A Theory of Stability in Matching with Incomplete Information*

Yi-Chun Chen† Gaoji Hu ‡

November 23, 2018

Abstract

This paper provides a framework for the study of two-sided matching markets with incomplete information. In those markets, two agents from opposite sides are willing to form a new partnership if both of them can benefit from it. We propose a blocking notion that describes how the two agents evaluate the new partnership (i.e., the potential blocking), and when they are both willing to participate in it. To evaluate the potential blocking under incomplete information, the two agents need to consider various possible types of their opponents. In our setting, an agent would ignore her/his opponent's types under which her/his opponent can never benefit from the potential blocking, and consider only the rest. Then, each of the two agents would further ignore some of her/his opponent's types under which her/his opponent, using the refined consideration sets, can never benefit from the potential blocking. With such refinements of the consideration sets, the two agents evaluate the potential blocking more precisely. They are willing to participate in the potential blocking if both of them, given their finest consideration sets, can benefit from it. We also propose a stability notion that captures, in addition to individual rationality and no blocking, the idea that absence of blocking conveys no further information. We compare four alternative blocking notions, i.e., (i) naive blocking, (ii) conservative blocking, (iii) blocking, and (iv) aggressive blocking. They are ranked according to how aggressive agents are when they refine their consideration sets. Among them, (i)-(iii) guarantee that the two blocking agents would obtain higher payoffs under the true types; and (i)-(ii) satisfy monotonicity with respect to information, i.e., it is easier to find blocking opportunities when agents' information is more precise.

JEL Classification: C78, D83

Keywords: two-sided matching, incomplete information, blocking, stability.

*We are especially grateful to Fuhito Kojima and Qingmin Liu for encouragement and stimulating discussions. We also thank David Ahn, In-Koo Cho, Laura Doval, Jiangtao Li, Bart Lipman, Xiao Luo, George Mailath, Luciano Pomatto, Andy Postlewaite, Mike Ostrovsky, Joel Sobel, Ning Sun, Xiang Sun, Satoru Takahashi, Qianfeng Tang, Yan Yan, Zaifu Yang and Yongchao Zhang for their comments and suggestions. Part of this paper was written while Hu was visiting Stanford University and he would like to thank the institution for its hospitality and support. This work is supported by Singapore MOE AcRF Tier 1 Grant. All remaining errors are our own.

†Department of Economics, National University of Singapore, S(117570). ecsycc@nus.edu.sg.

‡Nanyang Business School, Nanyang Technological University, S(639798). hu.gaoji@gmail.com

1 Introduction

Stability of matchings in two-sided markets has been connected to both equity and efficiency, two of most important objectives in economics.[1] A prevailing assumption in this literature is that the information is complete, i.e., the characteristics and preferences of all market participants are publicly known. However, incomplete information is ubiquitous in matching markets: a man may has imprecise information about his girl friend; firms may not know the productivity of their potential employees; professors may recruit incompetent PhD students based on their undergraduate transcripts; and etc. We consider a framework where workers are matched with firms, and depart from the prevailing assumption of complete information. We propose notions of (i) individual rationality, (ii) blocking and (iii) stability, and study their properties.

Stability under complete information requires individual rationality (i.e., each agent has a nonnegative payoff) and no blocking pair (i.e., no worker and firm would both prefer being matched with each other at some wage to staying with their current partners). When a firm has incomplete information about workers' types, however, she would not know whether she would prefer hiring another worker to keeping her current employee. Moreover, the firm may not even know whether keeping her current employee is better than becoming unmatched. In this situation, the notions of individual rationality, blocking and stability in the complete-information environment become inadequate.

In the incomplete-information setting, each agent is associated with a type, which determines agents' payoffs from a match. Moreover, an agent's information is described by a partition over possible type profiles. In our setting, a state of the market consists of an information structure (i.e., the true type profile and the partition profile) and an allocation (i.e., the prevailing matching and wage profile). We assume that agents can observe the prevailing allocation. A state is stable if (i) the allocation is individually rational with respect to the information structure, (ii) the allocation admits no blocking pair with respect to the information structure, and (iii) individual rationality and the absence of blocking convey *no* further information to the agents.

To be precise, we consider two decision making criteria for agents, i.e., max-min and Bayesian, and decribe the stability notions under both of them. The first criterion assumes that agents care about the worst-case payoff when they are not sure about the true type profile; the second criterion assumes that agents care about their (conditional) expected payoffs when they are not sure about the true type profile. Under the first criterion, a state is individually rational if for each agent, her/his payoff is nonnegative under all type profiles in the true partition cell, i.e., the partition cell that contains the

[1] See Balinski and Sönmez (1999) and Abdulkadiroğlu and Sönmez (2003) for how stability implies the elimination of justified envy, a basic fairness axiom. See Shapley and Shubik (1971) and Liu et al. (2014) for how stability leads to efficiency.

true type profile.

Following the literature of complete-information matching, a state is blocked if a worker and a firm are willing to form a new partnership, from which both of them can benefit. With incomplete information, it is crucial to describe how the two agents evaluate the new partnership (i.e., the potential blocking), and when they are both willing to participate in it. To evaluate the potential blocking under incomplete information, the two agents need to consider various possible type profiles. In our setting, an agent would ignore some of the possible type profiles under which her/his opponent is never willing to participate in the potential blocking; the reason is that the potential blocking would definitely fail under those type profiles. In other words, agents consider only a subset of the possible type profiles when they evaluate a potential blocking. Then, each of the two agents would further ignore some of her/his opponent's types under which her/his opponent, using the refined consideration sets, is never willing to participate in the potential blocking. With such refinements of the consideration sets, the two agents evaluate the potential blocking more precisely. We say that a state is blocked if there exist a worker-firm pair and a potential wage such that each of the two agents would obtain a higher payoff under all type profiles in her/his finest consideration set. Then, stability is described by individual rationality, no blocking, and no extra information from individual rationality and no blocking.

Requirements (i) and (ii) for stability extend their complete-information counterparts. However, requirement (iii) embodies a notion of "informational stability" which is specific to the incomplete-information setting.[2] The stability of market states under the Bayesian criterion is analogously defined, in which agents care about the expected payoffs instead of the worst-case payoffs.

We compare four alternative worst-case blocking notions, i.e., (i) naive blocking, (ii) conservative blocking, (iii) blocking, and (iv) aggressive blocking. They are ranked according to how aggressive agents are when they refine their consideration sets. Firstly, when a naive firm (worker) evaluate a potential blocking, she/he does not refine her/his consideration set, i.e., she/he considers all possible type profiles in her/his true partition cell. Secondly, a conservative firm (worker) refines her/his consideration set, by excluding a type profile, only if within the conservative worker's (firm's) consideration set which contains that type profile, the worker (the firm) would always obtain a lower payoff. Thirdly, the blocking notion which we adopt assumes that the agents are more aggressive than under conservative blocking: the intuition is that if a worker (a firm) would always obtain a lower payoff within his (her) consideration set, then the worker (the firm) is never willing to participate in the potential blocking. Finally, an aggressive firm (worker) refines her/his consideration set, by excluding a type profile, if within the aggressive worker's (firm's) consideration set which contains that type profile, the worker (the firm) would obtain a lower worst-case payoff.

[2]See Liu et al. (2014), Liu (2018), and Chen and Hu (2018) for more discussions on information stability.

The first three blocking notions are also ranked in the following sense: a state being naively blocked implies that it is conservatively blocked, which implies that it is blocked. Moreover, blocking implies complete-information blocking. See Figure 1 below.

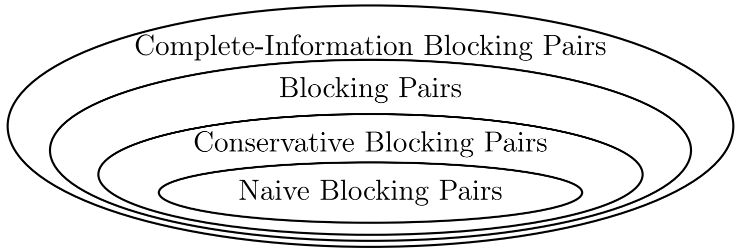

Figure 1: A Ranking of Blocking Notions.

We show that under any one of the blocking notions (i)-(iii), the two blocking agents would obtain higher payoffs under the true type profile. This property, termed as *improvement-at-the-truth*, is consistent with the complete-information blocking. Moreover, under any one of the blocking notions (i)-(ii), a property called *monotonicity* is satisfied. In particular, it is easier for agents to find blocking opportunities when their information is more precise, i.e., when agents have finer partitions. The comparison between different blocking notions are summarized in the table below. Notably, the notion(s) of aggressive blocking satisfy neither improvement-at-the-truth nor monotonicity.

Blocking Notions	Properties	
	Improvement at Truth	Monotonicity
Complete-Information Blocking	$\surd$	N.A.
Blocking	$\surd$	
Conservative Blocking	$\surd$	$\surd$
Naive Blocking	$\surd$	$\surd$

The rest of this section reviews the related literature. Section 2 introduces the model. Section 3 defines blocking and stability with incomplete information, and also documents some properties of blocking. Section 4 compares our blocking criteria with three alternative blocking notions: naive blocking, conservative blocking and aggressive blocking. Section 5 studies stability under the Bayesian criterion. Section 6 concludes.[3]

[3] See the supplementary appendix of the current paper for more results, examples and discussions. To be precise, Appendix C discusses a particular set of stable states. Appendix D investigates the convergence of learning-blocking paths under different stability notions. Appendix E studies the comparative statics when an outside worker (firm) enters a market which is already in its stable state. Examples in Appendix E also illustrates how some standard results in the complete-information setting, such as the lattice structure of stable matchings and the lone-wolf theorem, may fail when information is incomplete.

Related Literature

The seminal model of Gale and Shapley (1962), studying the marriage matching market and the college admission market, has been used in a large literature of two-sided matching. Many classical theories are surveyed in Roth and Sotomayor (1990) and more recently by, for example, Roth (2008) and Abdulkadiroglu and Sönmez (2013). In this literature, a prevailing assumption is that the information is complete.

Recently, Liu et al. (2014) studied stable matchings with one-sided incomplete information. Particularly, in the job market setting, workers have private types while firms possess some information about workers' types. They introduced an ex ante notion of stability to predict what kind of matching outcomes could arise when heterogeneous firm-specific information is not realized/specified. In a previous paper, Chen and Hu (2018), we formulated heterogeneous firm-specific information and studied a stability notion that is defined with respect to firms' information structure. In Liu et al. (2014) and the main part of Chen and Hu (2018), firms who are uncertain about workers' types care about the worst-case payoffs. The current paper extends the stability notions of both Liu et al. (2014) and Chen and Hu (2018), and also extends the Bayesian stability notion of Chen and Hu (2018), to the setting with two-sided incomplete information.

In another setting with one-sided incomplete information, Liu (2018) proposes a stability criterion that requires the Bayesian consistency of three beliefs: exogenously given prior beliefs, off-path beliefs conditional on counterfactual pairwise blockings, and on-path beliefs for stable matchings in the absence of such blockings. The current paper differs from Liu (2018): the stability criterion of Liu (2018) is ex ante in the sense that neither the true type profile nor the information partition is realized/fixed; in contrast, our stability notion is interim in the sense that, apart from the exogenously given prior belief, agents may also have access to additional exogenously given information about the true type profile. However, the Bayesian consistency of beliefs represents an ex ante version of information stability, which is implicitly introduced in Liu et al. (2014), and explicitly revisited in Chen and Hu (2018) and the current paper. Therefore, in terms of information stability, Liu (2018) and the current paper complement each other from different angles.

