

Efficiency of stable matchings with two-sided incomplete information

YI-CHUN CHEN

Department of Economics and Risk Management Institute, National University of Singapore

SAMUEL HO

Department of Economics, Yale University

We uncover an asymmetry in stable matching with two-sided incomplete information: Incomplete-information-stable matching outcomes must be efficient (ex post surplus-maximizing), when the agents' premuneration values are submodular but not necessarily when their values are supermodular. We show that efficiency of incomplete-information-stable matching hinges on a condition that we call *misassignment deduction*. The condition requires that for each potential blocking pair of a mismatch, one of the two agents' worst-case hypotheses contains only type assignments under which their status quo is a mismatch.

KEYWORDS. Stable matching, incomplete information, two-sided incomplete information, incomplete-information stability, premuneration values, supermodularity, submodularity, misassignment deduction.

JEL CLASSIFICATION. C78, D40, D82.

1. INTRODUCTION

A milestone result in the literature of two-sided matching is the equivalence between efficiency and stability, which dates back to [Koopmans and Beckmann \(1957\)](#) and [Shapley and Shubik \(1971\)](#). The result says that a matching is (ex post) surplus-maximizing if and only if the agents' surplus can be split in a way which admits no blocking pair. The equivalence is a consequence of linear-programming duality. To relate the dual constraints to stability, each agent on each side of the market must have complete information about the agents *on the other side*. Specifically, they must be able to evaluate how much they would gain from each potential blocking prospect. On one hand, the classical equivalence result does not require that the agents know the types of other agents on the same side. On the other hand, once there is cross-side incomplete information, the agents might not carry out a blocking prospect that is *de facto* mutually beneficial, so an inefficient matching outcome may become stable.

Yi-Chun Chen: ecsycc@nus.edu.sg

Samuel Ho: samuel.hochersien@yale.edu

We are especially grateful to Larry Samuelson for his guidance and encouragement. We also thank Gaoji Hu, Ryota Iijima, Qingmin Liu, Kai Hao Yang, and Ziwei Wang for their insightful comments as well as participants at Academia Sinica and the Yale Micro Theory Lunch for helpful suggestions. All errors are our own.

Liu et al. (2014) (hereafter, LMPS) shows that, strikingly, the efficiency of stable matching outcomes can be restored as long as the firms' types are commonly known to *everyone* in the market, and moreover, workers' premuneration values (match pay-offs before transfers) and match surplus are *either* all supermodular *or* all submodular in the agents' types. A matching outcome is blocked when a firm obtains a strictly higher payoff for *every* type of worker who strictly benefits from being rematched with the firm. Since the workers' values are supermodular (submodular) and each firm knows the other firms' type, a better (worse) firm can accurately determine a "separating wage" which only a worker of sufficiently high type will accept to form a block.

In this paper, we consider the matching setting of LMPS, yet we symmetrize their informational assumptions and notions of blocking under incomplete information.¹ Maintaining the assumption that each agent knows their own type and their partner's type, we assume that both firms and workers have complete information within their side of the market but incomplete information across the two sides. When agents on different sides of the market belong to different industries or sectors, it is conceivable that agents have better information about others on the same side than their potential partners on the other side. It is also conceivable that the agents' characteristics may be publicly observable and yet agents on the same side have better expertise to assess the relevance of their characteristics to a match. As a blocking pair typically concerns four agents (a pair of agents and their current partners), our assumptions imply that each agent is *only* uncertain about the type of their blocking partner. In this symmetric setting, we uncover a novel asymmetry: When agents' premuneration values are submodular, every stable matching outcome is efficient (Theorem 1). On the other hand, when the agents' premuneration values are supermodular, a stable matching outcome need not be efficient (Example 1).

To understand the asymmetry, we demonstrate that the efficiency of matching outcomes that are incomplete-information-stable hinges on a nonprimitive condition that we call *misassignment deduction*. The misassignment deduction condition requires that for any potential blocking pair of a "mismatch," at least one agent k has worst-case hypotheses that either never assign a type worse than the true type of the other agent k' , or are type assignments under which the matching remains a mismatch. Our main result shows that under this misassignment deduction condition, together with supermodularity or submodularity, every stable matching outcome is efficient (Theorem 2). This result clarifies the fact that inefficient stable matching outcomes do not arise due to incomplete information across sides or the agents' worst-case consideration per se, but rather due to a worst-case hypothesis under which the matching is no longer a mismatch.

Under one-sided incomplete information, the workers know the firms' types; hence, the misassignment deduction condition holds whether their values are supermodular or submodular. Under two-sided incomplete information, the misassignment deduction condition holds under submodularity but not under supermodularity. Intuitively, an asymmetry in this aspect is due to which agents would satisfy misassignment deduction. Under submodularity, among two pairs of agents who are mismatched, i.e., not

¹Remark 1 offers a more precise discussion of our notion of blocking and stability.

pairwise negatively assortative, the better firm and the better worker in the mismatch must have worst-case hypotheses which themselves violate pairwise negative assortativity. In contrast, the violation of pairwise positive assortativity need not be preserved under the worst-case hypotheses of the mismatched (under supermodularity) better firm or better worker. We demonstrate how the misassignment deduction condition can be restored under supermodularity by imposing additional primitive assumptions such as (i) having the distribution of firm types be common knowledge, or (ii) having the variation in workers' remuneration values across different firm types be small relative to the variation of firms' remuneration value across different worker types. Section 6.1 discusses the role of within-side information, and examples in Appendix A.5 demonstrate the necessity of within-side information for efficiency.

While economic settings with supermodularity have been studied extensively, submodular settings have also received attention in the literature. [Eeckhout \(2018\)](#) surveys various examples including low-skilled salespeople and coveted executives in superstores, matching among risk-averse agents ([Schulhofer-Wohl \(2006\)](#)), and ex ante surplus through optimally chosen joint actions ([Chade and Eeckhout \(2018\)](#)).² To the extent that stability serves as a descriptive notion of matching, our result suggests that transparency of within-side information can promote efficiency in submodular settings—but not necessarily in supermodular settings.

Our main result follows from our key observation that two conditions together suffice for a complete-information blocking pair to block a matching outcome under incomplete information (Lemma 3): The first condition requires that one agent k of the potential blocking pair deduces the type of the other agent k' , i.e., any hypothesis of k assigns to k' a type weakly greater than the true type of k' . The second condition requires that k' either also deduces the type of k , or can utilize a separating wage such that the type of k is sufficiently high were k to participate in the potential block. Crucially, we show that for some stable mismatch (if any exist), whoever satisfies misassignment deduction deduces the type of their blocking partner (Proposition 1). Our main result is implied by Proposition 1 in combination with the fact that under either supermodularity or submodularity, the other agent of the potential blocking pair will either deduce the agent's type or can use a separating wage.

The rest of the paper is organized as follows. Section 2 sets up the matching model. Section 3 introduces the two-sided incomplete-information environment along with the notions of within-side information, separating wages, and misassignment deduction. Section 4 presents our efficiency results. Section 5 proposes two primitive assumptions, each of which implies the misassignment deduction condition and thus restores the efficiency result under supermodularity. Finally, Section 6 discusses how our results also shed light on matching with one-sided incomplete information and relate to other literature.

²([Grossman and Maggi, 2000](#), p. 1259) also discuss examples such as performing arts, research activities, legal and consulting services, and architectural and fashion designs.

2. THE MODEL

We consider the following setup of two-sided matching, which is based on LMPS. Let I be a finite set of workers and J a finite set of firms. Denote a typical worker by i and a typical firm by j . We also denote an agent (that can be either a worker or a firm) by $k \in K = I \cup J$, and an agent, on the other side of k (unless otherwise stated), as $k' \in K$, namely, $k' \in I$ if $k \in J$ and $k' \in J$ if $k \in I$. The index of each agent k is publicly observed.

Define $\mu : K \rightarrow K \cup \{\emptyset\}$ as the matching function that assigns to each agent k at most one partner on the other side. An agent k is matched to k' if $\mu(k) = k'$ (or equivalently, $\mu(k') = k$) and unmatched if $\mu(k) = \emptyset$. Let $\bar{T} \subset W^{|I|} \times F^{|J|} \subset \mathbb{R}^{|K|}$ be a finite set of type assignments, namely, a type assignment $\mathbf{t} \in \bar{T}$ assigns each worker a type from W and each firm a type from F . We denote by $\mathbf{t}_k$ the type assigned to agent k under $\mathbf{t}$.

In a match between a worker of type w and firm of type f , the worker and the firm receive the premuneration values $\zeta(w, f)$ and $\zeta(f, w)$, respectively, before they receive/make wage transfers.³ The premuneration function $\zeta : \mathbb{R}^2 \rightarrow \mathbb{R}$ is commonly known. A wage profile associated with a matching function μ is a vector $\mathbf{p}$ under which $\mathbf{p}_{k, \mu(k)} = -\mathbf{p}_{\mu(k), k}$ denotes the transfer to k from $\mu(k)$. A matching $(\mu, \mathbf{p})$ consists of a matching function μ and a wage profile $\mathbf{p}$ associated with μ , both of which are assumed to be publicly observable. A *matching outcome* consists of a $(\mu, \mathbf{p})$ and a type assignment $\mathbf{t}$. Under a matching outcome $(\mu, \mathbf{p}, \mathbf{t})$, agent k receives $\zeta(\mathbf{t}_k, \mathbf{t}_{\mu(k)}) + \mathbf{p}_{k, \mu(k)}$ and we require that $\zeta(\mathbf{t}_k, \mathbf{t}_{\mu(k)}) = \mathbf{p}_{k, \mu(k)} = 0$ if k is unmatched.

Going forward, we maintain the following:

ASSUMPTION 0. *We impose the following assumptions throughout the paper:*

- (a) *Value Monotonicity:* ζ is strictly increasing on $\mathbb{R}^2$.
- (b) *Positive Surplus:* $\zeta(w, f) + \zeta(f, w) > 0$, $\forall w \in W, \forall f \in F$.

Assumption 0(a) is slightly stronger than Assumption 1 in LMPS to account for our symmetric two-sided incomplete-information setting. Assumption 0(b) rules out cases of the inefficiency of stable outcomes due to either an “over-match” or an “under-match”; see Remark 6 in Section 3.1 for details.

Our main results are concerned with the efficiency of a matching outcome, which we define as follows:

DEFINITION 1 (Efficiency). A matching outcome $(\mu, \mathbf{p}, \mathbf{t})$ is *efficient* if μ maximizes the total surplus under $\mathbf{t}$, namely, μ solves $\max_{\hat{\mu}} \sum_{i \in I} \zeta(\mathbf{t}_i, \mathbf{t}_{\hat{\mu}(i)}) + \zeta(\mathbf{t}_{\hat{\mu}(i)}, \mathbf{t}_i)$.

2.1 Private Information and Individual Rationality

Because we assume the matching $(\mu, \mathbf{p})$ is publicly observable, we henceforth fix an arbitrary $(\mu, \mathbf{p})$ which we will refer to as *the matching*. Agents know their own types, as well

³This formulation does not require that the premuneration functions be symmetric between worker and firm types, since we allow the sets W and F to be disjoint. In the examples that appear subsequently, however, the premuneration functions are indeed symmetric, for simplicity.

as the types of their partners. We summarize their private information in the following definition:

DEFINITION 2 (Private Information). *Private Information* for k at $\mathbf{t}$ is denoted by a subset of type assignments $\Pi_k(\mathbf{t}) \subset \bar{T}$, where $\Pi_k(\mathbf{t}) \subset \{\tilde{\mathbf{t}} \in \bar{T} : \tilde{\mathbf{t}}_k = \mathbf{t}_k, \tilde{\mathbf{t}}_{\mu(k)} = \mathbf{t}_{\mu(k)}\}$.

We will use the abbreviation (PI) as shorthand for the following implication: for any $\tilde{\mathbf{t}} \in \Pi_k(\mathbf{t})$, $\tilde{\mathbf{t}}_k = \mathbf{t}_k$, $\tilde{\mathbf{t}}_{\mu(k)} = \mathbf{t}_{\mu(k)}$.

Note that Definition 2 accommodates different assumptions about the private information that the agents may have access to. For example, if we were to assume one-sided incomplete information (where firms' types are commonly known), we would impose the following restriction: for each $\mathbf{t} \in \bar{T}$ and each agent k , $\Pi_k(\mathbf{t}) \subset \{\tilde{\mathbf{t}} \in \bar{T} : \tilde{\mathbf{t}}_j = \mathbf{t}_j \forall j \in J, \tilde{\mathbf{t}}_k = \mathbf{t}_k, \tilde{\mathbf{t}}_{\mu(k)} = \mathbf{t}_{\mu(k)}\}$. As another example, for a worker i who knows other workers' types, we would impose the following restriction: $\Pi_i(\mathbf{t}) \subset \{\tilde{\mathbf{t}} \in \bar{T} : \tilde{\mathbf{t}}_{i'} = \mathbf{t}_{i'} \forall i' \in I, \tilde{\mathbf{t}}_i = \mathbf{t}_i, \tilde{\mathbf{t}}_{\mu(i)} = \mathbf{t}_{\mu(i)}\}$. We invoke a similar idea to define within-side information in Section 3.1.

Since the agents know their own type as well as their partner's type, our notion of *individual rationality* coincides with the standard notion under complete information.

DEFINITION 3 (Individual Rationality). A matching outcome $(\mu, \mathbf{p}, \mathbf{t})$ is *individually rational* if $\zeta(\mathbf{t}_k, \mathbf{t}_{\mu(k)}) + \mathbf{p}_{k, \mu(k)} \geq 0$ for all $k \in K$.

2.2 Consideration Set and Blocking

We now define the set of type assignments which an agent regards as possible in evaluating a blocking opportunity. This is what we call a *consideration set*. While private information is exogenously given, information can be endogenously deduced through observing that a matching is “not blocked.” That is, certain type assignments may be regarded as “impossible” since they contradict the agent's observation that a status quo matching is unblocked. To accommodate such endogenous information, we now define the notion of a consideration set with respect to an arbitrary subset of type assignments $T \subset \bar{T}$:

DEFINITION 4 (Consideration Set). Fix a set of type assignments $T \subset \bar{T}$, and a type assignment $\mathbf{t} \in T$. Define agent k 's *T-consideration set* at $\mathbf{t}$ as

$$\alpha_k(\mathbf{t}; T) = \Pi_k(\mathbf{t}) \cap T.$$

In words, a type assignment belongs to agent k 's T -consideration set at $\mathbf{t}$ if it is consistent with both T and agent k 's private information $\Pi_k(\mathbf{t})$.

