Not all is lost: sorting and self-stabilizing sets

Esteban Peralta*

March 4, 2022

Summary

It is well known that in the absence of transfers, stable matchings are positive assortative when agents' preferences are strictly monotonic in each other's types and types are commonly known. Instead, monotonicity is consistent with the existence of stable matchings that exhibit negative sorting when types on one side of the market are private information. This paper sheds light on the scope and meaning of this consistency by showing that in most monotonic markets it cannot be concluded that a stable matching is negative, and not positive, assortative.

Keywords: incomplete information, sorting, stable matching.

*Department of Economics, University of Michigan, 611 Tappan Ave. (48109), Ann Arbor, MI, USA (eperalta@umich.edu). I am grateful to Ana Reynoso and Larry Samuelson for many insightful discussions of these topics. Financial support from the University of Michigan is greatly appreciated. The usual disclaimer applies.

1 Introduction

In matching markets without transfers in which agents' preferences are strictly increasing in the attribute or type of those on the other side *and* types are commonly known, stable matchings are positive assortative (see, e.g., Becker (1973)). This well-known relationship between stability and sorting has been important for applied work because it holds regardless of the vectors of agents' types.

Interestingly, however, Bikhchandani (2017) analyzes markets in which the type of workers is private information and shows by example that even with increasing preferences, there can be (incomplete-information) stable matchings that are negative, and not positive, assortative. This paper sheds light on the interpretation, and scope, of these examples.

The notion of incomplete-information stability studied by Bikhchandani (2017) adapts the one introduced by Liu *et al.* (2014) to markets without transferable utilities, and so presumes that firms use their information "cautiously" when evaluating any blocking opportunity. That is, given a vector of workers' types, firms are informed of the type of their own worker and are *not* willing to participate in a block with another worker if they can find a vector of workers' types consistent with their information at which both the type of the blocking worker is lower than the type of their own worker, *and* no other firm is willing to participate in a block. Thus, a matching is incomplete-information stable for a given vector of workers' types if and only if it can be *stabilized* (Liu *et al.*, 2014); namely, if and only if the vector belongs to a *set* with the property that for every vector in the set that prescribes a blocking opportunity, there is another vector in the set, consistent with the blocking firm's information, at which such opportunity disappears.

Given the permissive nature of incomplete-information stability, the negative, and not positive, assorted nature of some incomplete-information stable matching *at a given* vector of workers' types should come as no surprise, even within monotonic environments. On the contrary, the element of surprise should come, this paper argues, from the following observation: If *only one* firm has the *lowest-matched* type, then *no* stabilizing set can prescribe negative, and not positive, sorting at *all* of its members.

To be clear, the result *does not* say that every set of vectors that stabilizes a given non-empty matching satisfying the assumed unitary bound contains some vector at which the matching delivers positive sorting. Instead, it says that *every* stabilizing set containing *no* vector at which the matching delivers positive sorting *must* contain a vector at which the sorting is mixed.[1] Moreover, the result does not only concern the entire set of vectors at which a given matching is incomplete-information stable, the largest stabilizing set, but in fact *all* of its stabilizing subsets as well.

Since the only assumption in place concerns the *first-order* statistic of the distribution of *matched* firms' types, the observation above suggests that the presence of incomplete information cannot completely overturn the well-known relationship between stability and positive sorting, in most monotonic markets. What is more, the very permissive nature of incomplete-information stability entails that this is true pretty much regardless of the incomplete-information stability notion one looks at.

This paper contributes to the fast-growing literature studying stable

[1] It turns out that stabilizing sets that rule out positive assortativeness can either prescribe mixed sorting at all of its members, or a combination of negative and mixed sorting. See Section 5 for examples.

outcomes in the presence of incomplete information, but suggests that focusing attention to properties of a single vector of workers' types, rather than to those of self-stabilizing sets, might not offer a complete picture. This suggestion seems consistent with the intended interpretation of incomplete-information stability, as no outside observer who—knows that agents' preferences are increasing and—observes firms' types, and no blocking taking place, would be able to conclude that the matching she observes is negative, and not positive, assortative, *regardless of the underlying stabilizing set.*

The rest of the paper is organized as follows. Section 2 describes the environment and Section 3 the notions of complete- and incomplete-information stability. Section 4 contains the main result of the paper, and Section 5 offers some examples and discusses the related literature.

2 The Environment

Let $J = \{1, ..., |J|\} \subseteq \mathbb{N}$ and $I = \{1, ..., |I|\} \subseteq \mathbb{N}$ be the finite sets of firms and workers in a one-to-one labor market without transfers. I will write $i \in I$ for an individual worker and $j \in J$ for an individual firm. The finite sets of workers and firms are, respectively, $W := \{1, ..., K\} \subseteq \mathbb{N}$ and $F := \{1, ..., L\} \subseteq \mathbb{N}$. I denote by $\mathbf{w} \in W^{|I|}$ a vector of workers' types and by $\mathbf{f} \in F^{|J|}$ a vector of firms' types, writing $\mathbf{w}_i$ and $\mathbf{f}_j$ to denote, respectively, the type of worker i and firm j.

