy with assortative preferences for workers satisfies the SPC*.

Example 1. Consider an economy with three firms and three workers, $m = n = 3$, and two states of the world, $\Theta = \{1, 2\}$, which are equally likely. Figure 1 describes, for state $\theta \in \Theta$, the preferences. In this matrix notation, rows correspond to firms and columns correspond to workers: $u_{ij}(\theta) = (u_{ij}^f(\theta), u_{ij}^w)$. We assume that remaining unmatched generates a payoff of 0 for any market participant.

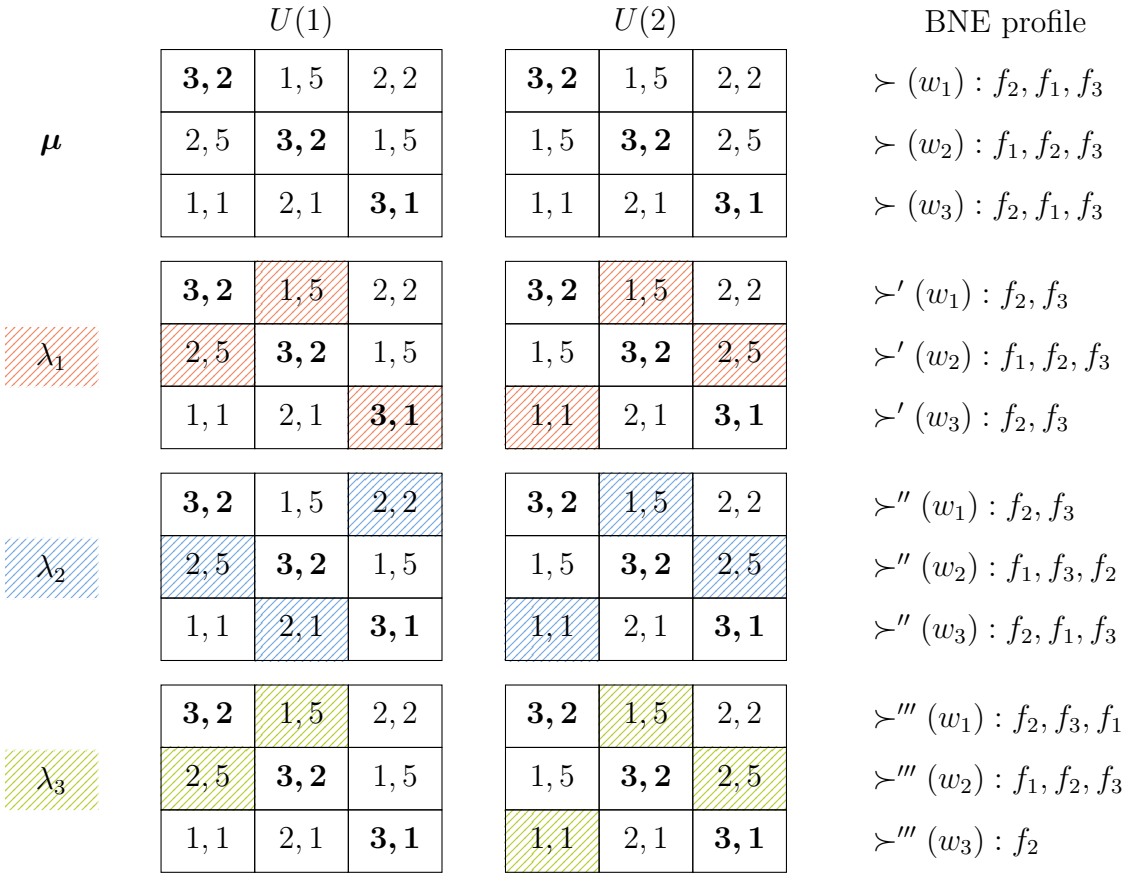

Figure 1: Economy with four equilibrium outcomes in which firms report truthfully

In both states, there is a unique complete-information stable matching highlighted in bold. Despite information being incomplete, agents have no uncertainty regarding the (complete-information) stable matching. The only difference between the two states appears in f_2's preferences—she ranks w_1 and w_3 differently across the states. In particular, if all participants knew the state to be θ, the resulting unique stable matching and unique equilibrium outcome of the DA mechanism would be the matching μ, $\mu(f_i) = w_i$, $i \in \{1, 2, 3\}$.

As stated, with firms truthfully reporting, the complete-information stable matching μ in each state is an equilibrium outcome of this game. We now show that it is not unique: there are three other unstable equilibrium outcomes, all supported by weakly undominated strategies. The resulting matchings in each state are depicted in Figure 1.

Consider the λ_1 equilibrium outcome. In the corresponding equilibrium, w_2 reports his preferences truthfully, while w_1 and w_3 *drop* f_1 from their preference list, declaring her unacceptable. In particular, truncation strategies are not best responses for both w_1 and w_3, unlike the complete-information setting, when there are always truncation best responses (see Roth and Vate, 1991).