Throughout the paper, we denote a typical (blocking) triplet as (k, k', p) which comprises (i) a pair of agents $k, k' \in K$ who are on opposite sides, and (ii) a blocking wage $p = (p_{k, k'}, p_{k', k})$, where $p_{k, k'} = -p_{k', k} \in \mathbb{R}$ stands for the transfer that agent k receives from agent k' . When evaluating such a block, however, agent k may refine their consideration set by restricting their attention to type assignments for which the other agent k' would strictly benefit in the block. We define this consideration set as follows:

DEFINITION 5 (Rational Consideration Set). Fix $\mathbf{t} \in \bar{T}$, and a blocking triple (k, k', p) . The T -rational consideration set of k' with respect to (k, k', p) is:

$$\beta_{k',k}(\mathbf{t}; T, p) = \{\tilde{\mathbf{t}} \in \alpha_{k'}(\mathbf{t}; T) : \zeta(\tilde{\mathbf{t}}_k, \tilde{\mathbf{t}}_{k'}) + p_{k,k'} > \zeta(\tilde{\mathbf{t}}_k, \tilde{\mathbf{t}}_{\mu(k)}) + \mathbf{p}_{k,\mu(k)}\}.$$

In words, the T -rational consideration set of an agent k' with respect to (k, k', p) consists of type assignments under which k will secure a *de facto benefit* from blocking with k' at p .

EXAMPLE 1. Consider the matching outcome $(\mu, \mathbf{p}, \mathbf{t})$ as depicted in Figure 1. Suppose $W = F = \{1, 2, 3\}$, and $\bar{T} = W^2 \times F^2$, and supermodular premuneration functions $\zeta(w, f) = \zeta(f, w) = wf + 2$.

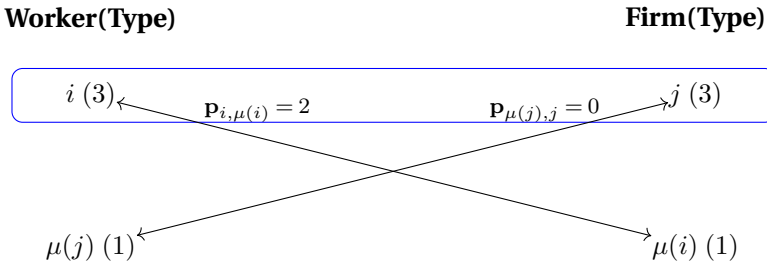


FIGURE 1. $(\mu, \mathbf{p}, \mathbf{t})$

In a potential blocking pair (i, j) with a wage -2 , the rational consideration set of firm j cannot contain type assignments where i is of a type 2 or worse than 2: If worker i were of type 2, i 's premuneration value increases by only 4 from the rematch with j which is exactly offset by i 's wage difference. $\diamond$

We are now ready to state our definition of blocking in this model.

DEFINITION 6. Say that the triplet (k, k', p) T -block the matching $(\mu, \mathbf{p})$ at $\mathbf{t} \in T$ if

$$\zeta(\tilde{\mathbf{t}}_k, \tilde{\mathbf{t}}_{k'}) + p_{k,k'} > \zeta(\tilde{\mathbf{t}}_k, \tilde{\mathbf{t}}_{\mu(k)}) + \mathbf{p}_{k,\mu(k)} \quad (1)$$

for all $\tilde{\mathbf{t}} \in \alpha_k(\mathbf{t}; T)$; and

$$\zeta(\tilde{\mathbf{t}}_{k'}, \tilde{\mathbf{t}}_k) + p_{k',k} > \zeta(\tilde{\mathbf{t}}_{k'}, \tilde{\mathbf{t}}_{\mu(k')}) + \mathbf{p}_{k',\mu(k')} \quad (2)$$

for all $\tilde{\mathbf{t}} \in \beta_{k',k}(\mathbf{t}; T, p)$. A matching outcome $(\mu, \mathbf{p}, \mathbf{t})$ is T -blocked if some triplet (k, k', p) T -block the matching $(\mu, \mathbf{p})$ at $\mathbf{t}$. We call (k, k') a T -blocking pair at $\mathbf{t}$ if for some p , (k, k', p) T -block the matching at $\mathbf{t}$.

We offer the following interpretation: A block must be initiated by at least one agent, say, k . The proposer k (which can either be a worker or a firm) is certain that a blocking with k' is strictly beneficial, i.e., strictly beneficial for all type assignments in their consideration set $\alpha_k(\mathbf{t}; T)$.⁴

If k' is also certain that the blocking is strictly beneficial, then the block takes place. Otherwise, upon receiving the blocking proposal from k , k' can refine their evaluation of the block by conditioning on the willingness of k to block, i.e., k' considers $\beta_{k',k}(\mathbf{t}; T, p)$. Definition 6 maintains flexibility of the consideration set refinement, provided that k' regards the willingness to block of k as k being able to secure a *de facto* benefit by being rematched with k' . This refinement amounts to requiring that the block be strictly beneficial to k' for all type assignments in k' 's rational consideration set. Any more permissive notion of blocking shrinks the set of stable matching outcomes that we are about to define, so our efficiency results (regarding stable outcomes) will still hold.

REMARK 1. Under one-sided incomplete information, i.e., when firms' types are public, a worker only considers the true type of the firm. As a result, the firm's rational consideration set includes only type assignments under which the worker benefits from the firm's true type. In other words, for $k = i$ and $k' = j$, the blocking notion in Definition 6 coincides with the one of LMPS. Under two-sided incomplete-information, our notion coincides with the outcome of a blocking game with due-diligence when agents play undominated strategies, and mutual disagreements on blocks can be sustained; we formally describe this game and continue this discussion in Section B.6 of [Chen and Ho \(2026\)](#).

REMARK 2. [Chen and Hu \(2023\)](#) proposes another “worst-case” notion of blocking which also extends LMPS' blocking notion to two-sided incomplete information. In their model, agents consider type assignments consistent with “common knowledge of mutual willingness” to block with their potential blocking partner,⁵ whereas in our blocking notion, agents rule out type assignments that do not support a *de facto* benefit of their blocking partner. In general, results in this paper do not carry over to the blocking notion of [Chen and Hu \(2023\)](#); see also Remark 5, and a more formal comparison of the notions in Section B.5 of [Chen and Ho \(2026\)](#). [Park \(2024\)](#) also considers a “worst-case” blocking notion under which our results do transfer; see Section 6.2 for further discussion.

⁴[Park \(2024\)](#) also considers a similar notion and interpretation of blocking; the proposer is certain that the block is strictly beneficial, and the proposee uses this deviating offer to evaluate the plausible type-assignments.

⁵([Chen and Hu, 2023](#), p. 307) describes the construction of agents' consideration sets to be built “[...] only upon the ‘common knowledge’ that each agent will rule out the type assignment according to which the other agent never considers joining the blocking combination... similar to the information structure resulted from iterative communication of willingness to join the blocking combination.”

2.3 Incomplete-Information Stability

In this section, we define our notion of incomplete-information stability, based on the T -blocking notion in Definition 6. For comparison, we first state the standard definition of complete-information stability.

DEFINITION 7 (Complete-Information Stability). The matching $(\mu, \mathbf{p})$ is *complete-information-stable* at $\mathbf{t}$ by (k, k', p) if (k, k', p) $\{\mathbf{t}\}$ -blocks the matching at $\mathbf{t}$. A matching outcome $(\mu, \mathbf{p}, \mathbf{t})$ is *complete-information-stable* if it is individually rational and not $\{\mathbf{t}\}$ -blocked at $\mathbf{t}$.

Suppose (k, k', p) block the matching at $\mathbf{t}$. By definition, $\mathbf{t} \in \alpha_k(\mathbf{t}; \{\mathbf{t}\})$, implying that inequality (1) is satisfied at $\tilde{\mathbf{t}} = \mathbf{t}$. This implies that $\mathbf{t} \in \beta_{k',k}(\mathbf{t}; \{\mathbf{t}\}, p)$, and that inequality (2) is also satisfied at $\tilde{\mathbf{t}} = \mathbf{t}$. Thus, Definition 7 corresponds to the standard definition of a complete-information block.

The following lemma shows that if a T -block is successful under less information, then it must continue to be successful under more information. This follows because both the consideration set and rational consideration set are monotonic (w.r.t. set-inclusion) in T . Consequently, a complete-information block is a necessary condition for a T -block.

LEMMA 1. Consider some $\mathbf{t} \in T' \subset T \subset \bar{T}$. If the matching at $\mathbf{t}$ is T -blocked by (k, k', p) , then the matching at $\mathbf{t}$ is T' -blocked by (k, k', p) .

PROOF. Let $\mathbf{t} \in T' \subset T$, and let (k, k', p) T -block the matching at $\mathbf{t}$. By definition, $\alpha_k(\mathbf{t}; T') \subset \alpha_k(\mathbf{t}; T)$. Thus, (1) is satisfied for all $\tilde{\mathbf{t}} \in \alpha_k(\mathbf{t}; T')$. Similarly, $\beta_{k',k}(\mathbf{t}; T', p) \subset \beta_{k',k}(\mathbf{t}; T, p)$, so (2) is satisfied for all $\tilde{\mathbf{t}} \in \beta_{k',k}(\mathbf{t}; T', p)$. Hence, (k, k', p) T' -block the matching at $\mathbf{t}$. $\square$

We now formally define incomplete-information stability, which follows the iterative fixed-point characterization proposed in LMPs.

DEFINITION 8 (Incomplete-Information Stability). Let T^0 be the set of all $\mathbf{t}$ such that the matching outcome $(\mu, \mathbf{p}, \mathbf{t})$ is individually rational. For $m \geq 1$, define

$$T^m := \{\mathbf{t} \in T^{m-1} : \text{the matching } (\mu, \mathbf{p}) \text{ is not } T^{m-1}\text{-blocked at } \mathbf{t}\}.$$

We say that the matching outcome $(\mu, \mathbf{p}, \mathbf{t})$ is (*incomplete-information*) *stable* if $\mathbf{t} \in T^\infty := \bigcap_{m=0}^\infty T^m$.

Going forward, the terms “stability” and “stable outcomes” by themselves will refer to incomplete-information stability and incomplete-information-stable outcomes. By Lemma 1, T^∞ contains all $\mathbf{t}$ at which the matching is complete-information-stable.

DEFINITION 9 (Self-Stabilizing Set). A set of type assignments $T \subset \bar{T}$ is *self-stabilizing* for the matching $(\mu, \mathbf{p})$ if $(\mu, \mathbf{p})$ is individually rational, and is not T -blocked, at any $\mathbf{t} \in T$.

The following result is analogous to Proposition 2 of LMPS, and states that T^∞ is the largest self-stabilizing set.

LEMMA 2. *For the matching $(\mu, \mathbf{p})$, the following hold:*

- (a) T^∞ is self-stabilizing; and
- (b) *If T is a self-stabilizing set, then $T \subset T^\infty$.*

PROOF. (a) holds by the definition of T^∞ . (b) is an immediate consequence of Lemma 1; see that any arbitrary T is such that $T \subset T^n$ for some $n \geq 0$. $\square$

Lemma 2 is useful in determining whether the matching is stable at a particular type assignment $\mathbf{t}$. Continuing from Example 1, we use Lemma 2 to show that stable matching outcomes need not be efficient.

EXAMPLE 1 (An Inefficient Stable Matching Outcome). Going forward, we describe a type assignment $\mathbf{t}$ in counterclockwise order of agents depicted in Figure 1, which we have reproduced below for ease of reference. For instance, starting with worker i in the top left of each subfigure, $\mathbf{t}(i, \mu(j), \mu(i), j) = (3, 1, 1, 3)$ in this example. Apparently, the matching $(\mu, \mathbf{p})$ at $\mathbf{t}$ is not efficient.

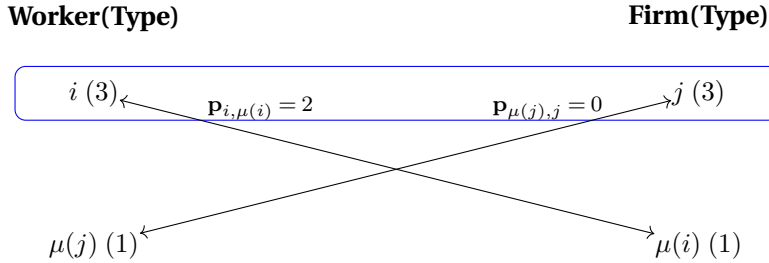


FIGURE 1. $(\mu, \mathbf{p}, \mathbf{t})$

Let $\mathbf{t}^i(i, \mu(j), \mu(i), j) = (3, 1, 1, 1)$, and $\mathbf{t}^j(i, \mu(j), \mu(i), j) = (1, 1, 1, 3)$. We show that $T = \{\mathbf{t}, \mathbf{t}^i, \mathbf{t}^j\}$ is self-stabilizing, and thus, incomplete-information-stable outcomes need not be efficient under two-sided incomplete information and supermodular premuneration values.

Note that at $\mathbf{t}^i$ and $\mathbf{t}^j$, the matching is complete-information-stable, and so by Lemma 1, the matching cannot be T -blocked at $\mathbf{t}^i$ and $\mathbf{t}^j$. Thus, it suffices to show that $(\mu, \mathbf{p}, \mathbf{t})$ is not T -blocked at $\mathbf{t}$ by (i, j) .⁶ We divide the argument into two cases.

Firstly, consider $p_{j,i} > -2$. See that $\zeta(\mathbf{t}_j^i, \mathbf{t}_i^i) + p_{j,i} > 3 - 2 = \zeta(\mathbf{t}_j^i, \mathbf{t}_{\mu(j)}^i) + \mathbf{p}_{j, \mu(j)}$, so j strictly benefits in the block at $\mathbf{t}^i$, so worker i rationally (and naively) considers $\mathbf{t}^i$, i.e., $\mathbf{t}^i \in \beta_{i,j}(\mathbf{t}; T, p)$ (and $\mathbf{t}^i \in \alpha_i(\mathbf{t}; T)$). However, when $p_{j,i} > -2$, we have $\zeta(\mathbf{t}_i, \mathbf{t}_j^i) - p_{j,i} < 3 + 2 = \zeta(\mathbf{t}_i, \mathbf{t}_{\mu(i)}) + \mathbf{p}_{i, \mu(i)}$, i.e., worker i does not strictly benefit in a block with j at

⁶We need not consider $(\mu(j), \mu(i))$; they do not form a complete-information blocking pair at $\mathbf{t}$.

$\mathbf{t}^i$. That is, inequality (2) is not satisfied for $\mathbf{t}^i \in \beta_{i,j}(\mathbf{t}; T, p)$ (and (1) is not satisfied for $\mathbf{t}^i \in \alpha_i(\mathbf{t}; T)$).

