First-order implications of stable allocations

Esteban Peralta*

July 21, 2022

Summary

This paper characterizes the set of stable allocations in labor markets in which workers' types are private information, as defined by Liu *et al.* (2014), when firms can only draw *first-order* inferences. The characterization reveals a rich set of implications stemming from stability, regardless of the number of inferences firms can draw, within familiar preference domains. The presence of incomplete information does not affect the sorting of singles, delivers a version of the "lone wolf" theorem, guarantees that the focus on upward or downward stability constraints involves no loss, and provides a novel characterization of complete-information stability.

Keywords: First-order inferences, incomplete information, matching, stability.

*Department of Economics, University of Michigan, 611 Tappan Ave. (48109), Ann Arbor, MI, USA (eperalta@umich.edu). I am indebted to Larry Samuelson for his constant suggestions, and grateful to Héctor Chade, Laura Doval, and Marek Pycia for comments on an earlier version. I also thank all participants in the 6th PKU-NUS Annual International Conference on Quantitative Finance and Economics and the 2022 Conference on Mechanism and Institution Design. Financial support from the University of Michigan is greatly appreciated. The usual disclaimer applies.

1 Introduction

The analysis of two-sided markets with heterogeneous agents has been largely carried out under the assumption that agents have complete information about each other's characteristics. This assumption proves convenient in understanding how these markets are organized, allowing the analysis to focus on the search for (complete-information) stable allocations (see, e.g., Shapley & Shubik (1971)), but it is not always a good approximation of how information is disseminated.

Recently, Liu *et al.* (2014) analyzed a one-to-one labor market with interdependent values in which workers' types are private information and propose an incomplete-information stability notion that prescribes that once a commonly known, individually rational allocation is in place and each firm is informed of the type of its own worker, a firm agrees to participate in a block if and only if it knows that the block is indeed profitable, once it accounts for the type of its own worker, the individually rational nature of the allocation, the blocking worker's willingness to participate, and the observation that no other block takes place.

Incomplete-information stability embeds a "worst-case" use of information that accounts for the analyst's ignorance about firms' beliefs, but assumes that firms can draw all possible finite orders of inferences about workers' types. This assumption is somewhat implausible (see, e.g., Alaoui *et al.* (2020) and Kneeland (2015)), but far from innocuous. For example, Liu *et al.* (2014) show that incomplete-information stable outcomes always exist and, within monotonic and strictly modular environments, are efficient and assortative. While the former holds true regardless of whether firms can draw higher-order inferences or not, the latter one does not. Thus, assortativeness and efficiency are implications of stability in those familiar environments, but only when firms can hold an arbitrarily large number of higher-order beliefs.

It seems reasonable to imagine that the analyst's uncertainty about firms' beliefs includes, in particular, doubts about whether, and how much, firms can engage in higher-order reasoning. If so, she might be interested in understanding what are the *first-order* implications of stability; namely, those steaming from stability, *regardless* of the number of inferences firms can draw.

This paper studies the weakening of incomplete-information stability–termed *first-order stability*—that is obtained when firms account for the type of their own worker, the individually rational nature of the allocation they commonly observe, and the workers' willingness to form a complete-information block with them, but *not* for the absence of *other* blocks. Intuitively, first-order stability weakens incomplete-information stability much in the same way that rationalizability refines the set of best responses in non-cooperative games. As a consequence, every property of first-order stability must be shared not only by its incomplete-information counterpart, but by every other of its refinements. I address the following question: what properties do first-order stable outcomes have in general, and familiar—i.e., monotonic and modular—environments?

1.1 Preview of results

Proposition 1 characterizes first-order stability and illustrates that the presence of incomplete information enlarges the set of stable outcomes *only if*, somewhere in their domain, firms' values are interdependent *and* workers' values exhibit either increasing or decreasing differences. Intuitively, for an allocation to be first-order stable at some state—profile of workers' types—at which it is *not* complete-information stable, every firm involved in a complete-information block must consider possible an alternative state at which the blocking worker says "yes" to *every* transfer she (the worker) accepts *at the true state*, but the firm says "no." The latter requires the firm's value from being matched to the blocking worker to be *strictly* higher at the true state, and the former demands the worker's value to exhibit either weak increasing or decreasing differences *at the two states*, depending on whether the worker's type is higher or lower at the alternative state, and her firm's type is higher or lower than the type of the blocking firm.

When looked at closely, the necessary disagreement required by first-order stability between any worker and firm involved in a complete-information block delivers important implications in general and familiar preference domains.

To start, Proposition 1 reveals the important, and opposite role played by the presence of private values on workers and firms' preferences. When firms' values are private, the set of first-order stable outcomes coincides with that of its complete-information counterpart (Corollary 1), and so the focus on the latter is without loss. Instead, the presence of private values on workers' preferences implies that they always say "yes," and so firms learn nothing from their willingness to accept. Thus, some interdependence on firms' values is necessary for the existence of *some* first-order stable outcome that is *not* complete-information stable, but when workers' values are private first-order stability is consistent with every complete-information block for which the blocking firm can say "no" and still be acceptable to the blocking worker.

Within familiar preference domains, the scope of first-order stability is surprisingly limited. I start with monotonicity. When agents' values are jointly increasing in workers' types, *no* first-order stable allocation can be complete-information blocked by (a firm and) a worker with a type that is equal or lower than the type of *some* unmatched worker (Corollary 2). This result is also true if agents' values are jointly decreasing in workers' types, for the same underlying reason: The selection of unmatched blocking workers is advantageous to the blocking firms, not adverse. If, on the other hand, agents' values are increasing with respect to firms' types, the type of every unmatched firm that forms a complete-information block with a matched worker must be lower than the type of the worker's firm (Corollary 3). The same result holds true when agents' values are decreasing in firms' types, if lower is replaced by higher.

Corollaries 2 and 3 deliver two important results. First, monotonicity implies that all unmatched agents are sorted above or below all those that are matched (Corollary 4), a property inherited from complete-information environments (see, e.g., Chiappori (2017).) Second, a version of the "lone wolf theorem" is obtained, when the match surplus created by every pair of types is non-negative (Corollary 5): If an agent is unmatched in every complete-information allocation, then she must be unmatched in *every* first-order stable allocation. These findings shed more light on the relationship between efficiency and stability in the presence of incomplete information; they reveal that not every inefficiency is consistent with stability, regardless of whether firms draw higher-order inferences or agents' values are modular.

Now we add modularity. When agents' values are jointly increasing in workers' types, *and* workers' values are *strictly* supermodular, an allocation is first-order stable *only if* it is *not-downward-blocked*; namely, only if *every* complete-information block involves a firm and a worker that is matched to a firm with a type that is either the *same or higher* than the type of the blocking firm (Corollary 7). One can replace downward by upward, with a similar interpretation, when workers' values are instead strictly submodular. Further, similar results are also true if agents' values are jointly *decreasing* in workers' types, when supermodular and submodular are "swapped." The intuition hinges, again, on workers' selection: Some complete-information blocks are not consistent with first order stability because they lead the corresponding *matched* blocking workers to select themselves in a way that is favorable to the blocking firms.

When all agents' values are increasing in *all* types, workers' values are *strictly* supermodular, and the match surplus is strictly supermodular, every first-order stable outcome is first-order worker-assortative (Liu *et al.*, 2014), and the following *local-upward* characterization of *complete*-information stability is obtained (Corollary 8): If an allocation is first-order stable, the allocation is in fact complete-information stable if and only if it is positive assortative and there is no *local-upward* block; namely, there is no complete-information block between a firm and a worker that is matched to a firm with a type that is either the same or higher than, but *adjacent* to, the type of the blocking firm. This result is also true if workers' values are *strictly* supermodular, the match surplus is strictly submodular, and agents' values are decreasing in firms' types, when local-upward is replaced by local-downward. These characterizations follow from the sufficiency of local complete-information stability constraints when the match surplus is monotonic and exhibits strict differences, an observation borrowed from Peralta (2022).

The results in this paper reveal that, within the standard class of monotonic and supermodular environments, the set of conclusions the analyst can draw about the nature of the allocation she observes, when no block takes place, is quite large, even if she cautiously assumes that firms are not sophisticated at all. Indeed, she can infer a great deal about the "form" of any underlying complete-information block, and thus sharpen her understanding of how complete- and incomplete-information stability relate to one another, regardless of the degree of sophistication. By understanding that first-order stability is *only* consistent with the existence of *upward/downward* complete-information blocks involving *matched* workers, the analyst would be able to rule out, at the outset, a potentially large number of "objections" and inefficiencies. Since all of this is true for *every* refinement of first-order stability, the results in this paper also uncover important, "hidden" properties of incomplete-information stability, in monotonic and modular environments, besides assortativeness and efficiency.