This equilibrium is appealing for workers—they uniformly prefer it to the one generating the (complete-information) stable outcome in each state. Thus, workers prefer being on the receiving side of DA with this equilibrium selected, see our discussion in Section 2.2.

What allows for this equilibrium to emerge? In state 1, were w_3 and f_3 absent, the resulting sub-market would have two stable matchings: one matching w_i with f_i; the other, more preferable to the workers, matching w_i with f_j, $i = 1, 2, j = 3 - i$. This multiplicity is generated by a *cycle* in the sub-market: f_1 prefers w_1, who prefers f_2, who prefers w_2, who prefers f_1. Worker w_1 would then benefit from "truncating" his preferences and reporting f_1 as unacceptable. Of course, w_3 is not absent and could attempt blocking such a matching with f_1. Why does he drop his claim for f_1 as well? Consider state 2. In the sub-market without w_1 and f_3, there is again a preference cycle yielding two stable matchings. Worker w_3 "truncating" and reporting f_1 as unacceptable then guarantees the workers' preferred stable matching. Of course, w_1 is present and could block this matching with f_1, but state 1 ensures that he does not.

Such inter-linked cycles may be generated through the underlying preferences, or through the reported preference profiles. For example, in the λ_2 equilibrium, under the reported preferences, there is a cycle involving all firms and workers in state 1, although there is no such cycle in the true preferences. Furthermore, the λ_2 and λ_3 equilibrium outcomes are

supported only by profiles that entail *permuting* firms in reported preferences. Importantly, the construction of these equilibria is not knife-edge: a small perturbation in match payoffs would not alter their description.

With complete information, when firms use their weakly dominant truth-telling strategy, the set of equilibrium outcomes of the firm-proposing DA coincides with the set of stable matchings and therefore inherits its lattice structure. As this example illustrates, with incomplete information, side-optimality and the existence of "extremal" equilibria break down. Indeed, f_1 and w_3 prefer the outcome λ_2 to λ_1, while f_3 and w_2 prefer λ_1 to λ_2.

Furthermore, in this example, the set of unmatched individuals varies across equilibria: λ_3 is the only equilibrium in which worker w_3 is unmatched in state 1. This stands in contrast to the complete-information setting, where the Rural Hospital Theorem (McVitie and Wilson, 1970) implies that the set of unmatched individuals is constant across stable matchings, and therefore DA equilibrium outcomes.

The equilibria all induce unique stable matchings for *reported* preferences in both states. Nonetheless, while corresponding cores are small, reports are not truthful.

The economy in this example does not satisfy the SPC* condition. However, without f_2, the resulting "sub-economy" effectively entails complete information, satisfies the SPC, and yields a unique equilibrium outcome. As we show, this fragility to the infusion of incomplete information through a small number of participants—here, one—is quite general. △

4. Stable Equilibrium Outcomes

In this section, we identify an important class of economies in which incomplete information does not hinder stability and results from the complete-information setting carry over. Nonetheless, in the next section, we show these conclusions are arguably fragile.[10]

[10]For presentation simplicity, we maintain our assumption of one-sided incomplete information throughout. In the Online Appendix, however, we illustrate that this section's results hold more generally, even when information is incomplete for both market sides.

Proposition 1. *Consider any economy $\mathcal{E}$ such that either (1) firms have (possibly state-specific) assortative preferences or (2) the SPC* is satisfied. Then, it admits a unique BNE outcome corresponding to the unique (complete-information) stable outcome in each state.*

Intuitively, consider first the case in which firms have assortative preferences. In each state, firms share the same rankings over workers that they report truthfully. Then, for any state, in the course of the firm-proposing DA, all active firms apply to the same worker. Through the DA steps, the sets of active firms are nested. Therefore, regardless of workers' reports, each worker makes only one choice in the course of the firm-proposing DA and these choices are decisive. In particular, misreporting cannot be beneficial to workers. The only potential impact of a worker misreporting is that he may forgo a firm he would otherwise match to and, consequently, be matched to an inferior firm, or no firm at all.

The intuition underlying the proposition's second part is more subtle. For any state θ, the corresponding top-top match pair $(w_{1|\theta}, f_{1|\theta})$ must be matched under any BNE (Lemma 1 in the Appendix). If not, since firm $f_{1|\theta}$ reports truthfully, worker $w_{1|\theta}$'s top reported firm is not his most preferred firm $f_{1|\theta}$, so that $w_{1|\theta}$ could profitably deviate by shifting $f_{1|\theta}$ to the top of his reported ranking.

Hence, if there exists an unstable BNE outcome, we can find a state together with an unmatched "sequential top-top pair" in this state of order at least two. Let $(w_{i|\theta}, f_{i|\theta})$ be an unmatched pair of the smallest possible order $i \geq 2$ among all such pairs and θ be the corresponding state.[11] Again, because firm $f_{i|\theta}$ reports truthfully, worker $w_{i|\theta}$ must be reporting his less desirable equilibrium partner as preferable to $f_{i|\theta}$. We can then construct a deviation strategy such that $w_{i|\theta}$ reports truthfully his preferences over firms $\{f : f \succeq_{w_{i|\theta}} f_{i|\theta}\}$ and maintains the equilibrium ranking for all other firms.