With two-sided incomplete information, the key subtlety, compared with the one-sided incomplete-information setting discussed in Liu et al. (2014) and Chen and Hu (2018), is that the worst-case criterion (similar for the Bayesian criterion) of a potential blocking pair involves interactions between the perception (i.e., willing or unwilling) of the worker and that of the firm. More precisely, when the potential blocking firm thinks about the worst possible type of the potential blocking worker, she takes into account the blocking worker's perception of the worst possible type of herself, which in turn takes into account herself's perception of the worst possible type of the blocking worker, etc. The potential blocking pair run through such an iterative thinking until any higher level of reasoning would not help the in evaluating each other.

Bikhchandani (2017) proposes a notion of stability which is similar to that of Liu et al. (2014) but which applies to a Bayesian setting with nontransferable utilities. Unlike Liu et al. (2014) and Bikhchandani (2017), Pomatto (2018) considers a noncooperative matching game and uses forward-induction reasoning to derive the set of stable outcomes that is identified in Liu et al. (2014).[4] Our stability notion is also related to the literature on the core, particularly the core in the incomplete-information problems. In our context, a coalition is simply a worker-firm pair. See Wilson (1978), Dutta and Vohra (2005), and the comprehensive discussions in Liu et al. (2014).

2 The Model

We consider the following setup of matching with incomplete information, which is based on Liu et al. (2014). The setup generalizes complete-information matching models studied by Shapley and Shubik (1971) and Crawford and Knoer (1981), and further the incomplete-information models studied by Liu et al. (2014) and Chen and Hu (2018).

There is a finite set I of workers to be matched with a finite set J of firms. Denote a generic worker by i, a generic firm by j, and a generic agent by k when we do not distinguish workers from firms. While each agent's index i or j is publicly observed, the agent's productivity is determined by the agent's private *type*. Let W be the finite set of worker types and F be the finite set of firm types. A type assignment for firms is a mapping $\mathbf{f} : J \to F$, and similarly a type assignment for workers is another mapping $\mathbf{w} : I \to W$. We denote by $\mathbf{t} := (\mathbf{w}, \mathbf{f})$ a generic type assignment, and denote by T a set of type assignments, i.e., $T \subset W^{|I|} \times F^{|J|}$.

A match between a worker of type $w \in W$ and a firm of type $f \in F$ gives rise to the *worker premuneration value* $\nu_{wf} \in \mathbb{R}$ and the *firm premuneration value* $\phi_{wf} \in \mathbb{R}$.[5] The sum of ν_{wf} and ϕ_{wf} is called the *surplus of the match*. Denote these values by $\nu_{\mathbf{w}(i),\mathbf{f}(\emptyset)}$ for unmatched worker i and $\phi_{\mathbf{w}(\emptyset),\mathbf{f}(j)}$ for unmatched firm j, both of which are set to be zero. The functions $\nu : W \times F \to \mathbb{R}$ and $\phi : W \times F \to \mathbb{R}$ are common knowledge among the agents. Given a match between worker i (of type $\mathbf{w}(i)$) and firm j (of type $\mathbf{f}(j)$) under some wage $p \in \mathbb{R}$, the worker's payoff and the firm's payoff are, respectively, $\nu_{\mathbf{w}(i),\mathbf{f}(j)} + p$ and $\phi_{\mathbf{w}(i),\mathbf{f}(j)} - p$.[6]

[4] Another stream of literature studies stable mechanisms (instead of stable matchings) which also involve incomplete information. See, for example, Roth (1989), Chakraborty et al. (2010), and Ehlers and Massó (2007, 2015).

[5] See Mailath et al. (2013, 2017) for discussions on premuneration values.

[6] If we adopt the practice fact that salaries must be rounded to the nearest dollar or penny, the analysis in this paper

A *matching* is a mapping $\mu : I \cup J \to I \cup J$ such that (i) $\mu(i) \in J \cup \{i\}$; (ii) $\mu(j) \in I \cup \{j\}$; and (iii) $\mu(i) = j$ if and only if $\mu(j) = i$ for all $i \in I$ and all $j \in J$. In words, each worker is either unmatched, denoted by $\mu(i) = i$, or assigned to a firm that employs him; and each firm is either unmatched, denoted by $\mu(j) = j$, or hires a worker. A *payment scheme* $\mathbf{p}$ associated with a matching μ is a vector that specifies a payment $\mathbf{p}_{i,\mu(i)} \in \mathbb{R}$ for each worker i, and a payment $\mathbf{p}_{\mu(j),j} \in \mathbb{R}$ for each firm j. To avoid nuisance cases, we associate zero payments with unmatched agents, by setting $\mathbf{p}_{ii} = \mathbf{p}_{jj} = 0$. Finally, an *allocation* $(\mu, \mathbf{p})$ consists of a matching μ and an associated payment scheme $\mathbf{p}$. We assume that the entire allocation is publicly observable.

We assume that the only fact that is common knowledge is that the agents' type assignment belongs to T.[7] Beyond the public information, each agent may also have her own private information about the type assignment. Specifically, for every agent k, we describe her private information by a partition Π_k over T. For any type assignment $\mathbf{t}$, write $\Pi_k(\mathbf{t})$ as the element of partition Π_k that contains $\mathbf{t}$. When the true type assignment is $\mathbf{t}$, agent k regards each type assignment $\mathbf{t}'$ in $\Pi_k(\mathbf{t})$ as possible. Denote the profile of partitions by Π, i.e., $\Pi := \left(\{\Pi_i\}_{i \in I}, \{\Pi_j\}_{j \in J}\right)$, which is assumed to be common knowledge. Say partition profile Π' is (weakly) *finer* than partition profile Π if, for each agent k, we have $\Pi'_k(\mathbf{t}) \subset \Pi_k(\mathbf{t})$ for every type assignment $\mathbf{t} \in T$. Say information is *complete* if every agent knows what is the true type assignment, i.e., $\Pi_k(\mathbf{t}) = \{\mathbf{t}\}$ for all $\mathbf{t} \in T$ and all k.

A *state* of the matching market, $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$, specifies an allocation $(\mu, \mathbf{p})$, a type assignment function $\mathbf{t}$ and a partition profile Π. In words, a state of the market describes who is matched with whom at which wage, and what each agent knows.

3 Stability with Incomplete Information

3.1 Individual Rationality

A state is said to be individually rational if each agent receives at least the payoff from remaining unmatched, which is assumed to be zero. Assume, as in Liu et al. (2014), that agents maximize their worst-case payoffs. The following definition of incomplete-information "individual rationality" naturally extends its complete-information counterpart.

will go through without any extra difficulty. This more practical restriction will be imposed in Supplementary Appendix D, where we study matching processes.

[7]In Supplementary Appendices, we will impose an additional assumption to study the path-to-stability problem and the related comparative statics. The assumption says that (i) that each agent knows her own type, and that (ii) within each matched pair, each agent knows her current partner's type. This assumption is also imposed in the literature; see Liu et al. (2014) and Chen and Hu (2018).

Definition 1. *A state* $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ *is said to be* ***individually rational*** *if for each agent and each type assignment that the agent considers possible, her/his payoff is nonnegative, i.e.,*

$$\nu_{\mathbf{t}'(i),\mathbf{t}'(\mu(i))} + \mathbf{p}_{i,\mu(i)} \geq 0 \text{ for all } \mathbf{t}' \in \Pi_i(\mathbf{t}) \text{ and all } i \in I \text{ and}$$
$$\phi_{\mathbf{t}'(\mu(j)),\mathbf{t}'(j)} - \mathbf{p}_{\mu(j),j} \geq 0 \text{ for all } \mathbf{t}' \in \Pi_j(\mathbf{t}) \text{ and all } j \in J.$$

3.2 Blocking

The notion of incomplete-information "blocking" naturally extends its complete-information counterpart and its counterpart with one-sided incomplete information; see Liu et al. (2014) and Chen and Hu (2018) for the latter. To be precise, a matching is *complete-information blocked* if there exist a worker and a firm such that both agents can benefit from being rematched with each other at some wage.

Consider a potential incomplete-information "blocking combination" $(i, j; p)$ for the state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$. We proceed to describe agents' consideration sets, which they use to evaluate the potential blocking. Particularly, for worker i to participate in the blocking combination $(i, j; p)$, he only needs to consider type assignments in $\Pi_i(\mathbf{t})$ under which firm j may be willing to participate in the blocking. Other type assignments in $\Pi_i(\mathbf{t})$ are irrelevant due to the firm's objection. Similarly, firm j only considers type assignments in $\Pi_j(\mathbf{t})$ under which worker i may be willing to participate in the blocking. Common knowledge of such consideration may lead to further refinements of their consideration sets, which is illustrated in the following example.

Example 1. *Consider a market with two workers, i.e.,* $I = \{\alpha, \beta\}$*, and two firms, i.e.,* $J = \{a, b\}$*. The set of possible type assignments is given by* $T = \{\mathbf{t}^{23,32}, \mathbf{t}^{21,34}, \mathbf{t}^{21,54}\}$*, i.e.,*

	α	β	a	b
$\mathbf{t}^{23,32}$:	2	3	3	2
$\mathbf{t}^{21,34}$:	2	1	3	4
$\mathbf{t}^{21,54}$:	2	1	5	4

In this example, a larger number means a better type. The premuneration values for workers and firms are given by the product form, i.e., $\nu_{w,f} = \phi_{w,f} = wf$*. Obviously, given any wage, every agent prefers a partner with a higher type to a partner with a lower type.*

Suppose firm a *hires worker* α *and firm* b *hires worker* β*. In other words, a matching* μ *is given by* $\mu(\alpha) = a$ *and* $\mu(\beta) = b$*. Suppose also that payments under* μ *are all zero, i.e.,* $\mathbf{p} = \mathbf{0}$*. Suppose further that that the first type assignment* $\mathbf{t}^{23,32}$ *is true, and that agents' partitions are given as if agents* *can observe their own types and their own partners' types, i.e.,*

$$\Pi_\alpha = \left\{\left\{\mathbf{t}^{23,32}, \mathbf{t}^{21,34}\right\}, \left\{\mathbf{t}^{21,54}\right\}\right\},$$
$$\Pi_\beta = \left\{\left\{\mathbf{t}^{23,32}\right\}, \left\{\mathbf{t}^{21,34}, \mathbf{t}^{21,54}\right\}\right\},$$
$$\Pi_a = \left\{\left\{\mathbf{t}^{23,32}, \mathbf{t}^{21,34}\right\}, \left\{\mathbf{t}^{21,54}\right\}\right\}, \text{ and}$$
$$\Pi_b = \left\{\left\{\mathbf{t}^{23,32}\right\}, \left\{\mathbf{t}^{21,34}, \mathbf{t}^{21,54}\right\}\right\}.$$

This completes the description of the market state $\left(\mu, \mathbf{p}, \mathbf{t}^{23,32}, \Pi\right)$.

It is straightforward to verify that if information is complete, then the following claims are true: (i) under $\mathbf{t}^{23,32}$, (β, a) *is the unique blocking pair; (ii) under* $\mathbf{t}^{21,34}$, (α, b) *is the unique blocking pair; and (iii) under* $\mathbf{t}^{21,54}$, (α, b) *is the unique blocking pair. For example, under* $\mathbf{t}^{21,54}$, *a blocking wage* p *could be* 3, *at which worker* α*'s new payoff is* 11 *and firm* b*'s new payoff is* 5, *both higher than their status quo payoffs.*

Now we consider an incomplete-information potential blocking combination $(\beta, a; 0)$. *Obviously, if agents consider all type assignments in* $\Pi_k\left(\mathbf{t}^{23,32}\right)$ *when they evaluate each other, then* $(\beta, a; 0)$ *is not a "reasonable blocking combination" for* $\left(\mu, \mathbf{p}, \mathbf{t}^{23,32}, \Pi\right)$. *The reason is that firm* a *would worry about the possible type assignment* $\mathbf{t}^{21,34}$. *To be precise, for the possible type* 1 *of worker* β, *firm* a *would obtain a lower payoff, i.e.,* 3, *than her status quo payoff, i.e.,* 6.