1.2 Organization of the paper and related literature

Section 2 describes the class of one-to-one matching markets studied by Liu *et al.* (2014), and Section 3 presents the standard notion of (complete-information) stability and the incomplete-information extension proposed by Liu *et al.* (2014). Section 4 offers a characterization of first-order stability, and Section 5 lays out its main implications in different classes of environments. The Appendix contains every proof omitted in the body of the paper.

This paper uncovers important properties of stable outcomes in decentralized markets with incomplete information, without making assumptions about the "depth of reasoning" agents engage in. In this sense, the results in this paper might contribute to the empirical (e.g., Alaoui *et al.* (2020) and Kneeland (2015)) and theoretical (e.g., Börgers & Li (2019) and Strzalecki (2014)) literature that seeks to understand what behavior can be predicted in situations in which agents can, for various reasons, only draw a limited number of inferences.

The paper builds heavily on the analysis in Liu *et al.* (2014), who extend the notion of stability introduced by Gale & Shapley (1962) and Shapley & Shubik (1971) to decentralized markets with interdependent values in which workers' types are private information and both the matching and the payment scheme are observable. Yet the notion introduced by Liu *et al.* (2014), incomplete-information stability, presumes that firms can draw an arbitrarily large number of inferences about the workers' types. A similar presumption is made by Forges (1994) and Holmström & Myerson (1983) in mechanism design problems. Instead, the present paper pays attention to the set of outcomes that are incomplete-information stable when firms can only draw first-order inferences. This focus is not entirely new. For example, Dutta & Vohra (2005) study core allocations in the presence of incomplete information that are credible in the sense that members of deviating coalitions must account for their mutual willingness to participate (see also Chade (2006).) More generally, the present paper is related to the literature that studies the core in the presence of incomplete information (see, e.g., Forges *et al.* (2002) and Myerson (2007), among others.) A key goal in this literature is to find incentive compatible ways in which coalitional members can share some or all of their information, an element that is absent in the present paper.

Most of the papers that followed the analysis in Liu *et al.* (2014) investigate refinements of incomplete-information stability—with and without transfers—by expressing agents' beliefs as an explicit part of the model (see, e.g., Alston (2020), Bikhchandani (2017), Jeong (2019), Liu (2020), and Pomatto (2018)). Instead, I study a weakening of incomplete-information stability.

A few number of papers have departed from the assumption that the matching is fixed, and thus a primitive part of the model. For example, Chen & Hu (2019) propose a random matching selection process that delivers incomplete-information stable outcomes in the sense that starting from any allocation and any information structure the process converges to an information structure for which the allocation is incomplete-information stable. See also Lazarova & Dimitrov (2017). On the other hand, Ehlers & Massó (2007), Chakraborty *et al.* (2010), Roth (1989), and Yenmez (2013), among others, analyze stability and/or incentive issues in centralized markets.

Finally, Chen & Hu (2018) have recently extended the analysis in Liu *et al.* (2014) to capture two-sided incomplete information and formulated various blocking notions that account for the agents' ability to draw inferences.

2 The environment

There is a finite set of workers, I, and a finite set of firms, J, with $i \in I$ and $j \in J$. There is also a finite set of types of workers, W, and a finite set of types of firms, F, where $W = \{w^1, w^2, ..., w^K\} \subseteq \mathbb{R}_+$, $F = \{f^1, f^2, ..., f^L\} \subseteq \mathbb{R}_+$, and w^k and f^l are increasing in their indices. A **state** is a vector $\mathbf{w} \in W^{|I|}$ of workers' types. I write $w \in W$ and $f \in F$ for generic elements of W and F, but also use $\mathbf{w}_i \in W$ to denote the type of worker i at state $\mathbf{w}$. Firms' types are commonly known among workers and firms, so that a vector of firms types $\mathbf{f} \in F^{|J|}$ is fixed throughout, denoting by $\mathbf{f}_j$ the type of firm j.

Value is generated by matches. Following Liu *et al.* (2014), I take as primitive agents' premuneration values; namely, the aggregate match value each agent receives in the absence of payments. Thus, a match between a worker of type $w \in W$ and a firm of type $f \in F$ gives rise to a premuneration value $\nu_{wf} \in \mathbb{R}$ for the worker and a premuneration value $\phi_{wf} \in \mathbb{R}$ for the firm. The sum of these premuneration values, $S_{wf} := \nu_{wf} + \phi_{wf}$, is the *surplus* of the match. Without loss, I assume that the premuneration value of unmatched agents is zero, and use the notation $f_\emptyset = \emptyset = \omega_\emptyset$, with the convention that $\emptyset < w$ and $\emptyset < f$ for every $\omega \in W$ and every $f \in F$.

Given a sate $\mathbf{w}$, a matching between worker i and firm j gives rise to *payoffs*

$$\pi_i^{\mathbf{w}} := \nu_{\mathbf{w}_i\mathbf{f}_j} + p \text{ and } \pi_j^{\mathbf{f}} := \phi_{\mathbf{w}_i\mathbf{f}_j} - p,$$

for i and j, respectively, where $p \in \mathbb{R}$ is the (possibly negative) payment from j to i.

A *matching* is a function $\mu : I \to J \cup \{\emptyset\}$, one-to-one on μ^{-1}, that assigns worker i to $\mu(i)$, where $\mu(i) = \emptyset$ means that i is unmatched. Similarly, $\mu^{-1}(j)$

denotes the assignment of firm j, where $\mu^{-1}(j) = \emptyset$ means that j is unmatched. I will use notation μ_i and μ_j to denote the (possibly empty) assignments of i and j, respectively.

A *payment scheme* $\mathbf{p}$ associated with a matching μ is a vector that specifies a payment $\mathbf{p}_{i,\mu_i} \in \mathbb{R}$ for each i and a payment $\mathbf{p}_{\mu_j,j} \in \mathbb{R}$ for each j. Without loss, I assume that $\mathbf{p}_{i,\emptyset} = \mathbf{p}_{\emptyset,j} = 0$.

An *allocation* is a pair $(\mu, \mathbf{p})$, consisting of a matching and a payment scheme. An *outcome* is a triplet $(\mu, \mathbf{p}, \mathbf{w})$, describing an allocation and a state.[1]

To capture firms' uncertainty about workers' types, I follow Liu *et al.* (2014) and assume that the true state is drawn from some distribution with common support $\Omega \subseteq W^{|I|}$.

3 Stability

3.1 Individual rationality

Definition 1. *An outcome* $(\mu, \mathbf{p}, \mathbf{w})$ *is individually rational if*

$$\nu_{\mathbf{w}_i \mathbf{f}_j} + \mathbf{p}_{i,\mu_i} \geq 0 \quad \textit{for every } i \in I, \textit{ and}$$
$$\phi_{\mathbf{w}_i \mathbf{f}_j} - \mathbf{p}_{\mu_j,j} \geq 0 \quad \textit{for every } j \in J.$$

I write $\sum^0$ for the set of individually rational outcomes. I denote by $\sum^0(\mu, \mathbf{p})$ the set of states at which allocation $(\mu, \mathbf{p})$ is individually rational.

3.2 Complete information

The following definition describes the well known notion of stability introduced by Shapley & Shubik (1971) for environments with complete information (see also Crawford & Knoer (1981).)

Definition 2. *An outcome* $(\mu, \mathbf{p}, \mathbf{w})$ *is* ***complete-information stable*** *if* $(\mu, \mathbf{p}, \mathbf{w}) \in \sum^0$ *and there is no complete-information block; i.e., there is no worker-firm pair*

[1]Notice that an outcome does not specify a vector of types for firms simply because firms' types are fixed throughout.

(i, j) *and payment* $p \in \mathbb{R}$ *such that:*

$$\nu_{\mathbf{w}_i \mathbf{f}_j} + p > \nu_{\mathbf{w}_i \mathbf{f}_{\mu_i}} + \mathbf{p}_{i,\mu_i} \text{ and } \phi_{\mathbf{w}_i \mathbf{f}_j} - p > \phi_{\mathbf{w}_{\mu_j} \mathbf{f}_j} - \mathbf{p}_{\mu_j,j}.$$

Notice that (i, j) forms a complete-information block, at state $\mathbf{w}$, if and only $S_{\mathbf{w}_i \mathbf{f}_j} > \pi_i^{\mathbf{w}} + \pi_j^{\mathbf{f}}$. For any $(\mu, \mathbf{p}, \mathbf{w})$, let

$$B_j(\mu, \mathbf{p}, \mathbf{w}) := \{i \in I : S_{\mathbf{w}_i \mathbf{f}_j} > \pi_i^{\mathbf{w}} + \pi_j^{\mathbf{f}}\}$$

denote the set of workers that form a complete-information block with j. I write $\mathcal{C}$ for the set of complete-information stable outcomes.