In any state θ in which worker $w_{i|\theta}$ is matched with firm $f \preceq_{w_{i|\theta}} f_{i|\theta}$ under the BNE, the deviation ensures that the worker either receives f or a preferable firm from $\{f : f \succeq_{w_{i|\theta}} f_{i|\theta}\}$.

Suppose that, for some state $\theta' \neq \theta$, the worker is supposed to match with firm $f \succ_{w_{i|\theta}} f_{i|\theta}$

[11] If there are multiple such pairs, we can pick any.

in equilibrium. Both the SPC* and the minimality of i are crucial for establishing that the proposed deviation cannot hurt worker $w_{i|\theta}$. The SPC* implies that in state θ', the "demanded" firm f belongs to a sequential pair of order at most $(i-1)$. By the minimality of i, in state θ', the corresponding sequential pair $(w_{i|\theta}, f)$ of order at most $(i-1)$ must be matched. The proposed deviation must then deliver a firm weakly preferable to f under the true preferences. Otherwise, the recursive use of minimality of i and stability for reported preferences yields a contradiction. Intuitively, since i is minimal, other workers' reports do not impede on worker $w_{i|\theta}$ getting his stable match.

In general, equilibria yielding unstable outcomes may exploit cycles in the *true preferences*, as in the motivating example's first equilibrium, or generate cycles through equilibrium reports involving permutations, as in the motivating example's second equilibrium.[12]

Do we restore stability if we prohibit permutations in reported preferences and focus on economies with no cycles in true preferences? The result below answers affirmatively.[13]

Proposition 2. *Restrict workers' strategy sets to dropping strategies and consider any economy $\mathcal{E}$ such that either (1) all markets in its support do not have preference cycles*[14] *or (2) the SPC* is satisfied. Then, it admits a unique BNE outcome corresponding to the unique (complete-information) stable outcome in each state.*

The proof of this result is in the Online Appendix.[15]

Common values in matching markets naturally translate to assortative preferences, where all agents on one market side (or both sides) rank agents on the other market side identically. Propositions 1 and 2 guarantee that in such markets, when common values are exhibited by either side of the market, there is a unique equilibrium outcome. We next show that the

[12] See also Examples 7 and 8 in the Online Appendix, which illustrate unstable equilibrium outcomes in two-state economies in which one state exhibits no (true) preference cycles.

[13] A dropping strategy for a worker declares some firms unacceptable, but ranks others truthfully.

[14] This condition is formally called the *aligned preferences condition* (Ferdowsian, Niederle, and Yariv, 2021). It is stronger than the SPC, weaker than requiring assortative preferences for either side, and different from the SPC*. Examples 3-5 in the Online Appendix illustrate connections between these conditions.

[15] Example 6 there illustrates that one cannot relax either condition to the SPC in both of this section's propositions.

addition of "minimal" private values, through a small fraction of added participants, can alter market outcomes dramatically.

5. Fragility Results

We now show that various properties displayed by equilibrium outcomes of complete-information economies are fragile to the infusion of "minute" uncertainty about others' preferences.

We first consider a complete-information economy satisfying the SPC that, from our results thus far, is potentially more resistant to the infusion of uncertainty. We refer to it as the *original economy*. For simplicity, we assume the original economy is balanced, with equal numbers of firms and workers, $m = n$.[16]

We augment the original economy by adding one state, and several workers and firms, with their preferences and others' preferences over them in each state. Each market in the augmented economy entails a unique stable matching and occurs with positive probability. Importantly, the preferences of workers and firms present in the original economy remain unchanged with respect to one another in both states. Furthermore, the preferences of workers in the augmented economy coincide across the two states, so that no worker is informed of the state.

Formally, an *augmented economy* of the original market $\mathcal{M} = (F, W, U)$ *with firms* F' *and workers* W' is an economy $\mathcal{E} = (F \cup F', W \cup W', \{U(\theta), U(\theta')\}, \{\theta, \theta'\}, \Psi)$ such that (1) both $U(\theta)$ and $U(\theta')$ coincide with U when restricted to $F \times W$; (2) each $w \in W'$ has identical match utilities in both states; and (3) Ψ is non-degenerate and both $U(\theta)$ and $U(\theta')$ are each associated with a unique stable matching.

Proposition 3. *Consider any balanced market $\mathcal{M}$ with at least two agents on each side satisfying the SPC. Then, we can construct an augmented economy $\mathcal{E}$ of $\mathcal{M}$ with one additional firm and one additional worker such that:*

1. *there is a BNE that supports unstable outcomes in both states;*

[16] Somewhat more involved arguments can be used to extend our analysis to imbalanced economies.

