Efficiency, sorting, and lower-order reasoning

Esteban Peralta*

Abstract

In labor markets in which workers' types are private information, Liu et al. (2014) show that all stable matchings are positive assortative and efficient within monotonic and supermodular domains. I show that this is only so when firms can make an arbitrarily large number of higher-order inferences, but identify an important subset of their domains within which second-order inferences suffice. The insufficiency of first-order inferences is discussed.

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

*Department of Economics, University of Michigan (eperalta@umich.edu)

Contents

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 attributes or types. This assumption is convenient because it delivers an equivalence between stability and efficiency (see, e.g., Shapley and Shubik (1971)) and, within preference domains in which the surplus is strictly supermodular and strictly increasing in agents' types, it implies that stable matchings are positive assortative (Becker (1973)).

Notably, Liu et al. (2014) introduce a notion of stability for labor markets in which workers' types are private information, dubbed incomplete-information stability, and show that incomplete-information stable matchings are assortative and efficient if firms' and workers' values are increasing ([i]), and both match surpluses and workers' values are strictly supermodular ([ii]).

Incomplete-information stability prescribes that once an array of workers' types is in place, and all agents commonly observe an individually rational allocation, each firm is informed of the type of its own worker and agrees to participate in a block if and only if the block is profitable at every array of workers' types that accounts for individual rationality of the matching it observes, the type of its own worker, the blocking worker's willingness to participate, and the fact that no other block takes place.

Incomplete-information stability presumes that firms make a '"autious" use of their information to account for the analyst's ignorance about firms' beliefs, and thus capture necessary conditions for stability. At the same time, however, incomplete-information stability assumes that after accounting for the type of its own worker each firm can make all possible inferences about the true array of workers' types. These include not only those that Liu et al. (2014) dub first-order inferences, namely those that account for the individually rational nature of the allocation and the willingness of workers to form blocks, but also all of those (higher-order inferences) coming from the common observation that no block takes place; i.e., first-order inferences that account for all firms' first-order inferences (second-order inferences), second-order inferences that account for all firms' second-order inferences (third-order inferences), and so on.

The assumption that agents can draw high-order inferences is standard in economic modelling, but somewhat inconsistent with the evidence that suggests that most agents cannot reason iteratively for more than two or three rounds (see, e.g., Ho et al. (1998), and Kneeland (2015)) and are not fully capable of engaging in contingent reasoning (see, e.g., Charness and Levin (2009) and Bayer and Renou (2016)). The assumption would not be problematic, one might argue, if its role in delivering important properties, like assortativeness and efficiency, was innocuous. Unfortunately, that's not the case: Mimicking the nontrivial impact of higher-order inferences in the equilibria of games (see, e.g., Rubinstein (1989)), stable allocations are consistent with failures of positive assortativeness and efficiency, within domains that satisfy [i] and [ii], when firms draw a fixed, albeit arbitrarily large number of inferences (Section 5). How confident should the analyst then be in concluding that the matching she observes is assortative and efficient? It seems reasonable to imagine that the analyst's uncertainty about firms' beliefs might include, in particular, doubts about whether, and how much, firms can engage in higher-order reasoning. If so, she might be interested in understanding whether, and when, assortativeness and efficiency are lower-order implications of stability. This paper brings good news: second-order inferences suffice when [ii] is strengthened by replacing the strict supermodularity of the surplus with the supermodularity of firms' values ([iii]) (Theorem 1).

To be clear, Theorem 1 is concerned with the number of inferences that are sufficient to conclude that any disassortative or inefficient allocation is *not* stable, but has no bearing on the number of inferences required to stabilize, or argue for the instability of, assortative and efficient ones (Section 7.2). Put another way, [i] and [iii] do *not* imply that second-order inferences capture all higher-order ones. Further, first-order inferences are generally insufficient to deliver assortativeness and efficiency, within domains that satisfy [i] and [iii], but they do imply that their failure is testable, *even* within domains that satisfy [i] and [ii] (Section 7.1).

The importance of searching for outcomes that are robust to the cognitive ability of the agents has long been recognized in centralized matching markets (see, e.g., Pathak and Sönmez (2008)). In fact, the same search has recently attracted a considerable amount of attention in mechanism design (see, e.g., Börgers and Li (2019) and Li (2017)). This paper hopes to shed some light to this search by showing that within interesting preference domains the assortativeness and efficiency of stable allocations are robust to the presence of incomplete information, independently not only of the particular beliefs firms might hold, as Liu et al. (2014) demonstrate, but also, to a very large extent, of their cognitive ability.[1] Interestingly, the preference domain considered by Theorem 1 have been the focus of related studies (see, e.g., Mailath et al. (2013) and Mailath et al. (2017)), and shown to be important for near-by goals. For example, Chen and Ho Cher Sien

[1]This contribution might, in turn, shed some light on the seemingly unrelated goal of understanding whether an agent's observed behavior is driven by some cognitive limitation or instead by her (high-order) beliefs about others' cognitive bounds (see, e.g., Alaoui and Penta (2016)). See Section 7.4 for a brief discussion.

(2020) show that within domains that satisfy [i] and [iii] all matchings that are stable in markets with two-sided uncertainty are both assortative and efficient.[2] Similarly, Dizdar and Moldovanu (2016) show that in suitable (one-dimensional) markets with two-sided uncertainty in which all match surpluses are strictly supermodular, the efficient matching is (ex-post) implementable if and only if both workers' and firms' values are supermodular.

Section 2 describes the class of one-to-one matching markets I consider, and Section 3 presents the standard notion of (complete-information) stability and the incomplete-information extension proposed by Liu et al. (2014). Section 4 describes the main result in Liu et al. (2014), and Section 5 offers an example that shows that their main result hinges on the assumption that firms can draw an arbitrarily large number of inferences. Section 6 offers the main result of the paper, namely that second-order inferences are sufficient within an interesting subdomain of the domain of preferences considered by Liu et al. (2014), and Section 7 offers a brief discussion about the scope of the main result, the (in)sufficiency of first-order inferences, the role of workers' selection, and a potentially interesting connection with the literature. The Appendix contains all the proofs.

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. Firms' types are commonly known by workers and firms. Thus, 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 $\mathbf{w}_i$ and $\mathbf{f}_j$ to denote the type of worker i and firm j when the state is $\mathbf{w}$ and the array of firms' types is $\mathbf{f}$.

Value is generated by matches. Following Liu et al. (2014), I take as primitive the 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. 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 $f \in F$.

[2] Chen and Ho Cher Sien (2020) also assume that the surplus is non-negative regardless of the agents' types, and that it is commonly known among workers and firms that higher-types earn and pay, respectively, higher wages. Whether first- or second-order suffice for their result is an interesting open question.

Given a state $\mathbf{w}$, 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 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. 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, and an *outcome* is a tuple $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$.

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 support $\Omega \subseteq W^{|I|}$.

3 Stability

3.1 Individual rationality

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

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

I write $\sum^0$ for the set of individually rational outcomes.

3.2 Complete information

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

Definition 2. An outcome $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ is **complete-information stable** if $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \sum^0$ and there is no complete-information block; i.e., there is no worker-firm pair (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 $(\mathbf{w},\mathbf{f})$ if and only if $S_{\mathbf{w}_i\mathbf{f}_j} > \pi_i^{\mathbf{w}} + \pi_j^{\mathbf{f}}$.[3] 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.

Definition 3. Fix any nonempty set $X \subseteq \sum^0$. An outcome $(\mu,\mathbf{p},\mathbf{w},\mathbf{f}) \in X$ 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 $(\mu,\mathbf{p},\tilde{\mathbf{w}},\mathbf{f}) \in X$ such that

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

In words: An individually rational outcome $(\mu,\mathbf{p},\mathbf{w},\mathbf{f}) \in X$ is X-blocked by (i,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 fact that j knows [a] the allocation $(\mu,\mathbf{p})$, [b] the type of its own worker, [c] that the allocation is individually rational, and [d] that i is willing to participate in the block.

Two comments about this definition are in order. First, all firms make a "cautious" use of their information. Indeed, a firm is *not* willing to participate in a complete-information block if there is (at least) *one* state consistent with [a]-[b]-[c]-[d] at which the block is not profitable. The intended interpretation, as argued in Liu et al. (2014), refers to an outside observer who presumes that all firms know [a]-[b]-[c]-[d], but is uncertain about the particular beliefs they might have. Under this interpretation, the definition above captures necessary conditions for stability.

Second, the definition above assumes that all firms use [a]-[b]-[c]-[d] to make inferences about the true array of workers' types. These inferences constitute their first-order inferences. Thus, the set of individually rational outcomes that are not $\sum^0$-blocked; i.e.,

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

[3] Because f is fixed throughout, from now on I will say that (i,j) forms a complete-information block at $\mathbf{w}$ instead.

can be interpreted as the set of outcomes that are stable when all firms can make first-order inferences. Of course, firms might know that all firms make first-order inferences, that all firms know that all firms make first-order inferences, and so on. For each order $k \geq 1$, the set of stable outcomes would then be

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

Liu et al. (2014) assume that all firms make all finite higher-order inferences, and so focus on $\sum =: \bigcap_{k \geq 1} \sum^k$, which they refer to as the set of incomplete-information stable outcomes. Liu et al. (2014) show that the set of incomplete-information stable outcomes is nonempty regardless of the arrays of agents' types, namely, that $\sum(\mathbf{w}, \mathbf{f}) \neq \emptyset$ for every $(\mathbf{w}, \mathbf{f})$. This is so, in fact, because $\sum^1(\mathbf{w}, \mathbf{f}) \neq \emptyset$ for every $(\mathbf{w}, \mathbf{f})$. Indeed, every property satisfied by—any outcome in—$\sum^k$ must be satisfied by—any outcome in—$\sum^s$, for every $s > k$. Hence, the existence of stable outcomes is independent of the number of inferences firms can make. The following picture describes the relationship between each of the different stability notions:

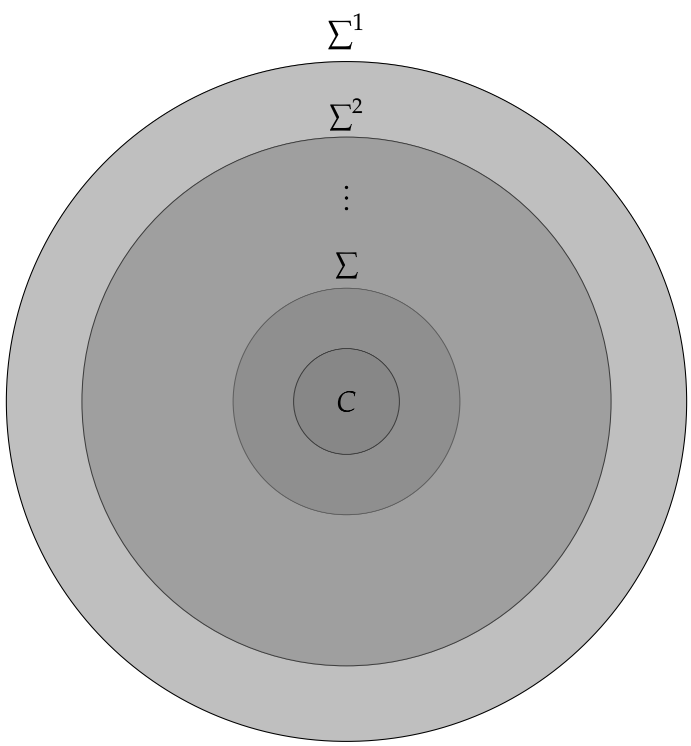

Figure 1: The set of stable outcomes for different number of inferences.

4 Assortativeness and efficiency

I start with the notions of positive assortativeness and efficiency that Liu et al. (2014) consider:

Definition 4. *A matching μ is positive assortative at $(\mathbf{w}, \mathbf{f})$ 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 $\mathbf{f}_{j'} > \mathbf{f}_j \Rightarrow \mathbf{w}_{\mu_{j'}} \geq \mathbf{w}_{\mu_j}$.*

I will refer to 1 and 2 in this definition as "worker assortativeness" and "firm assortativeness," respectively. The standard definition of efficiency is as follows:

Definition 5. *A matching μ is efficient at $(\mathbf{w}, \mathbf{f})$ if*

$$\textstyle\sum_{i \in I} S_{\mathbf{w}_i \mathbf{f}_{\mu_i}} = max_{\mu'} \sum_{i \in I} S_{\mathbf{w}_i \mathbf{f}_{\mu'_i}}.$$

The main result in Liu et al. (2014) shows that all incomplete-information stable matchings are positive assortative and efficient within preference domains that satisfy the following two payoff assumptions.