3.3 Incomplete information

The following blocking notion, introduced by Liu *et al.* (2014), extends the notion of complete-information block to markets in which workers' types are private information, but no dissolution or re-matching is observed.

Definition 3. *Fix any nonempty set* $X \subseteq \sum^0$. *An outcome* $(\mu, \mathbf{p}, \mathbf{w}) \in X$ *is* X***-blocked*** *if there exists* (i, j) *and* $p \in \mathbb{R}$ *such that:*

1. $\nu_{\mathbf{w}_i, \mathbf{f}_j} + p > \nu_{\mathbf{w}_i \mathbf{f}_{\mu_i}} + \mathbf{p}_{i,\mu_i}$, *and*

2. $\phi_{\tilde{\mathbf{w}}_i \mathbf{f}_j} - p > \phi_{\tilde{\mathbf{w}}_{\mu_j} \mathbf{f}_j} - \mathbf{p}_{\mu_j,j}$, *for every* $\tilde{\mathbf{w}} \in \Omega$ *with* $(\mu, \mathbf{p}, \mathbf{w}') \in X$ *such that:*

$$\tilde{\mathbf{w}}_{\mu_j} = \mathbf{w}_{\mu_j} \text{ and } \nu_{\tilde{\mathbf{w}}_i, \mathbf{f}_j} + p > \nu_{\tilde{\mathbf{w}}_i \mathbf{f}_{\mu_i}} + \mathbf{p}_{i,\mu_i}$$

To understand Definition 3, let $X = \sum^0$. An outcome $(\mu, \mathbf{p}, \mathbf{w})$ is $\sum^0$-blocked by worker i and firm j if and only if (i, j) forms a complete-information block at $\mathbf{w}$ *and at every other state* in Ω that is consistent with the signals j receives; namely, the type of its own worker, the individually rational nature of the allocation, and i's willingness to participate in the block.

Let $\sum^1$ denote the set of individually rational outcomes that are not $\sum^0$-blocked; i.e.,

$$\sum\nolimits^1 := \{(\mu, \mathbf{p}, \mathbf{w}) : (\mu, \mathbf{p}, \mathbf{w}) \in \sum\nolimits^0 \text{ and } (\mu, \mathbf{p}, \mathbf{w}) \text{ is not } \sum\nolimits^0\text{-blocked}\}.$$

$\sum^1$ is the set of first-order stable outcomes, and I write $\sum^1(\mathbf{w})$ for the set of allocations that are first-order stable at $\mathbf{w}$.

Two aspects of $\sum^1$ are worth emphasizing. First, $\sum^1$ embeds a "cautious" or "robust" use of information as it presumes that a firm follows a worst-case scenario decision rule; namely, firms are *not* willing to participate in a complete-information block if there is *at least* one state consistent with her signals for which the block is strictly profitable only to the blocking worker. The intended interpretation, as argued in Liu *et al.* (2014), refers to an outside observer who sees no block taking place, but is uncertain about firms' beliefs. Under this interpretation, any such observer might reasonably seek for *some* beliefs that explain the absence of blocks.

Second, the information updating in Definition 3—and embedded in $\sum^1$—presumes that, when contemplating a given potential complete-information block, a firm draws inferences about the type of the blocking worker from the worker's willingness to participate, and fact that no match is dissolved, but *not* from the absence of *other* complete-information blocks, including those between itself and the same worker at other states. These are what I call *first-order inferences.*

Of course, one could imagine firms drawing additional inferences from observing, not only that no match is dissolved, but also that no other complete-information block is formed. Liu *et al.* (2014) assumes that, indeed, firms draw second, third, and in fact all finite-order inferences from observing no re-matching taking place. Thus, they use $\sum^1$ as a "building block" and define, for every $k \geq 1$, the sets

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

Each of these sets describes an additional inference each firm draws—about workers' types—from observing, not only that no agent takes her outside option, but also that no firm and worker implement a complete-information block. Indeed, at every $k \geq 1$ every firm refines its information by ruling out states in which some firm considers possible that some firm considers possible that...some firm knows that it forms a complete-information block with a given worker. Thus, the set of (incomplete-information) stable outcomes studied by Liu *et al.* (2014) corresponds to the set $\sum =: \bigcap_{k\geq 1} \sum^k$. Intuitively, the difference between $\sum^1$ and $\sum$ resembles the difference between the notions of best response and rationalizability.

This paper is interested in the properties that stem from stability regardless of the number of inferences firms can draw. As a consequence, the focus will be on $\sum^1$. Importantly, Liu *et al.* (2014) show that the set of incomplete-information stable allocations is nonempty at every state; i.e., $\sum(\mathbf{w}) \neq \emptyset$ for every $\mathbf{w}$. Since $\sum \subseteq \sum^1$, it follows that the set of first-order stable allocations is nonempty at every state; i.e., $\sum^1(\mathbf{w}) \neq \emptyset$ for every $\mathbf{w}$.

4 First-order stability: characterization

Consider a pair (i, j) that, at some state $\mathbf{w}$, complete-information blocks the allocation $(\mu, \mathbf{p})$. By definition, this means that the surplus created by a match between i and j is strictly larger than the sum of the payoffs they currently receive; i.e.,

$$\mathbf{p}_{i,\mu_i} - \mathbf{p}_{\mu_j,j} < \phi_{\mathbf{w}_i\mathbf{f}_j} - \phi_{\mathbf{w}_{\mu_j}\mathbf{f}_j} + \nu_{\mathbf{w}_i\mathbf{f}_j} - \nu_{\mathbf{w}_i\mathbf{f}_{\mu_i}}. \tag{1}$$

Consider, in particular, the "smallest" transfer giving rise to this complete-information; namely,

$$p^{\epsilon}_{\mathbf{w}i} := \nu_{\mathbf{w}_i\mathbf{f}_{\mu_i}} + \mathbf{p}_{i,\mu_i} - \nu_{\mathbf{w}_i\mathbf{f}_j} + \epsilon. \tag{2}$$

The reader should think of ϵ as being "small," and so interpret $p^{\epsilon}_{\mathbf{w}_i}$ as the "smallest" transfer that worker i would accept, being of type $\mathbf{w}_i$, to leave her firm and match with j instead. For $(\mu, \mathbf{p}, \mathbf{w})$ to be in $\sum^1$, we need some alternative state $\mathbf{w}' \in \Omega$, consistent with j's information, at which the complete-information block, with respect to $p^{\epsilon}_{\mathbf{w}_i}$, is profitable to i, but not to j, for *every* $\epsilon > 0$. The former requires the worker's value to exhibit either increasing or decreasing differences, depending on the sign of $\mathbf{w}_i - \mathbf{w}'_i$ and $\mathbf{f}_{\mu_i} - \mathbf{f}_j$, and the latter demands (1) to *fail* when $\phi_{\mathbf{w}_i\mathbf{f}_j}$ is replaced by $\phi_{\mathbf{w}'_i\mathbf{f}_j}$. This is the content of (3) and (4) in the following characterization of first-order stability.

Proposition 1. *$(\mu, \mathbf{p}, \mathbf{w}) \in \sum^1$ if and only if $(\mu, \mathbf{p}, \mathbf{w}) \in \sum^0$ and for every (i, j) there exists $\mathbf{w}' \in \sum^0(\mu, \mathbf{p}) \cap \Omega$ with $\mathbf{w}'_{\mu_j} = \mathbf{w}_{\mu_j}$, such that:*

$$\mathbf{p}_{i,\mu_i} - \mathbf{p}_{\mu_j,j} \geq \phi_{\mathbf{w}'_i \mathbf{f}_j} - \phi_{\mathbf{w}_{\mu_j} \mathbf{f}_j} + \nu_{\mathbf{w}_i \mathbf{f}_j} - \nu_{\mathbf{w}_i \mathbf{f}_{\mu_i}}, \tag{3}$$

and

$$\nu_{\mathbf{w}'_i \mathbf{f}_j} + \nu_{\mathbf{w}_i \mathbf{f}_{\mu_i}} - \nu_{\mathbf{w}_i \mathbf{f}_j} - \nu_{\mathbf{w}'_i \mathbf{f}_{\mu_i}} \geq 0. \tag{4}$$

Intuitively, (4) can be interpreted as the "selection constraint" of the problem faced by each blocking firm. The following sections will exploit this interpretation and lay out the various implications that follow from (3) and (4) in general, and familiar environments. They will shed a good amount of light on the properties of first-order stable outcomes.