2. *any BNE outcome can be supported by weakly undominated strategies that induce a unique stable matching (for reported preferences) in each state;*

3. *the set of matched workers varies across equilibrium outcomes. Furthermore, workers disagree on which is preferred.*

Certainly, one can always augment a complete-information economy satisfying our restrictions by "appending" three workers and three firms that view all of the original economy's participants as least desirable and are viewed by them as least desirable as well. These added firms and workers can exhibit the same preferences over one another as in our motivating example. The conclusions of the proposition would then follow for *any* original complete-information economy, even absent the SPC assumption. The main point of the proposition is that instability can emerge in equilibrium even with minimal augmentation.

To illustrate our construction, the details of which appear in the Online Appendix, consider the motivating example. Workers $\{w_1, w_3\}$ and firms $\{f_1, f_3\}$ exhibit the same match utilities over one another in both states. Furthermore, the corresponding sub-market satisfies the SPC, with (f_1, w_1) as the top-top pair. The motivating example can then be thought of as an augmentation of this complete-information economy with f_2 and w_2.

Consider now an arbitrary SPC market and suppose the pairs (f_{n-1}, w_{n-1}) and (f_n, w_n) have the highest order (i.e., they constitute the last pairs matched in the recursive top-top matches' coupling). When their preferences over one another mimic those of $\{w_1, w_3\}$ and $\{f_1, f_3\}$ in our motivating example, we can replicate that construction by adding a pair similar to f_2 and w_2. In other cases, we construct examples analogous to our motivating example that allow the generalization.[17]

Our proof's construction is such that at most three workers need to misreport strategically in order to implement any BNE in the augmented economy. The proof is valid for

[17] With complete information, increasing competition on one side of the market cannot improve the matches of any participant on that market side and cannot harm the matches of participants on the other market side, see Roth and Sotomayor (1992). In the Online Appendix, we provide an incomplete-information example in which the addition of only one worker is beneficial for some workers.

any arbitrary probability of each state in the augmented economy. Furthermore, if worker w_n exhibits (non-justified) envy towards w_{n-1}, i.e. $u^w_{n-1,n} > u^w_{n,n}$, we can construct the augmented economy to entail no uncertainty regarding the unique stable matching. In that case, our construction also ensures that all unstable BNE outcomes ex-ante Pareto dominate the stable one for workers. Thus, unstable BNE outcomes can be more appealing for workers on efficiency grounds and require fairly limited strategizing relative to stable outcomes.

Naturally, one may wonder about the impact of such equilibrium multiplicity on overall outcomes of participants. Particularly when markets are large, as in many applications of DA, whether or not a substantial fraction of participants is affected would have important implications. We now show that, indeed, the addition of a small number of market participants can alter outcomes dramatically.

We start with an example of an $n \times n$ market satisfying the SPC, for which the addition of one firm and one worker allows for a BNE that substantially improves *all* workers' outcomes—in each state, the average rank increase for workers is roughly $n/2$.

Example 2. Consider a market with n firms $\{f_1, \ldots, f_n\}$ and n workers $\{w_1, \ldots, w_n\}$, with the following preferences. For all $i \in [n]$:

$$f_i : w_i \succ w_{i+1} \succ \ldots \succ w_n \succ w_{i-1} \succ w_{i-2} \succ \ldots \succ w_1;$$
$$w_i : f_{i-1} \succ f_{i-2} \succ \ldots \succ f_1 \succ f_i \succ f_{i+1} \succ \ldots \succ f_n.$$

This market exhibits no preference cycles and has a unique stable matching that matches w_i with f_i for all $i \in [n]$.

Consider now an augmentation with firm f and worker w and corresponding to two equally likely states, $\Theta = \{1, 2\}$.

Choose firm f's preferences as state-independent:

$$f : w \succ w_n \succ w_1 \succ w_2, w_3, \ldots, w_{n-1}.$$

Extend firm f_{n-1}'s preference in a state-dependent fashion. In state 1, the firm ranks w between w_{n-1} and w_n, while in state 2, she ranks w between w_n and w_{n-2}. All other firms retain state-independent preferences and rank w as their least favorite.

As for workers, worker w has the following preferences:

$$w : f_{n-1} \succ f \succ f_n \succ f_1,\, f_2,\, \ldots,\, f_{n-2}$$

and each of the original workers w_i ranks f right above f_i (for $i > 1$, below f_1).

Set w's utility from f_{n-1} to be sufficiently high so that his expected utility from matching with f_{n-1} in state 1 and f_n in state 2 is higher than his utility from matching with f in both states. Set w_n's utility from f to be sufficiently close to his utility from f_n, so that his expected utility from matching with f_{n-1} in state 1 and f_n in state 2 is higher than his utility from matching with f in both states. All unspecified utilities can be set arbitrarily.

In this economy, both states have the same unique stable matching μ such that, as before, $\mu(f_i) = w_i$ for any $i \in [n]$, and $\mu(f) = w$.