Assumption 1 (Monotonicity)**.** Workers' premuneration value ν_{wf} and firms' premuneration value ϕ_{wf} are increasing in w and f, with ν_{wf} strictly increasing in w and ϕ_{wf} strictly increasing in f.

Assumption 2 (Supermodularity)**.** The worker premuneration value ν_{wf} and the match surplus S_{wf} are strictly supermodular in w and f.

Liu et al. (2014) prove the following result:

Proposition 1 (Proposition 3 in Liu et al. (2014))**.** Under Assumptions 1 and 2, every outcome $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \sum$ is positive assortative and efficient.

The proof provided by Liu et al. (2014) proceeds in two steps. First, they show that Assumptions 1 and 2 imply that all incomplete-information stable matchings are positive assortative.[4] Second, they make use of the fact that positive assortativeness and efficiency are essentially the same, when Assumptions 1 and 2 are in place, to show that all incomplete-information stable matchings must also be efficient.[5] The first step, in particular, hinges on the following first-order implication of stability:

[4]The reader is invited to look at their Lemma B.5.

[5]This fact is stated in their Lemma B.1, and corresponds to Lemma 2 in Section 6 below.

Lemma 1. *Suppose that Assumptions 1 and 2 hold, and fix any* $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \sum^1$. *If* (i, j) *is a complete-information block at* $\mathbf{w}$, *then* $\mu_i \neq \emptyset$ *and* $\mathbf{f}_{\mu_i} \geq \mathbf{f}_j$.

The proof is in the Appendix. This result says that not every complete-information block is consistent with incomplete-information stability, regardless of the number of inferences firms can make, under Assumptions 1 and 2.[6] In particular, every complete-information block must involve a *matched* worker and the worker must be matched to a firm with a type that is *equal or smaller* than the type of the blocking firm. I will use this result heavily below.

5 Arbitrarily sophisticated firms

This paper is interested in understanding whether Proposition 1 is true for $\sum^k$, for some "small" value of k. Interestingly, the answer is "no." In fact, the following example illustrates that stability is consistent with failures of positive assortativeness and efficiency for an arbitrarily large number of inferences.

Example 1. *There are two workers,* $I = \{i_1, i_2\}$, *and two firms,* $J = \{j_1, j_2\}$. *The type of* j_1 *is 2, and the type of* j_2 *is 3. The type of worker* i_1 *is 5, and the type of worker* i_2 *is* $5 - s$, *with* $4 \geq s > 0$. *The premuneration value of both workers is* $\nu_{wf} = nwf$, *with* $2 \geq n > 1$, *and the premuneration value of both firms is* $\phi_{wf} = w(4 - f) + 5f$. *Let* $w^1 = 1$ *and* $w^K = 5$. *Thus, Assumptions 1 and 2 are satisfied.*[7] *Consider the following allocation:*

Worker indices:	i_1	i_2
Workers' payoffs, $\pi^{\mathbf{w}}$:	$10n$	$3n(5-s)+p$
Workers' types, $\mathbf{w}$:	5	$5-s$
Transfers, $\mathbf{p}$:	0	p
Firms' types, $\mathbf{f}$:	2	3
Firms' payoffs, $\pi^{\mathbf{f}}$:	20	$20-s-p$
Firms' indices:	j_1	j_2

Notice that $s > 0$ implies that the underlying matching is neither positive assortative nor efficient at $\mathbf{w}$. Consider

$$p := -2s_1 - 5n + ns, \tag{1}$$

[6]This result is not explicitly stated in Liu et al. (2014), but their Lemma 2 contains the same insight.

[7]This is clear for ν_{wf}, because $n > 1$, and notice that ϕ_{wf} is strictly increasing in w and f, when $w < 5$, which is the relevant range of workers' types. Further, S_{wf} is strictly supermodular, because $n > 1$.

where $s_1 > s$. The Appendix proves the following result:

Proposition 2. *For every $k \geq 1$ there is some $n \in (1, 2]$ such that $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \sum^k$.*

The proof constructs, for each $k \geq 1$, a sequence of of k states that can be used, when s is small enough, to argue that the allocation above belongs to $\sum^k$. Two observations about the construction are worth mentioning. First, the sequence associated with k is not just longer than the sequence associated with $k-1$, but in fact prescribes different values for its first $k-1$ terms. That is so, intuitively, because each additional inference changes, via stability, the implications of all lower-order ones. Second, stability imposes a non-trivial upper bound on n for third- and higher-order inferences, and the bound is decreasing in k.[8] In fact, the converse is also true; namely, for each value of $n \in (1, 2]$ there is an upper bound on the number of higher-order inferences needed to reach a contradiction (with the hypothesis that the allocation above is incomplete-information stable), and the bound is weakly decreasing in n. Thus, the closer n gets to 1 the higher k is, and viceversa.[9] For example, the allocation belongs to $\sum^3$ if and only if $n < 2$, to $\sum^4$ if and only if (approximately) $n < 1.23$, to $\sum^5$ if and only if (approximately) $n < 1.08$, and so on.

To illustrate these two observations, and in so doing the construction in the proof, suppose that one wants to argue that the allocation above belongs to $\sum^k$, for $k \in \{1, 2, 3, 4\}$. The idea is to define values for s_1, (s_1, s_2), (s_1, s_2, s_3), and (s_1, s_2, s_3, s_4) argue that the allocation belongs, respectively, to $\sum^1$, $\sum^2$, $\sum^3$, and $\sum^4$. The following table describes these values:

k	s_1	s_2	s_3	s_4
1	$\frac{n+1}{2}s$	//	//	//
2	ns	$\frac{n^2+n}{2}s$	//	//
3	$\frac{n}{2-n}s$	$\frac{n^2}{2-n}s$	$\frac{n^3+n^2}{4-2n}s$	//
4	$\frac{2n-n^2}{4-2n-n^2}s$	$\frac{n^2}{4-2n-n^2}s$	$\frac{n^3}{4-2n-n^2}s$	$\frac{n^4+n^3}{8-2n^2-4n}s$

Notice both that n decreases when k goes from 2 to 3, and from 3 to 4, for the values to be well defined, and that the value of s_1, s_2, and s_3 depends on $k \in \{1, 2, 3, 4\}$. The proof of Theorem 1 identifies the pattern behind these sequences to generalize the argument to any $k > 4$.

[8] By non-trivial I mean an upper bound that is strictly smaller than 2.

[9] When $n = 1$, the match surplus exhibits constant differences; that's why the restriction $n > 1$ is imposed.

6 Sufficiency of second-order inferences

A feature of Example 1 is that firms' premuneration values are strictly submodular; it turns out that the feature is a necessary ingredient of every example in which stability is consistent with failures of efficiency and positive assortativeness, when firms are able to draw more than second-order inferences.

Consider the following strengthening of Assumption 2:

Assumption 3. The worker premuneration value ν_{wf} is strictly supermodular and the firm premuneration value ϕ_{wf} is weakly supermodular, in w and f.

The following result, which constitutes the main result of the paper, says that under Assumptions 1 and 3 second-order inferences are sufficient to achieve assortativeness and efficiency:

Theorem 1. Under Assumptions 1 and 3, every outcome $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \sum^2$ is positive assortative and efficient.

The proof of Theorem 1 can be found in the Appendix, but proceeds by first showing that Assumptions 1 and 3 deliver positive assortativeness, and then using assortativeness to obtain efficiency. Intuitively, the argument for assortativeness goes as follows: Given that the surplus is strictly increasing and strictly supermodular, every failure of positive assortativeness gives rise to a complete-information block. By Lemma 1, no such complete-information block can involve an unmatched worker. Thus, every such failure must involve a "mismatch," namely, two workers and two firms of different types matched in a negative assorted way. Because of Lemma 1, the only complete-information block is formed by the "low-low" pair. Since the allocation belongs to $\sum^1$, there must be some alternative array of workers' types, consistent with the signals that the low-type firm receives, at which (a) the blocking opportunity is not longer profitable to the low-type firm and, given Lemma 1, (b) there is no blocking opportunity that involves the high-high pair. The result in Liu et al. (2014) is obtained, using the second-, third-, and higher-order inferences of the low-type firm because (a) requires, given that firms' values are increasing in workers' types, that the type of the low-type worker is reduced and, as a consequence, the failure of assortativeness is inherited by the alternative array. Theorem 1 shows, however, that under Assumption 3 it is impossible, to begin with, for (a) and (b) to be met. Plainly, the lower type of the low-type worker demanded by (a) makes the high-type worker attractive to the high-type firm at the alternative array, when firms' values are supermodular, regardless of the transfers that might be in place. Thus, Lemma 1 is contradicted.

Once positive assortativeness is established, the proof of Theorem 1 proceeds by showing that all incomplete-information stable matchings are efficient by means of the following result:[10]

Lemma 2. *Suppose that the surplus is strictly increasing in w and f, and strictly supermodular. Then, matching μ is efficient at $(\mathbf{w}, \mathbf{f})$ if and only if μ is positive assortative at $(\mathbf{w}, \mathbf{f})$, $S_{\mathbf{w}_i \mathbf{f}_{\mu_i}} \geq 0$ for every i, and there is no (i, j) with $\mu_i = \mu_j = \emptyset$ such that $S_{\mathbf{w}_i \mathbf{f}_j} > 0$.*

To see why the efficiency of every matching μ for which there exists $(\mathbf{p}, \mathbf{w}, \mathbf{f})$ such that $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \sum^2$ is a consequence of Lemma 2, notice that if $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \sum^2$, then $S_{\mathbf{w}_i \mathbf{f}_{\mu_i}} \geq 0$ for every i, by individual rationality. Since μ would be positive assortative, Lemma 2 implies that μ fails to be efficient if and only if there is some (i, j) with $\mu_i = \mu_j = \emptyset$ such that $S_{\mathbf{w}_i \mathbf{f}_j} > 0$. If that's the case, however, (i, j) must form a complete-information block, contradicting Lemma 1.

7 Discussion

7.1 Insufficiency of first-order inferences

The following example illustrates that first-order inferences are insufficient to deliver positive assortativeness and efficiency, even under Assumptions 1 and 3:

Example 2. *There are two workers, $I = \{i_1, i_2\}$, and two firms, $J = \{j_1, j_2\}$. The type of j_1 is 2, and the type of j_2 is 3. The type of worker i_1 is 5, and the type of worker i_2 is 4, with $W = \{1, 2, 3, 4, 5\}$. The premuneration value of both workers and both firms is the same, given by $\nu_{wf} = \phi_{wf} = wf$. Thus, Assumptions 1 and 3 are satisfied. Consider the following allocation:*

Worker indices:	i_1	i_2
Workers' payoffs, $\pi^{\mathbf{w}}$:	14	8
Workers' types, $\mathbf{w}$:	5	4
Transfers, $\mathbf{p}$:	4	-4
Firms' types, $\mathbf{f}$:	2	3
Firms' payoffs, $\pi^{\mathbf{f}}$:	6	16
Firms' indices:	j_1	j_2

[10] The proof of this Lemma is left to the reader, but Lemma B.1 in Liu et al. (2014) conveys the same result—with, however, stronger assumptions.