Secondly, consider $p_{j,i} < 0$. See that i strictly benefits in the block at $\mathbf{t}^j$, and thus j rationally (and naively) considers $\mathbf{t}^j$, i.e., $\mathbf{t}^j \in \beta_{j,i}(\mathbf{t}; T, p)$ (and $\mathbf{t}^j \in \alpha_j(\mathbf{t}; T)$). However, when $p_{j,i} < 0$, $\zeta(\mathbf{t}_j, \mathbf{t}_i^j) + p_{j,i} < 3 + 0 = \zeta(\mathbf{t}_j, \mathbf{t}_{\mu(j)}) - \mathbf{p}_{\mu(j),j}$, i.e., firm j does not strictly benefit in a block with i . That is, inequality (2) is not satisfied for $\mathbf{t}^j \in \beta_{j,i}(\mathbf{t}; T, p)$ (and (1) is not satisfied for $\mathbf{t}^j \in \alpha_j(\mathbf{t}; T)$).

Thus, there is no p that facilitates a T -block of $(\mu, \mathbf{p}, \mathbf{t})$ by (i, j) . $\diamond$

Example 1 sharply contrasts the efficiency of stable matchings with one-sided incomplete information. In particular, if firms' types are common knowledge, as is assumed in LMPS, then at $\mathbf{t}$, i knows j 's type and is willing to block with j at any wage $p_{i,j} > -4$. Then, for $p_{i,j} \leq 0$, i strictly benefits in the block only if i is of type 3. Thus, (i, j, p) T -block the matching at $\mathbf{t}$ for any $p_{i,j} \in (-4, 0]$.

2.4 Worst-Case Hypotheses and Deduction

In view of Example 1, we now focus on identifying conditions to restore the efficiency of stable matching outcomes. The condition we would like to propose is called “misassignment deduction,” which we will define in Section 3.2.

But first, let us begin with identifying conditions for which incomplete-information can be overcome, i.e., when does a complete-information block persist under incomplete-information? It should be the case that if an agent “deduces” the type of their potential blocking partner under complete-information, that they would participate in the potential block under the same conditions in which they would in complete-information. Since agents in our model are “pessimistic” in the sense that they participate in a block only if they expect a strict benefit in all cases they admit into consideration, we find it useful to focus on the set of *worst-case hypotheses*.

DEFINITION 10 (Worst-Case Hypotheses). For a potential blocking pair (k, k') , the set of *worst-case hypotheses* for agent k at $\mathbf{t}$ is given by

$$\mathcal{W}_{k,k'}(\mathbf{t}) = \arg \min_{\tilde{\mathbf{t}} \in \alpha_k(\mathbf{t}; T^\infty)} \tilde{\mathbf{t}}_{k'}.$$

In words, the set of worst-case hypotheses consists of type assignments that assign the worst type to k' among type assignments that k considers possible in evaluating a potential block with k' .

DEFINITION 11 (Deduction). For any pair (i, j) , we say that agent k *deduces* $T \subset \bar{T}$ at $\mathbf{t}$ if $\mathcal{W}_{k,k'}(\mathbf{t}) \subset T$.

Thus, in the context of evaluating a potential block with agent k' , agent k is said to deduce an event T if the “worst-case” implies the event T . Observe that if $\mathbf{t} \in \mathcal{W}_{k,k'}(\mathbf{t})$, then k deduces a set of type assignments that assign at least the type $\mathbf{t}_{k'}$ to agent k' . In

this case, we say that agent k deduces the type of agent k' at $\mathbf{t}$. We now arrive at the following lemma that will be used extensively in the rest of the paper.

To show that $\mathbf{t}$ is not in T^∞ , we need to only show that there exists a triplet (k, k', p) that satisfies conditions (a) and (b) of Lemma 3. That is, Lemma 3 provides sufficient conditions under which a complete-information block persists under incomplete-information.

LEMMA 3. *For any $\mathbf{t} \in T^\infty$, there does not exist a complete-information blocking triple (k, k', p) for the matching $(\mu, \mathbf{p})$ such that:*

- (a) k deduces the type of k' at $\mathbf{t}$, and
- (b) $\tilde{\mathbf{t}}_k \geq \mathbf{t}_k$ for every $\tilde{\mathbf{t}} \in \beta_{k',k}(\mathbf{t}; T^\infty, p)$.

PROOF. Suppose that there exists $\mathbf{t} \in T^\infty$ and a complete-information blocking triple (k, k', p) such that conditions (a) and (b) of Lemma 3 are satisfied. Since (k, k', p) form a complete-information block at $\mathbf{t}$, it follows that $p_{k,k'} > \zeta(\mathbf{t}_k, \mathbf{t}_{\mu(k)}) + \mathbf{p}_{k,\mu(k)} - \zeta(\mathbf{t}_k, \mathbf{t}_{k'})$. Then by condition (a), inequality (1) is satisfied for all $\tilde{\mathbf{t}} \in \alpha_k(\mathbf{t}; T^\infty)$. Next, by condition (b), $\tilde{\mathbf{t}} \in \beta_{k',k}(\mathbf{t}; T^\infty, p)$ implies $\tilde{\mathbf{t}}_k \geq \mathbf{t}_k$. Then, since (k, k', p) form a complete-information block at $\mathbf{t}$, it follows that $-p_{k',k} < \zeta(\mathbf{t}_{k'}, \mathbf{t}_k) - \zeta(\mathbf{t}_{k'}, \mathbf{t}_{\mu(k')}) - \mathbf{p}_{k',\mu(k')}$, and thus inequality (2) is satisfied for all $\tilde{\mathbf{t}} \in \beta_{k',k}(\mathbf{t}; T^\infty, p)$. Thus (k, k', p) T^∞ -block the matching at $\mathbf{t}$, which contradicts the fact that T^∞ is self-stabilizing. $\square$

Lemma 3 says that if both conditions (a) and (b) of Lemma 3 are satisfied, then (k, k', p) would T^∞ -block the matching $(\mu, \mathbf{p})$ at $\mathbf{t}$, yielding a contradiction of the premise $\mathbf{t} \in T^\infty$. Condition (a) of Lemma 3 corresponds to the information required by the proposing agent k in a potential block, which is exactly that proposer k deduces the type of their potential blocking partner k' . Condition (b) of Lemma 3 is associated with the responder k' in the potential block, and says that the wage under which k only strictly benefits from the block with k' if k were at least of type $\mathbf{t}_k$ (this is in addition to the fact that by the premise of the Lemma, p delivers a complete-information block at $\mathbf{t}$ between k and k'). Informally, Condition (b) of Lemma 3 requires that k' can “rationally deduce” that k is of at least $\mathbf{t}_k$, the true type of k . That is, under any type assignment in the rational consideration set of k' given p , k ’s type is at least $\mathbf{t}_k$. Note that Condition (b) of Lemma 3 is trivially satisfied if k' also deduces the type of k at $\mathbf{t}$. However, as we will subsequently demonstrate in Section 3.3, such p may exist even when k' does not deduce the type of k .

2.5 Misassignments

Recall that $(\mu, \mathbf{p})$ is publicly observed and fixed throughout the paper. For any *matched* agents k and k' , we say that (k, k') are *mismatched* at $\mathbf{t}$ if

$$\begin{aligned} & \zeta(\mathbf{t}_k, \mathbf{t}_{k'}) + \zeta(\mathbf{t}_{k'}, \mathbf{t}_k) + \zeta(\mathbf{t}_{\mu(k')}, \mathbf{t}_{\mu(k)}) + \zeta(\mathbf{t}_{\mu(k)}, \mathbf{t}_{\mu(k')}) \\ & > \zeta(\mathbf{t}_k, \mathbf{t}_{\mu(k)}) + \zeta(\mathbf{t}_{\mu(k)}, \mathbf{t}_k) + \zeta(\mathbf{t}_{\mu(k')}, \mathbf{t}_{k'}) + \zeta(\mathbf{t}_{k'}, \mathbf{t}_{\mu(k')}). \end{aligned}$$

In words, the two pairs of agents would produce a greater total surplus if they swapped their partners. We say that $\mathbf{t}$ is an (i, j) -misassignment if (i, j) are mismatched at $\mathbf{t}$. Let $M_{i,j}$ be the set of *stable* (i, j) -misassignments, namely,

$$M_{i,j} = \{\mathbf{t} \in T^\infty : (i, j) \text{ are mismatched at } \mathbf{t}\}.$$

Note that misassignments are defined only for matched agents, i.e., both $\mu(i)$ and $\mu(j)$ are in K . It follows that $M_{k,k'} = M_{i,j}$ for any $k, k' \in \{i, j, \mu(j), \mu(i)\}$, that is, if $\mathbf{t}$ is an (i, j) -misassignment, it must also be a (i) (j, i) -misassignment, a (ii) $(\mu(j), \mu(i))$ -misassignment, and a (iii) $(\mu(i), \mu(j))$ -misassignment. Moreover, an (i, j) -misassignment $\mathbf{t}$ must involve at least one complete-information blocking pair which is either (i, j) or $(\mu(j), \mu(i))$. In Section 4.3, we will invoke the notion of stable (i, j) -misassignments to formulate the “misassignment deduction condition” that is central to our results pertaining to the efficiency of stable matching outcomes.

Moving forward, we are primarily interested in the efficiency of stable outcomes under submodularity and supermodularity.

DEFINITION 12 (Submodularity and Supermodularity). We say that *submodularity* (*supermodularity*) holds if the premuneration functions $\zeta(w, f)$ and $\zeta(f, w)$ are strictly submodular (supermodular) in $w \in W, f \in F$.

To account for our two-sided incomplete-information setting, we have defined the case in which submodularity or supermodularity hold to be stronger than the case in which the analogous Assumptions 2 and 3 in LMPS hold.

Under submodularity or supermodularity, efficient matching outcomes necessarily cannot involve any misassignments which necessitates pairwise negative or positive assortativity, i.e., there are no (i, j) -mismatches for any pairs (i, j) . Thus, going forward, we will concentrate our analyses on the stability of a typical mismatch under two-sided incomplete information.

3. TWO-SIDED INCOMPLETE INFORMATION

One-sided incomplete information is typically modeled as common knowledge of firms' types; see, for instance, LMPS, Bikhchandani (2017), Liu (2020), Chen and Hu (2020), and Pomatto (2022). This implies, in particular, that firms also know other firms' types, i.e., firms have “within-side information.” This “within-side information” is not required in Koopmans and Beckmann (1957) and Shapley and Shubik (1971) for the efficiency of stable matching outcomes under complete-information across sides, but it becomes crucial for the efficiency of stable matching outcomes once there is one-sided (cross-side) incomplete information. Contrary to the finding of LMPS, in Appendix A.5.1, we construct an inefficient stable matching outcome in a simple economy with no within-side information among firms or workers, even though the workers have complete information about the firms.

3.1 Within-Side Information

To focus on the novelty of two-sided incomplete information, we start with formalizing within-side information and in Sections 3 and 4, suppose that agents on both sides of the market have within-side information. Later, in Section 6.1, we relax the assumption that workers have within-side information (but firms still do) to draw the connection to LMPS and the literature on one-sided incomplete information.

We say that a worker has within-side information if that worker knows other workers' types, and similarly, a firm has within-side information if that firm knows other firms' types. We formalize within-side information in the following definition. Hereafter, we refer to two-sided incomplete-information as incomplete information across the two sides.

DEFINITION 13 (Within-Side Information). Worker $i \in I$ has *within-side information* if $\tilde{\mathbf{t}}_{i'} = \mathbf{t}_{i'}$, $\forall \tilde{\mathbf{t}} \in \Pi_i(\mathbf{t})$, $\forall \mathbf{t} \in \bar{T}$, $\forall i' \in I$. Firm j has *within-side information* if $\tilde{\mathbf{t}}_{j'} = \mathbf{t}_{j'}$, $\forall \tilde{\mathbf{t}} \in \Pi_j(\mathbf{t})$, $\forall \mathbf{t} \in \bar{T}$, $\forall j' \in J$.

We will use the abbreviation (WI) as shorthand for the following implication: for any $\tilde{\mathbf{t}} \in \Pi_k(\mathbf{t})$, $\tilde{\mathbf{t}}_{\mu(k')} = \mathbf{t}_{\mu(k')}$, for all k' on the other side of the market of k . While a mismatch typically involves four agents, it follows from (PI) and (WI) that for any potential blocking pair, each agent is uncertain only about the other blocking agent's type.

We are now ready to state our first main result, which states that if premuneration functions are submodular, and agents have within-side information, then stable matching outcomes are efficient.

THEOREM 1 (Efficiency under Submodularity). *Suppose every agent has within-side information and submodularity holds. Then, every stable matching outcome is efficient.*

Submodularity applies to settings where agents are substitutable in creating match value, as illustrated by the examples discussed in the introduction. Together with Example 1, Theorem 1 establishes the core contribution of this paper: under two-sided incomplete information, stable matching outcomes are efficient when preferences are submodular, but may fail to be efficient under supermodularity. To prove Theorem 1 in subsequent sections, we will introduce building blocks to pinpoint the source of inefficiencies and how agents overcome incomplete information to block inefficient matching outcomes. The refined analysis also sheds light on how efficiency of stable matching outcomes might be restored under supermodularity in certain environments.

3.2 Misassignment Deduction

Recall that Lemma 3 provides sufficient conditions under which a complete-information block persists under incomplete information. While Lemma 3 is about complete-information blocks in general, we are interested in complete-information blocks that arise from inefficient matching outcomes. When agents' premuneration functions are supermodular or submodular, the main source behind inefficiencies in our model is due

to mismatches. To apply Lemma 3 in this case (to show that such matching outcomes are not stable), we would need to establish conditions (a) and (b) of Lemma 3 for either of those potential blocking pairs, when agents are mismatched.

In this subsection, we will work toward showing that if an agent always deduces an (i, j) -misassignment, then at some misassignment they deduce the type of their blocking partner, and thereby satisfy condition (a) of Lemma 3.

DEFINITION 14 (Misassignment Deduction). An agent $k \in \{i, j, \mu(j), \mu(i)\}$ satisfies (i, j) -Misassignment Deduction $((i, j)$ -MD) at $\mathbf{t} \in M_{i,j}$ if $\mathcal{W}_{k,k'}(\mathbf{t}) \subset M_{i,j}$ or $\mathbf{t} \in \mathcal{W}_{k,k'}(\mathbf{t})$.

In words, agent k satisfies (i, j) -MD at $\mathbf{t}$ either when k 's worst-case hypotheses are all (i, j) -misassignments, or when k has already deduced the type of k' at $\mathbf{t}$. When do agents satisfy (i, j) -MD? Consider a typical (i, j) -misassignment $\mathbf{t}$, which is illustrated in the left and right subfigures of Figure 2 for the case of submodularity and supermodularity, respectively. In the left subfigure, since $\mathbf{t}_i < \mathbf{t}_{\mu(j)}$ but $\mathbf{t}_{\mu(i)} < \mathbf{t}_j$, the matching outcome is not negatively assortative; similarly, the matching outcome $(\mu, \mathbf{p}, \mathbf{t})$ in the right subfigure is not positively assortative.