Consider the strategy profile such that workers w and w_n both drop f, and everyone else reports truthfully. It generates unstable outcomes $\lambda(1), \lambda(2) \neq \mu$ in the respective states such that (1) $\lambda(1)(w_1) = \lambda(2)(w_1) = f$; (2) for any $i \neq 1, n$, $\lambda(1)(w_i) = \lambda(2)(w_i) = f_{i-1}$; (3) $\lambda(1)(w) = \lambda(2)(w_n) = f_{n-1}$; and (4) $\lambda(2)(w) = \lambda(1)(w_n) = f_n$. These outcomes constitute unique stable matchings for the reported preferences.

The candidate profile constitutes a BNE in weakly undominated strategies. Indeed, worker w cannot get his most preferred f_{n-1} in state 2. In order to get his second-most preferred f in state 2, w needs to report f as acceptable. However, such a deviation precludes him from getting his most preferred f_{n-1} in state 1. By construction, it is not profitable.

Similarly, worker w_n cannot get firms $\{f_j\}_{j\in[n-2]}$ in state 1. In order to get f in state 1, w_n needs to report f to be more desirable than f_n. However, such a deviation precludes him from getting his most preferred f_{n-1} in state 2 and is thus unprofitable.

Finally, all other workers have no incentives to deviate since the generated matchings are unique stable for the reported preferences.

In the augmented economy, in either state, the average rank difference for workers is approximately $n/2$ when $\lambda(\theta)$ and μ are compared. $\triangle$

We now restrict attention to particular economies, those in which preferences are assortative on both sides of the market. While these preferences are certainly a polar extreme, they allow us to generalize the example above and illustrate the substantial impacts even a "small" infusion of uncertainty may have.

Proposition 4. *Consider any market $\mathcal{M}$ with n firms and n workers having assortative preferences. Let $k \leq n-2$. There is an economy $\mathcal{E}$ augmented with k firms and k workers such that the following holds. There is a BNE in weakly undominated dropping strategies that matches, identically in both states, $(n-k-1)$ original workers to original firms. According to the original preferences, each of these workers' matched firm has a rank lower by k relative to their stable partner.*

The proposition implies that, for any $\epsilon > 0$, when the economy is large enough, we can add a small volume of agents, accounting for an ϵ fraction of the economy's population, such that the resulting economy exhibits equilibrium multiplicity. In at least one of the "unstable" equilibria, a large fraction of the original workers, roughly $(1-\epsilon)$ of them, experiences a substantial improvement in the ranking of their matched partner. In fact, our construction guarantees that *all* workers in the augmented economy get higher expected payoffs in this equilibrium relative to the one implementing the stable outcome in each state.[18]

As mentioned in Section 2, our results rely on the uninformed workers being on the receiving side of the DA mechanism. Why wouldn't a market designer simply use the worker-proposing DA instead and guarantee stable outcomes? Importantly, if the designer uses the worker-proposing DA, a mirror image of our Example 1—and our entire analysis—in which

[18] As before, the proof also implies that, under the identified BNE, reported preferences are associated with a unique stable matching in each state.

the labels of workers and firms are switched, would replicate the current message of the paper. In fact, one can straightforwardly generate richer economies in which, with some probability, it is the workers who are uninformed and, with some probability, it is the firms who are uninformed, that would yield our negative results for *either* version of DA. In the Online Appendix, we show that *no quantile stable mechanism* is a panacea.[19] A modification of our Example 1 demonstrates that any such stable mechanism would yield multiple equilibrium outcomes in weakly undominated strategies.

Appendix – Proofs

Lemma 1. *Consider any economy $\mathcal{E}$. If firm f and worker w form a top-top match pair in state θ, they must be matched in this state under any BNE.*

Proof. Since firms report truthfully, in state θ, firm f reports worker w as her top choice. If they are not matched, then w's top-ranked firm is not his first true choice f. Therefore, w could profitably deviate by shifting f to the top of his ranking to strictly benefit in state θ without losing in other states. ∎

Proof of Proposition 1. We already discussed (1) in the text. Suppose (2) holds, the SPC* is satisfied.

Consider first any *two-state* economy with sequential top-top pairs $\{(f_{i|\theta}, w_{i|\theta})\}_{i\leq \min(m,n)}$ in state $\theta \in \{1,2\}$, defined in accordance with the SPC*. Since these two sequences correspond to unique stable matchings in the respective states, it is sufficient to prove that all top-top pairs must be matched under any BNE.

We apply induction on $k \leq \min(m,n)$, where k is the number of the first pairs $\{(f_{i|\theta}, w_{i|\theta})\}_{i\leq k}$ in the sequences above. The assertion holds for $k = 1$. Indeed, by Lemma 1, top-top matches must be matched under any BNE. Suppose that the assertion holds for $k \geq 1$. By symmetry

[19] Quantile stable mechanisms, which include DA mechanisms, guarantee that, for reported preferences, each participant receives a stable partner in the same quantile. See Teo and Sethuraman (1998) and references that follow.

with respect to states, it suffices to prove that $(f_{k+1|1}, w_{k+1|1})$ must be matched to each other in state 1 under any BNE.