Notice that the matching is neither positive assortative nor efficient at $\mathbf{w}$*, and that* (i_2, j_1) *forms a complete-information block at* $\mathbf{w}$*, but* (i_1, j_2) *does not. We can focus on the "smallest" transfer for which* (i_2, j_1) *forms a complete-information block at* $\mathbf{w}$*; i.e.,* $p^{\epsilon}_{\mathbf{w}_{i_2}} = \epsilon$*, where* $\epsilon > 0$*. Consider the vector of workers types* $\mathbf{w}' = (5, 3)$*, so that*

Worker indices:	i_1	i_2
Workers' payoffs, $\pi^{\mathbf{w}'}$:	14	5
Workers' types, $\mathbf{w}'$:	5	3
Transfers, $\mathbf{p}$:	4	−4
Firms' types, $\mathbf{f}$:	2	3
Firms' payoffs, $\pi^{\mathbf{f}}$:	6	13
Firms' indices:	j_1	j_2

At $\mathbf{w}'$*, the payoff of* i_2 *in the allocation (5) is strictly smaller than* $6 + \epsilon$*, the payoff she would get by matching with* j_1 *for* $p^{\epsilon}_{\mathbf{w}}$*, for every* $\epsilon > 0$*, but the payoff of* j_1 *from matching with* i_2*, at* $\mathbf{w}'$*, with respect to* $p^{\epsilon}_{\mathbf{w}_{i_2}}$ *is* $6 - \epsilon$*, weakly smaller than 6, the payoff* j_1 *obtains with* i_1*, for every* $\epsilon > 0$*. Thus,* $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \sum^1$*.*□

Intuitively, $\sum^1$ is consistent with failures of assortativeness and efficiency, under Assumptions 1 and 3, because it presumes that all firms account for the fact that the high-high pair cannot form a complete-information block, but does not assume that all firms know that all firms account for that. The latter is precisely what $\sum^2$ imposes over $\sum^1$. Indeed, notice that in the example above Lemma 1 fails at $\mathbf{w}'$.

Interestingly, however, the next result shows that first-order inferences do deliver assortativeness and efficiency for an interesting set of allocations, even under Assumptions 1 and 2:

Proposition 3. Suppose that Assumptions 1 and 2 hold, and fix any $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \sum^1$. If $\mathbf{p}_{j'} \geq \mathbf{p}_j$ for every j, j' such that $\mathbf{f}_{j'} > \mathbf{f}_j$, then $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$ is positive assortative and efficient.

The proof of Proposition 3 is in the Appendix, but the intuition comes directly from Lemma 1, because in every "mismatch" in which the high-type firm pays weakly more than the low-type firm, the high-high pair forms a complete-information block. Notice, in particular, that in Example 1 the transfer "paid" by the high-type firm, p, is smaller than 0, the transfer between the low-type firm and its workers. Indeed, for every $n \in (1, 2]$

$$p = -2s_1 + n(s - 5) < 0,$$

because $s \leq 4$ and $s_1 > s > 0$.

An interesting consequence of Proposition 3 is that the analyst would be able to test every failure of assortativeness and efficiency—because wages and firms' types are assumed to be observable—within the domains study by Liu et al. (2014), without knowing how sophisticated firms are.[11]

7.2 The scope of Theorem 1

The following example illustrates that Assumptions 1 and 3 do *not* imply that $\Sigma^2 = \Sigma$. Thus, more than second-order inferences may be needed, even under Assumptions 1 and 3, to stabilize a matching that is efficient and positive assortative.

Example 3. *There are two workers, $I = \{i_1, i_2\}$, and two firms, $J = \{j_1, j_2\}$. The type of j_1 is 2, and the type of j_2 is 3. The type of worker i_1 is 4, and the type of worker i_2 is 8, with $W = \{1, 2, 3, 4, 5\}$. The premuneration value of both workers and both firms is the same, given by $\nu_{wf} = \phi_{wf} = wf$. Thus, Assumptions 1 and 3 are satisfied. Consider the following allocation:*

Worker indices:	i_1	i_2
Workers' payoffs, $\pi^{\mathbf{w}}$:	8	20
Workers' types, $\mathbf{w}$:	4	8
Transfers, $\mathbf{p}$:	0	-4
Firms' types, $\mathbf{f}$:	2	3
Firms' payoffs, $\pi^{\mathbf{f}}$:	8	28
Firms' indices:	j_1	j_2

Notice that the matching is positive assortative and efficient at $\mathbf{w}$, and that (i_2, j_1) forms a complete-information block at $\mathbf{w}$, but (i_1, j_2) does not. We can focus on the "smallest" transfer for which (i_2, j_1) forms a complete-information block at $\mathbf{w}$; i.e., $p^{\epsilon}_{\mathbf{w}_{i_2}} = 4 + \epsilon$, where $\epsilon > 0$. Consider the vector of workers types $\mathbf{w}' = (4, 6)$, so that

[11] In some markets wages do seem to be positively assorted with respect to firms' types. See, e.g., Bloom et al. (2021) and Chen and Ho Cher Sien (2020).

Worker indices:	i_1	i_2
Workers' payoffs, $\pi^{\mathbf{w}'}$:	8	14
Workers' types, $\mathbf{w}'$:	4	6
Transfers, $\mathbf{p}$:	0	-4
Firms' types, $\mathbf{f}$:	2	3
Firms' payoffs, $\pi^{\mathbf{f}}$:	8	22
Firms' indices:	j_1	j_2

At $\mathbf{w}'$, the payoff of i_2 in the allocation (14) is strictly smaller than $16+\epsilon$, the payoff she would get by matching with j_1 for $p^{\epsilon}_{\mathbf{w}'}$, for every $\epsilon > 0$, but the payoff of j_1 from matching with i_2, at $\mathbf{w}'$, with respect to $p^{\epsilon}_{\mathbf{w}'}$ is $8-\epsilon$, weakly smaller than 8, the payoff j_1 obtains with i_1, for every $\epsilon > 0$. Thus, $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \sum^1$.

Once again, the matching is positive assortative and efficient at $\mathbf{w}'$, (i_2, j_1) forms a complete-information block at $\mathbf{w}'$, and (i_1, j_2) does not. Notice that $p^{\epsilon}_{\mathbf{w}'_{i_2}} = 2+\epsilon$, and consider the vector of workers types $\mathbf{w}'' = (4, 5)$, so that

Worker indices:	i_1	i_2
Workers' payoffs, $\pi^{\mathbf{w}''}$:	8	11
Workers' types, $\mathbf{w}''$:	4	5
Transfers, $\mathbf{p}$:	0	-4
Firms' types, $\mathbf{f}$:	2	3
Firms' payoffs, $\pi^{\mathbf{f}}$:	8	19
Firms' indices:	j_1	j_2

At $\mathbf{w}''$, the payoff of i_2 in the allocation (11) is strictly smaller than $12+\epsilon$, the payoff she would get by matching with j_1 for $p^{\epsilon}_{\mathbf{w}''}$, for every $\epsilon > 0$, but the payoff of j_1 from matching with i_2, at $\mathbf{w}''$, with respect to $p^{\epsilon}_{\mathbf{w}''}$ is $8-\epsilon$, weakly smaller than 8, the payoff j_1 obtains with i_1, for every $\epsilon > 0$. Thus, $(\mu, \mathbf{p}, \mathbf{w}', \mathbf{f}) \in \sum^1$, and so $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \sum^2$.

Even though the matching is positive assortative and efficient at $\mathbf{w}''$, it is not complete-information stable. In particular, (i_2, j_1) forms a complete-information block at $\mathbf{w}''$, but (i_1, j_2) does not. Notice that $p^{\epsilon}_{\mathbf{w}''_{i_2}} = 1+\epsilon$, and consider the vector of workers types $\mathbf{w}''' = (4, 4)$, so that

Worker indices:	i_1	i_2
Workers' payoffs, $\pi^{\mathbf{w}'''}$:	8	8
Workers' types, $\mathbf{w}'''$:	4	4
Transfers, $\mathbf{p}$:	0	-4
Firms' types, $\mathbf{f}$:	2	3
Firms' payoffs, $\pi^{\mathbf{f}}$:	8	16
Firms' indices:	j_1	j_2

At $\mathbf{w}'''$, the payoff of i_2 in the allocation (8) is strictly smaller than $9+\epsilon$, the payoff she would get by matching with j_1 for $p^{\epsilon}_{\mathbf{w}''}$, for every $\epsilon > 0$, but the payoff of j_1 from matching with i_2, at $\mathbf{w}''$, with respect to $p^{\epsilon}_{\mathbf{w}''}$ is $7-\epsilon$, weakly smaller than 8, the payoff j_1 obtains with i_1, for every $\epsilon > 0$. Thus, $(\mu, \mathbf{p}, \mathbf{w}'', \mathbf{f}) \in \sum^1$, so that $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \sum^3$. In fact, the allocation is complete-information stable at $\mathbf{w}'''$, so that, in particular, $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \sum$.□

Two follow-up comments on this example. First, the same sequence of states and allocations can be used to argue that $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \sum^3$ if $\mathbf{p} = (0, -3.99)$, but the allocation would not be complete-information stable at $\mathbf{w}'''$. In particular, (i_1, j_2) would form a complete-information block, thus violating Lemma 1, at $\mathbf{w}'''$. Thus, second-order inferences are not sufficient to argue for the instability of efficient and positive assortative allocations either. Second, the example also illustrates that "adding" the (weak) supermodularity of firms' premuneration values does not deliver positive assortative and efficiency because $\sum^2$ and $\mathcal{C}$ collapse.

7.3 Workers' selection

Arguably, Lemma 1 is the key driving force in Theorem 1 and Propositions 2 and 3, but holds true because $\sum^1$ presumes that all firms account for the selection of workers. The following example illustrates that the notion of incomplete-information stability that does not assume that firms account for the willingness of workers to participate in complete-information blocks is consistent with failures of positive assortativeness and efficiency, evem under Assumptions 1 and 3.

Example 4. *There are two workers, $I = \{i_1, i_2\}$, and two firms, $J = \{j_1, j_2\}$. The type of j_1 is 2, and the type of j_2 is 3. The type of worker i_1 is 5, and the type of worker i_2 is 4, with $W = \{1, 2, 3, 4, 5\}$. The premuneration value of both workers and both firms is the same, given by $\nu_{wf} = \phi_{wf} = wf$. Thus, Assumptions 1 and 3 are satisfied. Consider the following allocation:*

Worker indices:	i_1	i_2
Workers' payoffs, $\pi^{\mathbf{w}}$:	10	13
Workers' types, $\mathbf{w}$:	5	4
Transfers, $\mathbf{p}$:	0	1
Firms' types, $\mathbf{f}$:	2	3
Firms' payoffs, $\pi^{\mathbf{f}}$:	10	11
Firms' indices:	j_1	j_2

Notice that the matching is neither positive assortative nor efficient at $\mathbf{w}$, and that (i_1, j_2) forms a *complete-information block at* $\mathbf{w}$*, but* (i_2, j_1) *does not. We can focus on the "smallest" transfer for which* (i_1, j_2) *forms a complete-information block at* $\mathbf{w}$*; i.e.,* $p^{\epsilon}_{\mathbf{w}_{i_1}} = -5 + \epsilon$*, where* $\epsilon > 0$*. Consider the vector of workers types* $\mathbf{w}' = (2, 4)$*, so that*

Worker indices:	i_1	i_2
Workers' payoffs, $\pi^{\mathbf{w}'}$:	4	13
Workers' types, $\mathbf{w}'$:	2	4
Transfers, $\mathbf{p}$:	0	1
Firms' types, $\mathbf{f}$:	2	3
Firms' payoffs, $\pi^{\mathbf{f}}$:	4	11
Firms' indices:	j_1	j_2

The payoff of j_2 *from matching with* i_1*, at* $\mathbf{w}'$*, with respect to* $p^{\epsilon}_{\mathbf{w}_{i_1}}$*, is* $11 - \epsilon$*, strictly smaller than* 11*, the payoff* j_2 *obtains with* i_2*, for every* $\epsilon > 0$*. Notice, however, that* i_2 *is not strictly better-off, at* $\mathbf{w}'$*, with respect to* $p^{\epsilon}_{\mathbf{w}_{i_1}}$*. At* $\mathbf{w}'$*, the matching is positive assortative and efficient, and there are no complete-information blocks.*□

Notice that in the example firm j_2 does not account for the selection of worker i_2, and thus considers possible that the worker's type is 2, in which case the underlying matching is complete-information stable—and therefore positive assortative and efficient.

The reader should takeaway that whether the analyst should be confident in concluding that the allocation she observes is positive assortative and efficient, because of Theorem 1, depends on how confident she is that firms are sophisticated enough to infer that a worker would only accept a transfer that leaves her better-off (see, e.g., Esponda (2008)).

7.4 Cognitive and strategic bounds

Theorem 1 sheds light on the relationship between the underlying preference domain and the "cognitive load" required by assortativeness and efficiency and, perhaps, on the well-studied question of whether an agent's observed behavior is driven by some cognitive limitation or instead by her (high-order) beliefs about others' cognitive bounds (see, e.g., Alaoui and Penta (2016)). Besides a few exceptions (see, e.g., Alaoui et al. (2020) and Bayer and Renou (2016)), the literature has struggled to disentangle an agent's "cognitive" and "strategic" bounds.