5 First-order implications

5.1 Interdependent values

To start, Proposition 1 uncovers the important role played by the presence of interdependent values, defined as follows:

Definition 4. *Workers' premuneration values are private if ν_{wf} does not vary with f, and firms' premuneration values are private if ϕ_{wf} does not vary with w.*

It is easy to see that if workers' values are private, then (4) above is always satisfied. Intuitively, there is no selection whenever workers' values are private, and so firms learn nothing from their willingness to block. More precisely, every worker i that is part of a complete-information block at some state $\mathbf{w}$ would also be willing to form the block, for $p^{\epsilon}_{\mathbf{w}_i}$, in *every* other state. It follows that in environments in which workers' values are private it is "easiest" for firms to say "no," and so first-order stability grows larger.

The following result shows that, instead, $\sum^1$ "completely shrinks" when firms' values are the ones that are private:

Corollary 1. *If firms' values are private, then* $\sum^1 = \mathcal{C}$.

Proof. By definition, $\mathcal{C} \subseteq \sum^1$. The other direction follows from (3). Take any outcome $(\mu, \mathbf{p}, \mathbf{w}) \notin \mathcal{C}$, and any (i, j) such that $i \in B_j(\mu, \mathbf{p}, \mathbf{w})$. Thus, (1) follows. If, contrary to hypothesis, $(\mu, \mathbf{p}, \mathbf{w}) \in \sum^1$, then, Proposition 1 implies that there exists $\mathbf{w}' \in \sum^0(\mu, \mathbf{p}) \cap \Omega$, with $\mathbf{w}'_{\mu_j} = \mathbf{w}_{\mu_j}$, such that (3) holds. But then, (1) implies that

$$\phi_{\mathbf{w}_i \mathbf{f}_j} > \phi_{\mathbf{w}'_i \mathbf{f}_j}, \tag{5}$$

contradicting that firms' values are private. □

Corollary 1 implies that the presence of incomplete information "bites" only when firms' values vary with the type of workers. Said differently, focusing on complete-information stability is without loss in environments in which firms' values are private. Further, Corollary 1 implies that first-order stable outcomes are efficient if firms' values are private, and it is easy to find examples in which efficiency is lost when that is not the case.

5.2 Monotonic values

It turns out that when agents' values are either increasing or decreasing with respect to their types Proposition 1 implies that first-order stability sorts all unmatched agents, a well-known implications of complete-information stability, and satisfies a novel version of the "lone wolf" theorem.

5.2.1 Motonocinity with respect to workers' types

Assumption 1 (*Increasing values in workers' types*). *The worker premuneration value* ν_{wf} *and firm premuneration value* ϕ_{wf} *are increasing in* w*, with* ν_{wf} *strictly increasing in* w.

Assumption 1 is weaker than Assumption 1 in Liu *et al.* (2014), as it does not concern firms' types. In the presence of Assumption 1, an immediate observation follows: Saying "no" requires the blocking firms to conceive an alternative state at which the blocking worker's type is *strictly* lower than her true type. This follows from (5) above, when Assumption 1 is in place, as one must then have:

$$\mathbf{w}_i > \mathbf{w}'_i. \tag{6}$$

Armed with this observation, I now show that first-order stability is *only* consistent with complete-information blocks that involve workers whose types are strictly higher than the type of the *best* unmatched worker.

Corollary 2. *Suppose that Assumption 1 holds, and fix any* $(\mu, \mathbf{p}, \mathbf{w}) \in \sum^1$*. If* $i \in B_j(\mu, \mathbf{p}, \mathbf{w})$*, then* $\mathbf{w}_i > \mathbf{w}_{i'}$ *for every* i' *such that* $\mu_{i'} = \emptyset$*. Thus,* $\mu_i \neq \emptyset$*.*

The proof is in the appendix, but the intuition is simple for the particular case in which the blocking worker is unmatched: An unmatched blocking worker would only say "yes" to (2), given Assumption 1, at states in which her type is (weakly) higher than her true type. Indeed, (4) becomes $\nu_{\mathbf{w}'_i \mathbf{f}_j} - \nu_{\mathbf{w}_i \mathbf{f}_j} \geq 0$ whenever $\mu_i = \emptyset$. Yet in no such state the blocking firm would say "no", as indicated by (6). Notice that Corollary 2 holds true if agents' values are decreasing, when one replaces $\mathbf{w}_i > \mathbf{w}_{i'}$ with $\mathbf{w}_i < \mathbf{w}_{i'}$, because (6) becomes $\mathbf{w}_i < \mathbf{w}'_i$.

Why does Assumption 1 imply that i worker says "yes" only at states in which her type is weakly higher, intuitively? The reason is that under Assumption 1 the selection of an unmatched blocking worker is *favorable* to the blocking firm, not negative. To see this, notice that (2), which becomes $p^{\epsilon}_{\mathbf{w}_i} = -\nu_{\mathbf{w}_i \mathbf{f}_j} + \epsilon$, is *decreasing* in w in the presence of Assumption 1. Thus, the higher the worker's type, the *smaller* the worker's "reservation value."

5.2.2 Monotonicity with respect to firms' types

Assumption 2 (*Increasing values in firms' types*)**.** *The worker premuneration value* ν_{wf} *and firm premuneration value* ϕ_{wf} *are increasing in* f*, with* ϕ_{wf} *strictly increasing in* f*.*

Assumptions 1 and 2 are jointly equivalent to Assumption 1 in Liu *et al.* (2014). The following consequence of Proposition 1 follows:

Corollary 3. *Suppose that Assumption 2 holds, and fix any* $(\mu, \mathbf{p}, \mathbf{w}) \in \sum^1$. *If* $i \in B_j(\mu, \mathbf{p}, \mathbf{w})$ *and* $\mu_i \neq \emptyset$, *then* $\mu_j = \emptyset \Rightarrow f_j \leq f_{\mu_i}$.

The proof can be found in the Appendix, but the intuition goes as follows: If an unmatched firm forms a complete-information block with the worker of a (strictly) worse firm, then the required individual rationality of the allocation at the alternative state the blocking firm must consider possible entails that, at $\mathbf{w}$, the blocking worker is weakly worse-off with the blocking firm, contradicting that workers' values increase with firms' types.

Three comments are in order. First, the result is also true if agents' values are jointly decreasing in firms' types, when $\mathbf{f}_j \leq \mathbf{f}_{\mu_i}$ is replaced by $\mathbf{f}_j \geq \mathbf{f}_{\mu_i}$. Second, Corollary 2 implies that "$\mu_i \neq \emptyset$" is immediately satisfied if one adds Assumption 1. Third, and somewhat interestingly, we will see in Section 5.3 that Corollary 3 holds in fact for *all* blocking firms, matched or unmatched, when Assumption 1 is instead in place, and workers' values are strictly modular.

The next two subsections state the main implications of Corollaries 2 and 3.

5.2.3 Sorting singles

It is well known that complete-information stable outcomes sort all unmatched agents, when the surplus is strictly monotonic (see, e.g., Chiappori (2017).) That is, the type of every unmatched worker must be lower (resp. higher) than the type of every matched one, when the surplus is strictly increasing (resp. decreasing) on w, and the type of every unmatched firm must be lower (resp. higher) than the type of every matched one, when the surplus is strictly increasing (resp. decreasing) on f. Interestingly, the same property holds true for first-order stability, under Assumptions 1 and 2.

Corollary 4. *Fix any* $(\mu, \mathbf{p}, \mathbf{w}) \in \sum^1$.

1. *If Assumption 1 holds, then* $\mu_i \neq \emptyset \Rightarrow \mu_{i'} \neq \emptyset$ *for every* i' *with* $\mathbf{w}_{i'} > \mathbf{w}_i$.

2. *If Assumption 2 holds, then* $\mu_j \neq \emptyset \Rightarrow \mu_{j'} \neq \emptyset$ *for every* j' *with* $\mathbf{f}_{j'} > \mathbf{f}_j$.

The proof, which can be found in the Appendix, uses Corollary 2 to obtain 1. and Corollary 3 to prove 2. It is not hard to see that the same results are true if agents' values are instead decreasing in w and f, when the strict inequalities in 1. and 2. are reversed.

5.2.4 Lone wolfs

Corollaries 2 and 3 imply that every agent that is unmatched in every efficient matching must be unmatched in *every* first-order stable allocation, whenever the surplus created by every match is non-negative. To see this, let $\mathcal{E}(\mathbf{w})$ denote the set of (welfare) efficient matchings at state $\mathbf{w}$. The proof of the following result can be found in the Appendix.