Suppose, towards a contradiction, that $(f_{k+1|1}, w_{k+1|1})$ are not matched in state 1 under some equilibrium yielding matching $\lambda(\theta)$ in state $\theta \in \{1, 2\}$. Then, both $f_{k+1|1}$ and $w_{k+1|1}$ get less desirable assignments in state 1 under $\lambda(1)$. Because $f_{k+1|1}$ reports truthfully, this implies that worker $w_{k+1|1}$ reports the less desirable $\lambda(1)(w_{k+1|1})$ as preferable to $f_{k+1|1}$ in his equilibrium ranking Q. To reach a contradiction, it suffices to find a profitable deviation.

Construct a deviation Q' from Q by moving firms $\{f : f \succeq_{w_{k+1|1}} f_{k+1|1}\}$ to the top while keeping truthful ranking of these firms and Q's ranking of all remaining firms. Let $\lambda'(\theta)$ be the resulting stable matching in state $\theta \in \{1, 2\}$ for the reported preferences. This deviation delivers $\lambda'(1)(w_{k+1|1}) \succeq_{w_{k+1|1}} f_{k+1|1} \succ_{w_{k+1|1}} \lambda(1)(w_{k+1|1})$ to worker $w_{k+1|1}$ in state 1.

We now show that $\lambda'(2)(w_{k+1|1}) \succeq_{w_{k+1|1}} \lambda(2)(w_{k+1|1})$, i.e. worker $w_{k+1|1}$ cannot be worse off in state 2.

First, if $f_{k+1|1} \succeq_{w_{k+1|1}} \lambda(2)(w_{k+1|1})$, the proposed deviation Q' does not hurt worker $w_{k+1|1}$ in state 2, as desired.

Second, suppose that $\lambda(2)(w_{k+1|1}) \succ_{w_{k+1|1}} f_{k+1|1}$. Then, by the SPC*, for some $l \leq k$, $f_{l|2} = \lambda(2)(w_{k+1|1})$. Therefore, $w_{k+1|1} = w_{l|2}$ by the induction hypothesis.

Towards a contradiction, suppose that $f_{l|2} = \lambda(2)(w_{k+1|1}) \succ_{w_{k+1|1}} \lambda'(2)(w_{k+1|1})$ for truthful reports. By stability of $\lambda'(2)$ for the reported preferences, it must be the case that $w_{l_1|2} \coloneqq \lambda'(2)(f_{l|2}) \succ_{f_{l|2}} w_{l|2} = \lambda(2)(f_{l|2})$. Since $w_{l|2}$ is $f_{l|2}$'s favorite among all $\{w_{t|2}\}_{t \geq l}$, by our construction, we must have $l_1 < l$. By the induction hypothesis, $\lambda(2)(w_{l_1|2}) = f_{l_1|2}$.

Because $\lambda(2)$ is stable for the equilibrium reported preferences, under those, $f_{l_1|2} = \lambda(2)(w_{l_1|2}) \succ_{w_{l_1|2}} f_{l|2} = \lambda'(2)(w_{l_1|2})$. By stability of $\lambda'(2)$ for reported preferences, it then follows that $w_{l_2|2} \coloneqq \lambda'(2)(f_{l_1|2}) \succ_{f_{l_1|2}} w_{l_1|2} = \lambda(2)(f_{l_1|2})$. Again, by our construction, $l_2 < l_1$ and, therefore, by the induction hypothesis, $\lambda(2)(w_{l_2|2}) = f_{l_2|2}$.

We can continue recursively until we reach $l_T = 1$ for some $T \geq 1$, so that we have $w_{1|2} = \lambda'(2)(f_{l_{T-1}|2}) \succ_{f_{l_{T-1}|2}} w_{l_{T-1}|2} = \lambda(2)(f_{l_{T-1}|2})$ and $\lambda(2)(w_{1|2}) = f_{1|2}$ as above. Since

$\lambda(2)$ is stable for the equilibrium reported preferences, under those, $f_{1|2} = \lambda(2)(w_{1|2}) \succ_{w_{1|2}} f_{l_{T-1}|2} = \lambda'(2)(w_{1|2})$. Also, by our construction, $w_{1|2}$ is $f_{1|2}$'s favorite among all workers, and hence $w_{1|2} \succ_{f_{1|2}} \lambda'(2)(f_{1|2})$. This contradicts the stability of $\lambda'(2)$ under the reported preferences. Therefore, we must have $\lambda'(2)(w_{k+1|1}) \succeq_{w_{k+1|1}} \lambda(2)(w_{k+1|1})$, as desired.

Finally, for economies with more than two states, an identical proof works. Indeed, the same deviation is strictly beneficial in state 1 and weakly beneficial in any state $\theta \neq 1$. ∎

Proof of Proposition 4. Take any balanced market $\mathcal{M}$ such that, up to relabeling, all firms agree on $w_1 \succ w_2 \succ \ldots \succ w_n$ and all workers agree on $f_1 \succ f_2 \succ \ldots \succ f_n$. For any given non-degenerate distribution Ψ over the two states, $\Theta = \{1, 2\}$, consider this market augmented with firms $\{\mathbb{f}_i\}_{i\in[k]}$ and workers $\{\mathbb{w}_i\}_{i\in[k]}$.