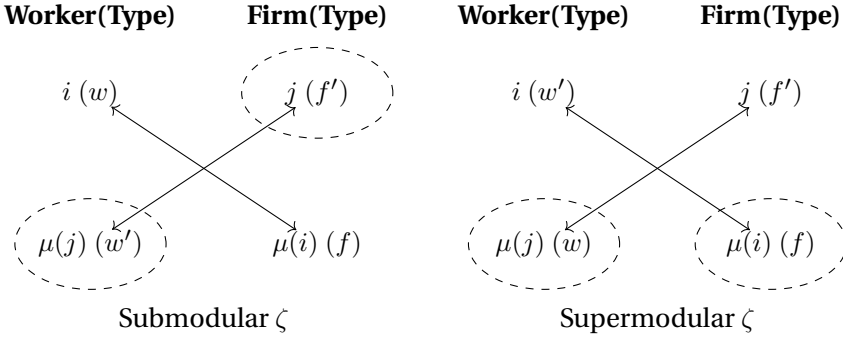


FIGURE 2. $(\mu, \mathbf{p}, \mathbf{t})$; Mismatches; $w < w'$ and $f < f'$

Note that firm $\mu(i)$ in the right subfigure of Figure 2 satisfies (i, j) -MD (equivalently $(\mu(j), \mu(i))$ -MD) at $\mathbf{t}$; by (WI), firm $\mu(i)$ knows that j is a better firm than $\mu(i)$, and thus any hypothesis where worker $\mu(j)$ is of a type worse than $\mu(j)$'s true type is also one where $\mu(j)$ is of a type worse than i 's, i.e., such hypotheses are also (i, j) -misassignments. Thus, $\mu(i)$ satisfies (i, j) -MD at $\mathbf{t}$. Similarly, in the left subfigure, one can observe that $\mu(j)$ satisfies (i, j) -MD at $\mathbf{t}$.

Equipped with the notion of (i, j) -MD, we now arrive at Proposition 1 which is a key building block of our efficiency results. Proposition 1 says that as long as there is a stable misassignment, we can find another stable misassignment $\mathbf{t}^*$ such that, if an agent $k \in \{i, j, \mu(j), \mu(i)\}$ has within-side information and satisfies (i, j) -MD, then agent k deduces the type of their potential blocking partner at $\mathbf{t}^*$.

PROPOSITION 1. *If $M_{i,j}$ is nonempty, there is some $\mathbf{t}^* \in M_{i,j}$ such that:*

- (a) For any $k \in \{i, \mu(j)\}$ that satisfies (i, j) -MD at $\mathbf{t}^*$, we have $\mathbf{t}^* \in \mathcal{W}_{k,k'}(\mathbf{t}^*)$.
- (b) For any $k \in \{j, \mu(i)\}$ that satisfies (i, j) -MD at $\mathbf{t}^*$ and that has within-side information, we also have $\mathbf{t}^* \in \mathcal{W}_{k,k'}(\mathbf{t}^*)$.

PROOF. To prove the result, we define four sets of type profiles:

$$M_{i,j}^1 = \arg \min_{\mathbf{t} \in M_{i,j}} \tilde{\mathbf{t}}_{\mu(i)}; M_{i,j}^2 = \arg \min_{\mathbf{t} \in M_{i,j}^1} \tilde{\mathbf{t}}_j; M_{i,j}^3 = \arg \min_{\mathbf{t} \in M_{i,j}^2} \tilde{\mathbf{t}}_{\mu(j)}; M_{i,j}^* = \arg \min_{\mathbf{t} \in M_{i,j}^3} \tilde{\mathbf{t}}_i.$$

Observe that $M_{i,j}^* \subset M_{i,j}^3 \subset M_{i,j}^2 \subset M_{i,j}^1 \subset M_{i,j}$; moreover, $M_{i,j}^*$ is nonempty if and only if $M_{i,j}$ is nonempty. The set $M_{i,j}^1$ includes the stable (i, j) -misassignments which minimize the type of $\mu(i)$ (the potential blocking partner of $\mu(j)$) over all (i, j) -misassignments; the set $M_{i,j}^2$ includes the type misassignments which minimize the type of j (the potential blocking partner of i) over all (i, j) -misassignments; and so on.⁷

Consider $\mathbf{t}^* \in M_{i,j}^*$. Steps 1 and 2 below prove part (a) of Proposition 1 and Steps 3 and 4 prove part (b) of Proposition 1.

STEP 1. If $\mu(j)$ satisfies (i, j) -MD at $\mathbf{t} \in M_{i,j}^1$, then $\mu(j)$ deduces $\mu(i)$'s type at $\mathbf{t}$.

Suppose to the contrary that $\mu(j)$ does not deduce the type of $\mu(i)$ at $\mathbf{t}$. Then $\mu(j)$ satisfies (i, j) -MD at $\mathbf{t}$ when $\mathcal{W}_{\mu(j), \mu(i)}(\mathbf{t}) \subset M_{i,j}$ by definition, and there exists $\tilde{\mathbf{t}} \in \mathcal{W}_{\mu(j), \mu(i)}(\mathbf{t})$ such that $\tilde{\mathbf{t}}_{\mu(i)} < \mathbf{t}_{\mu(i)}$. But this implies $\mathbf{t} \notin M_{i,j}^1$, a contradiction.

STEP 2. If i satisfies (i, j) -MD at $\mathbf{t} \in M_{i,j}^2$, then i deduces j 's type at $\mathbf{t}$.

Suppose to the contrary that i does not deduce the type of j at $\mathbf{t}$. Then i satisfies (i, j) -MD at $\mathbf{t}$ when $\mathcal{W}_{i,j}(\mathbf{t}) \subset M_{i,j}$ by definition. Since $\mathbf{t} \in M_{i,j}^2 \subset M_{i,j}^1$, i observes by (PI) that $\mu(i)$'s type is “minimized” at $\mathbf{t}$, which implies that $\mathcal{W}_{i,j}(\mathbf{t}) \subset M_{i,j}^1$. Since i does not deduce j 's type at $\mathbf{t}$, there exists $\tilde{\mathbf{t}} \in \mathcal{W}_{i,j}(\mathbf{t})$ such that $\tilde{\mathbf{t}}_j < \mathbf{t}_j$. But this implies $\mathbf{t} \notin M_{i,j}^2$, a contradiction.

STEP 3. If $\mu(i)$ has within-side information and satisfies (i, j) -MD at $\mathbf{t} \in M_{i,j}^3$, then $\mu(i)$ deduces $\mu(j)$'s type at $\mathbf{t}$.

Suppose to the contrary that $\mu(i)$ does not deduce the type of $\mu(j)$ at $\mathbf{t}$. Then $\mu(i)$ satisfies (i, j) -MD at $\mathbf{t}$ when $\mathcal{W}_{\mu(i), \mu(j)}(\mathbf{t}) \subset M_{i,j}$ by definition. Since $\mathbf{t} \in M_{i,j}^3 \subset M_{i,j}^2$, $\mu(i)$ observes by (PI) that $\mu(i)$'s type is “minimized” at $\mathbf{t}$ and by (WI) that j 's type is “minimized” given $\mu(i)$'s type, which implies that $\mathcal{W}_{\mu(i), \mu(j)}(\mathbf{t}) \subset M_{i,j}^2$.

Since $\mu(i)$ does not deduce $\mu(j)$'s type at $\mathbf{t}$, there exists $\tilde{\mathbf{t}} \in \mathcal{W}_{\mu(i), \mu(j)}(\mathbf{t}) \implies \tilde{\mathbf{t}}_{\mu(j)} < \mathbf{t}_{\mu(j)}$. This implies $\mathbf{t} \notin M_{i,j}^3$, a contradiction.

⁷We choose the order in constructing $M_{i,j}^n$ for $n = 1, 2, 3, *$ to single out part (a) of Proposition 1, which will be used later when we relax the assumption that workers have within-side information; see Section 6.1. Symmetrically, if firms do not have within-side information (while workers do), the order of (a) and (b) can be reversed, so that firms deduce the types of their potential blocking partner. When all agents have within-side information, the order is arbitrary.

STEP 4. If j has within-side information and satisfies (i, j) -MD at $\mathbf{t} \in M_{i,j}^*$, then j deduces i 's type at $\mathbf{t}$.

Suppose to the contrary that j does not deduce the type of i at $\mathbf{t}$. Then j satisfies (i, j) -MD at $\mathbf{t}$ when $\mathcal{W}_{j,i}(\mathbf{t}) \subset M_{i,j}$. Since $\mathbf{t} \in M_{i,j}^* \subset M_{i,j}^3$, j observes by (WI) that $\mu(i)$'s type is “minimized” at $\mathbf{t}$, by (PI) that j 's own type is “minimized” at $\mathbf{t}$ given j 's type, and by (PI) again that $\mu(j)$'s type is “minimized” at $\mathbf{t}$ given both $\mu(i)$ and j 's types, which implies that $\mathcal{W}_{j,i}(\mathbf{t}) \subset M_{i,j}^3$.

Since j does not deduce i 's type at $\mathbf{t}$, there exists $\tilde{\mathbf{t}} \in \mathcal{W}_{j,i}(\mathbf{t})$ such that $\tilde{\mathbf{t}}_i < \mathbf{t}_i$. This implies $\mathbf{t} \notin M_{i,j}^*$, a contradiction. $\square$

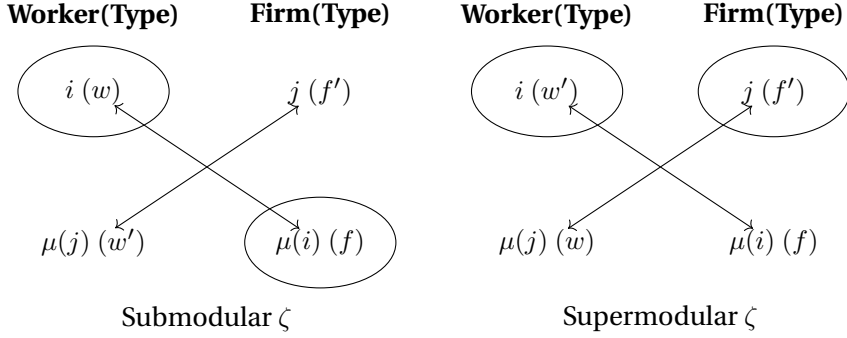
REMARK 3. Under one-sided incomplete information, LMPS invoke an induction argument to show that the type of the worse worker (of the two involved in a mismatch) can be deduced. In contrast, under two-sided incomplete information, as foreshadowed by condition (a) of Lemma 3, we need conditions under which an arbitrary agent in a mismatch can deduce the type of their potential blocking partner. Proposition 1 provides such conditions by constructing the sets $M_{i,j}^1$ to $M_{i,j}^*$. $M_{i,j}^1$ contains type assignments at which, if worker $\mu(j)$ satisfies (i, j) -MD, then worker $\mu(j)$ deduces the type of $\mu(i)$. Since $M_{i,j}^2$ is a subset of $M_{i,j}^1$, if both i and $\mu(j)$ satisfy (i, j) -MD (at $\mathbf{t} \in M_{i,j}^2$), then both deduce the type of their potential blocking partner j and $\mu(i)$ respectively, and so on.⁸

3.3 Separating Wages

In the previous subsection we investigated how condition (a) of Lemma 3 could be satisfied. In this subsection, we will investigate when condition (b) of Lemma 3 could be satisfied. Another immediate implication of agents having within-side information arises in situations in which they evaluate a block together; a firm j evaluating a block with a worker i knows the payoff that the worker i (for each possible type of worker i) is receiving in the current matching, and would receive in the block with firm j . Together with submodularity and supermodularity, this further implies that agents involved in an (i, j) -misassignment can utilize a “separating wage,” which we explain as follows.

Consider the firm j in the left subfigure of Figure 3. Here j will lose some remuneration value in a block with i since i is of a worse type than $\mu(j)$; however, under submodularity, the greater the type of j , the less j will lose. Thus, there exists some “separating wage” p at which j strictly benefits in the block with i if and only if j were of a type weakly greater than $\mathbf{t}_j$. Specifically, by invoking p , worker i can “rationally deduce” the type of j at $\mathbf{t}$ because i 's T^∞ -rational consideration set $\beta_{i,j}(\mathbf{t}; T^\infty, p)$ admits only type assignments $\tilde{\mathbf{t}}$ under which j 's type is at least $\mathbf{t}_j$. An analogous argument holds under supermodularity. Relying on within-side information, Lemma 4 formalizes the above observations (which were first made in LMPS).

⁸The argument in the proof of Proposition 1 also relates complete-information stability to incomplete-information stability. Specifically, suppose that we define: $C_{i,j} = \{\mathbf{t} \in T^\infty : (i, j) \text{ form a complete-information blocking pair at } \mathbf{t}\}$ as the set of type assignments that stabilize a complete-information block between (i, j) . Then when an agent $k \in \{i, j\}$ satisfies “ (i, j) -MD” with $M_{i,j}$ being replaced by $C_{i,j}$, they deduce the type of their potential blocking partner k' .

FIGURE 3. $(\mu, \mathbf{p}, \mathbf{t})$; Mismatches; $w < w'$ and $f < f'$

LEMMA 4 (Separating Wage). *Suppose (k, k') form a complete-information blocking pair at $\mathbf{t}$ and k' has within-side information. Suppose that either (a) submodularity holds and $\mathbf{t}_{\mu(k)} > \mathbf{t}_{k'}$, or (b) supermodularity holds and $\mathbf{t}_{\mu(k)} < \mathbf{t}_{k'}$, or (c) $\mu(k) = \emptyset$, then there exists p such that $\tilde{\mathbf{t}}_k \geq \mathbf{t}_k$ for every $\tilde{\mathbf{t}} \in \beta_{k',k}(\mathbf{t}; T^\infty, p)$ and (k, k', p) still complete-information block the matching at $\mathbf{t}$.*

In Figure 3, we circle in solid lines the agents who satisfy the conditions of Lemma 4 and thereby can utilize a separating wage. If, for example, (i, j) form a complete-information blocking pair at $\mathbf{t}$ in the left subfigure, then setting $k' = i$ in Lemma 4 implies that there exists p such that condition (b) of Lemma 3 is satisfied for (j, i, p) .

Intuitively, a separating wage is only effective for an agent k' if their potential blocking partner k participates only when k is of a high-type (but not when of a low-type). To this end, the agent k' must be able to accurately assess the change in payoffs that their potential blocking partner k would expect by participating in the block. Two conditions are essential for this: k' must have within-side information, and k' utilizes their rational consideration set in evaluating the block. We discuss the two conditions in the following remarks.