Corollary 5. *Suppose that* $S_{wf} \geq 0$ *for every* w *and* f*, and fix any state* $\mathbf{w}$.

1. *If Assumption 1 holds and* $\mu_i = \emptyset$ *in every* $\mu \in \mathcal{E}(\mathbf{w})$*, then* $\mu'_i = \emptyset$ *in every* $(\mu', \mathbf{p}') \in \sum^1(\mathbf{w})$.

2. *If Assumption 2 holds and* $\mu_j = \emptyset$ *in every* $\mu \in \mathcal{E}(\mathbf{w})$*, then* $\mu'_j = \emptyset$ *in every* $(\mu', \mathbf{p}') \in \sum^1(\mathbf{w})$.

Corollary 5 sheds light on when, and to what extent, the well-known relationship between efficiency and stability in complete-information environments is lost in the presence of incomplete information. When the true state is commonly known among workers and firms, stability grants efficiency "for free." We know that that is not the case in the presence of incomplete information. In fact, Liu *et al.* (2014) show that efficiency is only guaranteed when firms draw all possible inferences and, in addition to Assumption 1, one assumes that workers' values, *and* surpluses, are strictly supermodular. Even with these additional assumptions, however, first-order stable outcomes are not necessarily efficient. Nonetheless, Corollary 5 implies that first-order stable outcomes are *not* consistent with every inefficiency.

Interestingly, Corollary 5 resembles a version of the so called "lone wolf theorem", a result stating that, in the absence of transfers, every agent that is unmatched in *some* stable matching must in fact be unmatched in *all* of them (see, e.g., Klaus & Klijn (2010) and McVitie & Wilson (1970)). The result has a counterpart in the presence of transfers (see, e.g., Jagadeesan *et al.* (2020)), but to my knowledge all of its versions feature complete information. Corollary 5 shows that a similar conclusion is obtained when one moves, not across complete-information stable allocations, but rather from complete-information environments to those with incomplete information.

5.3 Monotonic *and* modular values

I start this section with the following definition:

Definition 5 (*Workers' supermodularity*)**.** *The worker premuneration value ν_{wf} is supermodular if for every (w', f') and (w, f) such that $w' > w$ and $f' > f$:*

$$\nu_{w'f'} - \nu_{wf'} \geq \nu_{w'f} - \nu_{wf}.$$

This definition concerns only the value of workers, not those of firms, and constitutes the first part of Assumption 2 in Liu *et al.* (2014). I will say that ν_{wf} is stricly supermodular when the weak inequality above is strict. Further, I will say that ν_{wf} is submodular when $\geq$ is replaced by $\leq$ in Definition 5, and strictly submodular when, in turn, $\leq$ is replaced by $<$.

Armed with this definition, one can see that Proposition 1 reveals yet another important consequence of first-order stability, when Assumption 1 is in place. To say "yes" in any alternative state that the blocking firm "uses" to say "no," the blocking worker's values *must* exhibit either increasing or decreasing differences between *the true and the alternative state*, depending on whether the type of the blocking firm is smaller or higher than the type of worker's firm, and the type of the blocking worker in the alternative state is higher or lower than her true type. This is formalized next:

Corollary 6. *Suppose that Assumption 1 holds, and fix any* $(\mu, \mathbf{p}, \mathbf{w}) \in \sum^1$*, and any* (i, j) *such that* $i \in B_j(\mu, \mathbf{p}, \mathbf{w})$*. Then, there exists* $\mathbf{w}' \in \sum^0(\mu, \mathbf{p}) \cap \Omega$ *with* $\mathbf{w}'_{\mu_j} = \mathbf{w}_{\mu_j}$ *such that:*

(a) $\mathbf{f}_j < \mathbf{f}_{\mu_i} \Rightarrow \nu_{\mathbf{w}'_i \mathbf{f}_j} + \nu_{\mathbf{w}_i \mathbf{f}_{\mu_i}} - \nu_{\mathbf{w}_i \mathbf{f}_j} - \nu_{\mathbf{w}'_i \mathbf{f}_{\mu_i}} \geq 0.$

(b) $\mathbf{f}_j > \mathbf{f}_{\mu_i} \Rightarrow \nu_{\mathbf{w}_i \mathbf{f}_j} - \nu_{\mathbf{w}'_i \mathbf{f}_j} + \nu_{\mathbf{w}'_i \mathbf{f}_{\mu_i}} - \nu_{\mathbf{w}_i \mathbf{f}_{\mu_i}} \leq 0.$

The proof of Corollary 6 is straightforward and therefore omitted, but notice that the result says that first-order stability *requires* workers' values to be supermodular (submodular) *between* states $\mathbf{w}$ and $\mathbf{w}'$, whenever the type of the blocking firm is lower (higher) than the type of the worker's firm.[2] Intuitively, Assumption 1 demands worker i to say "yes" to $p^{\epsilon}_{\mathbf{w}_i}$, being of a lower type—$\mathbf{w}'_i < \mathbf{w}_i$. If $\mathbf{f}_{\mu_i} > \mathbf{f}_j$, but i' value was submodular at $(\mathbf{w}'_i, \mathbf{w}_i, \mathbf{f}_{\mu_i}, \mathbf{f}_j)$, however, then $p^{\epsilon}_{\mathbf{w}_i}$ would be "too low" for i to say "yes."

Corollary 6 uncovers the different role played by the presence of increasing/decreasing differences in match surpluses and—those in some part of—workers' values. The former grants assortativeness in environments with complete information, but the latter is *required* for the existence of "new" stable allocations in environments in which information is incomplete. It follows immediately from Corollary 6 that the presence of either *strict* supermodularlity or *strict* submodularity in the *entire domain* of workers' values imposes a strong requirement on the "direction" of complete-information blocks that are consistent with first-order stability. The notion of "direction" can be formalized as follows:

Definition 6. *An outcome* $(\mu, \mathbf{p}, \mathbf{w})$ *is* ***not-downward-blocked*** *(NDB) if, for every* (i, j)*,* $i \in B_j(\mu, \mathbf{p}, \mathbf{w}) \Rightarrow \mu_i \neq \emptyset$ *and* $\mathbf{f}_{\mu_i} \geq \mathbf{f}_j$*.*

Instead, the outcome is said to be ***not-upward-blocked*** *(NUB)* whenever $\mathbf{f}_{\mu_i} \geq \mathbf{f}_j$ is replaced by $\mathbf{f}_{\mu_i} \leq \mathbf{f}_j$. We then reach the main result of this section:

Corollary 7. *Suppose that Assumption 1 holds, and fix any* $(\mu, \mathbf{p}, \mathbf{w}) \in \sum^1$*.*

[2] A similar conclusion holds when agents' values are instead decreasing, and so $\mathbf{w}'_i > \mathbf{w}_i$.

1. *If ν_{wf} is strictly supermodular, then $(\mu, \mathbf{p}, \mathbf{w})$ is not-downward-blocked.*

2. *If ν_{wf} is strictly submodular, then $(\mu, \mathbf{p}, \mathbf{w})$ is not-upward-blocked.*

The proof is in the Appendix, but the intuition is simple: For the given outcome to be first-order stable each blocking firm must consider possible an alternative state at which the blocking worker is of a strictly smaller type, but still willing to accept the firm's proposal. If the type of the blocking firm is higher (lower) than the type of the blocking worker's firm when workers' values are strictly supermodular (submodular), then only "high" types workers are willing to accept the firm's proposal. The underlying selection is then favorable to the blocking firms, not adverse. As a consequence, the blocking firms cannot refuse.

Corollary 7 implies that identifying whether an individually rational outcome is *not* $\sum^0$-blocked, when agents' values are increasing and workers' values are strictly supermodular (submodular), requires to consider *only* complete-information blocks for which the blocking worker is matched to some firm and this firm's type is weakly higher (lower) than that of the blocking firm. Thus, Corollary 7 refines the set of inferences that an outside observer can make, within these environments, about the nature of the allocation she observes, from observing that no block takes place. Indeed, any such observer might be interested in understanding whether the absence of blocks is driven by a lack of profitable blocking opportunities or instead by a purely informative motive, without making any assumption about the number of inferences firms are able to make.

Corollary 7 is also true if agents' values are decreasing, but in that case strict submodularity implies NDB and strict supermodularity implies NUB.

5.4 Monotonic and modular values, *and* modular surpluses

Liu *et al.* (2014) show that incomplete-information stable outcomes are efficient when agents' values are increasing, workers' values strictly supermodular, and the surplus created by every match is strictly supermodular. This section lays out what the three assumptions grant to first-order stability.