For the new firms, choose state-independent preferences:

$$\mathbb{f}_1 : \mathbb{w}_1 \succ w_n \succ w_1 \succ \text{rest},$$

$$\mathbb{f}_i,\ i \neq 1 : \mathbb{w}_i \succ w_i \succ \text{rest}.$$

Extend firm f_{n-1}'s preference in a state-dependent way. In state 1, the firm ranks $\mathbb{w}_1$ between w_{n-1} and w_n and views other new workers as least desirable. In state 2, she considers all new workers as least preferred, with $\mathbb{w}_1$ being the most desirable among them.

Other firms retain state-independent preferences. Each firm f_i, $i \leq n-k-1$, ranks all new workers as least favorite. Any firm f_i, $n-k \leq i \leq n-2$, ranks $\mathbb{w}_{n-i}$ between w_{n-1} and w_n and views other new workers as least desirable. The remaining firm f_n views all new workers as least preferred, with $\mathbb{w}_1$ being the most preferred among them.

As concern workers, added ones have:

$$\mathbb{w}_1 : f_{n-1} \succ \mathbb{f}_1 \succ f_n \succ \text{rest},$$

$$\mathbb{w}_i,\ i \neq 1 : f_{n-i} \succ \mathbb{f}_i \succ \text{rest}.$$

Any worker w_i, $i \leq k$, ranks all new firms as most preferred, with $\mathbb{f}_i$ being the most desirable among them. Each worker w_i, $k+1 \leq i \leq n-1$, ranks all new firms right above f_i (and below f_{i-1}). Finally, worker w_n ranks $\mathbb{f}_1$ between f_{n-1} and f_n and ranks other new firms as

least preferred.

Set $\mathtt{w}_1$'s utility from f_{n-1} to be sufficiently high so that his expected utility from matching with f_{n-1} in state 1 and f_n in state 2 is higher than his utility from matching with $\mathtt{f}_1$ in both states. Set w_n's utility from $\mathtt{f}_1$ to be sufficiently close to his utility from f_n, so that his expected utility from matching with f_{n-1} in state 1 and f_n in state 2 is higher than his utility from matching with $\mathtt{f}_1$ in both states. All unspecified utilities can be set arbitrarily.

In this economy, both states have the same unique stable matching μ such that $\mu(f_i) = w_i$ for any $i \in [n]$, and $\mu(\mathtt{f}_i) = \mathtt{w}_i$ for any $i \in [k]$.

Consider the strategy profile in which workers $\mathtt{w}_1$ and w_n both drop $\mathtt{f}_1$, worker w_{n-1} drops $\{f_{n-k}, f_{n-k+1}, \ldots, f_{n-2}\}$, and everyone else reports truthfully. It generates unstable outcomes $\lambda(1), \lambda(2) \neq \mu$ in the respective states such that (1) for any $i \neq 1$, $\lambda(1)(\mathtt{w}_i) = \lambda(2)(\mathtt{w}_i) = f_{n-i}$; (2) for any $i \in [k]$, $\lambda(1)(w_i) = \lambda(2)(w_i) = \mathtt{f}_i$; (3) for any $i \in \{k+1, k+2, \ldots, n-1\}$, $\lambda(1)(w_i) = \lambda(2)(w_i) = f_{i-k}$; (4) $\lambda(1)(\mathtt{w}_1) = \lambda(2)(w_n) = f_{n-1}$; and (5) $\lambda(2)(\mathtt{w}_1) = \lambda(1)(w_n) = f_n$. These outcomes constitute unique stable matchings for the reported preferences.

The candidate profile constitutes a BNE in weakly undominated strategies. Indeed, worker $\mathtt{w}_1$ cannot get his most preferred f_{n-1} in state 2. In order to get his second most preferred $\mathtt{f}_1$ in state 2, w needs to report $\mathtt{f}_1$ as acceptable. However, such a deviation precludes him from getting his most preferred f_{n-1} in state 1. By construction, it is unprofitable.

Similarly, worker w_n cannot get firms $\{f_1, f_2, \ldots, f_{n-2}\}$ in either state. Also, he cannot get firm f_{n-1} in state 1. In order to get $\mathtt{f}_1$ in state 1, w_n needs to report $\mathtt{f}_1$ as more desirable than f_n. However, such a deviation precludes him from getting his most preferred f_{n-1} among "achievable" firms in state 2 and is thus unprofitable.

Worker w_{n-1} cannot get firms $\{f_1, f_2, \ldots, f_{n-k-2}\}$ in either state. Consequently, he gets his most preferred f_{n-k-1} among "achievable" firms and, hence, has no incentives to deviate.