Lemma 1 implies that the role of high-order inferences in delivering assortativeness and efficiency is somewhat limited within domains that satisfy Assumptions 1 and 2. Indeed, in every mismatch the only relevant higher-order inferences are those made by the low-type firm, about the type of the low-type worker, given that there is no blocking opportunity that involves the high-high pair. It turns out that neither Lemma 1 nor this implication are necessary: Assortativeness and efficiency can be obtained within domains that satisfy Assumption 2 and a *weakening* of Assumption 1 within which Lemma 1 does *not* apply, so that a blocking opportunity can involve either the low-low or high-high type pairs (Peralta (2023)).[12] As a consequence, assortativeness and efficiency hinge, within those weaker domains, on a more elaborated, and somewhat more familiar, form of strategic thinking; i.e., on the ability of both the low- and high-type firms to draw higher-order inferences about the worker's type of each other. Interestingly, second-order inferences do *not* suffice under Assumption 3 and the weakening of Assumption 1.[13] Thus, three different, but logically related sets of preference domains deliver assortativeness and efficiency, but make an arguably substantially different use of high-order beliefs.

The relationship between the underlying preference domain and the "kind" of high-order inferences needed for assortativeness and efficiency discussed above could help, in principle, to distinguish between an agent's cognitive and strategic bound. Indeed, within domains that satisfy either Assumptions 1 and 2 or Assumptions 1 and 3 one could perhaps isolate the cognitive bound, because there is no relevant strategic interaction. In fact, one could perhaps measure how low or high that bound might be, by comparing the two iterations that are needed under Assumptions 1 and 3 with examples in which Assumptions 1 and 2 require a "large" number of iterations. Moreover, if assortativeness and efficiency are obtained under Assumptions 1 and 3—or under Assumptions 1 and 2 when they require a "small" number of iterations—but not under Assumption 2 and the weakening of Assumption 1, one might be able to say that behavior is mostly driven by a strategic bound.

[12]The weakening of Assumption 1 requires ν_{wf} to be strictly increasing in w and S_{wf} to be strictly increasing in both w and f.

[13]See Example 1 in Peralta (2023).

8 Appendix

8.1 Proof of Lemma 1

Take any $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \sum^1$, and any (i, j) that forms a complete-information block at $\mathbf{w}$. If $\mu_i = \emptyset$, then consider

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

where $\epsilon > 0$. Since unmatched agents get a payoff of 0, $p^{\epsilon}_{\mathbf{w}_i}$ is the "smallest" transfers for which (i, j) forms a complete-information block at $\mathbf{w}$. Since $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \sum^1$, there must be, for every $\epsilon > 0$, some $\mathbf{w}' \in \Omega$ with $(\mu, \mathbf{p}, \mathbf{w}', \mathbf{f}) \in \sum^0$ and $\mathbf{w}'_{\mu_j} = \mathbf{w}_{\mu_j}$ such that

$$\nu_{\mathbf{w}'_i \mathbf{f}_j} + p^{\epsilon}_{\mathbf{w}_i} > 0 \ \text{ and } \ \phi_{\mathbf{w}'_i \mathbf{f}_j} - p^{\epsilon}_{\mathbf{w}_i} \leq \phi_{\mathbf{w}_{\mu_j} \mathbf{f}_j} - \mathbf{p}_{\mu_j, j},$$

where the second inequality uses the fact that $\mathbf{w}'_{\mu_j} = \mathbf{w}_{\mu_j}$. Since W is finite, and $\Omega \subseteq W^{|I|}$, we can assume without loss that these two inequalities are true for every $\epsilon > 0$.

Notice that the inequality on the left holds for every $\epsilon > 0$ if and only if

$$\nu_{\mathbf{w}'_i \mathbf{f}_j} - \nu_{\mathbf{w}_i \mathbf{f}_j} + \epsilon \quad \geq \quad 0,$$

which is true if and only if $\mathbf{w}'_i \geq \mathbf{w}_i$, by Assumption 1. On the other hand, it is not hard to see that the inequality on the right implies, given that (i, j) forms a complete-information block at $\mathbf{w}$, that

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

which holds true if and only if $\mathbf{w}'_i < \mathbf{w}_i$, a contradiction. Thus, $\sum^1$ implies that we must have $\mu_i \neq \emptyset$ in every complete-information block (i, j).

Suppose now that $\mu_i \neq \emptyset$, but $\mathbf{f}_j > \mathbf{f}_{\mu_i}$. Consider

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

Since $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \sum^1$, there must be, for every $\epsilon > 0$, some $\mathbf{w}' \in \Omega$ with $(\mu, \mathbf{p}, \mathbf{w}', \mathbf{f}) \in \sum^0$ and $\mathbf{w}'_{\mu_j} = \mathbf{w}_{\mu_j}$ such that

$$\nu_{\mathbf{w'}_i \mathbf{f}_j} + p^{\epsilon}_{\mathbf{w}_i} > \pi^{\mathbf{w}}_i \;\text{ and }\; \phi_{\mathbf{w'}_i \mathbf{f}_j} - p^{\epsilon}_{\mathbf{w}_i} \leq \phi_{\mathbf{w}_{\mu_j} \mathbf{f}_j} - \mathbf{p}_{\mu_j, j},$$

Again, because W is finite, and $\Omega \subseteq W^{|I|}$, we can assume without loss that these two inequalities are true for every $\epsilon > 0$. The inequality on the right again implies that $\mathbf{w}'_i < \mathbf{w}_i$, and the inequality on the left is equivalent to

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

It is not hard to see that this inequality is true for every $\epsilon > 0$ if and only if

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

Since $\mathbf{f}_j > \mathbf{f}_{\mu_i}$ and ν_{wf} is strictly supermodular, because of Assumption 2, it follows that $\mathbf{w}'_i \geq \mathbf{w}_i$, a contradiction.□

8.2 Proof of Proposition 2

The proof constructs, for each $k \geq 1$, a sequence of states that can be used to argue that the allocation in the example, $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f})$, belongs to $\sum^k$. Consider, for each $k \geq 1$, the sequence

$$s^k_k := \frac{n^k + n^{k-1}}{2^k - \sum_{i=0}^{k-1} 2^i n^{k-1-i} + n^{k-1}} s,$$

and, for every $k \geq 2$, and each $t \in \{1, ..., k-1\}$,

$$s^k_t = \frac{2^{k-1} - \sum_{i=0}^{k-2} 2^i n^{k-2-i} + \mathbb{1}_{t>0} 2^{k-(t+1)} n^t + \mathbb{1}_{t-1>0} 2^{k-(t+1)} n^{t-1} + \mathbb{1}_{t-2>0} 2^{k-(t)} n^{t-2} + \mathbb{1}_{t-3>0} 2^{k-(t-1)} n^{t-3} + \ldots - 2^{k-2}}{2^{k-1} - \sum_{i=0}^{k-2} 2^i n^{k-2-i}} s,$$

where $\mathbb{1}$ denotes the indicator function.

These sequences prescribe the value $s^1_1 = \frac{n+1}{2} s$ for $k = 1$, the values $s^2_2 = \frac{n^2+n}{2} s$ and $s^2_1 = ns$, for $k = 2$, the values $s^3_3 = \frac{n^3+n^2}{4-2n} s$, $s^3_2 = \frac{n^2}{2-n} s$, and $s^3_1 = \frac{n}{2-n} s$, for $k = 3$, and so on. Hence, the sequence contains one term for $k = 1$, two terms for $k = 2$, three terms for $k = 3$, and so on.

The next two lemmas, whose proofs can be found at the end of this subsection, show that these sequences are well-defined for some $n \in (1, 2]$:

Lemma 3. *For every $k \geq 1$ there is some $n \in (1, 2]$ such that $2^k - \sum_{i=0}^{k-1} 2^i n^{k-1-i} + n^{k-1} > 0$.*

Lemma 4. *For every $k \geq 2$ and every $n \in (1, 2]$ such that $2^k - \sum_{i=0}^{k-1} 2^i n^{k-1-i} + n^{k-1} > 0$, we have*

$$2^{k-1} - \sum_{i=0}^{k-2} 2^i n^{k-2-i} > 0.$$

Define, for each $k \geq 1$ and each $t \in \{1, ..., k\}$, the sequence of states $\mathbf{w}_t^k = (5, 5 - s_t^k)$, and recall that $\mathbf{w} = (5, 5 - s)$. The following lemma constitutes the base case for the inductive argument of the proof:

Lemma 5. $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \sum^1$

Proof. Since $s > 0$, we have that μ is neither positive assortative nor efficient at $\mathbf{w}$. By Assumption 2, either (i_1, j_2) or (i_2, j_1) forms a complete-information block at $\mathbf{w}$. To see this, suppose that neither of them does. Then, $S_{\mathbf{w}_{i_1}\mathbf{f}_{j_2}} \leq \pi_{i_1}^{\mathbf{w}} + \pi_{j_2}^{\mathbf{f}}$ and $S_{\mathbf{w}_{i_2}\mathbf{f}_{j_1}} \leq \pi_{i_2}^{\mathbf{w}} + \pi_{j_1}^{\mathbf{f}}$. Adding them up gives:

$$\begin{aligned} S_{\mathbf{w}_{i_1}\mathbf{f}_{j_2}} + S_{\mathbf{w}_{i_2}\mathbf{f}_{j_1}} &\leq \pi_{i_1}^{\mathbf{w}} + \pi_{j_2}^{\mathbf{f}} + \pi_{i_2}^{\mathbf{w}} + \pi_{j_1}^{\mathbf{f}} \\ &= S_{\mathbf{w}_{i_1}\mathbf{f}_{j_1}} + S_{\mathbf{w}_{i_2}\mathbf{f}_{j_2}}. \end{aligned}$$

But since $s > 0$ this contradicts the strict supermodularity of the surplus, because $\mathbf{w}_{i_1} = 5 > \mathbf{w}_{i_2} = 5 - s$ and $\mathbf{f}_{j_1} = 2 < \mathbf{f}_{j_2} = 3$.

I now show that, using $s_1 := s_1^1$ in the transfer defined in (1), (i_1, j_2) does not form a complete-information block at $\mathbf{w}$. To see this, notice that (i_1, j_2) forms a complete-information block at $\mathbf{w}$ if and only if

$$\begin{aligned} n15 + 5 + 15 > 10n + 20 - s - p &= 10n + 20 - s + 2s_1^1 + 5n - ns \\ &= 10n + 20 - s + 2\frac{(n+1)}{2}s + 5n - ns \\ &= 10n + 20 - s + (n+1)s + 5n - ns \end{aligned}$$

where the second line used the fact that $s_1^1 = \frac{n+1}{2}s$, because $k = 1$. Since both sides are actually equal to one another, we reach a contradiction.

Next, I use $\mathbf{w}_1^1 = (5, 5 - s_1^1)$ to argue that individual rationality holds at both $\mathbf{w}$ and $\mathbf{w}_1^1 = (5, 5 - s_1^1)$, and that every transfer that leads (i_2, j_1) to form a complete-information block at $\mathbf{w}$ only leaves i_2 strictly better-off at $\mathbf{w}_1^1 = (5, 5 - s_1^1)$.

For individual rationality it suffices to show that the payoff of i_2 and j_1 is non-negative at both $\mathbf{w}$ and $\mathbf{w}_1^1 = (5, 5 - s_1^1)$, because $20 > 0$ and $10n > 0$, given that $n > 1$. The payoff of j_1 is equal to $20 + 5n$ at $\mathbf{w} = (5, 5 - s)$, and equals $20 + \frac{1-n}{2}s + 5n$ at $\mathbf{w}_1^1 = (5, 5 - s_1^1)$, both of which are positive regardless of $s \in (0, 4]$, for any $n > 1$. The payoff of i_2 is non-negative at $\mathbf{w} = (5, 5 - s)$ if and only if $s \leq \frac{10n}{3n+1}$, which is true for small enough s, for any $n > 1$, and is non-negative at $\mathbf{w}_1^1 = (5, 5 - s_1^1)$ if and only if $s \leq \frac{20n}{3n^2+3n+2}$, which is satisfied for small enough s, because $n \leq 2$.