I let $u_i(\mathbf{w}_i, \mathbf{f}_j)$ denote worker i's utility whenever she is of type $\mathbf{w}_i$ and is matched to firm j of type $\mathbf{f}_j$. Similarly, $v_j(\mathbf{w}_i, \mathbf{f}_j)$ represents firm j's utility when it is of type $\mathbf{f}_j$ and is matched to a worker of type $\mathbf{w}_i$. The dependence of u_i on $\mathbf{f}_j$ and of v_j on $\mathbf{w}_i$ captures the idea that types describe unobservable characteristics, such as productivity.[2]

2.1 Matchings and outcomes

A **matching** is a one-to-one map $\mu : I \cup J \to I \cup J$, where $\mu(i)$ denotes the firm matched to worker i, $\mu(j)$ denotes the worker matched to firm j, and $\mu(i) = j$ if and only if $\mu(j) = i$. I write μ_i and μ_j instead of $\mu(i)$ and $\mu(j)$. Moreover, I write $\mu_i = \emptyset$ instead of $\mu_i = i$ whenever i is unmatched (and similarly for any j). To deal with unmatched agents, I introduce a "dummy" type 0 and let $\mathbf{f}_{\mu_i} = 0 = \mathbf{w}_{\mu_j}$ whenever $\mu_i = \mu_j = \emptyset$.[3] Without loss, I assume that $u_i(\mathbf{w}_i, 0) = v_j(0, \mathbf{f}_j) = 0$ for every i, every j, every $\mathbf{w}$, and every $\mathbf{f}$.

An **outcome** is a triplet $(\mu, \mathbf{w}, \mathbf{f})$. A matching μ is **nonempty** if there exists i with $\mu_i \neq \emptyset$. An outcome $(\mu, \mathbf{w}, \mathbf{f})$ is nonempty if μ is nonempty.

2.2 Information

Firms' types are assumed to be commonly known among workers and firms, but workers have private information about their types.

To capture firms' uncertainty about workers' types, I follow Liu *et al.* (2014) and assume that the true vector of workers' types is drawn from some distribution with support $\Omega \subseteq W^{|I|}$. As in Liu *et al.* (2014), the underlying distribution will play no role.

[2] More generally, utilities might depend on the "identity" of the assignment; namely, on the indices i and j. This dependence would model situations where some observable characteristics, encapsulated by each agent's index, are of value to firms and workers.

[3] Since $0 < 1$, this ensures that this dummy type is strictly lower than every possible type of any firm and worker.

3 Stability

3.1 Complete-information stability

An outcome $(\mu, \mathbf{w}, \mathbf{f})$ is **individually rational** if $u_i(\mathbf{w}_i, \mathbf{f}_{\mu_i}) \geq 0$ for every i and $v_j(\mathbf{w}_{\mu_j}, \mathbf{f}_j) \geq 0$ for every j. Let $\sum^0$ denote the set of individually rational outcomes.

An outcome $(\mu, \mathbf{w}, \mathbf{f}) \in \sum^0$ is **complete-information blocked** by a pair (i, j) if $u_i(\mathbf{w}_i, \mathbf{f}_j) > u_i(\mathbf{w}_i, \mathbf{f}_{\mu_i})$ and $v_j(\mathbf{w}_i, \mathbf{f}_j) > v_j(\mathbf{w}_{\mu_j}, \mathbf{f}_j)$.

An outcome $(\mu, \mathbf{w}, \mathbf{f})$ is **complete-information stable** if $(\mu, \mathbf{w}, \mathbf{f}) \in \sum^0$ and $(\mu, \mathbf{w}, \mathbf{f})$ is not complete-information blocked.

3.2 Incomplete-information stability

The following blocking concept is proposed by Bikhchandani (2017), and adapts the one introduced by Liu *et al.* (2014) to markets without transfers.

Definition 1. Fix any nonempty set $X \subseteq \sum^0$. An outcome $(\mu, \mathbf{w}, \mathbf{f}) \in X$ is *X-blocked* if there is a pair (i, j) such that:

1. $u_i(\mathbf{w}_i, \mathbf{f}_j) > u_i(\mathbf{w}_i, \mathbf{f}_{\mu_i})$; *and*

2. *for every* $\mathbf{w}' \in \Omega$ *with* $(\mu, \mathbf{w}', \mathbf{f}) \in X$ *such that* $u_i(\mathbf{w}'_i, \mathbf{f}_j) > u_i(\mathbf{w}'_i, \mathbf{f}_{\mu_i})$ *and* $\mathbf{w}'_{\mu_j} = \mathbf{w}_{\mu_j}$*, we have*

$$v_j(\mathbf{w}'_i, \mathbf{f}_j) > v_j(\mathbf{w}_{\mu_j}, \mathbf{f}_j). \tag{1}$$

An outcome $(\mu, \mathbf{w}, \mathbf{f}) \in X$ *is X-stable if it is not X-blocked.*

To get an intuitive grasp, let $X = \sum^0$. Then, a firm forms a $\sum^0$-block with a given worker if and only if they form a complete-information block *after* the firm accounts for the individually rational nature of the matching, the type of its own worker, and the worker's willingness to participate.