Finally, other workers have no incentives to deviate since the generated matchings are unique stable for the reported preferences. ∎

References

Agarwal, Nikhil, "An Empirical Model of the Medical Match," *American Economic Review*, July 2015, *105* (7), 1939–1978.

Ashlagi, Itai, Yash Kanoria, and Jacob D Leshno, "Unbalanced Random Matching Markets: The Stark Effect of Competition," *Journal of Political Economy*, 2017, *125* (1), 69–98.

Bikhchandani, Sushil, "Stability with One-Sided Incomplete Information," *Journal of Economic Theory*, March 2017, *168*, 372–399.

Chakraborty, Archishman, Alessandro Citanna, and Michael Ostrovsky, "Two-Sided Matching with Interdependent Values," *Journal of Economic Theory*, January 2010, *145* (1), 85–105.

Chiappori, Pierre-André and Bernard Salanié, "The Econometrics of Matching Models," *Journal of Economic Literature*, September 2016, *54* (3), 832–861.

Clark, Simon, "The Uniqueness of Stable Matchings," *Contributions in Theoretical Economics*, January 2006, *6* (1), 1–28.

Coles, Peter and Ran Shorrer, "Optimal Truncation in Matching Markets," *Games and Economic Behavior*, 2014, *87*, 591–615.

Eeckhout, Jan, "On the Uniqueness of Stable Marriage Matchings," *Economics Letters*, 2000, *69* (1), 1–8.

Ehlers, Lars and Jordi Massó, "Incomplete Information and Singleton Cores in Matching Markets," *Journal of Economic Theory*, September 2007, *136* (1), 587–600.

Ferdowsian, Andrew, Muriel Niederle, and Leeat Yariv, "Decentralized Matching with Aligned Preferences," 2021. Unpublished.

Fernandez, Marcelo Ariel, "Deferred Acceptance and Regret-Free Truth-telling," 2020. Unpublished.

Gale, D. and L. S. Shapley, "College Admissions and the Stability of Marriage," *The American Mathematical Monthly*, January 1962, *69* (1), 9.

Holzman, Ron and Dov Samet, "Matching of Like Rank and the Size of the Core in the Marriage Problem," *Games and Economic Behavior*, November 2014, *88*, 277–285.

Hsieh, Yu-Wei, "Understanding Mate Preferences from Two-Sided Matching Markets: Identification, Estimation and Policy Analysis," 2012. Unpublished.

Immorlica, Nicole and Mohammad Mahdian, "Marriage, Honesty, and Stability," in "Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms" Society for Industrial and Applied Mathematics 1 2005, pp. 53–62.

_ **, Jacob Leshno, Irene Lo, and Brendan Lucier**, "Information Acquisition in Matching Markets: The Role of Price Discovery," 2020. Unpublished.

Legros, Patrick and Andrew Newman, "Co-Ranking Mates: Assortative Matching in Marriage Markets," *Economics Letters*, March 2010, *106* (3), 177–179.

Liu, Qingmin, "Stability and Bayesian Consistency in Two-Sided Markets," *American Economic Review*, August 2020, *110* (8), 2625–2666.

_ **, George J. Mailath, Andrew Postlewaite, and Larry Samuelson**, "Stable Matching with Incomplete Information," *Econometrica*, 2014, *82* (2), 541–587.

McKinney, C. Nicholas, Muriel Niederle, and Alvin E. Roth, "The Collapse of a Medical Labor Clearinghouse (and Why Such Failures Are Rare)," *American Economic Review*, May 2005, *95* (3), 878–889.

McVitie, D. G. and L. B. Wilson, "Stable Marriage Assignment for Unequal Sets," *BIT*, September 1970, *10* (3), 295–309.

Reny, Philip J., "A Simple Sufficient Condition for a Unique and Student-Efficient Stable Matching in the College Admissions Problem," *Economic Theory Bulletin*, January 2021.

Roth, Alvin E., "Two-Sided Matching with Incomplete Information about Others' Preferences," *Games and Economic Behavior*, June 1989, *1* (2), 191–209.

_ , "The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics," *Econometrica*, 2002, *70* (4), 1341–1378.

_ **and John H. Vande Vate**, "Incentives in Two-Sided Matching with Random Stable Mechanisms," *Economic Theory*, 1991, *1* (1), 31–44.

_ **and Marilda A. Oliveira Sotomayor**, *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis* Econometric Society Monographs, Cambridge: Cambridge University Press, 1992.

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

_ **and Xiaolin Xing**, "Jumping the Gun: Imperfections and Institutions Related to the Timing of Market Transactions," *American Economic Review*, 1994, *84* (4), 992–1044.

Teo, Chung-Piaw and Jay Sethuraman, "The Geometry of Fractional Stable Matchings and Its Applications," *Mathematics of Operations Research*, 1998, *23* (4), 874–891.

Van der Linden, Martin, "Deferred Acceptance is Minimally Manipulable," *International Journal of Game Theory*, 2019, *48* (2), 609–645.