In general, for any given $k \geq 1$ individual rationality is satisfied at $\mathbf{w}_t$, for every $t < k$, whenever it is satisfied at $\mathbf{w}_k$.

To argue that every transfer that leads (i_2, j_1) to form a complete-information block at $\mathbf{w}$ leaves i_2 strictly better-off at $\mathbf{w}_1^1 = (5, 5 - s_1^1)$, with respect to the payoff she obtains with j_2, we can focus, without loss, on the "smallest" such transfer; i.e.,

$$p_{\mathbf{w}}^{\epsilon} := 3n(5 - s) + p - 2n(5 - s) = n(5 - s) + p + \epsilon,$$

where p is defined in (1), and $\epsilon > 0$. Intuitively, the reader should think of ϵ as being "small" and, thus, of p_w^{ϵ} as the "smallest" blocking transfer between j_1 and i_2, at $\mathbf{w}$. We have to show that, at $\mathbf{w}_1^1 := (5, 5 - s_1^1)$, the payoff i_2 obtains with j_1 with respect to $p_{\mathbf{w}}^{\epsilon}$ is strictly higher than the payoff she obtains with j_2, for any $\epsilon > 0$; i.e., that

$$2n(5 - s_1^1) + n(5 - s) + p + \epsilon > 3n(5 - s_1^1) + p.$$

This is true, for any p, if and only if $n(s_1^1 - s) + \epsilon > 0$, and is satisfied for every $\epsilon > 0$ because $s_1^1 = \frac{n+1}{2}s > s$, given that $n > 1$. We also need that, at $\mathbf{w}_1^1 := (5, 5 - s_1^1)$, the payoff j_1 obtains matching with i_2 with respect to $p_{\mathbf{w}}^{\epsilon}$ is weakly smaller than the payoff it obtains with i_1, for any $\epsilon > 0$; namely, we need

$$2(5 - s_1^1) + 10 - n(5 - s) - p - \epsilon \leq 20,$$

to be true for every $\epsilon > 0$, which is equivalent to

$$-2s_1 - 5n + ns \leq p + \epsilon.$$

being true for every $\epsilon > 0$. The two sides are in fact equal to one another, for any $\epsilon > 0$, because p is defined as in (1)).□

Let the inductive hypothesis say that $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \sum^x$, for every $x \in \{2, ..., k - 1\}$, where $k > 2$. We have to prove the following claim:

Lemma 6. $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \sum^k$

Proof. Consider the sequence $\{\mathbf{w}_t^k\}_{t\in\{1,...,k\}}$. I use $\mathbf{w}_1^k$ to argue that $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \sum^1$, and $\mathbf{w}_t^k$ to argue that $(\mu, \mathbf{p}, \mathbf{w}_{t-1}^k, \mathbf{f}) \in \sum^1$, for every $t \in \{2, .., k\}$. By Lemmas 3 and 4, fix any $n \in (1, 2]$ such that $2^k - \sum_{i=0}^{k-1} 2^i n^{k-1-i} + n^{k-1} > 0$ and $2^{k-1} - \sum_{i=0}^{k-2} 2^i n^{k-2-i} > 0$, and notice that

$$s_1^k = \frac{2^{k-1} - \sum_{i=0}^{k-2} 2^i n^{k-2-i} + 2^{k-2}n - 2^{k-2}}{2^{k-1} - \sum_{i=0}^{k-2} 2^i n^{k-2-i}} s$$

We have $s_1^k > 0$, for any $n \in (1, 2]$ such that $2^{k-1} - \sum_{i=0}^{k-2} 2^i n^{k-2-i}$, by Lemmas 3 and 4. We use this value to define $\mathbf{w}_1^k := (5, 5 - s_1^k)$. Since $s > 0$, we have that μ is neither positive assortative nor efficient at $\mathbf{w}$. As before, Assumption 2 implies that either (i_1, j_2) or (i_2, j_1) forms a complete-information block at $\mathbf{w}$. Using $s_1 := s_1^k$ in the transfer defined in (1), (i_1, j_2) does not form a complete-information block at $\mathbf{w}$. To see this, notice that (i_1, j_2) forms a complete-information block at $\mathbf{w}$ if and only if

$$\begin{aligned}
n15 + 5 + 15 > 10n + 20 - s - p &= 10n + 20 - s + 2s_1^k + 5n - ns \\
&= 10n + 20 - s + 2[\frac{2^{k-1} - \sum_{i=0}^{k-2} 2^i n^{k-2-i} + 2^{k-2}n - 2^{k-2}}{2^{k-1} - \sum_{i=0}^{k-2} 2^i n^{k-2-i}} s] + 5n - ns \\
&= 10n + 20 - s + [\frac{2^k - 2\sum_{i=0}^{k-2} 2^i n^{k-2-i} + 2^{k-1}n - 2^{k-1}}{2^{k-1} - \sum_{i=0}^{k-2} 2^i n^{k-2-i}}]s + 5n - ns \\
&= 10n + 20 - s + [\frac{2^{k-1} - 2\sum_{i=0}^{k-2} 2^i n^{k-2-i} + 2^{k-1}n}{2^{k-1} - \sum_{i=0}^{k-2} 2^i n^{k-2-i}}]s + 5n - ns.
\end{aligned}$$

where the second line used the value of s_1^k above. It is easy to see that this strict inequality holds if and only if

$$1 + n - [\frac{2^{k-1} - 2\sum_{i=0}^{k-2} 2^i n^{k-2-i} + 2^{k-1}n}{2^{k-1} - \sum_{i=0}^{k-2} 2^i n^{k-2-i}}]s + 5n - ns \quad > \quad 0$$

which, in turn, is equivalent to

$$(1 - n) \sum_{i=0}^{k-2} 2^i n^{k-2-i} \quad > \quad 0.$$

This contradicts the assumption that $n > 1$.

Next, I use $\mathbf{w}_1^k = (5, 5 - s_1^k)$ to argue that individual rationality holds at both $\mathbf{w}$ and $\mathbf{w}_1^k =$

$(5, 5 - s_1^k)$, and that every transfer that leads (i_2, j_1) to form a complete-information block at $\mathbf{w}$ only leaves i_2 strictly better-off at $\mathbf{w}_1^k = (5, 5 - s_1^k)$.

For individual rationality it suffices, as before, to show that the payoff of i_2 and j_1 is non-negative at both $\mathbf{w}$ and $\mathbf{w}_1^k = (5, 5 - s_1^k)$, because $20 > 0$ and $10n > 0$, given that $n > 1$. At $\mathbf{w} = (5, 5 - s)$, the payoff of j_1 is equal to

$$20 + 5n + \tfrac{s(n-1)}{2^{k-1} - \sum_{i=0}^{k-2} 2^i n^{k-2-i}},$$

which is positive, for any $s \in (0, 4]$, because $n > 1$. At $\mathbf{w}_1^k = (5, 5 - s_1^k)$, the payoff of j_1 is equal to

$$20 + 5n + s[\frac{2^{k-2} - \sum_{i=0}^{k-2} 2^i n^{k-2-i} + 2^{k-2}n}{2^{k-1} - \sum_{i=0}^{k-2} 2^i n^{k-2-i}} - n] \quad = \quad 20 + 5n + s[\frac{(n-1)\sum_{i=0}^{k-2} 2^i n^{k-2-i} + 2^{k-2}(1-n))}{2^{k-1} - \sum_{i=0}^{k-2} 2^i n^{k-2-i}}].$$

Since $n > 1$ and $s \in (0, 4]$, this payoff is non-negative if and only if

$$(n-1)\sum_{i=0}^{k-2} 2^i n^{k-2-i} + 2^{k-2}(1-n) \geq 0.$$

But this is true because $(n-1)\sum_{i=0}^{k-2} 2^i n^{k-2-i} = (n-1)[n^{k-2} + ... + 2^{k-2}]$.

On the other hand, the payoff of i_2 at $\mathbf{w} = (5, 5 - s)$ is

$$15n - 3ns - 2s_1^k - 5n + ns \quad = \quad 10n + s[-2n + \frac{2\sum_{i=0}^{k-2} 2^i n^{k-2-i} - 2^k - 2^{k-1}n + 2^{k-1}}{2^{k-1} - \sum_{i=0}^{k-2} 2^i n^{k-2-i}}].$$

For any fixed $n \in (1, 2]$ such that $2^{k-1} - \sum_{i=0}^{k-2} 2^i n^{k-2-i} > 0$, delivered by Lemmas 3 and 4, the right-hand side of this expression is weakly positive for s small enough.

The payoff of i_2 at $\mathbf{w}_1^k = (5, 5 - s_1^k)$ is

$$15n - 3ns_1^k - 2s_1^k - 5n + ns,$$

which is equal to

$$10n + s[n - 3n\frac{2^{k-1} - \sum_{i=0}^{k-2} 2^i n^{k-2-i} + 2^{k-2}n - 2^{k-2}}{2^{k-1} - \sum_{i=0}^{k-2} 2^i n^{k-2-i}} - 2\frac{2^{k-1} - \sum_{i=0}^{k-2} 2^i n^{k-2-i} + 2^{k-2}n - 2^{k-2}}{2^{k-1} - \sum_{i=0}^{k-2} 2^i n^{k-2-i}}].$$

Again, the right-hand side is weakly positive, for small enough s, for any fixed $n \in (1, 2]$ such that $2^{k-1} - \sum_{i=0}^{k-2} 2^i n^{k-2-i} > 0$, granted by Lemmas 3 and 4.

To argue that every transfer that leads (i_2, j_1) to form a complete-information block at $\mathbf{w}$ leaves i_2 strictly better-off at $\mathbf{w}_1^k = (5, 5 - s_1^k)$, with respect to the payoff she obtains with j_2, we can again focus, without loss, on $p^{\epsilon}_{\mathbf{w}}$; i.e., we have to show that, for every $\epsilon > 0$

$$2n(5 - s^k_1) + n(5 - s) + p + \epsilon > 3n(5 - s^k_1) + p.$$

Given the expression for s^k_1 above, this strict inequality is equivalent to

$$\begin{aligned}
&\Leftrightarrow \quad ns^k_1 - sn + \epsilon > 0 \\
&\Leftrightarrow \quad ns^k_1 - 1 > 0 \\
&\Leftrightarrow \quad n(2^{k-1} - \sum_{i=0}^{k-2} 2^i n^{k-2-i} + n2^{k-2} - 2^{k-2}) - 2^{k-1} - \sum_{i=0}^{k-2} 2^i n^{k-2-i} > 0 \\
&\Leftrightarrow \quad [2^{k-1} - \sum_{i=0}^{k-2} 2^i n^{k-2-i}](n-1) + n[2^{k-2}(n-1)] > 0,
\end{aligned}$$

where the second line uses the fact that $s > 0$ and $\epsilon > 0$ is small enough. Clearly, the last inequality is true for any n such that $2^{k-1} - \sum_{i=0}^{k-2} 2^i n^{k-2-i} > 0$, given by Lemmas 3 and 4, because $n > 1$. Thus, it is left to show that, at $\mathbf{w}^k_1 := (5, 5 - s^k_1)$, the payoff j_1 obtains matching with i_2 with respect to $p^{\epsilon}_{\mathbf{w}}$ is weakly smaller than the payoff it obtains with i_1, for any $\epsilon > 0$; i.e., we need

$$2(5 - s^k_1) + 10 - n(5 - s) - p - \epsilon \leq 20,$$

to be true for every $\epsilon > 0$, where p is defined as in (1), with respect to s^k_1. The reader can easily check that this inequality is true, for $\epsilon > 0$, by construction.