However, we proceed to argue that the combination $(\beta, a; 0)$ *should be a blocking combination for* $\left(\mu, \mathbf{p}, \mathbf{t}^{23,32}, \Pi\right)$. *First of all, worker* β *is willing to participate in the blocking* $(\beta, a; 0)$ *because he knows that the true type assignment is* $\mathbf{t}^{23,32}$, *under which he will get a higher payoff if rematched with firm* a. *Then it suffices to check firm* a*'s perception. Firm* a *worries about the possible type assignment* $\mathbf{t}^{21,34}$. *She would think that if the true type assignment is* $\mathbf{t}^{21,34}$, *then worker* β *believes that* $\mathbf{t}^{21,54}$ *is also possible. To decide whether to participate in the blocking* $(\beta, a; 0)$, *worker* β *might worry about the type assignment* $\mathbf{t}^{21,54}$, *under which he gets a lower payoff if rematched with firm* a. *However, worker* β *would not consider* $\mathbf{t}^{21,54}$ *if the true type assignment were* $\mathbf{t}^{21,34}$, *because firm* a *would object to the blocking* $(\beta, a; 0)$. *Hence, worker* β *would consider only* $\mathbf{t}^{21,34}$, *under which worker* β *would object to the blocking* $(\beta, a; 0)$. *As a result, worker* β *would never benefit from the blocking* $(\beta, a; 0)$ *should* $\mathbf{t}^{21,34}$ *be the true type profile. Therefore, firm* a *does not need to consider* $\mathbf{t}^{21,34}$. *Considering only* $\mathbf{t}^{23,32}$, *firm* a *is willing to participate in the blocking* $(\beta, a; 0)$. *Therefore, the combination* $(\beta, a; 0)$ *should be a blocking combination for* $\left(\mu, \mathbf{p}, \mathbf{t}^{23,32}, \Pi\right)$.

The "blocking" notion which we are about to propose says that a state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is blocked by a combination $(i, j; p)$ if, at the true type assignment $\mathbf{t}$, both agents are willing to participate in the blocking $(i,j;p)$, conditioning on the willingness of their opponents.

Before we formulate agents' perception (willing or unwilling) and conditional perception, we introduce the join of two partitions. The *join* of two partitions is the coarsest common refinement of them; see Aumann (1976). We denote the join operator by $\vee$, i.e., the join of Π_i and Π_j is denoted by $\Pi_i \vee \Pi_j$. A generic element of $\Pi_i \vee \Pi_j$ is denoted by π. For each $\pi \in \Pi_i \cup \Pi_j \cup (\Pi_i \vee \Pi_j)$, we investigate i and j's perception to participate in the blocking $(i,j;p)$. Intuitively, we use Y for Yes (willing to participate in the blocking) and N for No (unwilling to participate in the blocking), and define an indicator correspondence, whose codomain is $\{\emptyset, \{Y\}, \{N\}, \{Y,N\}\}$, by two steps below.

First, for each $\pi \in \Pi_i \vee \Pi_j$, agent $k = i,j$ is willing to participate in the blocking if she/he can guarantee a higher payoff from the rematching $(i,j;p)$, i.e.,

$$\chi_i(\pi) := \begin{cases} \{Y\} & \text{if } \nu_{\mathbf{t}'(i),\mathbf{t}'(j)} + p > \nu_{\mathbf{t}'(i),\mathbf{t}'(\mu(i))} + \mathbf{p}_{i,\mu(i)} \text{ for all } \mathbf{t}' \in \pi, \\ \{N\} & \text{otherwise;} \end{cases} \tag{1}$$

$$\chi_j(\pi) := \begin{cases} \{Y\} & \text{if } \phi_{\mathbf{t}'(i),\mathbf{t}'(j)} - p > \phi_{\mathbf{t}'(\mu(j)),\mathbf{t}'(j)} - \mathbf{p}_{\mu(j),j} \text{ for all } \mathbf{t}' \in \pi, \\ \{N\} & \text{otherwise.} \end{cases} \tag{2}$$

For notational simplicity, the dependence of χ_i and χ_j on $(\mu, \mathbf{p}, \Pi)$ and $(i,j;p)$ is suppressed in the notation. Second, for each $k = i,j$ and each $\pi \in \Pi_k$,

$$\chi_k(\pi) := \bigcup_{\pi' \in \Pi_i \vee \Pi_j : \pi' \subset \pi} \chi_k(\pi'). \tag{3}$$

If $\chi_k(\pi) = \{Y\}$ for some $\pi \in \Pi_k$, then obviously agent k is definitely willing to participate in the blocking should the true type assignment be in π. In contrast, if $\chi_k(\pi) = \{N\}$ for some $\pi \in \Pi_k$, then we claim that agent k is definitely not willing to participate in the blocking at π. To see this, note that the finest information that worker i and firm j can have is represented by the join of their partitions, i.e., $\Pi_i \vee \Pi_j$. Thus, any of their nonempty consideration sets must be unions of elements in $\Pi_i \vee \Pi_j$. When $\chi_k(\pi)$ is a singleton $\{N\}$ for some $\pi \in \Pi_k$, we know that $\chi_k(\pi') = \{N\}$ for every $\pi' \in \Pi_i \vee \Pi_j$ such that $\pi' \subset \pi$. As a result, for any nonempty union of elements in $\Pi_i \vee \Pi_j$, the worst-case payoff for agent k from participating in the blocking is lower than her/his payoff from the status quo match. Another possibility is that for some $\pi \in \Pi_k$, we have $\chi_k(\pi) = \{Y,N\}$. In this case, agent k's perception is not determined at the set π: it is possible that conditioning on her/his opponent's willingness, she/he does not need to consider the type assignments that lead to N, which implies that agent k's perception may be Y at the set π.

The refinement of agents' consideration sets in Example 1 is formally represented by the iteration below. Define $\chi_i^{[0]} := \chi_i$, $\chi_j^{[0]} := \chi_j$, and recursively for $l = 1, 2, \ldots$ that

(i) for each $\pi \in \Pi_i \vee \Pi_j$, agent $k = i, j$ will not consider π if her/his opponent will definitely object the blocking on π, i.e.,

$$\chi_i^{[l]}(\pi) := \begin{cases} \chi_i^{[l-1]}(\pi) & \text{if } Y \in \chi_j^{[l-1]}(\pi'), \text{ where } \pi' \in \Pi_j \text{ and } \pi \subset \pi', \\ \emptyset & \text{otherwise,} \end{cases} \tag{4}$$

$$\chi_j^{[l]}(\pi) := \begin{cases} \chi_j^{[l-1]}(\pi) & \text{if } Y \in \chi_i^{[l-1]}(\pi'), \text{ where } \pi' \in \Pi_i \text{ and } \pi \subset \pi', \\ \emptyset & \text{otherwise;} \end{cases} \tag{5}$$

(ii) for each $k = i, j$ and each $\pi \in \Pi_k$,

$$\chi_k^{[l]}(\pi) := \bigcup_{\pi' \in \Pi_i \vee \Pi_j : \pi' \subset \pi} \chi_k^{[l]}(\pi'). \tag{6}$$

The iteration says that for some event $\pi \in \Pi_i \vee \Pi_j$, agent k knows that π will definitely not happen conditioning on her/his opponent's willingness of participating in the blocking $(i, j; p)$; and this knowledge about conditional information, regardless of her/his opponent's actual decision, is common. Let l^* be the smallest integer such that $\chi_k^{[l^*+1]} = \chi_k^{[l^*]}$ for both $k = i, j$. The following definition says that the state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is blocked by a combination $(i, j; p)$ if, at the true type assignment $\mathbf{t}$, both agents are willing to participate in the blocking conditioning on the willingness of their opponents.

Definition 2. *A state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is said to be **blocked** by $(i, j; p)$ if both i and j are willing to participate in the blocking, conditioning on the willingness of their opponents, i.e.,*

$$\chi_i^{[l^*]}(\Pi_i(\mathbf{t})) = \chi_j^{[l^*]}(\Pi_j(\mathbf{t})) = \{Y\}.$$

It is straightforward to see that if a state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is blocked by $(i, j; p)$, then for every $\mathbf{t}' \in [\Pi_i \vee \Pi_j](\mathbf{t})$, the state $(\mu, \mathbf{p}, \mathbf{t}', \Pi)$ is also blocked by $(i, j; p)$.

3.3 Stability

When information is complete, stable matching embodies the intuition that when "the agents have a very good idea of one another's preferences and have easy access to each other,. . . , we might expect that stable matching will be especially likely to occur" (Roth and Sotomayor, 1990, pp. 22). In this case, a stable state is simply a state that is individually rational and not blocked.

In contrast, with incomplete information, individual rationality and the absence of blocking pair are no longer sufficient to describe a "stable state."[8] The "stability" notion which we are about to propose not only requires individual rationality and no blocking pair but also necessitates that satisfying these two requirements provides no further information to agents. This latter requirement embodies a notion of information stability, which is specific to the incomplete-information environment.

To formulate information stability, we define a set of type assignments as follows:

$$N_{\mu,\mathbf{p},\Pi} := \{\mathbf{t} \in T : (\mu, \mathbf{p}, \mathbf{t}, \Pi) \text{ is individually rational and not blocked}\}.$$

Intuitively, by the public information $(\mu, \mathbf{p}, \Pi)$ and the absence of blocking, agents can infer that the true type assignment lies in $N_{\mu,\mathbf{p},\Pi}$. For notational convenience, we denote by $\mathcal{N}_{\mu,\mathbf{p},\Pi}$ the binary partition that is induced by $N_{\mu,\mathbf{p},\Pi}$, i.e., $\mathcal{N}_{\mu,\mathbf{p},\Pi} := \{N_{\mu,\mathbf{p},\Pi}, T \setminus N_{\mu,\mathbf{p},\Pi}\}$.

Upon observing the absence of blocking, each agent k aggregates the two pieces of information represented by Π_k and $\mathcal{N}_{\mu,\mathbf{p},\Pi}$. This aggregated information is represented by the join of the two partitions, i.e.,

$$H_{\mu,\mathbf{p}}(\Pi) := \Pi \vee \mathcal{N}_{\mu,\mathbf{p},\Pi}.$$

That is, at each type assignment $\mathbf{t}$, agent k knows that the true type assignment $\mathbf{t}$ lies in the set $[H_{\mu,\mathbf{p}}, (\Pi)]_k(\mathbf{t}) = \Pi_k(\mathbf{t}) \cap \mathcal{N}_{\mu,\mathbf{p},\Pi}(\mathbf{t})$. If $H_{\mu,\mathbf{p}}(\Pi) = \Pi$, then the fact of individual rationality and no blocking pair provides no further information to agents. In other words, $H_{\mu,\mathbf{p}}(\Pi) = \Pi$ amounts to requiring that $N_{\mu,\mathbf{p},\Pi}$ be a self-evident event. Indeed, in terms of knowledge operator, this is equivalent to saying that $N_{\mu,\mathbf{p},\Pi}$ is common knowledge at each type assignment in $N_{\mu,\mathbf{p},\Pi}$. See (Osborne and Rubinstein, 1994, pp. 73-74).

A state is said to be stable if it is individually rational and not blocked, and if no information can be inferred from the fact of individual rationality and no blocking.

Definition 3. *A state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is said to be* ***stable*** *if it satisfies the following three requirements:*

(i) $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ *is individually rational.*

(ii) $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ *is not blocked.*

(iii) $H_{\mu,\mathbf{p}}(\Pi) = \Pi$.