Define, for every $k \geq 1$:

$$\sum^k := \{(\mu, \mathbf{w}, \mathbf{f}) : (\mu, \mathbf{w}, \mathbf{f}) \in \sum^{k-1} \text{ and } (\mu, \mathbf{w}, \mathbf{f}) \text{ is not } \sum^{k-1}\text{-blocked}\}.$$

$\sum := \bigcap_{k \geq 0} \sum^k$ describes the set of outcomes that are incomplete-information stable in the sense of Liu *et al.* (2014) when transfers are not available.

Notice that the set of complete-information stable outcomes is contained in the set of incomplete-information stable ones, for any given vector of workers' types. From this, Bikhchandani (2017) concludes that, no mater the vector of agents' types that is drawn, an incomplete-information stable matching exists; i.e., $\sum(\mathbf{w}, \mathbf{f}) \neq \emptyset$ for every $(\mathbf{w}, \mathbf{f})$.[4]

3.3 Self-stabilizing sets

Liu *et al.* (2014) offer an equivalent way to describe the set of incomplete-information stable outcomes in terms of what they call *self-stabilizing sets.*

Definition 2. *A nonempty set of individually rational outcomes E is* self-stabilizing *if every outcome $(\mu, \mathbf{w}, \mathbf{f}) \in E$ is E-stable. The set E* stabilizes *a given outcome $(\mu, \mathbf{w}, \mathbf{f})$ if $(\mu, \mathbf{w}, \mathbf{f}) \in E$ and E is self-stabilizing. A set of vectors of workers' types $\Omega^* \subseteq \Omega$* stabilizes a matching *μ if $\{(\mu, \mathbf{w}', \mathbf{f}) : \mathbf{w}' \in \Omega^*)\}$ is a self-stabilizing set.*[5]

Intuitively, a set is self-stabilizing if for every outcome in the set that is complete-information blocked by a pair (i, j), one can find another out-

[4] This result follows directly from Proposition 1 in Liu *et al.* (2014).

[5] Notice that this definition fixes a (commonly known) vector of firms' types.

come *in the set*, consistent with j's information, at which the block is not profitable to j.

Following Proposition 2 in Liu *et al.* (2014), Bikhchandani (2017) shows that every self-stabilizing set is a subset of $\sum$. Hence, $\sum$ is the largest self-stabilizing set. This result is quite useful, because it provides an operative characterization of incomplete-information stability; namely, an outcome is incomplete-information stable if and only if it is a member of a self-stabilizing set. It follows that that a given matching μ is part of an incomplete-information stable outcome, for a given vector of firms' types, if and only if one can find a vector of workers' types such that the resulting set of outcomes is self-stabilizing. This latter characterization will be used below.

4 Main result

The following three subsections describe the class of environments and outcomes the main result of this paper will focus on. The last subsection presents the main result.

4.1 Increasing utilities

I will restrict attention to monotonic environments; namely, those in which agents' utilities increase, strictly, with the type of their match:

Definition 3. *Agents' utilities are* ***increasing*** *if* $u_i(\mathbf{w}_i, \mathbf{f}_j)$ *is strictly increasing in* $\mathbf{f}_j$ *for every* i *and every* $\mathbf{w}_i$ *and* $v_j(\mathbf{w}_i, \mathbf{f}_j)$ *is strictly increasing in* $\mathbf{w}_i$ *for every* j *and every* $\mathbf{f}_j$*.*

Notice that Definition 3 does *not* require utilities to increase, even weakly, with respect to an agent's own type.

4.2 Assortative outcomes

The following definition is standard:

Definition 4. *An outcome* $(\mu, \mathbf{w}, \mathbf{f})$ *is positive assortative (PAM) if, for every* i, i' *such that* $\mu_i \neq \emptyset$ *and* $\mu_{i'} \neq \emptyset$*, we have* $\mathbf{w}_i > \mathbf{w}_{i'} \Rightarrow \mathbf{f}_{\mu_i} \geq \mathbf{f}_{\mu_{i'}}$.

Negative assortativeness (NAM) can be defined analogously, replacing $\mathbf{f}_{\mu_i} \geq \mathbf{f}_{\mu_{i'}}$ by $\mathbf{f}_{\mu_i} \leq \mathbf{f}_{\mu_{i'}}$.[6] I will write $\mathcal{P}(\mu, \mathbf{f})$ and $\mathcal{N}(\mu, \mathbf{f})$ to denote the set of vectors of workers' types at which μ is positive and negative assortative, respectively. Two comments are in order.