REMARK 4. First, without within-side information, k' lacks knowledge of the type of $\mu(k)$ (the type of the incumbent match of k), and so k' does not know what k receives in the status quo, and thus does not know what the change in payoffs that k would expect in the block. A low-type k matched to a low-type $\mu(k)$ might expect a larger increase in payoff than a high-type k matched to a high-type $\mu(k)$. Thus, there is no wage the firm would employ to “separate” the two possibilities, as the high-type k only participates if the low-type k also would. For the same reason, it is also necessary in LMPS for firms to know other firms’ types. Appendix A.5.1 illustrates this via a counterexample.

REMARK 5. Second, the rational consideration set of an agent k' is based on the evaluation of the change in payoffs of k given the true type of k' . Thus, in alternative blocking notions, such as the one in Chen and Hu (2023) discussed in Remark 1, a separating wage might not be effective. In particular, Chen and Hu (2023) allows for k' to believe

that k might be overly optimistic, in the sense that k perceives k' to be much better than they actually are. An overly optimistic low-type k and a less optimistic high-type k may expect the same benefit from a rematch with k' , in which case no wage can separate the two worker types from participating in a block with k' . We continue this comparison formally in Section B.5 of [Chen and Ho \(2026\)](#).

4. EFFICIENCY

We present our main results in this section. Our argument consists of two steps. First, we present Lemma 5, which establishes the efficiency of stable matching outcomes under within-side information when there are no stable misassignments. In other words, Lemma 5 says that to establish the efficiency of stable matching outcomes when agents have within-side information, it suffices to rule out stable (i, j) -misassignments for any pair (i, j) .

LEMMA 5. *Suppose all agents have within-side information and either submodularity or supermodularity holds. If there are no stable misassignments, i.e., if $M_{i,j}$ is empty for every pair (i, j) , then every stable matching outcome is efficient.*

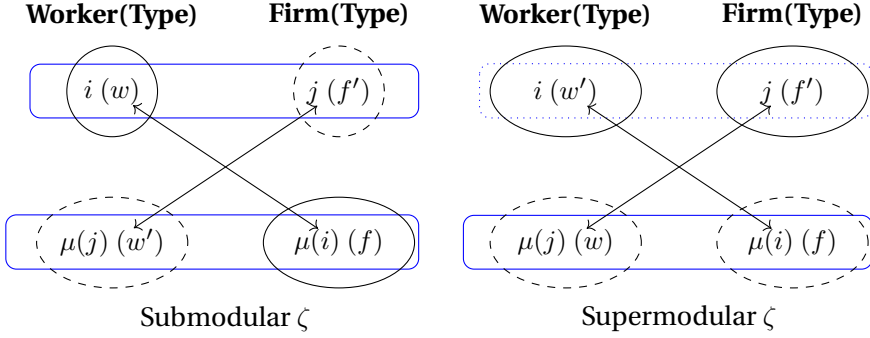
REMARK 6. Lemma 5 requires Assumption 0(b) (Positive Surplus). Without this assumption, there may be a stable matching that matches “too much” or “too little.” “Matching too little” refers to the case in which the surplus of a match can be arbitrarily low; hence, the inefficient matching such that all agents are unmatched will always be stable. “Matching too much” refers to the case in which leaving more agents unmatched and rematching a subset of agents amongst themselves produces a greater total surplus. Although this type of inefficiency arises from matched agents, it is not due to pairwise mismatches and thus not ruled out by $M_{i,j} = \emptyset$.

To elaborate, matching too little can occur under both submodularity and supermodularity. Consider a matching $(\mu, \mathbf{p})$, where we have only one unmatched worker i and one unmatched firm j , and the supermodular premuneration functions $\zeta(w, f) = \zeta(f, w) = wf - 3$ for $(w, f) \in \bar{T} = \{1, 2\}^2$ respectively. Then, $(\mu, \mathbf{p})$ is inefficient yet stable at $\mathbf{t} = (2, 2)$, since both agents are worried that the other is of type 1. A similar counterexample exists for the submodular case. Matching too much occurs only under submodularity since it is ruled out by individual rationality under supermodularity; see Section 4.1.3 of LMPS for an example and further details.

4.1 Mismatch under Submodularity or Supermodularity

Now that we have presented Lemma 5, we hereafter focus on eliminating stable misassignments, using the tools that we have developed thus far. To fix ideas, once again consider a typical (i, j) -misassignment as depicted in Figure 4.

Recall that in the right subfigure, $\mu(i)$ satisfies (i, j) -MD at $\mathbf{t}$, and symmetrically, so does $\mu(j)$. In the left subfigure, it follows from value monotonicity that any worst-case hypothesis of agent $k = j, \mu(j)$ (which at best assigns to $k' = i, \mu(i)$ their true type) is a misassignment, and thus k satisfies (k, k') -MD at $\mathbf{t}$. Lemma 6 below summarizes these observations.

FIGURE 4. $(\mu, \mathbf{p}, \mathbf{t})$; Mismatches; $w < w'$ and $f < f'$

LEMMA 6. *Under submodularity or supermodularity, for any $\mathbf{t} \in M_{i,j}$ and any agent $k \in \{i, j, \mu(j), \mu(i)\}$ that has within-side information, if the current partner of k , $\mu(k)$, is strictly better than k' , then k satisfies (k, k') -MD at $\mathbf{t}$.*

In Figure 4, circled by dashed lines are the agents k whose current partner is better than their blocking partner, while the agents circled by solid lines are those that are able to invoke a separating wage from Figure 3. Observe that in the submodular case, each potential blocking pair admits exactly one agent who satisfies (i, j) -MD (equivalently, $(\mu(j), \mu(i))$ -MD) at $\mathbf{t}$ and another agent that can use a separating wage. In the supermodular case, however, both agents in the top pair can only utilize a separating wage but do not satisfy (i, j) -MD at $\mathbf{t}$; in contrast, both agents in the bottom pair satisfy (i, j) -MD at $\mathbf{t}$. This asymmetry foreshadows the asymmetry in the respective efficiency results under submodularity and supermodularity that we will present in the next two subsections.

4.2 Efficiency under Submodularity - Proof of Theorem 1

Note that every agent in Figure 4 is circled by either dashed or solid lines; agents circled by solid lines can utilize a separating wage, while agents circled by dashed lines satisfy (i, j) -MD at $\mathbf{t}$. We are now ready to prove Theorem 1, which integrates Lemma 4 and Proposition 1.

Theorem 1 (Efficiency under Submodularity) *Suppose every agent has within-side information and submodularity holds. Then, every stable matching outcome is efficient.*

PROOF. By Lemma 5, it suffices to show that $M_{i,j}$ is empty for any pair (i, j) . Suppose to the contrary that $M_{i,j}$ is nonempty. By Proposition 1, there exists a stable (i, j) -misassignment $\mathbf{t}^*$ such that $k \in \{i, j, \mu(j), \mu(i)\}$ deduces the type of their potential blocking partner k' at $\mathbf{t}^*$ if k satisfies (i, j) -MD at $\mathbf{t}^*$. By definition of an (i, j) -misassignment, either (i, j) or $(\mu(j), \mu(i))$ form a complete-information blocking pair at $\mathbf{t}^*$. Note that the complete-information blocking pair must comprise one agent k who satisfies (k, k') -MD at $\mathbf{t}$ and the other agent k' that either also satisfies (k, k') -MD at $\mathbf{t}$, or can invoke a separating wage (recall Figure 4). Moreover, by Proposition 1, agent k must deduce the type of agent k' . It follows from Lemma 3 that $\mathbf{t}^* \notin T^\infty$, which is a contradiction. $\square$

Even though the building blocks—Proposition 1, and Lemmas 4, 5, and 6—all hold under both supermodularity and submodularity, the proof of Theorem 1 requires that when we examine either pair (i, j) or $(\mu(j), \mu(i))$ which are mismatched at $\mathbf{t}$, at least either agent in each pair (k, k') satisfies (i, j) -MD at $\mathbf{t}$. This is only true under submodularity. In particular, recall a typical mismatch under supermodularity as depicted in the right subfigure of Figure 4: If the complete-information blocking pair in question is the “top pair” (i, j) , then neither i nor j is circled by dashed lines, and so Lemma 6 does not apply. Indeed, in Example 1, since both agents $k \in \{i, j\}$ are worried about $\mathbf{t}^k$ which are not misassignments, they do not satisfy (i, j) -MD at $\mathbf{t}$. Thus, the proof of Theorem 1 does not apply to the case of supermodularity.

4.3 Efficiency under Supermodularity

Can we obtain the efficiency result under supermodularity? Recall that the matching outcome $(\mu, \mathbf{p}, \mathbf{t})$ in Example 1 is inefficient but stable. Observe that agent i entertains a worst-case hypothesis that agent j is of type 1, and such a hypothesis $\mathbf{t}^i(i, \mu(j), \mu(i), j) = (3, 1, 1, 1)$ is not a misassignment. Hence, worker i violates (i, j) -MD at $\mathbf{t}$. Similarly, firm j violates (i, j) -MD at $\mathbf{t}$. The observation that both i and j violate (i, j) -MD at $\mathbf{t}$ suggests the following non-primitive condition:

Misassignment Deduction (MD) condition: For every potential blocking pair (i, j) , at every $\mathbf{t} \in M_{i,j}$, at least either $k \in \{i, j\}$ satisfies (i, j) -MD.

THEOREM 2. *Suppose every agent has within-side information and supermodularity holds. If the MD condition is satisfied, then every stable matching outcome is efficient.*

Theorem 2 restores the efficiency of stable matching outcomes under supermodularity. Under the MD condition, for any potential blocking pair (i, j) , at any stable (i, j) -misassignment $\mathbf{t} \in M_{i,j}$, at least either agent satisfies (i, j) -MD at $\mathbf{t}$. Hence, once the MD condition is imposed, the same argument used to prove Theorem 1 can be used to prove Theorem 2. In Appendix A.5.3 we show that the MD condition is also necessary for the efficiency of stable matching outcomes in a simple economy with two workers matched to two firms.

5. APPLICATIONS: MD UNDER SUPERMODULARITY

In this section, we show that the MD condition can also be established from other primitive assumptions under supermodularity. This enables us to apply Theorem 2 to restore the efficiency of stable matching outcomes. Throughout this section, we make the following assumption, in addition to supermodularity and within-side information (WI) among all agents:

ASSUMPTION 1 (Unique Types). *For all $k \neq k'$, and for all $\mathbf{t} \in \bar{T}$, $\mathbf{t}_k \neq \mathbf{t}_{k'}$.*

Assumption 1, together with value monotonicity, implies that agents have strict preferences over agents on the other side of the market.

5.1 Firm-Type Permutations

We say that the set of type assignments $\bar{T}$ is a set of *firm-type permutations* if, for any $\mathbf{t}, \tilde{\mathbf{t}} \in \bar{T}$, there exists a one-to-one mapping $\iota : J \rightarrow J$ such that $\mathbf{t}_j = \tilde{\mathbf{t}}_{\iota(j)}$. Assumption 2 implies that the empirical distribution of firm types is common knowledge.

ASSUMPTION 2 (Firm-Type Permutations). $\bar{T}$ is a set of firm-type permutations.

Assumptions 1 and 2 may be contrasted with the weaker assumptions of Proposition 6 of LMPS. In particular, Proposition 6 of LMPS shows that under one-sided incomplete-information, if $\bar{T}$ is a set of worker-type permutations, and *either* different workers have different types, *or* different firms have different types, then complete-information and incomplete-information stability coincides. In Section B.3 of [Chen and Ho \(2026\)](#), we prove a counterpart of this result under two-sided incomplete-information: If $\bar{T}$ is both a set of worker-type permutations and a set of firm-type permutations, then complete-information and incomplete-information stability coincide.

Notably, [Piketty \(1993\)](#) shows that when the realized distribution of agents' types is publicly observable, a taxation mechanism exists that implements the first-best allocation under incomplete information. He further argues such information can be made public via a social planner with sufficient statistical knowledge of the market—for example, through census or labor-market data.

PROPOSITION 2. *Suppose that all agents have within-side information; also suppose that supermodularity, Assumption 1 (Unique Types), and Assumption 2 (Firm-Type Permutations) hold. Then, the MD condition is satisfied and every stable matching outcome is efficient.*

We show by induction that for any worker i whose employer is among the ℓ th-worst firms, i satisfies (i, j) -MD at any $\mathbf{t} \in M_{i,j}$, for any potential blocking partner j . The case for the worst firm is immediate: by Assumption 1, there is only one firm of the worst type. Hence, worker i must deduce that any other firm j is better than i 's current employer, i.e., i satisfies (i, j) -MD at $\mathbf{t}$. For the $(\ell + 1)$ th-step in the induction, by Assumption 2 there are exactly ℓ firms of types strictly worse than the $(\ell + 1)$ th-worst type $f^{\ell+1}$. By invoking the induction hypothesis, we then show that those firms must be either unmatched, or matched to employees worse than the firm of type f^ℓ . It follows that the best $L := (|J_\mu| - \ell)$ firms and workers are matched among themselves, where $|J_\mu|$ is the number of firms that are matched. Thus, if i 's employer has a type $f^{\ell+1}$ at an (i, j) -misassignment for some firm j , then (i) j 's employee $\mu(j)$ must be one of the best L workers, and (ii) by (WI), i observes (i) and concludes that j has a type strictly greater than $f^{\ell+1}$ and deduces the misassignment; see Appendix A.3 for the complete proof.⁹

⁹In fact, with the help of another intermediate result that states that agents must be sorted by their wages, we may apply a similar induction argument to establish the MD condition, even when workers' types are not unique; see Section B.2.1 of [Chen and Ho \(2026\)](#).

5.2 Negligible Worker-Surplus Variation

We now consider an economy in which the variation in workers' premuneration value across different firm types is small relative to the variation in firms' premuneration value across different worker types.

ASSUMPTION 3 (Negligible Worker-Surplus Variation). *For all $w' > w \in W$, $f' > f \in F$, we have $\zeta(w', f') - \zeta(w', f) < \zeta(f, w') - \zeta(f, w)$.*

Note that by value monotonicity, the variation in workers' premuneration values for any worker type across firms' types cannot be zero. Negligible Worker-Surplus Variation requires that relative to the change of firms' premuneration values, workers' premuneration values vary only a little across different firm types. Note that by supermodularity, the condition compares a larger variation of the workers' premuneration value (with w' as opposed to w) with the smaller variation of the firms' premuneration value (with f as opposed to f'). This is the case when, for instance, workers' premuneration values are a small constant fraction of the total surplus of a match, or when the workers' premuneration values do not differ much across different firm types. This assumption is reasonable in markets where job scopes are relatively uniform across the workers but not across the employers; for instance, job matching for entry-level or low-skilled workers whose impacts on firms' payoffs are scaled by the firm's size or type.