At this point, let the inductive hypothesis say that $(\mu, \mathbf{p}, \mathbf{w}^k_{t-1}, \mathbf{f}) \in \sum^1$, for every $t \in \{2, .., k-1\}$. I show that $\mathbf{w}^k_k := (5, 5 - s^k_k)$ can be used to argue that $(\mu, \mathbf{p}, \mathbf{w}^k_{k-1}, \mathbf{f}) \in \sum^1$, where $\mathbf{w}^k_{k-1} := (5, 5 - s^k_{k-1})$, and s^k_k and s^k_{k-1} are given by[14]

$$s^k_k := \frac{n^k + n^{k-1}}{2^k - \sum_{i=0}^{k-1} 2^i n^{k-1-i} + n^{k-1}} s$$

$$s^k_{k-1} := \frac{n^{k-1}}{2^{k-1} - \sum_{i=0}^{k-2} 2^i n^{k-2-i}} s.$$

Since $s \in (0, 4]$, it follows that s^k_k and s^k_{k-1} are positive for any $n \in (1, 2]$ given by Lemmas 3

[14]Notice that when $t = k-1$, we have

$$s^k_{k-1} = \frac{2^{k-1} - \sum_{i=0}^{k-2} 2^i n^{k-2-i} + \mathbb{1}_{k-1>0} 2^{k-(k-1+1)} n^{k-1} + \mathbb{1}_{k-1-1>0} 2^{k-(k-1+1)} n^{k-1-1} + \mathbb{1}_{k-1-2>0} 2^{k-(k-1)} n^{k-1-2} + \mathbb{1}_{k-1-3>0} 2^{k-(k-1-1)} n^{k-1-3} + \ldots - 2^{k-2}}{2^{k-1} - \sum_{i=0}^{k-2} 2^i n^{k-2-i}} s,$$

which is equal to $s^k_{k-1} = \frac{2^{k-1} - (n^{k-2} + 2n^{k-3} + 4n^{k-4} \ldots + 2^{k-2}) + \mathbb{1}_{k-1>0} n^{k-1} + \mathbb{1}_{k-2>0} n^{k-2} + \mathbb{1}_{k-3>0} 2n^{k-3} + \mathbb{1}_{k-4>0} 2^2 n^{k-4} + \ldots - 2^{k-2}}{2^{k-1} - \sum_{i=0}^{k-2} 2^i n^{k-2-i}} s,$

which is equal to $s^k_{k-1} := \frac{n^{k-1}}{2^{k-1} - \sum_{i=0}^{k-2} 2^i n^{k-2-i}} s$

and 4. Thus, for any such value of n we have that μ is neither positive assortative nor efficient at $\mathbf{w}^k_{k-1}$. As before, Assumption 2 implies that either (i_1, j_2) or (i_2, j_1) forms a complete-information block at $\mathbf{w}^k_{k-1}$. Using $s_1 := s^k_1$ in the transfer defined in (1), (i_1, j_2) does not form a complete-information block at $\mathbf{w}^k_{k-1}$. To see this, notice that (i_1, j_2) forms a complete-information block at $\mathbf{w}^k_{k-1}$ if and only if

$$n15 + 5 + 15 > 10n + 20 - s - p \quad = \quad 10n + 20 - s^k_{k-1} + 2s^k_1 + 5n - ns.$$

Simplifying and using the expressions for s^k_k and s^k_{k-1} above, this inequality is equivalent to:

$$0 \quad > \quad s[\frac{2^k - 2\sum_{i=0}^{k-2} 2^i n^{k-2-i} + 2^{k-1}n - 2^{k-1} - n^{k-1} - n2^{k-1} + n\sum_{i=0}^{k-2} 2^i n^{k-2-i}}{2^{k-1} - \sum_{i=0}^{k-2} 2^i n^{k-2-i}}]$$

Clearly, this inequality holds if and only if

$$2^k - 2\sum_{i=0}^{k-2} 2^i n^{k-2-i} + 2^{k-1}n - 2^{k-1} - n^{k-1} - n2^{k-1} + n\sum_{i=0}^{k-2} 2^i n^{k-2-i} < 0.$$

Suppose, contrary to hypothesis, that the inequality does not hold, so that

$$2^k - 2\sum_{i=0}^{k-2} 2^i n^{k-2-i} + 2^{k-1}n - 2^{k-1} - n2^{k-1} + n\sum_{i=0}^{k-2} 2^i n^{k-2-i} \geq n^{k-1}.$$

By the induction hypothesis, (i_1, j_2) does not form a complete-information block at $\mathbf{w}^k_{k-2}$, which means, the reader can easily check, that

$$2^k - 2\sum_{i=0}^{k-2} 2^i n^{k-2-i} + 2^{k-1}n - 2^{k-1} - n2^{k-1} + n\sum_{i=0}^{k-2} 2^i n^{k-2-i} < n^{k-2}.$$

But then, $n^{k-2} > n^{k-1}$, contradicting that $n > 1$, because $k > 1$.

Next, I use $\mathbf{w}^k_k = (5, 5 - s^k_k)$ to argue that individual rationality holds at both $\mathbf{w}^k_{k-1}$ and $\mathbf{w}^k_k$. As before, it is sufficient to show that the payoff of i_2 and j_1 is non-negative. The payoff of the latter at $\mathbf{w}^k_{k-1}$ is

$$\begin{aligned} 20 - s^k_1 + 2s^k_1 + 5n - ns \quad &= \quad 20 + 5n + s^k_1 - ns \\ &= \quad 20 + s^k_1 + n(5 - s). \end{aligned}$$

This value is positive, because $s_1^k > 0$, given that $s \in (0, 4]$. At $\mathbf{w}_k^k$, the payoff of j_1 is

$$20 - s_k^k + 2s_1^k + 5n - ns \quad = \quad 20 + n(5 - s) + 2s_1^k - s_k^k.$$

This payoff is weakly positive, for small enough s, for any fixed n such that $2^k - \sum_{i=0}^{k-1} 2^i n^{k-1-i} + n^{k-1} > 0$ and $2^{k-1} - \sum_{i=0}^{k-2} 2^i n^{k-2-i} > 0$, delivered by Lemmas 3 and 4.

On the other hand, the payoff of i_2 at $\mathbf{w}_{k-1}^k$ is

$$15n - 3ns_{k-1}^k - 2s_1^k - 5n + ns \quad = \quad 10n + s[\frac{-3n^k - 2^k + 2\sum_{i=0}^{k-2} 2^i n^{k-2-i} + 2^{k-1} - n\sum_{i=0}^{k-2} 2^i n^{k-2-i}}{2^{k-1} - 2\sum_{i=0}^{k-2} 2^i n^{k-2-i}}].$$

As before, the right-hand side is weakly positive, for small enough s, for any fixed n such that $2^{k-1} - \sum_{i=0}^{k-2} 2^i n^{k-2-i} > 0$ that Lemmas 3 and 4 deliver.

Finally, the payoff of i_2 at $\mathbf{w}_k^k$ is

$$15n - 3ns_k^k - 2s_1^k - 5n + ns \quad = \quad 10n + s[n - \frac{-3(n^{k+1} + n^k)}{2^k - \sum_{i=0}^{k-1} 2^i n^{k-1-i} + n^{k-1}} - \frac{2^k + 2\sum_{i=0}^{k-2} 2^i n^{k-2-i} + 2^{k-1}n - 2^{k-1}}{2^{k-1} - \sum_{i=0}^{k-2} 2^i n^{k-2-i}}].$$

Once more, the right-hand side is weakly positive, for small enough s, for any fixed n such that $2^k - \sum_{i=0}^{k-1} 2^i n^{k-1-i} + n^{k-1} > 0$ and $2^{k-1} - \sum_{i=0}^{k-2} 2^i n^{k-2-i} > 0$, delivered by Lemmas 3 and 4.

To argue that every transfer that leads (i_2, j_1) to form a complete-information block at $\mathbf{w}_{k-1}^k$ leaves i_2 strictly better-off at $\mathbf{w}_k^k = (5, 5 - s_k^k)$, with respect to the payoff she obtains with j_2, we can again focus, without loss, on $p^{\epsilon}_{\mathbf{w}_{k-1}^k}$; i.e., we have to show that, for every $\epsilon > 0$

$$2n(5 - s_k^k) + n(5 - s_{k-1}^k) + p + \epsilon > 3n(5 - s_k^k) + p. \tag{2}$$

This is equivalent to

$$n(s_k^k - s_{k-1}^k) + \epsilon > 0.$$

Since $n > 1$ and $s_k^k > s_{k-1}^k$, for every $\epsilon > 0$, we are done.

Thus, it is left to show that, at $\mathbf{w}_k^k := (5, 5 - s_k^k)$, the payoff j_1 obtains matching with i_2 with respect to $p^{\epsilon}_{\mathbf{w}_{k-1}^k}$ is weakly smaller than the payoff it obtains with i_1, for any $\epsilon > 0$; i.e., we need

$$2(5 - s_k^k) + 10 - n(5 - s_{k-1}^k) - p - \epsilon \leq 20,$$

to be true for every $\epsilon > 0$, where p is defined as in (1), with respect to s_1^k. This inequality is equivalent to

$$2(s_1^k - s_k^k) + ns(\frac{n^{k-1}}{2^{k-1} - \sum_{i=0}^{k-2} 2^i n^{k-2-i} > 0} - 1) - \epsilon \leq 0.$$

Suppose, contrary to hypothesis, that this is not true, so that

$$2(s_1^k - s_k^k) + ns(\frac{n^{k-1}}{2^{k-1} - \sum_{i=0}^{k-2} 2^i n^{k-2-i} > 0} - 1) - \epsilon > 0.$$

At this point we use the induction hypothesis; namely, the fact that every transfer that leads (i_2, j_1) to form a complete-information block at $\mathbf{w}_{k-2}^k$ leaves j_1 weakly worse-off at $\mathbf{w}_{k-1}^k = (5, 5 - s_{k-1}^k)$, with respect to the payoff she obtains with j_2 and $p^{\epsilon}_{\mathbf{w}_{k-2}^k}$, which means that

$$2(5 - s_{k-1}^k) + 10 - n(5 - s_{k-2}^k) - p - \epsilon \leq 20,$$

where (the reader can easily check) $s_{k-2}^k = \frac{n^{k-2}}{2^{k-1} - \sum_{i=0}^{k-2} 2^i n^{k-2-i} > 0}$. This inequality is equivalent to

$$2(s_1^k - s_{k-1}^k) + ns(\frac{n^{k-2}}{2^{k-1} - \sum_{i=0}^{k-2} 2^i n^{k-2-i} > 0} - 1) - \epsilon \leq 0.$$

Given the hypothesis above, we then have

$$\begin{aligned}
ns\frac{n^{k-1} - n^{k-2}}{2^{k-1} - \sum_{i=0}^{k-2} 2^i n^{k-2-i}} &> 2(s_k^k - s_{k-1}^k) \\
&\Leftrightarrow \\
ns_{k-1}^k - s_{k-1}^k &> 2s_k^k - 2s_{k-1}^k \\
&\Leftrightarrow \\
(n+1)s_{k-1}^k &> 2s_k^k \\
&\Leftrightarrow \\
\frac{n^k + n^{k-1}}{2^k - 2\sum_{i=0}^{k-2} 2^i n^{k-2-i}} &> \frac{n^k + n^{k-1}}{2^k - \sum_{i=0}^{k-1} 2^i n^{k-1-i} + n^{k-1}} \\
&\Leftrightarrow \\
2\sum_{i=0}^{k-2} 2^i n^{k-2-i} + n^{k-1} &> \sum_{i=0}^{k-1} 2^i n^{k-1-i}.
\end{aligned}$$

This last expression is equivalent to

$$2(n^{k-2}+2n^{k-3}+...+2^{k-2})+n^{k-1} \quad > \quad n^{k-1}+2n^{k-2}+4n^{k-3}+...+2^{k-1},$$

which (the reader can easily check) is in turn equivalent to $0>0$, a contradiction.