First, notice that $\mathcal{P}(\mu, \mathbf{f}) \cap \mathcal{N}(\mu, \mathbf{f}) \neq \emptyset$ for every μ, as every nonempty matching is both positive and negative assortative at every vector of agents' types at which either every matched worker or every matched firm has the same type.[7] Second, complete-information stable outcomes are positive assortative in monotonic environments.

4.3 Lowest-matched-firm-type outcomes

For any pair $(\mu, \mathbf{f})$ where μ is nonempty, let $\underline{J}(\mu, \mathbf{f})$ denote the set of firms holding the lowest type in $\mathbf{f}$, among those that are matched in μ; i.e.:

$$\underline{J}(\mu, \mathbf{f}) := \{j : \mathbf{f}_j = min\{\mathbf{f}_{\tilde{j}} : \mu_{\tilde{j}} \neq \emptyset\}\}.$$

We can then define:

[6] Notice that these definitions of PAM and NAM account only for matched agents. This is why asking for "every j, j' such that $\mu_j \neq \emptyset$ and $\mu_{j'} \neq \emptyset$, $\mathbf{f}_j > \mathbf{f}_{j'} \Rightarrow \mathbf{w}_{\mu_j} \geq \mathbf{w}_{\mu_{j'}}$" in PAM is redundant.

[7] Notice that empty outcomes are both PAM and NAM since, by convention, $\mathbf{f}_{\mu_i} = 0$ whenever $\mu_i = \emptyset$.

Definition 5. *A pair* $(\mu, \mathbf{f})$ *where* μ *is nonempty satisfies the lowest-matched-firm-type (LMFT) property if* $\underline{J}(\mu, \mathbf{f})$ *is a singleton.*

In words, LMFT requires that there is only one matched firm holding the *lowest-matched* type; i.e., the lowest type in $\mathbf{f}$, among all matched firms. Notice that LMFT is a joint condition on μ and $\mathbf{f}$, both commonly known among workers and firms, but does not depend on the vector of workers' types.

I view LMFT as a weak requirement, as it concerns only the "left-end tail", or first-order statistic, of the distribution of firms' types, among those that are matched. In particular, LMFT is weaker than requiring that there are no two firms with the lowest type in $\mathbf{f}$, as it concerns only the set of matched firms, and much weaker than requiring different firms to be of different types.[8]

4.4 Sorting and stabilizing sets

The following result constitutes the main finding of the paper:

Proposition 1. *Suppose that utilities are increasing, and fix any pair* $(\mu, \mathbf{f})$ *and any set* $\Omega^* \subseteq \Omega$ *that stabilizes* μ*. If* $(\mu, \mathbf{f})$ *satisfies the LMFT property, then*

$$\Omega^* \subseteq \mathcal{N}(\mu, \mathbf{f}) \Rightarrow \Omega^* \cap \mathcal{P}(\mu, \mathbf{f}) \neq \emptyset.$$

The proof of Proposition 1 can be found in the appendix, but a rough intuition of its content goes as follows: the presence of increasing utilities implies that if a matching is negative, but not positive, assortative for a given vector of workers' types, then a complete-information block must be

[8] In fact, LMFT does not even rule out nonempty outcomes at which every firm has the same type.

formed by one of the highest-type matched firms and one of the highest-type matched workers. The assumed negative assortative nature of the matching at the given vector of workers' types implies that some of those workers is matched to a firm with the lowest-matched type, but by hypothesis only one such worker exists. If the vector of workers' types belongs to a set that stabilizes the matching, then the set must contain another vector at which the blocking firm does not form a block with the blocking worker. By monotonicity, the type of the blocking worker must be lower at this new vector. If the matching is once again negative, but not positive, assortative at the new vector, then another firm with the *same* type than the original blocking firm must form a block with the *same* blocking worker, and so the same argument can be repeated, lowering the type of the blocking worker even further. Eventually, one must reach a vector of workers' types *in the set* at which the matching delivers positive sorting. Whenever, on the other hand, two or more firms have the lowest-matched type, the "anchor" used to carry out this inductive argument does not exist, and so a "loop" in the sequence of vectors of workers' types required by incomplete-information stability can be created, with each vector in the sequence delivering negative, but not positive, sorting. This is illustrated by Example 1 below.

A few comments are in order. First, the result can be extended to positive and negative sorting of *all* agents, matched or not, when all types are acceptable to all agents and there is at most one unmatched worker.[9]

Second, the result does not imply that every stabilizing set must deliver positive sorting at some of its members. Indeed, Section 5 shows by

[9]Further, one can show that these two assumption are also "tight" in the sense that the result does not necessarily holds when either of them is relaxed.

example(s) that stabilizing sets can prescribe either mixed sorting at all of its members, or a combination of mixed and negative sorting.