PROPOSITION 3. *Suppose that all agents have within-side information; also suppose that supermodularity, Assumption 1 (Unique Types), and Assumption 3 (Negligible Worker-Surplus Variation) hold. Then, the MD condition is satisfied and every stable matching outcome is efficient.*

Under negligible worker-surplus variation, workers' premuneration value becomes a secondary consideration relative to their wage. Under complete information, if a better worker, say i , is compensated less than a worse worker, say $\mu(j)$, then worker i finds it profitable to match with firm j , which would also be willing to pay more to hire a better worker. Under incomplete information, if negligible worker-surplus variation holds, stable matching outcomes must feature a similar property. Namely, $(\mu, \mathbf{p})$ must exhibit *wage monotonicity*, that is, where matched worker cannot be paid higher than another matched worker, if both the former worker and their employer have strictly lower types than the latter worker and their employer; see Lemma 8 in Appendix A.4. To see why the MD condition holds under wage monotonicity, recall the right subfigure of Figure 3. If worker i is paid no more than worker $\mu(j)$, then worker i , knowing that they have a higher type than $\mu(j)$, must deduce that j is strictly better than $\mu(i)$, which implies an (i, j) -misassignment; otherwise, wage monotonicity is violated. Symmetrically, if worker i is paid more than worker $\mu(j)$, then j satisfies (i, j) -MD at $\mathbf{t}$.

6. DISCUSSION

6.1 One-Sided Within-Side Information

So far, we have assumed within-side information for both firms and workers. Section A.5 provides various examples outlining the role and necessity of within-side information with respect to efficiency of stable matchings. In this section, we will demonstrate that the idea of MD can replace within-side information to retain the efficiency of stable matching outcomes. In particular, we study the extent to which efficiency still persists under stable outcomes, if only firms have within-side information while workers do not. Under one-sided incomplete information where firms' types are common knowledge, it is implied that firms also have within-side information, but workers may not. Because our matching model is symmetric, the discussion applies if workers retain the within-side information but firms do not.

Observe first that the MD condition holds when, specifically, both workers i and $\mu(j)$ satisfy (i, j) -MD at every $\mathbf{t} \in M_{i,j}$; in Figure 5, we have circled by dashed lines the agents that satisfy (i, j) -MD at $\mathbf{t}$, and in solid lines the agents that can utilize a separating price. We can think of two instances in which both of the workers satisfy (i, j) -MD at any $\mathbf{t} \in M_{i,j}$: The first instance is when every worker knows the (ordinal) rankings among workers and among firms. The second is when firms' types are common knowledge, every worker knows and thereby deduces every firms' types; hence, the MD condition also holds.

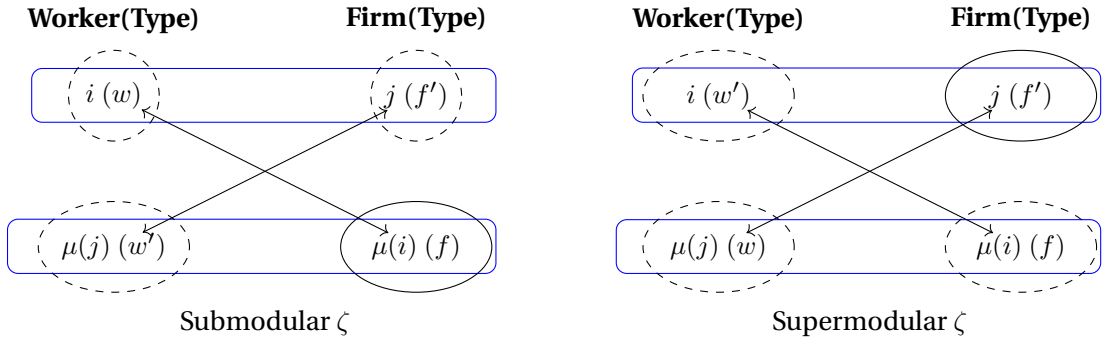


FIGURE 5. $(\mu, \mathbf{p}, \mathbf{t})$; Mismatches; $w < w'$ and $f < f'$, Firms' types are public.

PROPOSITION 4. *Suppose firms have within-side information; also suppose that either submodularity or supermodularity holds. If for some pair (i, j) , $M_{i,j}$ is not empty, then either worker i or $\mu(j)$ does not satisfy (i, j) -MD at some $\mathbf{t} \in M_{i,j}$.*

PROOF. Suppose the contrary, that $M_{i,j}$ is nonempty, but both i and $\mu(j)$ satisfies (i, j) -MD at all $\mathbf{t} \in M_{i,j}$. By Proposition 1(a), there exists some $\mathbf{t}^* \in M_{i,j}$ such that i and $\mu(j)$ deduce the type of j and $\mu(i)$ at $\mathbf{t}^*$, respectively. Thus, workers $k = i, \mu(j)$ satisfy condition (a) of Lemma 3. Moreover, by Lemma 4(a) (under submodularity) or Lemma

4(b) (under supermodularity), and Lemma 6, firms $k' = j, \mu(i)$ satisfy condition (b) of Lemma 3. Since $\mathbf{t}^*$ is an (i, j) -misassignment, either (i, j) or $(\mu(j), \mu(i))$ form a complete-information blocking pair, but this contradicts Lemma 3. □

Proposition 4 serves as an intermediate step in proving results for markets where one side is sufficiently informed about the market (satisfy misassignment deduction), while the other only has within-side information. The extreme example is the case of one-sided incomplete information.

COROLLARY 1 (One-Sided Incomplete Information). *Suppose firms' types are common knowledge. Then, under either submodularity or supermodularity, every stable matching outcome is efficient.*

Under one-sided incomplete information, each worker i (and $\mu(j)$) satisfies (i, j) -MD at every $\mathbf{t} \in M_{i,j}$ for each firm j , since workers know the type of each firm j . Thus, Proposition 4 implies that there are no stable (i, j) -misassignments.

Finally, to conclude that the absence of stable misassignments implies the efficiency of stable matching outcomes, we will require an analogous version of Lemma 5 that does not rely on workers having within-side information. In Section B.1 of [Chen and Ho \(2026\)](#), we verify that when firms' types are common knowledge, the conclusion still holds, that is, that $M_{i,j}$ being empty implies that stable matching outcomes are efficient. Thus, Propositions 3 and 4 of LMPS (Corollary 1) follow from our Proposition 4.¹⁰

Alternatively, we may define misassignments to account for unmatched agents and accordingly define a (stronger) notion of misassignment deduction. We call these notions *u-misassignments* and *uMD*, respectively. For example, in Figure 4, if the better firm j (in both subfigures) is unmatched, i.e., $\mu(j) = \emptyset$, then both subfigures would depict a *u-misassignment* (by Assumption 0(a) (Monotonicity)). Observe that unlike the case in which firm j is matched, here j necessarily satisfies (i, j) -uMD at any $\mathbf{t}$ that assigns to j a greater type than it does to $\mu(i)$, since by (WI), j observes that worker i is matched, and worker i 's partner $\mu(i)$ has a worse type than j ; furthermore, j knows that the complete-information blocking pair (which must exist among *u-mismatched* agents) at such $\mathbf{t}$ must be (i, j) .

The notion of assortativity defined in LMPS actually amounts to “for all pairs (i, j) , the set of stable *u-misassignments* is empty.” Given an analogous version of Proposition 4, i.e., that the set of stable (i, j) -*u-misassignments* is empty, Corollary 1 directly follows from Lemma B.1 of LMPS. However, *uMD* is also a stronger requirement than MD, which matters if we would like to assume, or verify, that an agent k satisfies (i, j) -uMD (at some $\mathbf{t}$), or if we are concerned about efficiency only among matched agents.

¹⁰Note that Lemma 5 does not hold when workers have no within-side information and firms' types are not common knowledge; see Appendix A.5.2.

6.2 Related Literature

Like LMPS, [Chen and Hu \(2020\)](#) and [Pomatto \(2022\)](#) also analyze worst-case-based (prior-free) notions of stable matching in one-sided incomplete-information models but, unlike this paper, do not study efficiency of stable outcomes. [Bikhchandani \(2017\)](#) analyzes stability with one-sided incomplete-information and nontransferable utility.

[Liu \(2020\)](#) develops a matching model under one-sided incomplete information and endogenizes agents' on-path and off-path beliefs via a notion called Bayesian-consistency. Assuming that the agents and an outside observer share a common prior, [Liu \(2020\)](#) identifies assumptions on preferences under which stable matchings with Bayesian-consistent beliefs are Bayesian-efficient conditional on the outside observer's information. [Wang \(2023\)](#) also formulates endogenous beliefs, but unlike the equilibrium notion of [Liu \(2020\)](#), he proposes a notion akin to rationalizability and characterizes the corresponding set of stable outcomes.

[Chen and Hu \(2023\)](#) investigate various notions of stability under two-sided incomplete information based on refining agents' consideration sets, but they do not pursue an efficiency result. [Liu \(2017, 2023\)](#) extend the analysis of [Liu \(2020\)](#) to two-sided incomplete information and, like [Liu \(2020\)](#), obtain results regarding Bayesian-efficiency of stable matchings under a common prior.

[Park \(2024\)](#) studies the efficiency of stable matchings under two-sided incomplete information while maintaining within-side information. His information structure is asymmetric: firms have no prior information about individual workers' types, whereas each worker has partial information about each firm's type, represented by an ordered partition of the firm-type space. Accordingly, while Park's efficiency criterion remains ex-post surplus maximization, his stability notion is prior-free and indexed by a profile of workers' information partitions; moreover, the partition profile enters through the associated notions of blocking, stability, and partial assortativity. In particular, when every worker's partition is finest, Park's stability notion reduces to the one-sided incomplete-information notion of [Liu et al. \(2014\)](#); when every worker's partition is coarsest, it corresponds to the two-sided symmetric incomplete-information case considered in an earlier working-paper version [Chen and Ho \(2020\)](#).

APPENDIX: PROOFS AND EXAMPLES

Throughout the appendix, we write (PI) and (WI) as the shorthand for private information and within-side information, respectively.

A.1 Proof of Lemma 4

Lemma 4 (Separating Wage) *Suppose (k, k') form a complete-information blocking pair at $\mathbf{t}$ and k' has within-side information. Suppose that either (a) Submodularity holds and $\mathbf{t}_{\mu(k)} > \mathbf{t}_{k'}$, or (b) Supermodularity holds and $\mathbf{t}_{\mu(k)} < \mathbf{t}_{k'}$, or (c) $\mu(k) = \emptyset$, then there exists p such that $\tilde{\mathbf{t}}_k \geq \mathbf{t}_k$ for every $\tilde{\mathbf{t}} \in \beta_{k',k}(\mathbf{t}; T^\infty, p)$ and (k, k', p) still complete-information block the matching at $\mathbf{t}$.*

PROOF. By Lemmata 2 and O.1 of LMPS, there exists p such that:

$$\zeta(t, \mathbf{t}_{k'}) + p_{k,k'} > \zeta(t, \mathbf{t}_{\mu(k)}) + \mathbf{p}_{k,\mu(k)} \iff t \geq \mathbf{t}_k. \quad (3)$$

Denote $\hat{t}$ as the greatest type (on the side of k) such that $\hat{t} < \mathbf{t}_k$. Consider p such that

$$p_{k,k'} > \zeta(\mathbf{t}_k, \mathbf{t}_{\mu(k)}) + \mathbf{p}_{k,\mu(k)} - \zeta(\mathbf{t}_k, \mathbf{t}_{k'}), \quad (4)$$

$$p_{k,k'} \leq \zeta(\hat{t}, \mathbf{t}_{\mu(k)}) + \mathbf{p}_{k,\mu(k)} - \zeta(\hat{t}, \mathbf{t}_{k'}), \quad (5)$$

$$p_{k,k'} < \zeta(\mathbf{t}_{k'}, \mathbf{t}_k) - \zeta(\mathbf{t}_{k'}, \mathbf{t}_{\mu(k')}) - \mathbf{p}_{k',\mu(k')}. \quad (6)$$

By (3), there exists p such that (4) and (5) is simultaneously satisfied. By the premise that (k, k') form a complete-information blocking pair, there exists p such that (4) and (6) are satisfied. Thus, the right-hand side of (4) must be strictly less than the right-hand sides of (5) and (6), implying there exists $p_{k,k'}$ that satisfies all three inequalities. By (4) and (6), p satisfies Condition (b) of Lemma 3, By (4) and (5), (WI) implies $\tilde{\mathbf{t}} \in \beta_{k',k}(\mathbf{t}; T^\infty, p) \implies \tilde{\mathbf{t}}_k \geq \mathbf{t}_k$, and thus p satisfies condition (b) of Lemma 3. $\square$

A.2 Proof of Lemma 5

Lemma 5 *Suppose all agents have within-side information and submodularity or supermodularity holds. If there are no stable misassignments, namely that $M_{i,j}$ is empty for any pair (i, j) , then every stable matching outcome is efficient.*

Since any match produces a positive surplus (Assumption 0(b)), and the the premuneration functions are monotonic (Assumption 0(a)), any stable matching outcome $(\mu, \mathbf{p}, \mathbf{t})$ cannot have an unmatched worker and an unmatched firm, i.e., μ must match the maximal number of agents. Then, under supermodularity (resp. submodularity), any inefficient matching outcome must either (i) violate positive (resp. negative) assortativity (among matched agents), or (ii) the best $\min\{|I|, |J|\}$ firms are not matched to the best $\min\{|I|, |J|\}$ workers. (i) holds for stable matching outcomes since $M_{i,j}$ is empty for any pair (i, j) , while (ii) holds by Lemma 7 below.

LEMMA 7. *Suppose all agents have within-side information. Then, for any $\mathbf{t} \in T^\infty$ and any two agents k, k' on the same side, if $\mathbf{t}_{k'} > \mathbf{t}_k$, then $\mu(k) \neq \emptyset \implies \mu(k') \neq \emptyset$.*

PROOF. We prove this for workers only, the case for firms is symmetric.

Suppose to the contrary, that for some type assignment $\mathbf{t} \in T^\infty$, there exists workers $i, i' \in I$ such that $\mathbf{t}_{i'} > \mathbf{t}_i$ but $\mu(i) \neq \emptyset$ and $\mu(i') = \emptyset$. We will consider the potential blocking pair $(i', \mu(i))$.