When information is complete, i.e., if $\Pi_k(\mathbf{t}) = \{\mathbf{t}\}$ for every $\mathbf{t} \in T$ and every k, then Π is a fixed point of $H_{\mu,\mathbf{p}}(\cdot)$ regardless of $(\mu, \mathbf{p})$. Hence, Definition 3 reduces to the standard definition of stable matching. In this case, a stable state exists (see Theorem 2 of Crawford and Knoer (1981)).

[8]See Example 1 of Chen and Hu (2018) for an illustration.

3.4 Properties of Blocking

In this subsection, we document two simple properties of blocking. Particularly, we first discuss whether or not the blocking agents would obtain higher payoffs if they are rematched. This connects the blocking agents' perception under the true type assignment $\mathbf{t}$ to their perception under the limit consideration correspondence, i.e., $\chi_k^{[l^*]}\left(\Pi_k\left(\mathbf{t}\right)\right)$. Then, we discuss whether or not conditioning on the opponents' willingness is informative. This connects the blocking agents perception under the limit consideration correspondence, i.e., $\chi_k^{[l^*]}\left(\Pi_k\left(\mathbf{t}\right)\right)$, to their perception under the initial consideration correspondence, i.e., $\chi_k\left(\Pi_k\left(\mathbf{t}\right)\right)$.

The following proposition says that if a state is blocked, then satisfying the blocking combination will deliver higher payoffs to both agents under all type assignments in the set $\left[\Pi_i \vee \Pi_j\right](\mathbf{t})$, particularly under the true type assignment $\mathbf{t}$.

Proposition 1. *If $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is blocked by $(i, j; p)$, then for each $\mathbf{t'} \in \left[\Pi_i \vee \Pi_j\right](\mathbf{t})$,*

$$\nu_{\mathbf{t'}(i),\mathbf{t'}(j)} + p > \nu_{\mathbf{t'}(i),\mathbf{t'}(\mu(i))} + \mathbf{p}_{i,\mu(i)} \text{ and } \phi_{\mathbf{t'}(i),\mathbf{t'}(j)} - p > \phi_{\mathbf{t'}(\mu(j)),\mathbf{t'}(j)} - \mathbf{p}_{\mu(j),j}.$$

Proof of Proposition 1. Suppose to the contrary that for some $\mathbf{t'} \in \left[\Pi_i \vee \Pi_j\right](\mathbf{t})$, the first inequality is violated, i.e.,

$$\nu_{\mathbf{t'}(i),\mathbf{t'}(j)} + p \leq \nu_{\mathbf{t'}(i),\mathbf{t'}(\mu(i))} + \mathbf{p}_{i,\mu(i)}.$$

Then worker i is not willing to participate in the blocking if he considers the set $\left[\Pi_i \vee \Pi_j\right](\mathbf{t})$, i.e., $\chi_i\left(\left[\Pi_i \vee \Pi_j\right](\mathbf{t})\right) = \{N\}$ by (1). Note that $\left[\Pi_i \vee \Pi_j\right](\mathbf{t}) \subset \Pi_i(\mathbf{t})$, which implies by (3) that

$$N \in \chi_i\left(\Pi_i(\mathbf{t})\right).$$

On the other hand, since $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is blocked by $(i, j; p)$, we know by the definition of blocking that

$$\chi_i^{[l^*]}\left(\Pi_i(\mathbf{t})\right) = \{Y\}.$$

Therefore, the set $\left[\Pi_i \vee \Pi_j\right](\mathbf{t})$ is not considered by worker i, i.e., $\chi_i^{[l^*]}\left(\left[\Pi_i \vee \Pi_j\right](\mathbf{t})\right) = \emptyset$. By (4), this is true only if

$$\chi_j^{[l^*-1]}\left(\Pi_j(\mathbf{t})\right) = \{N\} \text{ or } \chi_j^{[l^*-1]}\left(\Pi_j(\mathbf{t})\right) = \emptyset. \tag{7}$$

Since the indicator correspondence is decreasing in l, we have $\chi_j^{[l^*]}\left(\Pi_j(\mathbf{t})\right) \subset \chi_j^{[l^*-1]}\left(\Pi_j(\mathbf{t})\right)$, which implies by (7) that $\chi_j^{[l^*]}\left(\Pi_j(\mathbf{t})\right) \in \{\{N\}, \emptyset\}$. However, $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ being blocked by $(i, j; p)$ implies that $\chi_j^{[l^*]}\left(\Pi_j(\mathbf{t})\right) = \{Y\}$, a contradiction.

The argument for $\phi_{\mathbf{t'}(i),\mathbf{t'}(j)} - p > \phi_{\mathbf{t'}(\mu(j)),\mathbf{t'}(j)} - \mathbf{p}_{\mu(j),j}$ is symmetric and thus omitted. $\square$

Proposition 1 has two implications: Firstly, blocking implies higher payoff under the true type assignment, even if agents do not know the truth. Secondly, any incomplete-information blocking combination must also be a complete-information blocking combination, i.e., the set of incomplete-information blocking combinations is a subset of complete-information blocking combinations.

Recall that given any state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ and any blocking combination $(i, j; p)$ for the state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$, we have $\chi_k^{[l^*]}(\Pi_k(\mathbf{t})) \subset \chi_k(\Pi_k(\mathbf{t}))$ for $k = i, j$ because some type assignments are ignored in the iteration (4)-(6). In Example 1, this set inclusion is strict for firm a when the potential blocking combination is $(\beta, a; 0)$, i.e.,

$$\chi_a^{[l^*]}\left(\Pi_a\left(\mathbf{t}^{23,32}\right)\right) = \{Y\} \subsetneq \{Y, N\} = \chi_a\left(\Pi_a\left(\mathbf{t}^{23,32}\right)\right).$$

However, it is also possible that $\chi_k^{[l^*]}(\Pi_k(\mathbf{t})) = \chi_k(\Pi_k(\mathbf{t}))$, i.e., conditioning on the opponent's willingness is not informative for an agent k. Given the initial indicator correspondence $\chi(\cdot)$, the next proposition, which is self-proving, provides a sufficient condition for $\chi_k^{[l^*]}(\Pi_k(\mathbf{t})) = \chi_k(\Pi_k(\mathbf{t}))$.

Proposition 2. *Fix a state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ and a potential blocking combination $(i, j; p)$. If $Y \in \chi_k(\pi)$ for every $\pi \in \Pi_k$ and each $k = i, j$, then*

$$\chi_k^{[l^*]}(\Pi_k(\mathbf{t})) = \chi_k(\Pi_k(\mathbf{t})).$$

The condition in Proposition 2 is imposed on the initial indicator correspondence $\chi(\cdot)$, and thus not primitive. We will provide primitive conditions for $\chi_k^{[l^*]}(\Pi_k(\mathbf{t})) = \chi_k(\Pi_k(\mathbf{t}))$ in Section 4.1.

4 Blocking Notions

In this section, we investigate three alternative blocking notions. Section 4.1 studies *naive blocking*, in which agents are the most conservative in the sense that they consider all type assignments in the true partition cell. We provides primitive conditions such that blocking is equivalent to naive blocking, which also serve as primitive conditions such that conditioning on opponents' willingness is not informative, i.e., $\chi^{[l^*]} = \chi$. In Section 4.2, we provide an alternative blocking notion, termed as *conservative blocking*. In this case, agents are more aggressive when they refine their consideration sets than in naive blocking, but they are more conservative than in blocking (Definition 2). Finally, in Section 4.3, we discuss another alternative blocking notion, termed as *aggressive blocking*, where agents are more aggressive than in blocking.

The blocking notions, except for aggressive blocking, are also ranked in the following sense: a state being naively blocked implies that it is conservatively blocked, which implies that it is blocked. Moreover, blocking implies complete-information blocking. We show that under any one of the these three blocking notions, the two blocking agents would obtain higher payoffs under the true type profile (improvement-at-the-truth). This is consistent with the complete-information blocking. Finally, under naive blocking or conservative blocking, the monotonicity property is satisfied: it is easier for agents to find blocking opportunities when their information is more precise, i.e., when agents have finer partitions.

4.1 Naive Blocking

A state is naively blocked if there exist a firm, a worker, and a potential wage such that upon rematching, both agents would receive a higher payoff under all type assignments that they consider possible.

Definition 4. *A state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is said to be* ***naively blocked*** *by $(i, j; p)$ if*

$$\nu_{\mathbf{t}'(i),\mathbf{t}'(j)} + p > \nu_{\mathbf{t}'(i),\mathbf{t}'(\mu(i))} + p \text{ for all } \mathbf{t}' \in \Pi_i(\mathbf{t}), \text{ and}$$
$$\phi_{\mathbf{t}'(i),\mathbf{t}'(j)} - p > \phi_{\mathbf{t}'(\mu(j)),\mathbf{t}'(j)} - p \text{ for all } \mathbf{t}' \in \Pi_j(\mathbf{t}).$$

The following proposition says that if a state is naively blocked, then it is blocked. The proof will be given in Section 4.2.2.

Proposition 3. *If a state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is naively blocked by $(i, j; p)$, then it is blocked by $(i, j; p)$.*

Obviously, the converse is not true: the state in Example 1 is blocked by $(\beta, a; 0)$ but not naively blocked by $(\beta, a; 0)$. We now provide sufficient conditions such that if a state is blocked, then it is naively blocked. These conditions also serve as primitive conditions such that conditioning on opponents' willingness is not informative, i.e., $\chi^{[l^*]} = \chi$.

Assumption 1. *(One-Dimensional Type) $W \subset \mathbb{R}$ and $F \subset \mathbb{R}$.*

Assumption 2. *(Increasing and Continuous Utility) The premuneration functions $\nu_{w,f}$ and $\phi_{w,f}$ are strictly increasing and continuous in w and f.*

Assumption 3. *(Non-Transferable Utility) No transfer is permitted in the model.*

Assumption 4. *(Knowledge within One Side) It is common knowledge that each worker knows the types of all workers and each firm knows the types of all firms.*

All these primitive conditions exist in the literature: Assumptions 1-2 are imposed in Liu et al. (2014); Assumptions 1-3 are imposed in Bikhchandani (2017); and Assumption 4 is imposed in Pomatto

(2018). Under Assumption 3, the market states have no payment components. Under Assumption 4, the worker type assignment $\mathbf{w}$ and the firm type assignment $\mathbf{f}$ need to be treated separately. Thus, we write type profiles as $(\mathbf{w}, \mathbf{f})$ instead of $\mathbf{t}$ to distinguish two sides of the market. A state is then denoted by $(\mu, \mathbf{w}, \mathbf{f}, \Pi)$. The following proposition says that under Assumptions 1-4, naive blocking is the same as blocking.

Proposition 4. *Under Assumptions 1-4, $(\mu, \mathbf{w}, \mathbf{f}, \Pi)$ is blocked if and only if it is naively blocked.*

Proof of Proposition 4. The sufficiency is by Proposition 3.

Suppose that $(\mu, \mathbf{w}, \mathbf{f}, \Pi)$ is blocked by (i, j). First of all, by Assumption 4, the join partition must be a singleton at every type profile, particularly at the true type profile $(\mathbf{w}, \mathbf{f})$, i.e.,

$$[\Pi_i \vee \Pi_j]\left(\mathbf{w}', \mathbf{f}'\right) = \left\{\left(\mathbf{w}', \mathbf{f}'\right)\right\} \text{ for all } \left(\mathbf{w}', \mathbf{f}'\right) \in T. \tag{8}$$

We proceed to argue that $\chi_j^{[l^*]}\left(\Pi_j\left(\mathbf{w}, \mathbf{f}\right)\right) = \chi_j\left(\Pi_j\left(\mathbf{w}, \mathbf{f}\right)\right)$ and $\chi_i^{[l^*]}\left(\Pi_i\left(\mathbf{w}, \mathbf{f}\right)\right) = \chi_i\left(\Pi_i\left(\mathbf{w}, \mathbf{f}\right)\right)$. Then it follows from $\chi_j^{[l^*]}\left(\Pi_j\left(\mathbf{w}, \mathbf{f}\right)\right) = \chi_i^{[l^*]}\left(\Pi_i\left(\mathbf{w}, \mathbf{f}\right)\right) = \{Y\}$ that $\chi_j\left(\Pi_j\left(\mathbf{w}, \mathbf{f}\right)\right) = \chi_i\left(\Pi_i\left(\mathbf{w}, \mathbf{f}\right)\right) = \{Y\}$, which is the desired condition for naive blocking.