Third, the result could prove useful to any outside observer who sees no blocking taking place, and knows that utilities are increasing. Indeed, both the matching and the vector of firms' types are observable, and so whether LMFT holds or not can be readily assessed. If it holds, then any such observer would be able to conclude that the matching she observes cannot be negative, and not positive, assortative at all vectors of workers' types in the underlying stabilizing set, *whatever this set happens to be.* At this point, one could argue: We sort of knew this. After all, stabilizing sets at which some of its members feature positive assortativeness are relatively easy to come by.[10] Thus, the argument would go, any outsider could "always" imagine a vector of workers' types at which the matching she observes is positive assortative. To this argument I would respond that such an outsider knows better: Self-stabilizing sets are commonly known among firms and workers (see Bikhchandani (2014)), and so readily available to her. As a consequence, she should restrict her inquiry to the members of the set she has access to. This is the whole point of the present paper; namely, unless the outsider elicits workers' private information, she should be interested in properties of stabilizing sets, not on those of isolated vectors of workers' types where the matching she observes happens to be incomplete-information stable.

The following example illustrates that Proposition 1 is not necessarily true when LMFT fails.

Example 1. *There are three workers and three firms,* $I = \{i_1, i_2, i_3\}$ *and*

[10] This is so because every complete-information stable outcome is incomplete-information stable, *and* positive assortative when utilities are increasing.

$J = \{j_1, j_2, j_3\}$. The utility of every i is given by $u_i(\mathbf{w}_i, \mathbf{f}_j) = \mathbf{f}_j$ and the utility of every j by $v_j(\mathbf{w}_i, \mathbf{f}_j) = \mathbf{w}_i$. Thus, utilities are increasing.[11] *Suppose that $\mathbf{f}_{j_1} = \mathbf{f}_{j_2} = 2$, and $\mathbf{f}_{j_3} = 3$, and consider the following two vectors of workers types:*

Worker indices:	i_1	i_2	i_3
Worker types, $\mathbf{w}$:	4	2	2
Firm types, $\mathbf{f}$:	2	2	3
Firm indices:	j_1	j_2	j_3

Worker indices:	i_1	i_2	i_3
Worker types, $\mathbf{w}'$:	2	4	2
Firm types, $\mathbf{f}$:	2	2	3
Firm indices:	j_1	j_2	j_3

Imagine the matching μ that assigns i_1 to j_1, i_2 to j_2 and i_3 to j_3. Clearly, LFMT fails since j_1 and j_2 are both matched and have the lowest type (2) among all matched firms. Further, μ is negative, but not positive, assortative at both $\mathbf{w}$ and $\mathbf{w}'$. Moreover, the set $E = \{(\mu, \mathbf{w}, \mathbf{f}), (\mu, \mathbf{w}', \mathbf{f})\}$ is self-stabilizing. To see this, notice that j_3 is the only firm with a complete-information block at both $\mathbf{w}$ and $\mathbf{w}'$, and it can use each of these vectors to justify its unwillingness to form the complete-information block at the other.

5 Discussion

5.1 Alternative sorting patterns

Proposition 1 shows that there is no stabilizing set prescribing negative, and not positive, sorting at all of its members. A natural follow-up question is whether some other sorting pattern is inconsistent with incomplete-information stability within monotonic environments. As it turns out, the answer is "no." In particular, the following two examples illustrate that one

[11] The particular utilities are irrelevant as long as they increasing.

can find stabilizing sets that prescribe either mixed assortativeness at all of its members, or a combination of mixed and negative, but not positive, sorting. Example 2 illustrates the former, and Example 3 the latter.

Example 2. *There are four workers and four firms,* $I = \{i_1, i_2, i_3, i_4\}$ *and* $J = \{j_1, j_2, j_3, j_4\}$. *The utility of every* i *is given by* $u_i(\mathbf{w}_i, \mathbf{f}_j) = \mathbf{f}_j$ *and the utility of every* j *by* $v_j(\mathbf{w}_i, \mathbf{f}_j) = \mathbf{w}_i$. *Thus, utilities are increasing. Suppose that* $\mathbf{f}_{j_1} = 1$, $\mathbf{f}_{j_2} = 2$, $\mathbf{f}_{j_3} = 3$, *and* $\mathbf{f}_{j_4} = 4$, *and consider the following two vectors of workers types:*

Worker indices:	i_1	i_2	i_3	i_4
Worker types, $\mathbf{w}$:	2	1	2	3
Firm types, $\mathbf{f}$:	1	2	3	4
Firm indices:	j_1	j_2	j_3	j_4

Worker indices:	i_1	i_2	i_3	i_4
Worker types, $\mathbf{w}'$:	1	1	4	3
Firm types, $\mathbf{f}$:	1	2	3	4
Firm indices:	j_1	j_2	j_3	j_4