5.4.1 First-order-worker assortativeness

When Assumption 1 is satisfied, and both ν_{wf} and S_{wf} are strictly supermodular, Lemma B.3 in Liu *et al.* (2014) shows that first-order stable outcomes cannot prescribe a failure of positive assortativeness involving a matched worker with the lowest possible type w^1. More precisely, no first-order stable outcome can prescribe two workers, i and i', with $\mathbf{w}_{i'} > \mathbf{w}_i = w^1$ and $\mathbf{f}_{\mu_{i'}} < \mathbf{f}_{\mu_i}$, a property Liu *et al.* (2014) call *first-order-worker assortativeness.*[3] One can use Corollary 7 to understand why. If two such workers i and i' were to exist, then the strict supermodularity of the match surplus would imply that either $(i, \mu_{i'})$ or (i', μ_i) forms a complete-information block. By Corollary 7, the latter cannot be the case. But then, (6) implies that $\mu_{i'}$ must consider possible that worker i is of a type *strictly* lower than w^1; impossible.

5.4.2 Characterizing complete-information stability

Interestingly, the three assumptions above provide a novel, *local* characterization of *complete*-information stability.

For any outcome $(\mu, \mathbf{p}, \mathbf{w})$, define the relation $\xrightarrow{(\mu,\mathbf{p},\mathbf{w})}$ on $J \times J$ as:

$$j \xrightarrow{(\mu,\mathbf{p},\mathbf{w})} j' \text{ if and only if } \mu_{j'} \neq \emptyset \text{ and } \mu_{j'} \in B_j(\mu, \mathbf{p}, \mathbf{w}).$$

In words, $j \xrightarrow{(\mu,\mathbf{p},\mathbf{w})} j'$ means that firm j forms a block with the worker of firm j'.[4]

Define, for every j, the sets

$$U(j) := \{j' \in J : \mathbf{f}_{j'} = \mathbf{f}_j \text{ or } \mathbf{f}_{j'} = min_{f:f>\mathbf{f}_j}\mathbf{f}\}, \text{ and}$$

$$D(j) := \{j' \in J : \mathbf{f}_{j'} = \mathbf{f}_j \text{ or } \mathbf{f}_{j'} = max_{f:f<\mathbf{f}_j}\mathbf{f}\}.$$

That is, $U(j)$ $(D(j))$ is the set of firms that have either the same type than j or the higher (lower), but *consecutive* one. I invite the read to think of any j and $j' \in U(j) \cup D(j)$ as being "local" with respect to one another.

[3] See their Definition B.1.

[4] Recall that $B_j(\mu, \mathbf{p}, \mathbf{w})$ is the set of workers with whom firm j forms a complete-information block of μ, when the outcome is $(\mu, \mathbf{p}, \mathbf{w})$.

Definition 7. *An outcome* $(\mu, \mathbf{p}, \mathbf{w})$ *is positive assortative (PAM) if*

1. *for every* i, i' *with* $\mu_i \neq \emptyset$ *we have* $\mathbf{w}_{i'} > \mathbf{w}_i \Rightarrow \mathbf{f}_{\mu_{i'}} \geq \mathbf{f}_{\mu_i}$, *and*

2. *for every* j, j' *with* $\mu_j \neq \emptyset$ *we have* $\mathbf{f}_{j'} > \mathbf{f}_j \Rightarrow \mathbf{w}_{\mu_{j'}} \geq \mathbf{w}_{\mu_j}$.

One can analoguosly define $(\mu, \mathbf{p}, \mathbf{w})$ to be negative assortative (NAM) by reversing both strict inequalities. Notice that these definitions make use of the fixed vector of firms' types. Thus, I write $\mathcal{P}(\mathbf{w})$ and $\mathcal{N}(\mathbf{w})$ for the sets of matchings that are, respectively, positive and negative at state $\mathbf{w}$.

Armed with these definitions, we can now state the main result of this section:

Corollary 8. *Fix any* $(\mu, \mathbf{p}, \mathbf{w}) \in \sum^1$, *and suppose that Assumptions 1 and 2 hold, and that* ν_{wf} *and* S_{wf} *are strictly supermodular. Then,*

$$(\mu, \mathbf{p}, \mathbf{w}) \in \mathcal{C} \Leftrightarrow \mu \in \mathcal{P}(\mathbf{w}) \text{ and } j \overset{(\mu,\mathbf{p},\mathbf{w})}{\not\longrightarrow} j' \text{ for every } j \text{ and every } j' \in U(j).$$

The proof is in the appendix, but notice that Corollary 8 takes the content of Corollary 7 one important step further; namely, the presence of increasing surpluses exhibiting strict differences happens to guarantee that it is without loss to focus on *local* upward/downward complete-information blocks, when dealing with assortative outcomes (Peralta, 2022).

Assumptions 1 and 2 are necessary to guarantee that all complete-information stable outcomes are positive assortative. A similar result would hold if, instead, values are decreasing, workers' values and surpluses are strictly submodular, and surpluses decrease with firms' types, if PAM is replaced by NAM, and $U(j)$ by $D(j)$. Further, the main insight of Corollary 8 does not hinge on the supermodularity of ν_{wf} and S_{wf}. For example, one can replace "ν_{wf} is strictly supermodular" with "ν_{wf} is strictly submodular," and $U(j)$ with $D(j)$.

Corollary 8 drastically simplifies the task of checking whether a given first-order, assortative outcome is in fact complete-information or not. This would offer a particular leeway in environments in which no two firms have the same type, as then one would only have to check *at most one* stability constraint per firm. In turn, these simplifications could prove useful to any outside observer.

Even when an outside observer cannot infer that no complete-information block exists from observing that no such block forms, she can actually infer that, *if* some complete-information block does exist, then: i) *all* of them must be "going in the same direction"; upward or downward, depending on whether surpluses and workers' values are (strictly) supermodular or submodular, and ii) *at least one* of them must be *local*.

Corollary 8 holds not only for first-order stability, but for *every* stability notion sandwiched between complete-information and first-order stability, including incomplete-information stability. In fact, I view Corollary 8 as particularly insightful for any outside observer who is confident that firms draw all possible inferences, and so focus on incomplete-information stability. This is so because first-order stable allocations are not necessarily assortative, when the three assumptions in question are in place, but incomplete-information stable outcomes are.[5]

[5] As Liu *et al.* (2014) demonstrate, incomplete-information stable outcomes are always negative assortative, when agents' values are increasing, and surpluses and workers' values are strictly submodular, only if, in addition, the surplus created by every pair of types is strictly positive. See their Proposition 4.

6 Appendix

Proof of Proposition 1

Take any (i, j) such that $i \in B_j(\mu, \mathbf{p}, \mathbf{w})$, and define, for any $\epsilon > 0$:

$$p^{\epsilon}_{\mathbf{w}i} := \nu_{\mathbf{w}_i \mathbf{f}_{\mu_i}} + \mathbf{p}_{i,\mu_i} - \nu_{\mathbf{w}_i \mathbf{f}_j} + \epsilon.$$

Since $i \in B_j(\mu, \mathbf{p}, \mathbf{w})$, it follows that (i, j) forms a complete-information block, at $\mathbf{w}$, for $p^{\epsilon}_{\mathbf{w}_i}$. Suppose that there exists $\mathbf{w}' \in \sum^0(\mu, \mathbf{p}) \cap \Omega$, with $\mathbf{w}'_{\mu_j} = \mathbf{w}_{\mu_j}$, such that (3) and (4) hold. Clearly, (1) is equivalent to j saying "no" to *every* transfer i would accept to form a complete-information block with j, being of type $\mathbf{w}_i$; namely, (3) is equivalent to

$$\phi_{\mathbf{w}'_i \mathbf{f}_j} - p^{\epsilon}_{\mathbf{w}_i} \leq \phi_{\mathbf{w}_{\mu_j}} - \mathbf{p}_{\mu_j, j}, \tag{7}$$

for every $\epsilon > 0$. Similarly, (4) is equivalent to i saying "yes" to $p^{\epsilon}_{\mathbf{w}_i}$ for *every* $\epsilon > 0$, at state $\mathbf{w}'$. That is, (4) is equivalent to

$$\nu_{\mathbf{w}'_i \mathbf{f}_j} + p^{\epsilon}_{\mathbf{w}_i} > \nu_{\mathbf{w}'_i \mathbf{f}_{\mu_i}} + \mathbf{p}_{i,\mu_i}, \tag{8}$$

for every $\epsilon > 0$. This showed the "if" part.