Consider some $\tilde{\mathbf{t}} \in \arg \min_{\mathbf{t} \in T^\infty, \hat{\mathbf{t}}_{i'} = \mathbf{t}_{i'}, \hat{\mathbf{t}}_i = \mathbf{t}_i} \hat{\mathbf{t}}_{\mu(i)}$. By Assumption 0(a) (Monotonicity) and individual rationality, $(i', \mu(i))$

form a complete-information blocking pair at $\tilde{\mathbf{t}}$. By our choice of $\tilde{\mathbf{t}}$, $\tilde{\mathbf{t}} \in \mathcal{W}_{i', \mu(i)}(\tilde{\mathbf{t}})$, i.e., i' deduces $\mu(i)$'s type at $\tilde{\mathbf{t}}$, satisfying condition (a) of Lemma 3 for $k = i'$. Since $\mu(i') = \emptyset$, by Lemma 4(c), there exists p that satisfies condition (b) of Lemma 3 for $k' = \mu(i)$. Thus, $\tilde{\mathbf{t}} \notin T^\infty$, a contradiction. $\square$

A.3 Proof of Proposition 2

Proposition 2 *Suppose that Assumptions 1 (Unique Types) and 2 (Firm-Type Permutations) hold. Then, every stable matching outcome is efficient.*

Let us denote the set of matched workers and firms by $I_\mu = \{i \in I : \mu(i) \in J\}$ and $J_\mu = \{j \in J : \mu(j) \in I\}$. By Assumptions 1 and 2, there exists a set of firm types (relabelling types if necessary) $\bar{F} = \{f^1, f^2, \dots, f^{|J_\mu|}\}$ with $f^\ell > f^{\ell-1}$ for all ℓ such that each type $f \in \bar{F}$ is assigned to exactly one matched firm $j \in J_\mu$. Note that by Assumption 0 (Monotonicity and Positive Surplus), any stable matching outcome must match the maximal number of agents. If there are more workers than firms so that all firms are matched, then it is immediate that it is commonly known that f^ℓ is ℓ th-worst firm type. If there are more firms than workers so that some firms are not matched, then by Lemma 7, only the best firms are matched under stability. Thus, Assumptions 1 and 2 imply that it is common knowledge that under stability, f^ℓ is exactly the $(|J| - |I| + \ell)$ th-worst firm type (which is also the ℓ th-worst firm type among matched firms).

We would like to show that every worker i will satisfy (i, j) -MD at every $\mathbf{t} \in M_{i,j}$, which implies the MD condition. We proceed with the following induction hypothesis.

$H(\ell)$: For any $\mathbf{t} \in M_{i,j}$, if $\mathbf{t}_{\mu(i)} = f$ and $f \leq f^\ell$, then i satisfies (i, j) -MD at $\mathbf{t}$.

We have established $H(1)$ in Section 5.1. By induction, the following two claims imply that $H(\ell)$ holds for each ℓ . Hence, the MD condition holds.

CLAIM 1. *Suppose that $H(\ell)$ holds. Then for any worker i , firm j , and any $\mathbf{t} \in M_{i,j}$, we have $\mathbf{t}_{\mu(i)} > f^\ell$.*

PROOF. Suppose the contrary, that for some worker i and firm j , for some $\mathbf{t} \in M_{i,j}$, $\mathbf{t}_{\mu(i)} = f \leq f^\ell$. By construction of $M_{i,j}^*$ from Proposition 1 (specifically, $M_{i,j}^1 = \arg \min_{\tilde{\mathbf{t}} \in M_{i,j}} \tilde{\mathbf{t}}_{\mu(i)}$), $M_{i,j}^* \cap \{\tilde{\mathbf{t}} \in T^\infty : \tilde{\mathbf{t}}_{\mu(i)} \leq f^\ell\}$ must be non-empty. Thus, we may assume that $\mathbf{t} \in M_{i,j}^*$.

Since $\mathbf{t} \in M_{i,j}^*$, either (i, j) or $(\mu(j), \mu(i))$ can form a complete-information blocking pair at $\mathbf{t}$. Firstly, by $H(\ell)$, i satisfies (i, j) -MD at $\mathbf{t}$. There are two cases; either $\mathbf{t}_j > \mathbf{t}_{\mu(i)}$ or $\mathbf{t}_j < \mathbf{t}_{\mu(i)}$. In the former, by Lemma 6, $\mu(j)$ also satisfies (i, j) -MD at $\mathbf{t}$. In the latter, $\mathbf{t}_j < \mathbf{t}_{\mu(i)} \leq f^\ell$, so by $H(\ell)$, $\mu(j)$ also satisfies (i, j) -MD at $\mathbf{t}$.

Thus, in either case, Proposition 1 implies condition (a) of Lemma 3 for $k = i, \mu(j)$ for either potential blocking pair that forms a complete-information block at $\mathbf{t}$. Then, from Figure 4, the other agent $k' = j, \mu(i)$ in either potential blocking pair either also deduces (i, j) -MD at $\mathbf{t}$ or can utilize a separating wage to satisfy condition (b) of Lemma 3. Thus, $\mathbf{t} \notin T^\infty$ which is a contradiction. $\square$

CLAIM 2. *If $H(\ell)$ holds, then $H(\ell + 1)$ holds.*

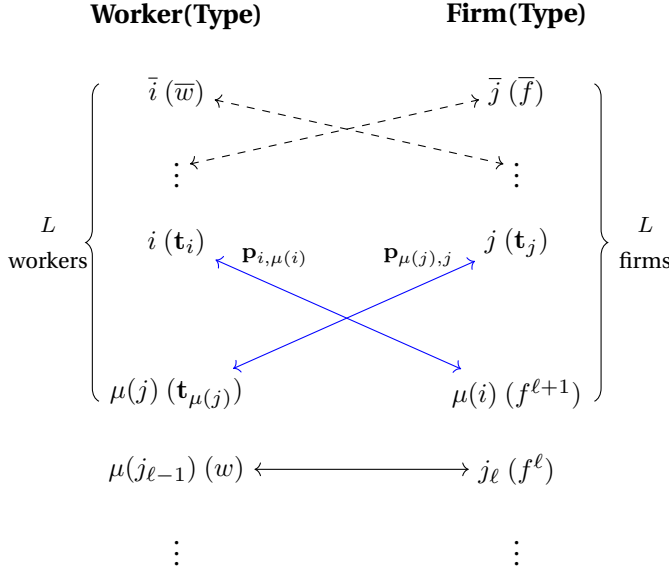


FIGURE 6. $\bar{w}, \bar{f}$ are the greatest types assigned by $\mathbf{t} \in M_{i,j}$; the best agents are matched.

PROOF. Firstly, by Claim 1, at any $\mathbf{t} \in T^\infty$, the best $L := (|J_\mu| - \ell)$ workers are matched to the best L firms, where the L th-best firm has type $f^{\ell+1}$; see Figure 6. Otherwise, a firm of type $f^{\ell+1}$ or greater, say j' , must be matched to a worker that is not one of the L best, but this implies that one of the L best workers, say i' , must be matched to a firm that is not one of the L best firms. However, this implies that $\mathbf{t} \in M_{i',j'}$ and $\mathbf{t}_{\mu(i')} \leq f^\ell$, which contradicts Claim 1.

Now suppose that for some pair (i, j) and $\mathbf{t} \in M_{i,j}$, we have $\mathbf{t}_{\mu(i)} = f^{\ell+1}$. Then by the argument above, at $\mathbf{t}$, $\mu(j)$ must be among the L -best workers; moreover, by (WI), i observes this and thus deduces that j has at least type $f^{\ell+1}$ at $\mathbf{t}$.¹¹ Since $\mathbf{t}_{\mu(i)} = f^{\ell+1}$ and there is only one firm of type $f^{\ell+1}$ (Assumption 1), i deduces that j is of at least type $f^{\ell+2}$. That is, i satisfies (i, j) -MD at $\mathbf{t}$. $\square$

A.4 Proof of Lemma 8

LEMMA 8. Suppose Assumptions 1 (Unique Types) and 3 (Negligible Worker-Surplus Variation) hold, then $(\mu, \mathbf{p})$ exhibits wage monotonicity.

PROOF. Consider any matched worker i and matched firm j such that $\mu(i) \neq j$ and $\mathbf{p}_{i,\mu(i)} \leq \mathbf{p}_{\mu(j),j}$. It suffices to show that the set of type assignments $N_{i,j} := \{\tilde{\mathbf{t}} \in T^\infty :$

¹¹Formally, let $w^{L(\mathbf{t})} = \max\{w \in W : |\{i' \in I_\mu : \mathbf{t}_{i'} \geq w\}| \geq L\}$ be the L th-best worker type. Note that for each $w \in W$, the sets $|\{i' \in I_\mu : \mathbf{t}_{i'} \geq w\}| = |\{i' \in I_\mu : \tilde{\mathbf{t}}_{i'} \geq w\}|$ remain the same across $\Pi_{i'}(\mathbf{t})$ by (WI), for each worker $i' \in I$.

$\tilde{t}_{\mu(i)} > \tilde{t}_j\}$ must be empty, i.e., that a strictly better firm must pay a higher wage.¹² Suppose to the contrary that $N_{i,j} \neq \emptyset$. Then, the following sets must also be nonempty:

$$N_{i,j}^1 = \arg \min_{\tilde{t} \in N_{i,j}} \tilde{t}_i, \quad N_{i,j}^2 = \arg \min_{\tilde{t} \in N_{i,j}^1} \tilde{t}_j, \quad N_{i,j}^* = \arg \min_{\tilde{t} \in N_{i,j}^2} \tilde{t}_{\mu(j)}.$$

We first establish the following claim:

CLAIM 3. *At any $\mathbf{t} \in N_{i,j}^*$, agent $k = i, j, \mu(i)$ deduces the type of agent $k' = j, i, \mu(j)$, respectively.*

PROOF. First, for $k = i$, observe that $\tilde{t} \in \mathcal{W}_{i,j}(\mathbf{t}) \implies \tilde{t}_j \leq t_j < t_{\mu(i)} = \tilde{t}_{\mu(i)}$, where the weak inequality is by definition of $\mathcal{W}_{i,j}(\mathbf{t})$, and the equality is by (PI). Also by (PI), we have $\tilde{t}_i = t_i$ and thereby $\tilde{t} \in N_{i,j}^1$. Then it follows from the definition of $N_{i,j}^2$ that $\tilde{t}_j \geq t_j$, thus $\mathbf{t} \in \mathcal{W}_{i,j}(\mathbf{t})$. Second, for $k = j, \mu(i)$, by (WI) and (PI), we have $\mathcal{W}_{k,k'}(\mathbf{t}) \subset N_{i,j}$. Then, by the definitions of $N_{i,j}^1$ (for $k = j$) or $N_{i,j}^*$ (for $k = \mu(i)$), and $\mathcal{W}_{k,k'}(\mathbf{t})$, we have $\tilde{t} \in \mathcal{W}_{k,k'}(\mathbf{t}) \implies \tilde{t}_{k'} \geq t_{k'}$, thus $\mathbf{t} \in \mathcal{W}_{k,k'}(\mathbf{t})$. $\square$

To reach a contradiction, consider $\mathbf{t} \in N_{i,j}^* \subset T^\infty$ and the following cases (i) and (ii). First, suppose (i) $t_i < t_{\mu(j)}$. For this case, $\mathbf{t}$ is an (i, j) -misassignment, and thus either (i, j) or $(\mu(j), \mu(i))$ form a complete-information blocking pair at $\mathbf{t}$. By Claim 3, condition (a) of Lemma 3 will hold for whichever pair forms a complete-information block at $\mathbf{t}$ (at any p). From Figure 4, the other agent in any of these potential blocking pairs either also deduces (i, j) -MD at $\mathbf{t}$ or can utilize a separating wage, so condition (b) of Lemma 3 holds. Thus, Lemma 3 implies $\mathbf{t} \notin T^\infty$, which is a contradiction.

Next, suppose (ii) $t_i > t_{\mu(j)}$. Then, by Assumption 3 (Negligible Worker-Surplus Variation), we have

$$[\zeta(\mathbf{t}_i, \mathbf{t}_{\mu(i)}) - \zeta(\mathbf{t}_i, \mathbf{t}_j)] < [\zeta(\mathbf{t}_j, \mathbf{t}_i) - \zeta(\mathbf{t}_j, \mathbf{t}_{\mu(j)})].$$

Now since $[p_{\mu(j),j} - p_{i,\mu(i)}] \geq 0$ by the initial assumption, we have

$$[\zeta(\mathbf{t}_j, \mathbf{t}_i) - \zeta(\mathbf{t}_j, \mathbf{t}_{\mu(j)})] + p_{\mu(j),j} > \zeta(\mathbf{t}_i, \mathbf{t}_{\mu(i)}) - \zeta(\mathbf{t}_i, \mathbf{t}_j) + p_{i,\mu(i)}. \quad (7)$$

Then, for $p_{i,j}$ between RHS and LHS of (7), we have

$$\zeta(\mathbf{t}_i, \mathbf{t}_j) + p_{i,j} > \zeta(\mathbf{t}_i, \mathbf{t}_{\mu(i)}) + p_{i,\mu(i)} \text{ and } \zeta(\mathbf{t}_j, \mathbf{t}_i) - p_{i,j} > \zeta(\mathbf{t}_j, \mathbf{t}_{\mu(j)}) - p_{\mu(j),j}.$$

Thus, (i, j, p) complete-information block the matching at $\mathbf{t}$. By Claim 3, since i and j deduce the type of each other, the conditions for Lemma 3 hold. This implies that $\mathbf{t} \notin T^\infty$, which is a contradiction. $\square$

¹²This is an implication stronger than that implied by Wage Monotonicity.

A.5 Examples of Inefficient Stable Matching Outcomes

A.5.1 Without Within-Side Information among Firms We construct an example assuming supermodularity and all four agents are matched. A similar example can also be constructed under submodularity or when some agents are unmatched. Consider an economy with two workers, two firms, and a matching function μ that match workers i and $\mu(j)$ to firms $\mu(i)$ and j , respectively. Premuneration values are $\zeta(w, f) = \zeta(f, w) = wf + 2$, $\mathbf{p}_{i, \mu(i)} = 2$, and $\mathbf{p}_{\mu(j), j} = 0$; $\bar{T} = \{\mathbf{t}, \tilde{\mathbf{t}}\}$ where $\mathbf{t}(i, \mu(j), \mu(i), j) = (2, 1, 2, 3)$ and $\tilde{\mathbf{t}}(i, \mu(j), \mu(i), j) = (1, 1, 1, 3)$. Thus, $(\mu, \mathbf{p})$ is individually rational at both $\mathbf{t}, \tilde{\mathbf{t}}$. Suppose every agent has complete information except for j , for whom $\Pi_j(\mathbf{t}) = \{\bar{T}\}$, namely, that j still knows j 's own type and the type of j 's partner $\mu(j)$ but not the type of $\mu(i)$ or i . In particular, firms' types are common knowledge as is the case with one-sided incomplete information.