Imagine the matching μ *that assigns* i_1 *to* j_1, i_2 *to* j_2, i_3 *to* j_3, *and* i_4 *to* j_4. *Notice that* μ *exhibits mixed sorting at both* $\mathbf{w}$ *and* $\mathbf{w}'$. *In particular,* (i_1, j_2) *is the only complete-information block at* $\mathbf{w}$, *and* (i_3, j_4) *the only complete-information block at* $\mathbf{w}'$. *Moreover,* j_2 *can appeal to* $\mathbf{w}'$ *to say "no" to* i_1, *and* j_4 *can appeal to* $\mathbf{w}$ *to say "no" to* i_3. *Thus, the set* $E = \{(\mu, \mathbf{w}, \mathbf{f}), (\mu, \mathbf{w}', \mathbf{f})\}$ *is self-stabilizing.*[12]

Example 3. *There are four workers and four firms,* $I = \{i_1, i_2, i_3, i_4\}$ *and* $J = \{j_1, j_2, j_3, j_4\}$. *The utility of every* i *is given by* $u_i(\mathbf{w}_i, \mathbf{f}_j) = \mathbf{f}_j$ *and the utility of every* j *by* $v_j(\mathbf{w}_i, \mathbf{f}_j) = \mathbf{w}_i$. *Thus, utilities are increasing. Suppose that* $\mathbf{f}_{j_1} = 1$, $\mathbf{f}_{j_2} = 2$, $\mathbf{f}_{j_3} = 3$, *and* $\mathbf{f}_{j_4} = 4$, *and consider the following three vectors of workers types:*

[12] Further, notice that LMFT holds in this example.

Worker indices:	i_1	i_2	i_3	i_4
Worker types, $\mathbf{w}$:	2	1	2	3
Firm types, $\mathbf{f}$:	1	2	3	4
Firm indices:	j_1	j_2	j_3	j_4

Worker indices:	i_1	i_2	i_3	i_4
Worker types, $\mathbf{w}'$:	1	1	4	3
Firm types, $\mathbf{f}$:	1	2	3	4
Firm indices:	j_1	j_2	j_3	j_4

Worker indices:	i_1	i_2	i_3	i_4
Worker types, $\mathbf{w}''$:	4	4	4	3
Firm types, $\mathbf{f}$:	1	2	3	4
Firm indices:	j_1	j_2	j_3	j_4

Imagine again the matching μ that assigns i_1 to j_1, i_2 to j_2, i_3 to j_3, and i_4 to j_4. Notice that μ exhibits mixed sorting at both $\mathbf{w}$ and $\mathbf{w}'$, but negative, and not positive, sorting at $\mathbf{w}''$. In particular, j_4 forms a complete-information block with i_1, i_2, and i_3 at $\mathbf{w}''$, and again (i_1, j_2) is the only complete-information block at $\mathbf{w}$, and (i_3, j_4) the only complete-information block at $\mathbf{w}'$. As before, j_2 can appeal to $\mathbf{w}'$ to say "no" to i_1, and j_4 can appeal to $\mathbf{w}$ to say "no" to i_3. Moreover, j_4 can appeal to $\mathbf{w}$ to say "no" to i_1, i_2, and i_3. Thus, the set $E = \{(\mu, \mathbf{w}, \mathbf{f}), (\mu, \mathbf{w}', \mathbf{f}), (\mu, \mathbf{w}'', \mathbf{f})\}$ is self-stabilizing.[13]

5.2 Literature

The earliest attempt to embed incomplete information in the theory of matching markets goes back to Roth (1989), who studies stable mechanisms in the presence of preference uncertainty and shows that some results that hold under complete information do not hold in the presence of incomplete information.[14] Unlike the private values model analyzed by

[13] Further, notice that LMFT holds in this example.

[14] Another earlier attempt is that of Ehlers & Massó (2007), who also analyze a model with preference uncertainty and examine when complete-information stable mechanisms satisfy an ordinal notion of incentive compatibility (see also Ehlers & Massó (2015)).

Roth (1989), Chakraborty *et al.* (2010) examines a one-sided incomplete-information model with interdependent values and shows that the existence of a stable—and strategy-proof—mechanism depends on whether the mechanism makes the matching public or not. Both of these papers analyze stability in centralized markets. In contrast, this paper contributes to the literature that studies stability in decentralized matching markets with incomplete information, and interdependent values.