We first argue that $\chi_j^{[1]}\left(\Pi_j\left(\mathbf{w}, \mathbf{f}\right)\right) = \chi_j\left(\Pi_j\left(\mathbf{w}, \mathbf{f}\right)\right)$ and $\chi_i^{[1]}\left(\Pi_i\left(\mathbf{w}, \mathbf{f}\right)\right) = \chi_i\left(\Pi_i\left(\mathbf{w}, \mathbf{f}\right)\right)$. Proposition 1 implies that

$$\nu_{\mathbf{w}(i),\mathbf{f}(j)} > \nu_{\mathbf{w}(i),\mathbf{f}(\mu(i))} \text{ and } \phi_{\mathbf{w}(i),\mathbf{f}(j)} > \phi_{\mathbf{w}(\mu(j)),\mathbf{f}(j)}.$$

Since agents prefer higher types (Assumption 2), the first inequality implies that

$$\mathbf{f}(j) > \mathbf{f}(\mu(i)). \tag{9}$$

For each type assignment $(\mathbf{w}', \mathbf{f}') \in \Pi_j(\mathbf{w}, \mathbf{f})$, since firm j knows the types of all firms (Assumption 4), we have $\mathbf{f}' = \mathbf{f}$. Thus, for each $(\mathbf{w}', \mathbf{f}') \in \Pi_j(\mathbf{w}, \mathbf{f})$, inequation (9) and Assumption 2 imply that

$$\nu_{\mathbf{w}'(i),\mathbf{f}'(j)} - \nu_{\mathbf{w}'(i),\mathbf{f}'(\mu(i))} > 0.$$

Therefore, worker i is willing to participate in the blocking at the set $[\Pi_i \vee \Pi_j]\left(\mathbf{w}', \mathbf{f}'\right)$, i.e.,

$$\chi_i\left([\Pi_i \vee \Pi_j]\left(\mathbf{w}', \mathbf{f}'\right)\right) = \chi_i\left(\left\{\left(\mathbf{w}', \mathbf{f}'\right)\right\}\right) = \{Y\},$$

where the first equality follows from (8). Since $(\mathbf{w}', \mathbf{f}') \in \Pi_i(\mathbf{w}', \mathbf{f}')$, we know that

$$Y \in \chi_i\left(\Pi_i(\mathbf{w}', \mathbf{f}')\right).$$

Hence, firm j cannot ignore any type assignment in $\Pi_j(\mathbf{w}, \mathbf{f})$ in the iteration (4)-(5), i.e.,

$$\chi_j^{[1]}\left(\Pi_j\left(\mathbf{w}, \mathbf{f}\right)\right) = \chi_j\left(\Pi_j\left(\mathbf{w}, \mathbf{f}\right)\right).$$

By a symmetric argument, we know that for worker i,

$$\chi_i^{[1]}\left(\Pi_i\left(\mathbf{w}, \mathbf{f}\right)\right) = \chi_i\left(\Pi_i\left(\mathbf{w}, \mathbf{f}\right)\right).$$

Then, continuing the same argument for $l = 2, \ldots,$ we know that

$$\chi_j^{[l]}\left(\Pi_j\left(\mathbf{w}, \mathbf{f}\right)\right) = \chi_j\left(\Pi_j\left(\mathbf{w}, \mathbf{f}\right)\right) \text{ and } \chi_i^{[l]}\left(\Pi_i\left(\mathbf{w}, \mathbf{f}\right)\right) = \chi_i\left(\Pi_i\left(\mathbf{w}, \mathbf{f}\right)\right),$$

particularly for l^*. This completes the proof. $\square$

4.2 Conservative Blocking

In this subsection, we first introduce the notion of conservative blocking and document some properties of it. Then we compare conservative blocking with blocking and naive blocking. Finally, we study the monotonicity properties of conservative blocking.

4.2.1 Conservative Blocking: Definition

Consider a potential blocking combination $(i, j; p)$ for the state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$. When worker i evaluates a potential blocking firm j at some potential wage p, a type assignment $\mathbf{t}'$ is relevant for worker i if

$$\mathbf{t}' \in \Pi_i\left(\mathbf{t}\right) \text{ and } \max_{\mathbf{t}'' \in \Pi_j(\mathbf{t}')} \left[\phi_{\mathbf{t}''(i), \mathbf{t}''(j)} - p\right] - \left[\phi_{\mathbf{t}''(\mu(j)), \mathbf{t}''(j)} - \mathbf{p}_{\mu(j), j}\right] > 0, \tag{10}$$

i.e., firm j may get better off under $\mathbf{t}'$. Any type assignment that violates the inequality is definitely irrelevant because the "blocking pair" can never be formed due to the firm's objection: she cannot benefit even in the best case. Note that worker i is conservative in the sense that he considers as many type assignments as possible, as long as firm j may benefit from the blocking under them. Similarly, when firm j evaluates worker i at p, a type assignment $\mathbf{t}'$ is relevant for firm j if

$$\mathbf{t}' \in \Pi_j\left(\mathbf{t}\right) \text{ and } \max_{\mathbf{t}'' \in \Pi_i(\mathbf{t}')} \left[\nu_{\mathbf{t}''(i), \mathbf{t}''(j)} + p\right] - \left[\nu_{\mathbf{t}''(i), \mathbf{t}''(\mu(i))} + \mathbf{p}_{i, \mu(i)}\right] > 0. \tag{11}$$

Since agents commonly know that others care only about the relevant type assignments, potential blocking agents consider even less type assignments than specified in (10)-(11). Generally, agents will conservatively refine their consideration sets whenever possible. Define conservative consideration sets $C^{[0]} = \Pi$, and recursively for all $\mathbf{t}' \in T$ and $l = 1, 2, \ldots$ that

$$C_i^{[l]}\left(\mathbf{t}'\right) := \left\{ \mathbf{t}'' \in \Pi_i\left(\mathbf{t}'\right) : \begin{array}{c} C_j^{[l-1]}\left(\mathbf{t}''\right) \neq \emptyset \text{ and} \\ \max_{\tilde{\mathbf{t}} \in C_j^{[l-1]}(\mathbf{t}'')} \left[\phi_{\tilde{\mathbf{t}}(i),\tilde{\mathbf{t}}(j)} - p\right] - \left[\phi_{\tilde{\mathbf{t}}(\mu(j)),\tilde{\mathbf{t}}(j)} - \mathbf{p}_{\mu(j),j}\right] > 0 \end{array} \right\} \text{ and} \tag{12}$$

$$C_j^{[l]}\left(\mathbf{t}'\right) := \left\{ \mathbf{t}'' \in \Pi_j\left(\mathbf{t}'\right) : \begin{array}{c} C_i^{[l-1]}\left(\mathbf{t}''\right) \neq \emptyset \text{ and} \\ \max_{\tilde{\mathbf{t}} \in C_i^{[l-1]}(\mathbf{t}'')} \left[\nu_{\tilde{\mathbf{t}}(i),\tilde{\mathbf{t}}(j)} + p\right] - \left[\nu_{\tilde{\mathbf{t}}(i),\tilde{\mathbf{t}}(\mu(i))} + \mathbf{p}_{i,\mu(i)}\right] > 0 \end{array} \right\}, \tag{13}$$

where, for simplicity, the dependence of $C_i^{[l]}$ and $C_j^{[l]}$ on $(\mu, \mathbf{p}, \Pi)$ and $(i, j; p)$ is suppressed in the notation.

Since $C_j^{[1]}\left(\mathbf{t}''\right) \subset C_j^{[0]}\left(\mathbf{t}''\right)$ for all $\mathbf{t}'' \in T$, we know that $C_j^{[1]}\left(\mathbf{t}''\right) \neq \emptyset$ only if $C_j^{[0]}\left(\mathbf{t}''\right) \neq \emptyset$, and that

$$\max_{\tilde{\mathbf{t}} \in C_j^{[1]}(\mathbf{t}'')} \left[\phi_{\tilde{\mathbf{t}}(i),\tilde{\mathbf{t}}(j)} - p\right] - \left[\phi_{\tilde{\mathbf{t}}(\mu(j)),\tilde{\mathbf{t}}(j)} - \mathbf{p}_{\mu(j),j}\right] > 0$$

only if

$$\max_{\tilde{\mathbf{t}} \in C_j^{[0]}(\mathbf{t}'')} \left[\phi_{\tilde{\mathbf{t}}(i),\tilde{\mathbf{t}}(j)} - p\right] - \left[\phi_{\tilde{\mathbf{t}}(\mu(j)),\tilde{\mathbf{t}}(j)} - \mathbf{p}_{\mu(j),j}\right] > 0.$$

Therefore, $C_i^{[2]}\left(\mathbf{t}'\right) \subset C_i^{[1]}\left(\mathbf{t}'\right)$ for all $\mathbf{t}' \in T$. Similarly, $C_j^{[2]}\left(\mathbf{t}'\right) \subset C_j^{[1]}\left(\mathbf{t}'\right)$ for all $\mathbf{t}' \in T$. By induction, $\Pi_k^{[l]}\left(\mathbf{t}'\right)$ is decreasing in l in the set-inclusion sense for all $\mathbf{t}' \in T$ and $k = 1, 2$. Let l^* be the smallest integer such that $\Pi_k^{[l^*+1]}\left(\mathbf{t}'\right) = \Pi_k^{[l^*]}\left(\mathbf{t}'\right)$ for all $\mathbf{t}' \in T$ and $k = i, j$.

If $C_i^{[l]}\left(\mathbf{t}'\right) = \emptyset$ for some l and some $\mathbf{t}' \in T$, then worker i believes that firm j would never prefer him at the wage p, should $\mathbf{t}'$ be the true type assignment. In other words, at the type assignment $\mathbf{t}'$, worker i considers no type assignment conditioning on firm j's willingness. Furthermore, if $C_i^{[l^*]}\left(\mathbf{t}\right) = \emptyset$, then worker i believes that firm j would never prefer him at the wage p. Symmetric argument applies for firms. The following definition says that a state is blocked if for both agents, (i) their opponents may benefit from blocking, and (ii) they would benefit under all type assignments in their conservative consideration sets.

Definition 5. *A state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is said to be **conservatively blocked** by $(i, j; p)$ if $C_i^{[l^*]}\left(\mathbf{t}\right) \neq \emptyset$, $C_j^{[l^*]}\left(\mathbf{t}\right) \neq \emptyset$ and*

$$\nu_{\mathbf{t}'(i),\mathbf{t}'(j)} + p > \nu_{\mathbf{t}'(i),\mathbf{t}'(\mu(i))} + \mathbf{p}_{i,\mu(i)} \text{ for all } \mathbf{t}' \in C_i^{[l^*]}\left(\mathbf{t}\right) \text{ and} \tag{14}$$

$$\phi_{\mathbf{t}'(i),\mathbf{t}'(j)} - p > \phi_{\mathbf{t}'(\mu(j)),\mathbf{t}'(j)} - \mathbf{p}_{\mu(j),j} \text{ for all } \mathbf{t}' \in C_j^{[l^*]}\left(\mathbf{t}\right). \tag{15}$$

As we can see from the iteration (12)-(13), the consideration sets of worker i and firm j may be refined at each type profile, should it be the true type profile. However, it is not obvious that to what extent they can refine their consideration sets. The following proposition says that the conservative consideration sets, particularly the limit ones, must be measurable with respect to $\Pi_i \vee \Pi_j$.[9] Therefore, should $\mathbf{t}'$ be the true type profile, the smallest possible consideration set is $[\Pi_i \vee \Pi_j](\mathbf{t}')$.