For the "only if" part, the existence of some $\mathbf{w}' \in \sum^0(\mu, \mathbf{p}) \cap \Omega$, with $\mathbf{w}'_{\mu_j} = \mathbf{w}_{\mu_j}$, such that (7) and (8) is required by $\sum^1$. But since we established that (7) is equivalent to (3), and (8) is equivalent to (4), the "only if" part follows. □

Proof of Corollary 2

Fix any $(\mu, \mathbf{p}, \mathbf{w}) \in \sum^1$ and suppose, contrary to hypothesis, that there exists (i, j) such that $i \in B_j(\mu, \mathbf{p}, \mathbf{w})$, but $\mathbf{w}_i \leq \mathbf{w}_{i'}$ for some i' such that $\mu_{i'} = \emptyset$. Since values are increasing, we must then have that $i' \in B_j(\mu, \mathbf{p}, \mathbf{w})$. To see this, suppose otherwise. Then,

$$S_{\mathbf{w}_i \mathbf{f}_j} > \pi^{\mathbf{w}}_i + \phi^{\mathbf{f}}_j \quad \text{and} \quad S_{\mathbf{w}_{i'} \mathbf{f}_j} \leq \pi^{\mathbf{w}}_{i'} + \phi^{\mathbf{f}}_j.$$

But since $\mu_{i'} = \emptyset$, we have $\pi_{i'}^{\mathbf{w}} = 0$. Hence,

$$S_{\mathbf{w}_{i'}\mathbf{f}_j} - S_{\mathbf{w}_i\mathbf{f}_j} + \pi_i^{\mathbf{w}} < 0.$$

Since $(\mu, \mathbf{p}, \mathbf{w}) \in \sum^0$, it follows that $S_{\mathbf{w}_{i'}\mathbf{f}_j} - S_{\mathbf{w}_i\mathbf{f}_j} < 0$. Yet $\mathbf{w}_{i'} \geq \mathbf{w}_i$, so that we contradict Assumption 1.

Thus, $i' \in B_j(\mu, \mathbf{p}, \mathbf{w})$ must follow. Yet since $\mu_{i'} = \emptyset$ and $(\mu, \mathbf{p}, \mathbf{w}) \in \sum^1$, one can use Proposition 1 to conclude that (4) becomes:

$$\nu_{\mathbf{w}'_{i'}\mathbf{f}_j} - \nu_{\mathbf{w}_{i'}\mathbf{f}_j} \geq 0.$$

At this point, one uses (6)—applied to worker i'—to infer that $\mathbf{w}'_{i'} < \mathbf{w}_{i'}$, and contradict Assumption 1. □

Proof of Corollary 3

Fix any $(\mu, \mathbf{p}, \mathbf{w}) \in \sum^1$, and any $i \in B_j(\mu, \mathbf{p}, \mathbf{w})$ such that $\mu_i \neq \emptyset$ and $\mu_j = \emptyset$. Suppose, contrary to hypothesis, that $\mathbf{f}_j > \mathbf{f}_{\mu_i}$. Since $(\mu, \mathbf{p}, \mathbf{w}) \in \sum^1$, there exists $\mathbf{w}' \in \Omega \in \sum^0(\mu, \mathbf{p})$ such that $\phi_{\mathbf{w}'_i\mathbf{f}_j} - p^{\epsilon}_{\mathbf{w}_i} \leq 0$, where

$$p^{\epsilon}_{\mathbf{w}i} := \nu_{\mathbf{w}_i\mathbf{f}_{\mu_i}} + \mathbf{p}_{i,\mu_i} - \nu_{\mathbf{w}_i\mathbf{f}_j} + \epsilon.$$

Since $\mathbf{f}_j > \mathbf{f}_{\mu_i}$, Assumption 2 implies that $\phi_{\mathbf{w}'_i\mathbf{f}_{\mu_i}} - p^{\epsilon}_{\mathbf{w}_i} < 0$. But then

$$\phi_{\mathbf{w}'_i\mathbf{f}_{\mu_i}} - \mathbf{p}_{i,\mu_i} - \nu_{\mathbf{w}_i\mathbf{f}_{\mu_i}} + \nu_{\mathbf{w}_i\mathbf{f}_j} \leq 0.$$

Since $(\mu, \mathbf{p}, \mathbf{w}) \in \sum^1$, it follows that $(\mu, \mathbf{p}, \mathbf{w}') \in \sum^0$. Hence, $\phi_{\mathbf{w}'_i\mathbf{f}_{\mu_i}} - \mathbf{p}_{i,\mu_i} \geq 0$. Hence, we must have $\nu_{\mathbf{w}_i\mathbf{f}_{\mu_i}} \geq \nu_{\mathbf{w}_i\mathbf{f}_j}$. But since $\mathbf{f}_j > \mathbf{f}_{\mu_i}$, this contradicts that ν_{wf} weakly increases with f. □

Proof of Corollary 4

Fix any $(\mu, \mathbf{p}, \mathbf{w}) \in \sum^1$. I start with 1., so assume Assumption 1 is satisfied. Suppose, contrary to hypothesis, that there are two workers i, i', with $\mathbf{w}_{i'} > \mathbf{w}_i$, such that $\mu_i \neq \emptyset$ and $\mu_{i'} = \emptyset$. It then follows that $(\mu, \mathbf{p}, \mathbf{w})$ is complete-information blocked by (i', μ_i). To see why, suppose not. Then, $S_{\mathbf{w}_{i'}\mathbf{f}_{\mu_i}} \leq \pi^{\mathbf{f}}_{\mu_i}$, because $\pi^{\mathbf{w}}_{i'} = 0$. But $\pi^{\mathbf{f}}_{\mu_i} = S_{\mathbf{w}_i\mathbf{f}_{\mu_i}} - \pi^{\mathbf{w}}_i$. Hence,

$$\pi^{\mathbf{w}}_i \leq S_{\mathbf{w}_i\mathbf{f}_{\mu_i}} - S_{\mathbf{w}_{i'}\mathbf{f}_{\mu_i}} < 0,$$

where the strict inequality follows from Assumption 1, because $\mathbf{w}_{i'} > \mathbf{w}_i$. But then, individual rationality fails, contradicting that $(\mu, \mathbf{p}, \mathbf{w}) \in \sum^1$. Hence, $(\mu, \mathbf{p}, \mathbf{w})$ is complete-information blocked by (i', μ_i). Since $\mu_{i'} = \emptyset$, however, this contradicts Corollary 2.

Let's now prove 2. Suppose, contrary to hypothesis, that there are two firms j, j', with $\mathbf{f}_{j'} > \mathbf{f}_j$, such that $\mu_j \neq \emptyset$ and $\mu_{j'} = \emptyset$. The same argument displayed above can be used to show that Assumption 2 entails that $(\mu, \mathbf{p}, \mathbf{w})$ is complete-information blocked by (μ_j, j'). Since $\mu_{j'} = \emptyset$, we contradict Corollary 3. □

Proof of Corollary 5

I start with 1. Fix any state $\mathbf{w}$, and suppose that there exists i such that $\mu_i = \emptyset$ in every $\mu \in \mathcal{E}(\mathbf{w})$. The following claims are in order.

***Claim 1*:** $\mu_j \neq \emptyset$ **for every** j **in every** $\mu \in \mathcal{E}(\mathbf{w})$**.**

Proof. Suppose not, and take any $\mu \in \mathcal{E}(\mathbf{w})$, and any j such that $\mu_j = \emptyset$. Since $\mu_i = \emptyset$, by hypothesis, consider the matching $\tilde{\mu}$ that results from "adding" the match between i and j to μ; i.e., $\tilde{\mu}_j = \mu_{\hat{j}}$ for every $\hat{j}$ such that $\mu_{\hat{j}} = \emptyset$, and

$$\tilde{\mu}_{\hat{i}} = \begin{cases} \mu_{\hat{i}} & \hat{i} \neq i \\ j & \hat{i} = i \end{cases}$$

Since $S_{wf} \geq 0$ for every w and every f, we must have $S_{\mathbf{w}_i\mathbf{f}_j} = 0$, since otherwise μ would not be efficient. But then, $\tilde{\mu}$ is efficient, contradicting that i is unmatched in every efficient matching at $\mathbf{w}$. □

Suppose, contrary to hypothesis, that there exists some allocation $(\mu', \mathbf{p}') \in \sum^1(\mathbf{w})$ with $\mu'_i \neq \emptyset$. Since $\mu_j \neq \emptyset$ for every j in every $\mu \in \mathcal{E}(\mathbf{w})$, Claim 1 implies that there must be some $i' \neq i$ such that $\mu'_{i'} = \emptyset$, but $\bar{\mu}_{i'} \neq \emptyset$ in some $\bar{\mu} \in \mathcal{E}(\mathbf{w})$. We must have that