We show the inefficient outcome $(\mu, \mathbf{p}, \mathbf{t})$ is stable; by Lemma 1, (i) it suffices to show that $\bar{T}$ is self-stabilizing, (ii) one can verify that $(\mu, \mathbf{p}, \tilde{\mathbf{t}})$ is complete-information-stable so by (i) we need only verify that $(\mu, \mathbf{p}, \mathbf{t})$ is not $\bar{T}$ -blocked, and (iii), since only (i, j) can form a complete-information blocking pair at $\mathbf{t}$, by (ii) we need only verify that (i, j) cannot $\bar{T}$ -block $(\mu, \mathbf{p}, \mathbf{t})$ (in particular, we do not need to check potential $\bar{T}$ -blocks by $(\mu(j), \mu(i))$).

Suppose, to the contrary, that (i, j, p) or (j, i, p) $\bar{T}$ -block the matching at $\mathbf{t}$. Note that worker i has complete-information, thus p must be such that i benefits at $\mathbf{t}$, i.e., $\zeta(\mathbf{t}_i, \mathbf{t}_j) + p_{i,j} = (6 + 2) + p_{i,j} > (4 + 2) + 2 = \zeta(\mathbf{t}_i, \mathbf{t}_{\mu(i)}) + \mathbf{p}_{i, \mu(i)}$, which simplifies to $p_{i,j} > 0$.

But, for any $p_{i,j} > 0$, i benefits at $\tilde{\mathbf{t}}$, i.e., $\zeta(\tilde{\mathbf{t}}_i, \tilde{\mathbf{t}}_j) + p_{i,j} > (3 + 2) + 0 = (1 + 2) + 2 = \zeta(\tilde{\mathbf{t}}_i, \tilde{\mathbf{t}}_{\mu(i)}) + \mathbf{p}_{i, \mu(i)}$. But this implies $\tilde{\mathbf{t}} \in \beta_{j,i}(\mathbf{t}; \bar{T}, p)$, i.e., j rationally considers $\tilde{\mathbf{t}}$.

However, j does not strictly benefit at $\tilde{\mathbf{t}}$, i.e., $\zeta(\tilde{\mathbf{t}}_j, \tilde{\mathbf{t}}_i) + p_{j,i} < (3 + 2) - 0 = (3 + 2) - 0 = \zeta(\tilde{\mathbf{t}}_j, \tilde{\mathbf{t}}_{\mu(j)}) + \mathbf{p}_{j, \mu(j)}$. Thus inequality (2) is not satisfied for j , for $\tilde{\mathbf{t}} \in \beta_{j,i}(\mathbf{t}; \bar{T}, p)$. Additionally, recall that $\beta_{j,i}(\mathbf{t}; \bar{T}, p) \subset \alpha_j(\mathbf{t}; \bar{T})$. Thus there is no p such that (i, j, p) or (j, i, p) $\bar{T}$ -block $(\mu, \mathbf{p}, \mathbf{t})$.

A.5.2 No Stable Misassignments, but Workers do not have Within-Side Information

Consider an economy and matching $(\mu, \mathbf{p})$ where there is one firm j matched to a worker $\mu(j)$ and is paying $\mathbf{p}_{\mu(j), j} = -4$, and one other worker i who is unmatched. Premuneration functions $\zeta(w, f) = \zeta(f, w) = wf$ are the product of types. Let $\bar{T} = \{\mathbf{t}, \mathbf{t}^i, \mathbf{t}^j\}$, where $\mathbf{t}(i, \mu(j), j) = (3, 2, 2)$, $\mathbf{t}^i(i, \mu(j), j) = (3, 4, 1)$, and $\mathbf{t}^j(i, \mu(j), j) = (1.5, 2, 2)$. Assume that agents do not have within-side information, so $\Pi_k(\mathbf{t}) = \{\mathbf{t}, \mathbf{t}^k\}$ for $k = i, j$, i.e., agent k cannot distinguish between $\mathbf{t}, \mathbf{t}^k$ (while $\mu(j)$ has complete-information).

Firstly, since there is only one matched pair, $M_{i,j}$ is empty. We now show that the inefficient outcome $(\mu, \mathbf{p}, \mathbf{t})$ is stable. By Lemma 1, it suffices to show that $\bar{T}$ is self-stabilizing, and since $(\mu, \mathbf{p}, \mathbf{t}^k)$ is complete-information-stable for $k = i, j$, it only remains to verify that (i, j) are not a $\bar{T}$ -blocking pair of $(\mu, \mathbf{p}, \mathbf{t})$.

Suppose to the contrary that (i, j, p) $\bar{T}$ -block $(\mu, \mathbf{p}, \mathbf{t})$. Since $\mathbf{t}^i \in \alpha_i(\mathbf{t}; \bar{T})$, i requires $\zeta(\mathbf{t}_i, \mathbf{t}_j^i) + p_{i,j} = 3 + p_{i,j} > 0 \iff p_{i,j} > -3$. At $p_{i,j} > -3$, however, $\zeta(\mathbf{t}_i^j, \mathbf{t}_j) + p_{i,j} > 3 - 3 = 0$, i.e., $\mathbf{t}^j \in \beta_{j,i}(\mathbf{t}; \bar{T}, p)$. But then $\zeta(\mathbf{t}_j, \mathbf{t}_i^j) - p_{i,j} < 6 < 8 = \zeta(\mathbf{t}_j, \mathbf{t}_{\mu(j)}) - \mathbf{p}_{\mu(j), j}$, i.e., j does not benefit at $\mathbf{t}^j$.

Similarly, suppose (j, i, p) $\bar{T}$ -block $(\mu, \mathbf{p}, \mathbf{t})$; since $\mathbf{t}^j \in \alpha_j(\mathbf{t}; \bar{T})$, j requires $\zeta(\mathbf{t}_j, \mathbf{t}_i^j) - p_{i,j} = 3 - p_{i,j} > 8 = \zeta(\mathbf{t}_j, \mathbf{t}_{\mu(j)}) - \mathbf{p}_{\mu(j),j} \iff p_{i,j} < -5$. At $p_{i,j} < -5$, we have $\zeta(\mathbf{t}_j^i, \mathbf{t}_i) - p_{i,j} > 3 + 5 = \zeta(\mathbf{t}_j^i, \mathbf{t}_{\mu(j)}^i) - \mathbf{p}_{\mu(j),j}$, i.e., $\mathbf{t}^i \in \beta_{i,j}(\mathbf{t}; \bar{T}, p)$. However, since $\zeta(\mathbf{t}_i, \mathbf{t}_j^i) + p_{i,j} < 3 - 5 < 0$, worker i does not benefit at $\mathbf{t}^i$. Thus, there is no p such that (i, j, p) or (j, i, p) $\bar{T}$ -block $(\mu, \mathbf{p}, \mathbf{t})$.

A.5.3 Necessity of the MD Condition Consider an economy with two workers, two firms, and a matching function μ that match workers i and $\mu(j)$ to firms $\mu(i)$ and j , respectively. Consider $\bar{T} \subset \mathbb{R}^4$, and premuneration functions ζ that satisfy supermodularity, and a matching $(\mu, \mathbf{p})$ of an economy with four agents $i, j, \mu(j), \mu(i)$ who all have within-side information, and wages $\mathbf{p}_{i,\mu(i)} = \mathbf{p}_{j,\mu(j)} = 0$. With respect to each $\mathbf{t} \in M_{i,j}$, at which $(\mu, \mathbf{t})$ is non-positive assortative, (re)label the “better” agents as i and j , i.e., $\mathbf{t}_i > \mathbf{t}_{\mu(j)}$ and $\mathbf{t}_j > \mathbf{t}_{\mu(i)}$. We only focus on (i, j) , since $(\mu(j), \mu(i))$ deduce an (i, j) -misassignment at $\mathbf{t}$ by Lemma 6.

PROPOSITION 5. *If neither of the better agents $k \in \{i, j\}$ satisfy (i, j) -MD at $\mathbf{t}$, that is, for both $k = i, j$, there exists $\mathbf{t}^k \in \mathcal{W}_{k,k'}(\mathbf{t}) \setminus M_{i,j}$, then $\mathbf{t} \in T^\infty$.*

PROOF. It suffices to show that $T := \{\mathbf{t}^i, \mathbf{t}^j, \mathbf{t}\}$ is self-stabilizing.

For $k = i, j$, since $\mathbf{t}^k$ is a worst-case hypothesis of agent k at $\mathbf{t}$ and agent k knows k 's own type and the type of $\mu(k')$ by (WI), we have $\mathbf{t}_k^k = \mathbf{t}_k > \mathbf{t}_{\mu(k')} = \mathbf{t}_{\mu(k')}^k$. Moreover, since $\mathbf{t}^k \notin M_{i,j}$, we must have $\mathbf{t}_{\mu(k)}^k \geq \mathbf{t}_{k'}^k$. If (k, k', p) T -block $(\mu, \mathbf{p}, \mathbf{t})$, agent $k = i, j$ satisfies inequality (1) at all $\tilde{\mathbf{t}} \in \alpha_k(\mathbf{t}; T)$, including $\tilde{\mathbf{t}} = \mathbf{t}^k$. Since $\mathbf{p}_{k,\mu(k)} = 0$, it follows from (1) that $p_{k,k'} > \zeta(\mathbf{t}_k^k, \mathbf{t}_{\mu(k)}^k) - \zeta(\mathbf{t}_k^k, \mathbf{t}_{k'}^k)$. Then, since $\mathbf{t}_{\mu(k)}^k \geq \mathbf{t}_{k'}^k$, we have $p_{k,k'} > 0$.

On the other side, $k' = j, i$ must satisfy (2) for all $\tilde{\mathbf{t}} \in \beta_{k',k}(\mathbf{t}; T, p)$, which implies that at least one of the follow two cases must be true, but

will derive a contradiction to $p_{k,k'} > 0$.

- (i) $\mathbf{t}^{k'} \notin \beta_{k',k}(\mathbf{t}; T, p)$;
- (ii) $\mathbf{t}^{k'} \in \beta_{k',k}(\mathbf{t}; T, p)$ and (2) is satisfied for $\mathbf{t}^{k'}$.

In Case (i), since $\mathbf{p}_{k,\mu(k)} = \mathbf{p}_{k',\mu(k')} = 0$, we have $p_{k',k} \geq \zeta(\mathbf{t}_k^{k'}, \mathbf{t}_{k'}^{k'}) - \zeta(\mathbf{t}_k^{k'}, \mathbf{t}_{\mu(k)}^{k'})$. Since

agent k' knows k' 's own type and the type of $\mu(k)$ by (WI) in contemplating the worst-case hypothesis $\mathbf{t}^{k'}$ at $\mathbf{t}$, it follows that $\mathbf{t}_{k'}^{k'} = \mathbf{t}_{k'} > \mathbf{t}_{\mu(k)} = \mathbf{t}_{\mu(k)}^{k'}$ and hence $p_{k',k} > 0$, which contradicts $p_{k,k'} > 0$.

In Case (ii), since (2) is satisfied for $\mathbf{t}^{k'}$ and $\mathbf{p}_{k,\mu(k)} = \mathbf{p}_{k',\mu(k')} = 0$, we have $p_{k',k} > \zeta(\mathbf{t}_k^{k'}, \mathbf{t}_{\mu(k)}^{k'}) - \zeta(\mathbf{t}_k^{k'}, \mathbf{t}_{k'}^{k'})$. Again, since $\mathbf{t}^{k'} \notin M_{i,j}$, we also have $\mathbf{t}_k^{k'} \leq \mathbf{t}_{\mu(k')}^{k'}$. Hence, we also have $p_{k',k} > 0$, which again contradicts $p_{k,k'} > 0$. $\square$

REFERENCES

Bikhchandani, Sushil (2017), “Stability with one-sided incomplete information.” *Journal of Economic Theory*, 168, 372–399. [12, 25]

Chade, Hector and Jan Eeckhout (2018), “Matching information.” *Theoretical Economics*, 13(1), 377–414. [3]

Chen, Yi-Chun and Samuel Ho (2020), “Efficiency of stable matching with two-sided incomplete information.” Working paper, Department of Economics, National University of Singapore, and Department of Economics, Yale University. [25]

Chen, Yi-Chun and Samuel Ho (2026), “Efficiency of stable matching with two-sided incomplete information.” Working paper, Department of Economics, National University of Singapore, and Department of Economics, Yale University. Available at SSRN: <https://www.ssrn.com/abstract=3496454>. [7, 18, 21, 24]

Chen, Yi-Chun and Gaoji Hu (2020), “Learning by matching.” *Theoretical Economics*, 15(1), 29–56. [12, 25]

Chen, Yi-Chun and Gaoji Hu (2023), “A theory of stability in matching with incomplete information.” *American Economic Journal: Microeconomics*, 15(1), 288–322. [7, 17, 25]

Eeckhout, Jan (2018), “Sorting in the labor market.” *Annual Review of Economics*, 10(1), 1–29. [3]

Grossman, Gene M. and Giovanni Maggi (2000), “Diversity and trade.” *American Economic Review*, 90(5), 1255–1275. [3]

Koopmans, Tjalling C. and Martin Beckmann (1957), “Assignment problems and the location of economic activities.” *Econometrica*, 25(1), 53–76. [1, 12]

Liu, Qingmin (2017), “Rational expectations stability and competitive equilibrium in matching with incomplete information.” Working paper, Department of Economics, Columbia University. [25]

Liu, Qingmin (2020), “Stability and bayesian consistency in two-sided markets.” *American Economic Review*, 110(8), 2625–66. [12, 25]

Liu, Qingmin (2023), “Cooperative analysis of incomplete information.” Working paper, Department of Economics, Columbia University. [25]

Liu, Qingmin, George J. Mailath, Andrew Postlewaite, and Larry Samuelson (2014), “Stable matching with incomplete information.” *Econometrica*, 82(2), 541–587. [1, 25]

Park, Changwoo (2024), “Stable matchings under two-sided asymmetric incomplete information.” Working paper, Department of Economics, University of Rochester. [7, 25]

Piketty, Thomas (1993), “Implementation of first-best allocations via generalized tax schedules.” *Journal of Economic Theory*, 61(1), 23–41. [21]

Pomatto, Luciano (2022), “Stable matching under forward-induction reasoning.” *Theoretical Economics*, 17(4), 1619–1649. [12, 25]

Schulhofer-Wohl, Sam (2006), “Negative assortative matching of risk-averse agents with transferable expected utility.” *Economics Letters*, 92(3), 383–388. [3]

Shapley, Lloyd S. and Martin Shubik (1971), “The assignment game i: The core.” *International Journal of Game Theory*, 1(1), 111–130. [1, 12]

Wang, Ziwei (2023), “Rationalizable stability in matching with incomplete information.” Working paper, Guanghua School of Management, Peking University. [25]