Proposition 5. *Fix a state* $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ *and a potential blocking combination* $(i, j; p)$*. For each* $\mathbf{t}'$ *and each* $l = 1, 2, \ldots,$ *the conservative consideration sets* $C_i^{[l]}(\mathbf{t}')$ *and* $C_j^{[l]}(\mathbf{t}')$ *are measurable w.r.t.* $\Pi_i \vee \Pi_j$.

Proof of Proposition 5. Pick an arbitrary $\mathbf{t}'$ such that $C_i^{[1]}(\mathbf{t}') \neq \emptyset$. Let $\mathbf{t}'' \in C_i^{[1]}(\mathbf{t}')$. We proceed to show that for all $\mathbf{t}''' \in [\Pi_i \vee \Pi_j](\mathbf{t}'')$, we have $\mathbf{t}''' \in C_i^{[1]}(\mathbf{t}')$. This is true because $\mathbf{t}''' \in [\Pi_i \vee \Pi_j](\mathbf{t}'')$ implies that $\Pi_j(\mathbf{t}'') \neq \emptyset \iff \Pi_j(\mathbf{t}''') \neq \emptyset$, and that

$$\max_{\tilde{\mathbf{t}} \in \Pi_j(\mathbf{t}'')} \left[\phi_{\tilde{\mathbf{t}}(i), \tilde{\mathbf{t}}(j)} - p\right] - \left[\phi_{\tilde{\mathbf{t}}(\mu(j)), \tilde{\mathbf{t}}(j)} - \mathbf{p}_{\mu(j), j}\right] > 0$$
$$\iff \max_{\tilde{\mathbf{t}} \in \Pi_j(\mathbf{t}''')} \left[\phi_{\tilde{\mathbf{t}}(i), \tilde{\mathbf{t}}(j)} - p\right] - \left[\phi_{\tilde{\mathbf{t}}(\mu(j)), \tilde{\mathbf{t}}(j)} - \mathbf{p}_{\mu(j), j}\right] > 0$$

Therefore, $C_i^{[1]}(\mathbf{t}')$ is measurable w.r.t. $\Pi_i \vee \Pi_j$. Similar argument applies to $C_j^{[1]}(\mathbf{t}')$. Induction completes the proof. $\square$

We close this subsection by the following proposition, which is a counterpart of Proposition 1.

Proposition 6. *If* $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ *is conservatively blocked by* $(i, j; p)$*, then*

$$\nu_{\mathbf{t}(i), \mathbf{t}(j)} + p > \nu_{\mathbf{t}(i), \mathbf{t}(\mu(i))} + \mathbf{p}_{i, \mu(i)} \text{ and } \phi_{\mathbf{t}(i), \mathbf{t}(j)} - p > \phi_{\mathbf{t}(\mu(j)), \mathbf{t}(j)} - \mathbf{p}_{\mu(j), j}.$$

Proof of Proposition 6. Suppose to the contrary that

$$\nu_{\mathbf{t}(i), \mathbf{t}(j)} + p \leq \nu_{\mathbf{t}(i), \mathbf{t}(\mu(i))} + \mathbf{p}_{i, \mu(i)}.$$

Then (14) implies that $\mathbf{t} \notin C_i^{[l^*]}(\mathbf{t})$, which by (12) implies either that $C_j^{[l^*]}(\mathbf{t}) = \emptyset$ or that

$$\max_{\tilde{\mathbf{t}} \in C_j^{[l^*]}(\mathbf{t})} \left[\phi_{\tilde{\mathbf{t}}(i), \tilde{\mathbf{t}}(j)} - p\right] - \left[\phi_{\tilde{\mathbf{t}}(\mu^{-1}(j)), \tilde{\mathbf{t}}(j)} - \mathbf{p}_{\mu^{-1}(j), j}\right] \leq 0. \tag{16}$$

Since $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is blocked by $(i, j; p)$, we have $C_j^{[l^*]}(\mathbf{t}) \neq \emptyset$. Therefore, (16) holds, which is a contradiction to (12). Hence, $\nu_{\mathbf{t}(i), \mathbf{t}(j)} + p > \nu_{\mathbf{t}(i), \mathbf{t}(\mu(i))} + \mathbf{p}_{i, \mu(i)}$. The argument for $\phi_{\mathbf{t}(i), \mathbf{t}(j)} - p > \phi_{\mathbf{t}(\mu^{-1}(j)), \mathbf{t}(j)} - \mathbf{p}_{\mu^{-1}(j), j}$ is symmetric and thus omitted. $\square$

[9]We say that a set is measurable w.r.t. $\Pi_i \vee \Pi_j$ if it is measurable w.r.t. the σ-algebra generated by all unions of sets in $\Pi_i \vee \Pi_j$.

4.2.2 Naive Blocking, Conservative Blocking, and Blocking

Now we proceed to compare Naive Blocking, Conservative Blocking, and Blocking. The following proposition says that if a state is naively blocked, then it is conservatively blocked. The converse is not true as is illustrated in Example 1, where the state is conservatively blocked by $(\beta, a; 0)$ but not naively blocked by it.

Proposition 7. *If a state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is naively blocked by $(i, j; p)$, then it is conservatively blocked by $(i, j; p)$.*

Proof of Proposition 7. It suffices to show that $C_i^{[l^*]}(\mathbf{t}) \neq \emptyset$ and $C_j^{[l^*]}(\mathbf{t}) \neq \emptyset$. To see this, we show that $\mathbf{t} \in C_i^{[l^*]}(\mathbf{t})$ and $\mathbf{t} \in C_j^{[l^*]}(\mathbf{t})$. Since $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is naively blocked by $(i, j; p)$, we have

$$\nu_{\mathbf{t}(i),\mathbf{t}(j)} + p > \nu_{\mathbf{t}(i),\mathbf{t}(\mu(i))} + \mathbf{p}_{i,\mu(i)} \text{ and } \phi_{\mathbf{t}(i),\mathbf{t}(j)} - p > \phi_{\mathbf{t}(\mu(j)),\mathbf{t}(j)} - \mathbf{p}_{\mu(j),j}. \tag{17}$$

Initially, both agents consider the true type assignment, i.e., $\mathbf{t} \in \Pi_i(\mathbf{t})$ and $\mathbf{t} \in \Pi_j(\mathbf{t})$. Moreover, the true type assignment $\mathbf{t}$ is not ignored in the iteration (12)-(13), because of (17). It follows from induction that $\mathbf{t}$ is never ignored in the iteration (12)-(13). This completes the proof. □

The following proposition says that if a state is conservatively blocked, then it is blocked.

Proposition 8. *If a state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is conservatively blocked by $(i, j; p)$, then it is blocked by $(i, j; p)$.*

Proof of Proposition 8. Suppose $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is conservatively blocked by $(i, j; p)$. Then by Definition 5, we have $C_i^{[l^*]}(\mathbf{t}) \neq \emptyset$, $C_j^{[l^*]}(\mathbf{t}) \neq \emptyset$ and

$$\begin{aligned}
&\nu_{\mathbf{t}'(i),\mathbf{t}'(j)} + p > \nu_{\mathbf{t}'(i),\mathbf{t}'(\mu(i))} + \mathbf{p}_{i,\mu(i)} \text{ for all } \mathbf{t}' \in C_i^{[l^*]}(\mathbf{t}) \text{ and} \\
&\phi_{\mathbf{t}'(i),\mathbf{t}'(j)} - p > \phi_{\mathbf{t}'(\mu(j)),\mathbf{t}'(j)} - \mathbf{p}_{\mu(j),j} \text{ for all } \mathbf{t}' \in C_j^{[l^*]}(\mathbf{t}).
\end{aligned}$$

Since, by Proposition 5, $C_i^{[l^*]}(\mathbf{t})$ and $C_j^{[l^*]}(\mathbf{t})$ are both measurable w.r.t. $\Pi_i \vee \Pi_j$, we can represent the blocking conditions by indicator correspondences defined in Section 3.2, i.e.,

$$\chi_i\left(C_i^{[l^*]}(\mathbf{t})\right) = \chi_j\left(C_j^{[l^*]}(\mathbf{t})\right) = \{Y\}. \tag{18}$$

On the one hand, Proposition 6 implies that both agents benefit from blocking at the true type profile $\mathbf{t}$. Therefore, according to (12) and (13), the true type profile is always considered by both agents, i.e., $\mathbf{t} \in C_i^{[l]}(\mathbf{t})$ and $\mathbf{t} \in C_j^{[l]}(\mathbf{t})$ for all l. By Proposition 5, we have $[\Pi_i \vee \Pi_j](\mathbf{t}) \subset C_i^{[l^*]}(\mathbf{t})$ and

$[\Pi_i \vee \Pi_j](\mathbf{t}) \subset C_j^{[l^*]}(\mathbf{t})$. Thus, (18) implies that

$$\chi_i([\Pi_i \vee \Pi_j](\mathbf{t})) = \chi_j([\Pi_i \vee \Pi_j](\mathbf{t})) = \{Y\},$$

which in turn implies that $[\Pi_i \vee \Pi_j](\mathbf{t})$ is never ignored in the iteration (4)-(5), i.e.,

$$Y \in \chi_i^{[l^*]}(\Pi_i(\mathbf{t})) \text{ and } Y \in \chi_j^{[l^*]}(\Pi_j(\mathbf{t})). \tag{19}$$

On the other hand, we proceed to show that

$$\chi_i^{[l^*]}(\Pi_i(\mathbf{t})) \subset \chi_i\left(C_i^{[l^*]}(\mathbf{t})\right) \text{ and } \chi_j^{[l^*]}(\Pi_j(\mathbf{t})) \subset \chi_j\left(C_j^{[l^*]}(\mathbf{t})\right). \tag{20}$$

It follows from induction in $l = 0, 1, \dots$ that

$$\chi_i^{[l]}(\Pi_i(\mathbf{t})) \subset \chi_i\left(C_i^{[l]}(\mathbf{t})\right) \text{ and } \chi_j^{[l]}(\Pi_j(\mathbf{t})) \subset \chi_j\left(C_j^{[l]}(\mathbf{t})\right).$$

Hence, (20) is true. To sum up, (18)-(20) imply that

$$\chi_i^{[l^*]}(\Pi_i(\mathbf{t})) = \chi_j^{[l^*]}(\Pi_j(\mathbf{t})) = \{Y\}.$$

That is, $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is blocked by $(i, j; p)$. □

The converse of Propsition 8 is not true due to the following example.

Example 2 (Example 1 Revisited)**.** *Consider the alternative set of possible type assignments with one more element* $\mathbf{t}^{01,54}$*, i.e.,* $T = \left\{\mathbf{t}^{23,32}, \mathbf{t}^{21,34}, \mathbf{t}^{21,54}, \mathbf{t}^{01,54}\right\}$.