The proofs of Lemmas 3 and 4 conclude the proof of Proposition 2.□

Proof of Lemma 3

I show this by induction. If $k=1$, then

$$2^k-\sum_{i=0}^{k-1}2^i n^{k-1-i}+n^{k-1}=2,$$

for every $n\in(1,2]$. Let the inductive hypothesis say that the desired result holds for every value k takes in $\{1,...,x\}$, for some $x>1$, but fails for $k=x+1$. Then, it does so for $n=(1+\epsilon)$, and every $\epsilon>0$; i.e.,

$$\begin{aligned}
2^k-\sum_{i=0}^{k-1}2^i n^{k-1-i} &= 2^k-\sum_{i=0}^{k-2}2^i n^{k-1-i}-2^{k-1}n^{k-1-(k-1)} \\
&= 2^{k-1}-\sum_{i=0}^{k-2}2^i(1+\epsilon)^{k-1-i} \\
&= 2^{k-1}-(1+\epsilon)^{k-1}\sum_{i=0}^{k-2}2^i(1+\epsilon)^{-i} \\
&= \frac{2^{k-1}}{(1+\epsilon)^{k-1}}-[1+\frac{2}{(1+\epsilon)}+\frac{2^2}{(1+\epsilon)^2}+...+\frac{2^{k-2}}{(1+\epsilon)^{k-2}}] \\
&= \frac{2^{x}}{(1+\epsilon)^{x}}-[1+\frac{2}{(1+\epsilon)}+\frac{2^2}{(1+\epsilon)^2}+...+\frac{2^{x-1}}{(1+\epsilon)^{x-1}}]\leq 0,
\end{aligned}$$

where the second equality follows from the fact that $n=1+\epsilon$, the fourth inequality from the fact that $\epsilon>0$, and the last one from the assumption that $k=x+1$. Hence,

$$\frac{2^{x}}{(1+\epsilon)^{x}} \quad \leq \quad 1+\frac{2}{(1+\epsilon)}+\frac{2^2}{(1+\epsilon)^2}+...+\frac{2^{x-1}}{(1+\epsilon)^{x-1}}. \tag{3}$$

Since $\frac{2^{x-1}}{(1+\epsilon)^{x-1}}=\frac{2^{x}}{(1+\epsilon)^{x}}\frac{2^{-1}}{(1+\epsilon)^{-1}}$, it follows that

$$\frac{2^{x-1}}{(1+\epsilon)^{x-1}} \quad \leq \quad 1+\frac{2}{(1+\epsilon)}+\frac{2^2}{(1+\epsilon)^2}+...+\frac{2^{x-2}}{(1+\epsilon)^{x-2}}, \tag{4}$$

for every ϵ, contradicting the inductive hypothesis.□

Proof of Lemma 4

The claim is true for $k = 2$, because both inequalities are positive for every n. Indeed, when $k = 2$

$$2^k - \sum_{i=0}^{k-1} 2^i n^{k-1-i} + n^{k-1} = 4 - (n+2) + n = 2 > 0,$$

and

$$2^{k-1} - \sum_{i=0}^{k-2} 2^i n^{k-2-i} = 2 - 1 = 1 > 0.$$

Hence, suppose that the claim fails for some $k > 1$. Then,

$$\begin{aligned}
2^k - \sum_{i=0}^{k-1} 2^i n^{k-1-i} + n^{k-1} &= 2^{k-1}2 - \sum_{i=0}^{k-2} 2^i n^{k-1-i} - 2^{k-1} + n^{k-1} > 0 \\
&\Leftrightarrow \\
\frac{2^{k-1} + n^{k-1}}{n} &> \sum_{i=0}^{k-2} 2^i n^{k-2-i} \geq 2^{k-1} \\
&\Leftrightarrow \\
2^{k-1} + n^{k-1} &> n2^{k-1} \\
&\Leftrightarrow \\
n^{k-1} &> 2^{k-1}(n-1) > 2^{k-1} \\
&\Leftrightarrow \\
n &> 2
\end{aligned}$$

where the second line uses $n > 1$ and the fact that the desired inequality fails for $k > 1$, the fourth line uses $n > 1$ and $k > 1$, and the fifth one is obtained because $k > 1$. Since this contradicts that $n \in (1, 2]$, we are done.□

8.3 Proof of Theorem 1

I first show that Assumptions 1 and 3 imply that every incomplete-information stable outcome is positive assortative, and then argue that that implies, in turn, that they must also be efficient.

Assume first that the failure corresponds to worker assortativeness, so that there are two workers, i and i', with $\mu_i \neq \emptyset$, such that

$$\mathbf{w}_{i'} > \mathbf{w}_i$$
$$\mathbf{f}_{\mu_{i'}} < \mathbf{f}_{\mu_i}.$$

We can have either $\mu_{i'} = \emptyset$ or $\mu_{i'} \neq \emptyset$. If $\mu_{i'} = \emptyset$, then (i', μ_i) forms a complete-information block at $\mathbf{w}$, because otherwise

$$\begin{aligned} S_{\mathbf{w}_{i'}\mathbf{f}_{\mu_i}} &\leq \pi^{\mathbf{w}}_{i'} + \pi^{\mathbf{f}}_{\mu_i} \\ &= \pi^{\mathbf{f}}_{\mu_i} \\ &= S_{\mathbf{w}_i\mathbf{f}_{\mu_i}} - \pi^{\mathbf{w}}_i \end{aligned}$$

where the second line uses the fact that the payoff of unmatched agents is zero, and the third the fact that the sum of the payoffs of any two agents that are matched exhausts the surplus they create. Since the surplus is strictly increasing in w, by Assumption 1, it follows that

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

contradicting that the given allocation is individually rational. But if (i', μ_i) forms a complete-information block at $\mathbf{w}$, one can invoke Lemma 1 to reach a contradiction. Hence, it must be that $\mu_{i'} \neq \emptyset$. At this point, one uses the fact that the surplus is strictly supermodular, given Assumption 3, to argue that either $(i, \mu_{i'})$ or (i', μ_i) must form a complete-information block at $\mathbf{w}$. To see this, suppose not, so that

$$S_{\mathbf{w}_{i'}\mathbf{f}_{\mu_i}} \leq \pi^{\mathbf{w}}_{i'} + \pi^{\mathbf{f}}_{\mu_i} \text{ and } S_{\mathbf{w}_i\mathbf{f}_{\mu_{i'}}} \leq \pi^{\mathbf{w}}_i + \pi^{\mathbf{f}}_{\mu_{i'}}.$$

Adding up, side to side, one obtains

$$S_{\mathbf{w}_{i'}\mathbf{f}_{\mu_i}} + S_{\mathbf{w}_i\mathbf{f}_{\mu_{i'}}} \leq S_{\mathbf{w}_{i'}\mathbf{f}_{\mu_{i'}}} + S_{\mathbf{w}_i\mathbf{f}_{\mu_i}},$$

which contradicts that the surplus is strictly supermodular, because $\mathbf{w}_{i'} > \mathbf{w}_i$ and $\mathbf{f}_{\mu_i} > \mathbf{f}_{\mu_{i'}}$. Yet (i', μ_i) cannot form a complete-information block at $\mathbf{w}$ without contradicting either Lemma

1 or the hypothesis that the given outcome belongs to $\sum^2$. Thus, $(i, \mu_{i'})$ must form a complete-information block at $\mathbf{w}$. Consider

$$p^{\epsilon}_{\mathbf{w}} := \pi^{\mathbf{w}}_i - \nu_{\mathbf{w}_i \mathbf{f}_{\mu_{i'}}} + \epsilon,$$

$\epsilon > 0$. Intuitively, the reader should think of ϵ as being "small" and, thus, of $p^{\epsilon}_{\mathbf{w}}$ as the "smallest" blocking transfer between i and $\mu_{i'}$, at $\mathbf{w}$. Since the outcome belongs to $\sum^1$, there must be, for every $\epsilon > 0$, some $\tilde{\mathbf{w}} \in \Omega$ with $(\mu, \mathbf{p}, \tilde{\mathbf{w}}, \mathbf{f}) \in \sum^0$ and $\tilde{\mathbf{w}}_{i'} = \mathbf{w}_{i'}$ such that

$$\phi_{\tilde{\mathbf{w}}i}\mathbf{f}_{\mu_{i'}} - p^{\epsilon}_{\mathbf{w}_i} \leq \phi_{\mathbf{w}_i \mathbf{f}_{\mu_{i'}}} - \mathbf{p}_{\mu_{i'},i'} \text{ and } \nu_{\tilde{\mathbf{w}}_i,\mathbf{f}_j} + p^{\epsilon}_{\mathbf{w}_i} > \nu_{\tilde{\mathbf{w}}_i \mathbf{f}_{\mu_i}} + \mathbf{p}_{i,\mu_i}. \quad (5)$$

Since $\Omega \subseteq W^{|I|}$ and W is finite, we must in fact have some $\tilde{\mathbf{w}} \in \Omega$ with $(\mu, \mathbf{p}, \tilde{\mathbf{w}}, \mathbf{f}) \in \sum^0$ and $\tilde{\mathbf{w}}_{i'} = \mathbf{w}_{i'}$ such that (5) is satisfied for every $\epsilon > 0$. Notice that

$$\begin{aligned}
\phi_{\tilde{\mathbf{w}}_i}\mathbf{f}_{\mu_{i'}} - p^{\epsilon}_{\mathbf{w}_i} &\leq \phi_{\mathbf{w}_{i'}\mathbf{f}_{\mu_{i'}}} - \mathbf{p}_{\mu_{i'},i'} \\
&\Leftrightarrow \\
\phi_{\tilde{\mathbf{w}}_i}\mathbf{f}_{\mu_{i'}} - \pi^{\mathbf{w}}_i + \nu_{\mathbf{w}_i \mathbf{f}_{\mu_{i'}}} - \epsilon &\leq \phi_{\mathbf{w}_{i'}\mathbf{f}_{\mu_{i'}}} - \mathbf{p}_{\mu_{i'},i'} \\
&\Leftrightarrow \\
\phi_{\tilde{\mathbf{w}}_i}\mathbf{f}_{\mu_{i'}} - \nu_{\mathbf{w}_i \mathbf{f}_{\mu_i}} - \mathbf{p}_{i,\mu_i} + \nu_{\mathbf{w}_i \mathbf{f}_{\mu_{i'}}} - \epsilon &\leq \phi_{\mathbf{w}_{i'}\mathbf{f}_{\mu_{i'}}} - \mathbf{p}_{\mu_{i'},i'} \\
&\Leftrightarrow \\
\mathbf{p}_{\mu_{i'},i'} - \mathbf{p}_{i,\mu_i} - \epsilon &\leq \phi_{\mathbf{w}_{i'}\mathbf{f}_{\mu_{i'}}} - \phi_{\tilde{\mathbf{w}}_i \mathbf{f}_{\mu_{i'}}} + \nu_{\mathbf{w}_i \mathbf{f}_{\mu_i}} - \nu_{\mathbf{w}_i \mathbf{f}_{\mu_{i'}}}.
\end{aligned}$$

For ϵ small enough, this last inequality is equivalent to

$$\mathbf{p}_{\mu_{i'},i'} - \mathbf{p}_{i,\mu_i} \leq \phi_{\mathbf{w}_{i'}\mathbf{f}_{\mu_{i'}}} - \phi_{\tilde{\mathbf{w}}_i \mathbf{f}_{\mu_{i'}}} + \nu_{\mathbf{w}_i \mathbf{f}_{\mu_i}} - \nu_{\mathbf{w}_i \mathbf{f}_{\mu_{i'}}}. \quad (6)$$

Since $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \sum^2$, we can assume, without loss of generality, that $(\mu, \mathbf{p}, \tilde{\mathbf{w}}, \mathbf{f}) \in \sum^1$. Hence, Lemma 1 implies that (i', μ_i) cannot form a complete-information block at $\tilde{\mathbf{w}}$; i.e.,

$$\begin{aligned}
S_{\tilde{\mathbf{w}}_{i'}\mathbf{f}_{\mu_i}} &\leq \pi^{\tilde{\mathbf{w}}}_{i'} + \pi^{\mathbf{f}}_{\mu_i} \\
&\Leftrightarrow \\
\nu_{\tilde{\mathbf{w}}_{i'}\mathbf{f}_{\mu_i}} - \nu_{\tilde{\mathbf{w}}_{i'}\mathbf{f}_{\mu_{i'}}} + \phi_{\tilde{\mathbf{w}}_{i'}\mathbf{f}_{\mu_i}} - \phi_{\tilde{\mathbf{w}}_i \mathbf{f}_{\mu_i}} &\leq \mathbf{p}_{\mu_{i'},i'} - \mathbf{p}_{i,\mu_i},
\end{aligned}$$

where the second line uses the fact that $\mathbf{p}_{\mu_{i'},i'} = \mathbf{p}_{i',\mu_{i'}}$ and $\mathbf{p}_{i,\mu_i} = \mathbf{p}_{\mu_i,i}$. Since $\tilde{\mathbf{w}}_{i'} = \mathbf{w}_{i'}$, this is equivalent to

$$\nu_{\mathbf{w}_{i'}\mathbf{f}_{\mu_i}} - \nu_{\mathbf{w}_{i'}\mathbf{f}_{\mu_{i'}}} + \phi_{\mathbf{w}_{i'}\mathbf{f}_{\mu_i}} - \phi_{\tilde{\mathbf{w}}_i\mathbf{f}_{\mu_i}} \quad \leq \quad \mathbf{p}_{\mu_{i'},i'} - \mathbf{p}_{i,\mu_i}. \tag{7}$$