Liu *et al.* (2014) analyze the same model studied in this paper, but allow for monetary transfers. Like the present paper, though, Liu *et al.* (2014) assume that the prior information of all firms is identical. Instead, Chen & Hu (2019) makes firms' information explicit by describing it as arbitrary partitions. They do so to study convergence to a notion of stability they propose that turns out to be equivalent to incomplete-information stability.[15] In markets with transfers, other important contributions are those by Alston (2020), Bikhchandani (2017), Jeong (2019), Liu (2020), and Pomatto (2018). Unlike Liu *et al.* (2014) and Chen & Hu (2019), these papers enlarge the description of the model to account, explicitly, for agents' beliefs.[16] In a recent working paper, Chen & Hu (2018) generalize the theory of stability in the presence of incomplete information to markets with two-sided uncertainty. They allow for arbitrary partitions, relax the assumption that firms know the type of their match, and analyze several stability concepts that accommodate the different inferences that a potential blocking worker-firm pair can make about each other's type from the

[15] Lazarova & Dimitrov (2017) also study markets with two-sided incomplete information and, like Chen & Hu (2019), address the question of "path to stability" for a stability concept that can be interpreted as capturing a best-case notion of blocking.

[16] Further, Liu (2020) and Pomatto (2018) study stability under incomplete information from a noncooperative viewpoint.

observation that both are willing to participate.

The present paper builds from Bikhchandani (2017), who study the notion of incomplete-information stability proposed by Liu *et al.* (2014) in markets without transfers. In fact, the very question answered by the present paper arises from an example in Bikhchandani (2017) in which it is shown that incomplete-information stable matchings can be negative assortative in monotonic environments. Besides a few differences in the set of primitives of both papers, the main difference with the analysis in Bikhchandani (2017) is that the present paper provides a general understanding of when, and *what it means* to say that, an incomplete-information stable matching is negative assortative in monotonic environments, beyond those covered by the example in Bikhchandani (2017). In particular, the main result of this paper does not require the market to be balanced, all types to be acceptable by all agents, all agents to be of different types, or any specific form for agents' utilities.

Appendix

Proof of Proposition 1

Proof. Fix any $(\mu, \mathbf{f})$ that satisfies the LMFT property, and take any set $\Omega^* \subseteq \Omega$ that stabilizes μ and satisfies $\Omega^* \subseteq \mathcal{N}(\mu, \mathbf{f})$. For any $\mathbf{w}' \in \Omega^*$, define $\overline{k}_{\mathbf{w}'}$ and $\underline{k}^{\mathbf{w}'}$ to be, respectively, the highest and lowest worker-type in $\mathbf{w}'$, among all matched workers; i.e.,

$$\overline{k}_{\mathbf{w}'} := max\{\mathbf{w}'_{i'} : \mu_{i'} \neq \emptyset\} \text{ and } \underline{k}^{\mathbf{w}'} := min\{\mathbf{w}'_{i'} : \mu_{i'} \neq \emptyset\}.$$

Since $(\mu, \mathbf{f})$ satisfies the LMFT property, $\underline{J}(\mu, \mathbf{f})$ is a singleton. Thus, let $\underline{J}(\mu, \mathbf{f}) = \{j\}$ and $i = \mu_j$.

LMFT implies that i has the highest type, among all matched workers, in every member of Ω^*. This is shown next:

Claim 1: $\mathbf{w}'_i = \overline{k}_{\mathbf{w}'}$ at every $\mathbf{w}' \in \Omega^*$.

Proof. Suppose, instead, that $\mathbf{w}'_i < \overline{k}_{\mathbf{w}'}$ in some $\mathbf{w}' \in \Omega^*$. By construction, there must then be some i' with $\mathbf{w}'_{i'} = \overline{k}_{\mathbf{w}'}$ such that $\mu_{i'} \neq \emptyset$. We would then have $\mathbf{w}'_{i'} > \mathbf{w}'_i$. Since $\mathbf{w}' \in \mathcal{N}(\mu, \mathbf{f})$ by hypothesis, however, it would then follow that $\mathbf{f}_{\mu_{i'}} \leq \mathbf{f}_{\mu_i} = \mathbf{f}_j$. Yet since $\mu_{i'} \neq j$, this would contradict that $(\mu, \mathbf{f})$ satisfies the LMFT property. □

Let $\bar{l}$ be the highest firm-type, in $\mathbf{f}$, among all matched firms; i.e.,

$$\bar{l} := max\{\mathbf{f}_{j'} : \mu_{j'} \neq \emptyset\}.$$

Define $\bar{J}(\mu, \mathbf{f}) := \{j : \mathbf{f}_j = max\{\mathbf{f}_{\tilde{j}} : \mu_{\tilde{j}} \neq \emptyset\}\}$. Hence, $\mathbf{f}_{j'} = \bar{l}$ for every $j' \in \bar{J}(\mu, \mathbf{f})$.

The following claim shows that, in every member of Ω^*, there must some firm in $\bar{J}(\mu, \mathbf{f})$ matched to a worker of the highest type, among those that are matched; i.e.,

Claim 2: In every $\mathbf{w}' \in \Omega^*$, there exists $\hat{j} \in \bar{J}(\mu, \mathbf{f})$ such that $\mathbf{w}'_{\mu_{\hat{j}}} = \underline{k}^{\mathbf{w}'}$.