	α	β	a	b
$\mathbf{t}^{23,32}$:	2	3	3	2
$\mathbf{t}^{21,34}$:	2	1	3	4
$\mathbf{t}^{21,54}$:	2	1	5	4
$\mathbf{t}^{01,54}$:	0	1	5	4

The partitions of worker β *and firm* a *are given as follows:*

$$\Pi_\beta = \left\{\left\{\mathbf{t}^{23,32}\right\}, \left\{\mathbf{t}^{21,34}, \mathbf{t}^{21,54}, \mathbf{t}^{01,54}\right\}\right\}, \text{ and}$$
$$\Pi_a = \left\{\left\{\mathbf{t}^{23,32}, \mathbf{t}^{21,34}\right\}, \left\{\mathbf{t}^{21,54}, \mathbf{t}^{01,54}\right\}\right\}.$$

Other agents' partitions are not relavant when we consider the potential blocking combination $(\beta, a; 0)$ *for* $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$*, and thus omitted. It is straightforward to verify that if information is complete, then under* $\mathbf{t}^{01,54}$*, the worker-firm pair* (β, a) *is the unique blocking pair.*

Similar arguments as in Example 1 shows that $\chi_a\left(\Pi_a\left(\mathbf{t}^{23,32}\right)\right) = \{Y\}$*. Therefore,* $(\beta, a; 0)$ *is a blocking combination for* $(\mu, \mathbf{p}, \mathbf{t}^{23,32}, \Pi)$*. However, conservative consideration sets defined by (12)-(13) are given as*

$$C_\beta^{[l^*]}(\mathbf{t}^{23,32}) = C_\beta^{[0]}(\mathbf{t}^{23,32}) = \left\{\mathbf{t}^{23,32}\right\} \text{ and}$$
$$C_a^{[l^*]}(\mathbf{t}^{23,32}) = C_a^{[0]}(\mathbf{t}^{23,32}) = \left\{\mathbf{t}^{23,32}, \mathbf{t}^{21,34}\right\}.$$

Therefore, $(\beta, a; 0)$ *is not a conservative blocking combination for* $(\mu, \mathbf{p}, \mathbf{t}^{23,32}, \Pi)$*.*

Propositions 7 and 8 have two implications. First, they constitute a proof for Proposition 3, i.e., naive blocking implies blocking. Second, because of Proposition 1 and Propositions 7 and 8, Proposition 6 can be strengthened as follows.

Proposition 6′. *If* $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ *is conservatively blocked by* $(i, j; p)$*, then for each* $\mathbf{t}' \in [\Pi_i \vee \Pi_j](\mathbf{t})$*,*

$$\nu_{\mathbf{t}'(i),\mathbf{t}'(j)} + p > \nu_{\mathbf{t}'(i),\mathbf{t}'(\mu(i))} + \mathbf{p}_{i,\mu(i)} \text{ and } \phi_{\mathbf{t}'(i),\mathbf{t}'(j)} - p > \phi_{\mathbf{t}'(\mu(j)),\mathbf{t}'(j)} - \mathbf{p}_{\mu(j),j}.$$

We now provide a sufficient condition for the notion of conservative blocking being the same as the notion of blocking. The following proposition is self-proving.

Proposition 9. *Fix a state* $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ *and a potential blocking combination* $(i, j; p)$*. If for each* $\pi \in \Pi_i \vee \Pi_j$*,*

$$\chi_i(\pi) = \{N\} \text{ implies that } \nu_{\mathbf{t}'(i),\mathbf{t}'(j)} + p < \nu_{\mathbf{t}'(i),\mathbf{t}'(\mu(i))} + \mathbf{p}_{i,\mu(i)} \text{ for all } \mathbf{t}' \in \pi \text{ and}$$
$$\chi_j(\pi) = \{N\} \text{ implies that } \phi_{\mathbf{t}'(i),\mathbf{t}'(j)} - p < \phi_{\mathbf{t}'(\mu(j)),\mathbf{t}'(j)} - \mathbf{p}_{\mu(j),j} \text{ for all } \mathbf{t}' \in \pi,$$

then $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ *is conservatively blocked by* $(i, j; p)$ *if and only if it is blocked by* $(i, j; p)$*.*

Alternatively, as a corollary of Proposition 4, Assumptions 1-4 also serve as a set of sufficient conditions such that conservative blocking is the same as blocking. They are primitive conditions.

4.2.3 Monotonicity

In this subsection, we study the monotonicity properties of conservative blocking. The monotonicity property of naive blocking follows as a corollary.[10] The following proposition says that a conservative blocking opportunity is more likely to exist if agents have more precise information, i.e., finer partitions.

Proposition 10. *Suppose that $\hat{\Pi}$ is a finer partition profile than Π. If a state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is conservatively blocked by $(i, j; p)$, then the state $\left(\mu, \mathbf{p}, \mathbf{t}, \hat{\Pi}\right)$ is also conservatively blocked by $(i, j; p)$.*

Proof of Proposition 10. Suppose $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is conservatively blocked by $(i, j; p)$. Since $\hat{\Pi}$ is finer than Π, we know that for each $\mathbf{t}' \in T$,

$$\hat{\Pi}_i\left(\mathbf{t}'\right) \subset \Pi_i\left(\mathbf{t}'\right) \text{ and } \hat{\Pi}_j\left(\mathbf{t}'\right) \subset \Pi_j\left(\mathbf{t}'\right).$$

Note that $C_i^{[l]}\left(\mathbf{t}'\right)$ defined in (12) is increasing in both $\Pi_i\left(\mathbf{t}'\right)$ and $C_j^{[l-1]}\left(\mathbf{t}''\right)$, and that $C_j^{[l]}\left(\mathbf{t}'\right)$ defined in (13) is increasing in both $\Pi_j\left(\mathbf{t}'\right)$ and $C_i^{[l-1]}\left(\mathbf{t}''\right)$. It follows from induction that for each $\mathbf{t}' \in T$ and each $l = 1, 2, \ldots,$

$$\hat{C}_i^{[l]}\left(\mathbf{t}'\right) \subset C_i^{[l]}\left(\mathbf{t}'\right) \text{ and } \hat{C}_j^{[l]}\left(\mathbf{t}'\right) \subset C_j^{[l]}\left(\mathbf{t}'\right).$$

Particularly, we have that for each $\mathbf{t}' \in T$,

$$\hat{C}_i^{[l^*]}\left(\mathbf{t}'\right) \subset C_i^{[l^*]}\left(\mathbf{t}'\right) \text{ and } \hat{C}_j^{[l^*]}\left(\mathbf{t}'\right) \subset C_j^{[l^*]}\left(\mathbf{t}'\right),$$

where the dependence of l^* on $\hat{\Pi}$ and Π is suppressed in the notation.

On the other hand, since $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is blocked by $(i, j; p)$, we know by Proposition 6 that

$$\nu_{\mathbf{t}(i),\mathbf{t}(j)} + p > \nu_{\mathbf{t}(i),\mathbf{t}(\mu(i))} + \mathbf{p}_{i,\mu(i)} \text{ and } \phi_{\mathbf{t}(i),\mathbf{t}(j)} - p > \phi_{\mathbf{t}(\mu(j)),\mathbf{t}(j)} - \mathbf{p}_{\mu(j),j}.$$

Therefore, the true type assignment $\mathbf{t}$ is not eliminated from the consideration sets $\hat{C}_i^{[l]}(\mathbf{t})$ and $\hat{C}_j^{[l]}(\mathbf{t})$ for all $l = 1, 2, \ldots$. In particular, $\mathbf{t} \in \hat{C}_i^{[l^*]}(\mathbf{t})$ and $\mathbf{t} \in \hat{C}_j^{[l^*]}(\mathbf{t})$. Therefore, the limit consideration sets are nonempty for both agents, i..e, $\hat{C}_i^{[l^*]}(\mathbf{t}) \neq \emptyset$ and $\hat{C}_j^{[l^*]}(\mathbf{t}) \neq \emptyset$. It follows from Definition 5 that $\left(\mu, \mathbf{p}, \mathbf{t}, \hat{\Pi}\right)$ is conservatively blocked. □

The following example says that blocking does not satisfy monotonicity.

Example 3 (Example 2 Revisited). *Recall that in Example 2, the partitions of worker β and firm a*

[10] The monotonicity properties of naive stability and conservative stability are discussed in Appendix B.

are given by

$$\Pi_\beta = \left\{\left\{\mathbf{t}^{23,32}\right\}, \left\{\mathbf{t}^{21,34}, \mathbf{t}^{21,54}, \mathbf{t}^{01,54}\right\}\right\} \text{ and}$$
$$\Pi_a = \left\{\left\{\mathbf{t}^{23,32}, \mathbf{t}^{21,34}\right\}, \left\{\mathbf{t}^{21,54}, \mathbf{t}^{01,54}\right\}\right\},$$

and the state $(\mu, \mathbf{p}, \mathbf{t}^{23,32}, \Pi)$ *is blocked by* $(\beta, a; 0)$. *Now we consider a finer partition profile in which*

$$\hat{\Pi}_\beta = \left\{\left\{\mathbf{t}^{23,32}\right\}, \left\{\mathbf{t}^{21,34}, \mathbf{t}^{01,54}\right\}, \left\{\mathbf{t}^{21,54}\right\}\right\} \text{ and}$$
$$\hat{\Pi}_a = \left\{\left\{\mathbf{t}^{23,32}, \mathbf{t}^{21,34}\right\}, \left\{\mathbf{t}^{01,54}\right\}, \left\{\mathbf{t}^{21,54}\right\}\right\}.$$

Similar arguments as in Example 2 would show that $(\mu, \mathbf{p}, \mathbf{t}^{23,32}, \hat{\Pi})$ *is not blocked by* $(\beta, a; 0)$.

4.3 Aggressive Blocking

In this subsection, we discuss aggressive blocking. Consider a potential blocking combination $(i, j; p)$ for the state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$. Since it is publicly known that agents care about the worst case of their potential payoffs, agents could further restrict their attention, compared with (12)-(13). Formally, we define aggressive consideration sets $C^{(0)} = \Pi$ and recursively for $l = 1, 2, \ldots$ and all $\mathbf{t}' \in T$ that

$$C_i^{(l)}\left(\mathbf{t}'\right) := \left\{\mathbf{t}'' \in C_i^{(l-1)}\left(\mathbf{t}'\right) : \begin{array}{c} C_j^{(l-1)}\left(\mathbf{t}''\right) \neq \emptyset \text{ and} \\ \min_{\tilde{\mathbf{t}} \in C_j^{(l-1)}(\mathbf{t}'')} \left[\phi_{\tilde{\mathbf{t}}(i),\tilde{\mathbf{t}}(j)} - p\right] - \left[\phi_{\tilde{\mathbf{t}}(\mu(j)),\tilde{\mathbf{t}}(j)} - \mathbf{p}_{\mu(j),j}\right] > 0 \end{array}\right\} \tag{21}$$

$$C_j^{(l)}\left(\mathbf{t}'\right) := \left\{\mathbf{t}'' \in C_j^{(l-1)}\left(\mathbf{t}'\right) : \begin{array}{c} C_i^{(l-1)}\left(\mathbf{t}''\right) \neq \emptyset \text{ and} \\ \min_{\tilde{\mathbf{t}} \in C_i^{(l-1)}(\mathbf{t}'')} \left[\nu_{\tilde{\mathbf{t}}(i),\tilde{\mathbf{t}}(j)} + p\right] - \left[\nu_{\tilde{\mathbf{t}}(i),\tilde{\mathbf{t}}(\mu(i))} + \mathbf{p}_{i,\mu(i)}\right] > 0 \end{array}\right\}, \tag{22}$$

where the dependence of $C_i^{(l)}$ and $C_j^{(l)}$ on $(\mu, \mathbf{p}, \Pi)$ and $(i, j; p)$ is suppressed in the notation.