If we put (6) and 7 together, we get

$$\nu_{\mathbf{w}_{i'}\mathbf{f}_{\mu_i}} - \nu_{\mathbf{w}_{i'}\mathbf{f}_{\mu_{i'}}} + \phi_{\mathbf{w}_{i'}\mathbf{f}_{\mu_i}} - \phi_{\tilde{\mathbf{w}}_i\mathbf{f}_{\mu_i}} \quad \leq \quad \phi_{\mathbf{w}_{i'}\mathbf{f}_{\mu_{i'}}} - \phi_{\tilde{\mathbf{w}}_i\mathbf{f}_{\mu_{i'}}} + \nu_{\mathbf{w}_i\mathbf{f}_{\mu_i}} - \nu_{\mathbf{w}_i\mathbf{f}_{\mu_{i'}}}.$$

Re-arranging, we obtain

$$\nu_{\mathbf{w}_{i'}\mathbf{f}_{\mu_i}} - \nu_{\mathbf{w}_{i'}\mathbf{f}_{\mu_{i'}}} + \nu_{\mathbf{w}_i\mathbf{f}_{\mu_{i'}}} - \nu_{\mathbf{w}_i\mathbf{f}_{\mu_i}} + \phi_{\mathbf{w}_{i'}\mathbf{f}_{\mu_i}} - \phi_{\tilde{\mathbf{w}}_i\mathbf{f}_{\mu_i}} + \phi_{\tilde{\mathbf{w}}_i\mathbf{f}_{\mu_{i'}}} - \phi_{\mathbf{w}_{i'}\mathbf{f}_{\mu_{i'}}} \quad \leq \quad 0.$$

Clearly,

$$\nu_{\mathbf{w}_{i'}\mathbf{f}_{\mu_i}} - \nu_{\mathbf{w}_{i'}\mathbf{f}_{\mu_{i'}}} + \nu_{\mathbf{w}_i\mathbf{f}_{\mu_{i'}}} - \nu_{\mathbf{w}_i\mathbf{f}_{\mu_i}} > 0,$$

because $\mathbf{w}_{i'} > \mathbf{w}_i$ and $\mathbf{f}_{\mu_i} > \mathbf{f}_{\mu_{i'}}$, and ν_{wf} is strictly supermodular, by Assumption 3. Hence, we must have

$$\phi_{\mathbf{w}_{i'}\mathbf{f}_{\mu_i}} - \phi_{\tilde{\mathbf{w}}_i\mathbf{f}_{\mu_i}} + \phi_{\tilde{\mathbf{w}}_i\mathbf{f}_{\mu_{i'}}} - \phi_{\mathbf{w}_{i'}\mathbf{f}_{\mu_{i'}}} \quad < \quad 0.$$

Yet this contradicts that ϕ_{wf} is weakly supermodular, by Assumption 3, because $\mathbf{w}_{i'} > \tilde{\mathbf{w}}_i$ and $\mathbf{f}_{\mu_i} > \mathbf{f}_{\mu_{i'}}$. This completes the proof that all outcomes in $\sum^2$ must be worker assortative.

If the failure of positive assortativeness is due to a failure of firm assortativeness, there would be two firms, j and j', with $\mu_j \neq \emptyset$, such that

$$\mathbf{f}_{j'} > \mathbf{f}_j$$
$$\mathbf{w}_{\mu_{j'}} < \mathbf{w}_{\mu_j}.$$

As before, we can have either $\mu_{j'} = \emptyset$ or $\mu_{j'} \neq \emptyset$, but the latter would imply a failure of worker assortativeness, so assume that $\mu_{j'} = \emptyset$ is the case.

The strict monotonicity of S_{wf} with respect to f imposed by Assumption 1 implies that (μ_j, j') must form a complete-information block at $\mathbf{w}$. To see this suppose not; i.e., assume that

$$S_{\mathbf{w}_{\mu_j}\mathbf{f}_{j'}} \leq \pi^{\mathbf{w}}_{\mu_j} + \pi^{\mathbf{f}}_{j'} = \pi^{\mathbf{w}}_{\mu_j},$$

where the equality follows from $\mu_{j'} = \emptyset$. Since $\pi^{\mathbf{w}}_{\mu_j} = S_{\mathbf{w}_{\mu_j}\mathbf{f}_j} - \pi^{\mathbf{f}}_j$, $\mathbf{f}_{j'} > \mathbf{f}_j$, and S_{wf} strictly increases with f, it follows that $\pi^{\mathbf{f}}_j < 0$, contradicting individual rationality.

We can now consider

$$p^{\epsilon}_{\mathbf{w}_{\mu_j}} := \nu_{\mathbf{w}_{\mu_j}\mathbf{f}_j} + \mathbf{p}_{\mu_j,j} - \nu_{\mathbf{w}_{\mu_j}\mathbf{f}_{j'}} + \epsilon.$$

Since $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \sum^1$, there must be, for every $\epsilon > 0$, some $\mathbf{w}' \in \Omega$, with $(\mu, \mathbf{p}, \mathbf{w}', \mathbf{f}) \in \sum^0$ such that $\nu_{\mathbf{w}'_{\mu_j}\mathbf{f}_{j'}} + p^{\epsilon}_{\mathbf{w}_{\mu_j}} > \pi^{\mathbf{w}'}_{\mu_j}$ and $\phi_{\mathbf{w}'_{\mu_j}\mathbf{f}_{j'}} - p^{\epsilon}_{\mathbf{w}_{\mu_j}} \leq 0$, because $\mu_{j'} = \emptyset$. Since $\Omega \subseteq W^{|I|}$ and W is finite, we must in fact have one such $\mathbf{w}'$ for every $\epsilon > 0$.

Notice that $\phi_{\mathbf{w}'_{\mu_j}\mathbf{f}_{j'}} - p^{\epsilon}_{\mathbf{w}_{\mu_j}} \leq 0$ implies that

$$\phi_{\mathbf{w}_{\mu_j}\mathbf{f}_{j'}} > \phi_{\mathbf{w}'_{\mu_j}\mathbf{f}_{j'}}, \tag{8}$$

whereas the former implies that

$$\nu_{\mathbf{w}'_{\mu_j}\mathbf{f}_{j'}} + \nu_{\mathbf{w}_{\mu_j}\mathbf{f}_j} - \nu_{\mathbf{w}_{\mu_j}\mathbf{f}_{j'}} - \nu_{\mathbf{w}'_{\mu_j}\mathbf{f}_j} \geq 0. \tag{9}$$

Given that $\mathbf{f}_{j'} > \mathbf{f}_j$, the strict supermodularity of ν_{wf} granted by Assumption 3 implies that (9) is satisfied if and only if $\mathbf{w}'_{\mu_j} \geq \mathbf{w}_{\mu_j}$. Yet (8) implies, by Assumption 1, that $\mathbf{w}'_{\mu_j} < \mathbf{w}_{\mu_j}$, a contradiction.

As argued in Section 6, one can show that every $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \sum^2$ is efficient by using Lemma 2.□

8.4 Proof of Proposition 3

Take any $(\mu, \mathbf{p}, \mathbf{w}, \mathbf{f}) \in \sum^1$ such that $\mathbf{p}_{j'} \geq \mathbf{p}_j$ for every j, j' such that $\mathbf{f}_{j'} > \mathbf{f}_j$, but suppose, contrary to hypothesis that μ is not positive assortative at $\mathbf{w}$. If there is a failure of worker assortativeness, there are two workers, i and i', with $\mu_i \neq \emptyset$, such that

$$\mathbf{w}_{i'} > \mathbf{w}_i$$
$$\mathbf{f}_{\mu_{i'}} < \mathbf{f}_{\mu_i}.$$

We must have $\mu_{i'} \neq \emptyset$, because otherwise the same argument made in the proof of Theorem 1 would lead to a contradiction with the assumption that the given outcome belongs to $\sum^1$. As in the proof of Theorem 1, it follows by Lemma 1 that, at $\mathbf{w}$, $(i, \mu_{i'})$ forms a complete-information block, but (i', μ_i) does not. Hence, we have

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

which is equivalent to

$$\nu_{\mathbf{w}_i \mathbf{f}_{\mu_{i'}}} - \nu_{\mathbf{w}_i \mathbf{f}_{\mu_i}} + \phi_{\mathbf{w}_i \mathbf{f}_{\mu_{i'}}} - \phi_{\mathbf{w}_{i'} \mathbf{f}_{\mu_{i'}}} > \mathbf{p}_{i,\mu_i} - \mathbf{p}_{\mu_{i'},i'}.$$

But since ν_{wf} is weakly increasing in f and ϕ_{wf} is weakly increasing in w, by Assumption 1, the left-hand side of this strict inequality is weakly negative, contradicting that $\mathbf{p}_{i,\mu_i} - \mathbf{p}_{\mu_{i'},i'}$, because $\mathbf{f}_{\mu_i} > \mathbf{f}_{\mu_{i'}}$.□

References

Alaoui, L., Janezic, K. A., and Penta, A. (2020). Reasoning about others' reasoning. *Journal of Economic Theory*, 189:105091.

Alaoui, L. and Penta, A. (2016). Endogenous depth of reasoning. *The Review of Economic Studies*, 83(4):1297–1333.

Bayer, R. C. and Renou, L. (2016). Logical omniscience at the laboratory. *Journal of Behavioral and Experimental Economics*, 64:41–49.

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

Bloom, N., Ohlmacher, S. W., Tello-Trillo, C. J., and Wallskog, M. (2021). Pay, productivity and management. Technical report, National Bureau of Economic Research.

Börgers, T. and Li, J. (2019). Strategically simple mechanisms. *Econometrica*, 87(6):2003–2035.

Charness, G. and Levin, D. (2009). The origin of the winner's curse: a laboratory study. *American Economic Journal: Microeconomics*, 1(1):207–236.

Chen, Y.-C. and Ho Cher Sien, S. (2020). Efficiency of stable matching with two-sided incomplete information. *Efficiency of Stable Matching with Two-Sided Incomplete Information (February 18, 2020)*.

Crawford, V. P. and Knoer, E. M. (1981). Job matching with heterogeneous firms and workers. *Econometrica: Journal of the Econometric Society*, pages 437–450.

Dizdar, D. and Moldovanu, B. (2016). On the importance of uniform sharing rules for efficient matching. *Journal of Economic Theory*, 165:106–123.

Esponda, I. (2008). Behavioral equilibrium in economies with adverse selection. *American Economic Review*, 98(4):1269–1291.

Ho, T.-H., Camerer, C., and Weigelt, K. (1998). Iterated dominance and iterated best response in experimental" p-beauty contests". *The American Economic Review*, 88(4):947–969.

Kneeland, T. (2015). Identifying higher-order rationality. *Econometrica*, 83(5):2065–2079.

Li, S. (2017). Obviously strategy-proof mechanisms. *American Economic Review*, 107(11):3257–3287.

Liu, Q., Mailath, G. J., Postlewaite, A., and Samuelson, L. (2014). Stable matching with incomplete information. *Econometrica*, 82(2):541–587.

Mailath, G. J., Postlewaite, A., and Samuelson, L. (2013). Pricing and investments in matching markets. *Theoretical Economics*, 8(2):535–590.

Mailath, G. J., Postlewaite, A., and Samuelson, L. (2017). Premuneration values and investments in matching markets. *The Economic Journal*, 127(604):2041–2065.

Pathak, P. A. and Sönmez, T. (2008). Leveling the playing field: Sincere and sophisticated players in the boston mechanism. *American Economic Review*, 98(4):1636–1652.

Peralta, E. (2023). Efficiency, sorting, and selection. *Working paper, University of Michigan.*

Rubinstein, A. (1989). The electronic mail game: Strategic behavior under" almost common knowledge". *The American Economic Review*, pages 385–391.

Shapley, L. S. and Shubik, M. (1971). The assignment game i: The core. *International Journal of game theory*, 1:111–130.