Proof. Suppose, contrary to hypothesis, that there exists $\mathbf{w}' \in \Omega^*$ at which no firm $j' \in \bar{J}(\mu, \mathbf{f})$ is matched to a worker of type $\underline{k}^{\mathbf{w}'}$. By construction, there exists a worker i' of type $\underline{k}^{\mathbf{w}'}$ such that $\mu_{i'} \neq \emptyset$. By hypothesis, moreover, $\mathbf{f}_{\mu_{i'}} < \bar{l}$. But since $\mu_{j'} \neq \emptyset$ for every $j' \in \bar{J}(\mu, \mathbf{f})$, we have $\mathbf{f}_{j'} > \mathbf{f}_{\mu_{i'}}$ and $\mathbf{w}'_{\mu_{j'}} > \mathbf{w}'_{i'} = \underline{k}^{\mathbf{w}'}$, for every $j' \in \bar{J}(\mu, \mathbf{f})$, contradicting that $\mathbf{w}' \in \mathcal{N}(\mu, \mathbf{f})$. □

I show the desired result by induction, using Claims 1 and 2. If there exists some $\mathbf{w}' \in \Omega^*$ such that $\overline{k}_{\mathbf{w}'} = 1$, then $\overline{k}_{\mathbf{w}'} = \underline{k}^{\mathbf{w}'}$. But then, the very definitions of $\overline{k}_{\mathbf{w}'}$ and $\underline{k}^{\mathbf{w}'}$ imply that $\mathbf{w}'_{\tilde{i}} = \underline{k}^{\mathbf{w}'}$ for every $\tilde{i}$ such that $\mu_{\tilde{i}} \neq \emptyset$. Hence, $\mathbf{w}' \in \mathcal{P}(\mu, \mathbf{f})$, as desired.

Suppose, as induction hypothesis, that the desired result is true if $\overline{k}_{\mathbf{w}'} \leq t$ in any $\mathbf{w}' \in \Omega^*$. Take, then, any $\mathbf{w}' \in \Omega^*$ and assume that $\overline{k}_{\mathbf{w}'} = t + 1$. If $\underline{k}^{\mathbf{w}'} = t + 1$, then we are again done, so suppose that $\underline{k}^{\mathbf{w}'} < t + 1$, and use Claim 2 to pick any $\hat{j} \in \bar{J}(\mu, \mathbf{f})$ such that $\mathbf{w}'_{\mu_{\hat{j}}} = \underline{k}^{\mathbf{w}'}$. Since Ω^* stabilizes μ, the presence of increasing utilities and Claim 1 jointly entail that there must be some $\mathbf{w}'' \in \Omega^*$ such that $\mathbf{w}''_i \leq \mathbf{w}'_{\mu_{\hat{j}}} = \underline{k}^{\mathbf{w}'} < t + 1$. But then, $\mathbf{w}''_i \leq t$. Hence, the induction hypothesis delivers the desired result. □

References

Alston, Max. 2020. On the non-existence of stable matches with incomplete information. *Games and Economic Behavior*, **120**, 336–344.

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

Bikhchandani, Sushil. 2014. *Two-sided matching with incomplete information*. Tech. rept. mimeo.

Bikhchandani, Sushil. 2017. Stability with one-sided incomplete information. *Journal of Economic Theory*, **168**, 372–399.

Chakraborty, Archishman, Citanna, Alessandro, & Ostrovsky, Michael. 2010. Two-sided matching with interdependent values. *Journal of Economic Theory*, **145**(1), 85–105.

Chen, Yi-Chun, & Hu, Gaoji. 2018. A Theory of Stability in Matching with Incomplete Information. *Available at SSRN 3384439*.

Chen, Yi-Chun, & Hu, Gaoji. 2019. Learning by matching. *Theoretical Economics*.

Ehlers, Lars, & Massó, Jordi. 2007. Incomplete information and singleton cores in matching markets. *Journal of Economic Theory*, **136**(1), 587–600.

Ehlers, Lars, & Massó, Jordi. 2015. Matching markets under (in) complete information. *Journal of economic theory*, **157**, 295–314.

Jeong, Byeonghyeon. 2019. *Essays on Market Design and Auction Theory*. Ph.D. thesis, UCLA.

Lazarova, Emiliya, & Dimitrov, Dinko. 2017. Paths to stability in two-sided matching under uncertainty. *International Journal of Game Theory*, **46**(1), 29–49.

Liu, Qingmin. 2020. Stability and Bayesian Consistency in Two-Sided Markets. *American Economic Review*, **110**(8), 2625–66.

Liu, Qingmin, Mailath, George J, Postlewaite, Andrew, & Samuelson, Larry. 2014. Stable matching with incomplete information. *Econometrica*, **82**(2), 541–587.

Pomatto, Luciano. 2018. Stable matching under forward-induction reasoning.

Roth, Alvin E. 1989. Two-sided matching with incomplete information about others' preferences. *Games and Economic Behavior*, **1**(2), 191–209.