***Claim 2*: $S_{\mathbf{w}_i \mathbf{f}_{\mu'_i}} < S_{\mathbf{w}_{i'} \mathbf{f}_{\mu'_i}}$.**

Proof. Since $\mu_i = \emptyset$ in every $\mu \in \mathcal{E}(\mathbf{w})$, we must have that $S_{\mathbf{w}_i \mathbf{f}_{\mu_{\hat{i}}}} < S_{\mathbf{w}_{i'} \mathbf{f}_{\mu_{\hat{i}}}}$ for every $\hat{i}$ such that $\mu_{\hat{i}} \neq \emptyset$ in any $\mu \in \mathcal{E}(\mathbf{w})$. Hence, Assumption 1 implies that $\mathbf{w}_i < \mathbf{w}_{\hat{i}}$ for every $\hat{i}$ such that $\mu_{\hat{i}} \neq \emptyset$ in any $\mu \in \mathcal{E}(\mathbf{w})$. Since $\bar{\mu}_{i'} \neq \emptyset$, it follows that $S_{\mathbf{w}_i \mathbf{f}_{\mu'_i}} < S_{\mathbf{w}_{i'} \mathbf{f}_{\mu'_i}}$, as desired. □

It must then follow that $(\mu', \mathbf{p}', \mathbf{w})$ is complete-information blocked by $(i', \mu'_{i'})$. Indeed, Claim 2 entails that

$$S_{\mathbf{w}_{i'} \mathbf{f}_{\mu'_{i'}}} > S_{\mathbf{w}_i \mathbf{f}_{\mu'_i}} = \pi_i^{\mathbf{w}} + \pi_{\mu'_i}^{\mathbf{f}} \geq \pi_{i'}^{\mathbf{w}} + \pi_{\mu'_i}^{\mathbf{f}},$$

where the last inequality follows by individual rationality and the fact that worker i' is unmatched under μ', so that $\pi_{i'}^{\mathbf{w}} = 0$. Since $\mu'_{i'} = \emptyset$, we can use Assumption 1 to invoke Corollary 2, and reach a contradiction.

I now prove 2. Fix any state $\mathbf{w}$, and suppose that there exists j such that $\mu_j = \emptyset$ in every $\mu \in \mathcal{E}(\mathbf{w})$. The analogue of Claim 1 holds; namely:

***Claim 3*: $\mu_i \neq \emptyset$ for every i in every $\mu \in \mathcal{E}(\mathbf{w})$.**

The proof of Claim 3 is identical to that of Claim 2, and therefore omitted. Suppose, again contrary to hypothesis, that there exists some allocation $(\mu', \mathbf{p}') \in \sum^1(\mathbf{w})$ with $\mu'_j \neq \emptyset$. Since $\mu_i \neq \emptyset$ for every i in every $\mu \in \mathcal{E}(\mathbf{w})$, Claim 3 implies that there must be some $j' \neq j$ such that $\mu'_{j'} = \emptyset$, but $\bar{\mu}_{j'} \neq \emptyset$ in some $\bar{\mu} \in \mathcal{E}(\mathbf{w})$. We must then have that

***Claim 4*: $S_{\mathbf{w}_{\mu'_j} \mathbf{f}_j} < S_{\mathbf{w}_{\mu'_j} \mathbf{f}_{j'}}$.**

The proof of Claim 4 is similar to that of Claim 2, when one instead uses Assumption 2 to obtain $\mathbf{f}_{j'} > \mathbf{f}_j$. Hence, an argument similar to the one used above entails that $(\mu', \mathbf{p}', \mathbf{w})$ is complete-information blocked by (μ'_j, j'). But since $\mu'_{j'} = \emptyset$, we can invoke Corollary 3 to reach a contradiction. □

Proof of Corollary 7

Take any $(\mu, \mathbf{p}, \mathbf{w}) \in \sum^1$. If $(\mu, \mathbf{p}, \mathbf{w}) \in \mathcal{C}$, then the desired conclusion follows. Hence, suppose that $(\mu, \mathbf{p}, \mathbf{w}) \notin \mathcal{C}$, and consider any complete-information block (i, j). By Corollary 2, $\mu_i \neq \emptyset$. If ν_{wf} is strictly supermodular, then Corollary 6 implies that $\mathbf{f}_j \leq \mathbf{f}_{\mu_i}$. A similar argument applies if, instead, ν_{wf} is strictly submodular. □

Proof of Corollary 8

The "only if" part of the equivalence is immediate since complete-information stability requires the absence of all complete-information blocks, and complete-information stable outcomes are positive assortative when the match surplus is strictly increasing and strictly supermodular.

Suppose, contrary to hypothesis, that the "if" part of the equivalence fails; i.e., that there exists $(\mu, \mathbf{p}, \mathbf{w}) \in \sum^1 \setminus \mathcal{C}$, with $\mu \in \mathcal{P}(\mathbf{w})$, such that

$$j \overset{(\mu,\mathbf{p},\mathbf{w})}{\not\longrightarrow} j' \text{ for every } j \text{ and every } j' \in U(j).$$

Here's where we reach a contradiction with Proposition 1 in Peralta (2022). Indeed, $(\mu, \mathbf{p}, \mathbf{w}) \in \sum^0$ and Corollary 2 implies that $(\mu, \mathbf{p}, \mathbf{w})$ cannot be complete-information blocked by unmatched workers. By hypothesis, $(\mu, \mathbf{p}, \mathbf{w})$ cannot be complete-information blocked by unmatched firms either. Thus, *all* complete-information blocks must either be upward, but global, by Corollary 7, or involve a failure of equal treatment of equal firms. The latter is impossible, by hypothesis, since $U(j)$ and $D(j)$ contain all firms with j's type. But the latter is also impossible, since Proposition 1 in Peralta (2022) entails that no global complete-information block can exist. □

References

Alaoui, Larbi, Janezic, Katharina A, & Penta, Antonio. 2020. Reasoning about others' reasoning. *Journal of Economic Theory*, **189**, 105091.

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

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

Börgers, Tilman, & Li, Jiangtao. 2019. Strategically simple mechanisms. *Econometrica*, **87**(6), 2003–2035.

Chade, Hector. 2006. Matching with noise and the acceptance curse. *Journal of Economic Theory*, **129**(1), 81–113.

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.*

Chiappori, Pierre-André. 2017. *Matching with transfers.* Princeton University Press.

Crawford, Vincent P, & Knoer, Elsie Marie. 1981. Job matching with heterogeneous firms and workers. *Econometrica: Journal of the Econometric Society*, 437–450.

Dutta, Bhaskar, & Vohra, Rajiv. 2005. Incomplete information, credibility and the core. *Mathematical Social Sciences*, **50**(2), 148–165.

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

Forges, Françoise. 1994. Posterior efficiency. *Games and Economic Behavior*, **6**(2), 238–261.

Forges, Françoise, Minelli, Enrico, & Vohra, Rajiv. 2002. Incentives and the core of an exchange economy: a survey. *Journal of Mathematical Economics*, **38**(1-2), 1–41.

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

Holmström, Bengt, & Myerson, Roger B. 1983. Efficient and durable decision rules with incomplete information. *Econometrica: Journal of the Econometric Society*, 1799–1819.

Jagadeesan, Ravi, Kominers, Scott Duke, & Rheingans-Yoo, Ross. 2020. Lone wolves in competitive equilibria. *Social Choice and Welfare*, **55**(2), 215–228.

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

Klaus, Bettina, & Klijn, Flip. 2010. Smith and Rawls share a room: stability and medians. *Social Choice and Welfare*, **35**(4), 647–667.

Kneeland, Terri. 2015. Identifying Higher-Order Rationality. *Econometrica*, **83**(5), 2065–2079.

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.

McVitie, David G, & Wilson, Leslie B. 1970. Stable marriage assignment for unequal sets. *BIT Numerical Mathematics*, **10**(3), 295–309.

Myerson, Roger B. 2007. Virtual utility and the core for games with incomplete information. *Journal of Economic Theory*, **136**(1), 260–285.

Peralta, Esteban. 2022. Local stability constraints, forthcoming, Economic Theory Bulletin. Available at https://sites.lsa.umich.edu/eperalta/research/.

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.

Shapley, Lloyd S, & Shubik, Martin. 1971. The assignment game I: The core. *International Journal of game theory*, **1**(1), 111–130.

Strzalecki, Tomasz. 2014. Depth of reasoning and higher order beliefs. *Journal of Economic Behavior & Organization*, **108**, 108–122.

Yenmez, M Bumin. 2013. Incentive-compatible matching mechanisms: consistency with various stability notions. *American Economic Journal: Microeconomics*, **5**(4), 120–41.