Since $C_j^{(1)}\left(\mathbf{t}''\right) \subset C_j^{(0)}\left(\mathbf{t}''\right)$ for all $\mathbf{t}'' \in T$, we know that

$$\min_{\tilde{\mathbf{t}} \in C_j^{(0)}(\mathbf{t}'')} \left[\phi_{\tilde{\mathbf{t}}(i),\tilde{\mathbf{t}}(j)} - p\right] - \left[\phi_{\tilde{\mathbf{t}}(\mu(j)),\tilde{\mathbf{t}}(j)} - \mathbf{p}_{\mu(j),j}\right] > 0$$

implies that

$$\min_{\tilde{\mathbf{t}} \in C_j^{(1)}(\mathbf{t}'')} \left[\phi_{\tilde{\mathbf{t}}(i),\tilde{\mathbf{t}}(j)} - p\right] - \left[\phi_{\tilde{\mathbf{t}}(\mu(j)),\tilde{\mathbf{t}}(j)} - \mathbf{p}_{\mu(j),j}\right] > 0.$$

Therefore, the set $C_i^{(2)}(\mathbf{t}')$ defined by (21) can be simplified as

$$C_i^{(2)}(\mathbf{t}') = \left\{\mathbf{t}'' \in C_i^{(1)}(\mathbf{t}') : C_j^{(1)}(\mathbf{t}'') \neq \emptyset\right\} \tag{23}$$

for all $\mathbf{t}' \in T$. Similarly, for firm j we have

$$C_j^{(2)}(\mathbf{t}') = \left\{\mathbf{t}'' \in C_j^{(1)}(\mathbf{t}') : C_i^{(1)}(\mathbf{t}'') \neq \emptyset\right\} \tag{24}$$

for all $\mathbf{t}' \in T$. We claim that for each $k = i, j$, we must have $C_k^{(3)}(\mathbf{t}') = C_k^{(2)}(\mathbf{t}')$ for all $\mathbf{t}' \in T$, which implies that $C_k^{(l)}(\mathbf{t}') = C_k^{(2)}(\mathbf{t}')$ for all $\mathbf{t}' \in T$ and all $l \geq 3$. See Appendix A for a formal proof. Therefore, agents refine their consideration sets at most twice.

A state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is aggressively blocked by a combination $(i, j; p)$ if both agents $k = i, j$ obtain a higher payoff under the true type assignment, and if both agents evaluate their opponents via the consideration sets $C_k^{(2)}(\mathbf{t})$.

Definition 6. *A state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is said to be **aggressively blocked** by $(i, j; p)$ if*

$$\nu_{\mathbf{t}(i),\mathbf{t}(j)} + p > \nu_{\mathbf{t}(i),\mathbf{t}(\mu(i))} + \mathbf{p}_{i,\mu(i)} \text{ and } \phi_{\mathbf{t}(i),\mathbf{t}(j)} - p > \phi_{\mathbf{t}(\mu(j)),\mathbf{t}(j)} - \mathbf{p}_{\mu(j),j}, \tag{25-1}$$

$$\nu_{\mathbf{t}'(i),\mathbf{t}'(j)} + p > \nu_{\mathbf{t}'(i),\mathbf{t}'(\mu(i))} + \mathbf{p}_{i,\mu(i)} \text{ for all } \mathbf{t}' \in C_i^{(2)}(\mathbf{t}), \text{ and} \tag{26}$$

$$\phi_{\mathbf{t}'(i),\mathbf{t}'(j)} - p > \phi_{\mathbf{t}'(\mu(j)),\mathbf{t}'(j)} - \mathbf{p}_{\mu(j),j} \text{ for all } \mathbf{t}' \in C_j^{(2)}(\mathbf{t}). \tag{27}$$

We first compare aggressive blocking and blocking, and then discuss the additional condition (25-1). The following example shows the difference between aggressive blocking and blocking.

Example 4 (Example 1 Revisited)**.** *Replace the type assignment $\mathbf{t}^{21,54}$ by $\mathbf{t}^{01,54}$, i.e., T is now given by*

	α	β	a	b
$\mathbf{t}^{23,32}$:	2	3	3	2
$\mathbf{t}^{21,34}$:	2	1	3	4
$\mathbf{t}^{01,54}$:	0	1	5	4

The partitions of worker β and firm a are given as follows:

$$\Pi_\beta = \left\{\left\{\mathbf{t}^{23,32}\right\}, \left\{\mathbf{t}^{21,34}, \mathbf{t}^{01,54}\right\}\right\}, \text{ and}$$
$$\Pi_a = \left\{\left\{\mathbf{t}^{23,32}, \mathbf{t}^{21,34}\right\}, \left\{\mathbf{t}^{01,54}\right\}\right\}.$$

Other agents' partitions are not relavant when we consider the potential blocking combination $(\beta, a; 0)$ for the state $(\mu, \mathbf{p}, \mathbf{t}^{23,32}, \Pi)$, and thus omitted. Clearly, $(\beta, a; 0)$ is not a blocking combination for

$(\mu, \mathbf{p}, \mathbf{t}^{23,32}, \Pi)$*: firm* a *worries about the counterfactual type assignment* $\mathbf{t}^{21,34}$*, and she cannot ignore it since worker* β *might be willing to participate in the blocking under* $\mathbf{t}^{21,34}$*. However,* $(\mu, \mathbf{p}, \mathbf{t}^{23,32}, \Pi)$ *is aggressively blocked by* $(\beta, a; 0)$*: condition (25-1) is satisfied and* $C_\beta^{(2)}(\mathbf{t}^{23,32}) = C_a^{(2)}(\mathbf{t}^{23,32}) = \emptyset$*.*

Condition (25-1) has two roles: (i) it rules out the counterfactual cases where the state is blocked by agents but they do not believe it, i.e., (26)-(27) are trivially satisfied when $C_i^{(2)} = \emptyset$ and $C_j^{(2)} = \emptyset$; and (ii) it ensures that the blocking agents will indeed benefit from being matched with each other. The following example documents a situation in which both roles fail when we drop condition (25-1).

Example 5 (Example 4 Revisited)**.** *Let* $\mathbf{t}^{21,34}$ *be the true type assignment instead of* $\mathbf{t}^{23,32}$*. Under the true type assignment, both agents get worse off from being matched with each other.*

It is straightforward to verify that $C_\beta^{(2)}(\mathbf{t}^{21,34}) = C_a^{(2)}(\mathbf{t}^{21,34}) = \emptyset$*. If we drop condition (25-1), then* $(\mu, \mathbf{p}, \mathbf{t}^{21,34}, \Pi)$ *is aggressively blocked by* $(\beta, a; 0)$*: two agents who aim to guarantee higher worst-case payoffs will end up with, however, lower actual payoffs.*

One may consider an alternative notion of aggressive blocking based on the consideration sets $C_k^{(2)}(\mathbf{t})$: a state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is aggressively blocked by $(i, j; p)$ if

$$C_i^{(2)}(\mathbf{t}) \neq \emptyset \text{ and } C_j^{(2)}(\mathbf{t}) \neq \emptyset, \tag{25-2}$$

and (26)-(27) are satisfied. In this case, role (i) of condition (25-1) is kept while role (ii) is relaxed, i.e., we no longer require that both agents benefit under the true type assignment.

Definition 7. *A state* $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ *is said to be* ***aggressively* blocked*** *by* $(i, j; p)$ *if (25-2) and (26)-(27) are satisfied.*

The following example documents a situation in which the state is aggressively* blocked; however, the blocking agents will obtain worse payoffs from being rematched with each other.

Example 6. *Consider a market with two workers, i.e.,* $I = \{\alpha, \beta\}$*, and two firms, i.e.,* $J = \{a, b\}$*. The set of possible type assignments is given by* $T = \{\mathbf{t}^{24,31}, \mathbf{t}^{34,21}, \mathbf{t}^{32,23}, \mathbf{t}^{12,43}, \mathbf{t}^{13,42}\}$*, i.e.,*

	α	β	a	b
$\mathbf{t}^{24,31}$:	2	4	3	1
$\mathbf{t}^{34,21}$:	3	4	2	1
$\mathbf{t}^{32,23}$:	3	2	2	3
$\mathbf{t}^{12,43}$:	1	2	4	3
$\mathbf{t}^{13,42}$:	1	3	4	2

In this example, a larger number means a better type. The premuneration values for workers and firms are given by the product form, i.e., $\nu_{w,f} = \phi_{w,f} = wf$. *Obviously, every agent prefers a partner with a higher type to a partner with a lower type.*

Suppose firm a *hires worker* α *and firm* b *hires worker* β. *In other words, a matching* μ *is given by* $\mu(\alpha) = a$ *and* $\mu(\beta) = b$. *Suppose that payments under* μ *are all zero, i.e.,* $\mathbf{p} = \mathbf{0}$. *Suppose also that the type assignment* $\mathbf{t}^{32,23}$ *is true, and that agents' partitions are given as if they can observe their own types and their own partners' types. Then, for worker* β *and firm* a, *the partition profiles are as follows:*

$$\begin{aligned}\Pi_\beta &= \left\{\left\{\mathbf{t}^{24,31}, \mathbf{t}^{34,21}\right\}, \left\{\mathbf{t}^{32,23}, \mathbf{t}^{12,43}\right\}, \left\{\mathbf{t}^{13,42}\right\}\right\}, \text{ and} \\ \Pi_a &= \left\{\left\{\mathbf{t}^{24,31}\right\}, \left\{\mathbf{t}^{34,21}, \mathbf{t}^{32,23}\right\}, \left\{\mathbf{t}^{12,43}, \mathbf{t}^{13,42}\right\}\right\}.\end{aligned}$$

The description of the other two agents' partitions is irrelevant to the arguments below, and thus omitted.

Now we consider a potential aggressive blocking combination* $(\beta, a; 0)$ *for the state* $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$. *It is straightforward to verify that if information is complete, then the following claims are true: (i) under* $\mathbf{t}^{32,23}$, $(\beta, a; 0)$ *is not a blocking combination; and (ii) under every other type profile,* $(\beta, a; 0)$ *is a blocking combination. As a result, the consideration sets of both worker* β *and firm* a *are refined. Particularly,*

$$\begin{aligned}C_\beta^{(2)} = C_\beta^{(1)} &= \left\{\left\{\mathbf{t}^{24,31} \qquad\qquad\right\}, \left\{\qquad\qquad \mathbf{t}^{12,43}\right\}, \left\{\mathbf{t}^{13,42}\right\}\right\}, \text{ and} \\ C_a^{(2)} = C_a^{(1)} &= \left\{\left\{\mathbf{t}^{24,31}\right\}, \left\{\mathbf{t}^{34,21} \qquad\qquad\right\}, \left\{\qquad\qquad \mathbf{t}^{13,42}\right\}\right\}.\end{aligned}$$

Then, for worker β, *he considers* $C_\beta^{(2)}(\mathbf{t}^{32,23}) = \left\{\mathbf{t}^{12,43}\right\}$, *and he prefers the rematching* $(\beta, a; 0)$ *under* $\mathbf{t}^{12,43}$. *Similarly, for firm* a, *she considers* $C_j^{(2)}(\mathbf{t}^{32,23}) = \left\{\mathbf{t}^{34,21}\right\}$, *and she prefers the rematching* $(\beta, a; 0)$ *under* $\mathbf{t}^{34,21}$.

5 Bayesian Stability

Up to now, we have been making the assumption that agents evaluate the prospect of a blocking opportunity according to the worst possible scenario, which follows from Liu et al. (2014). However, we can instead adopt a Bayesian perspective in defining, individual rationality, blocking and stability, and compare them with their counterparts in Section 3. Consider a Bayesian setting in which we fix the agents' common prior λ, which has full support on T. The notation $\tilde{\mathbf{t}}$ is for a random variable drawn from λ. For a subset T' of T, we use the notation $\mathbb{E}_{\tilde{\mathbf{t}}}\left(g(\tilde{\mathbf{t}})|T'\right)$ to denote the conditional expectation of a function g when type assignments are restricted to T'.

5.1 Bayesian Individual Rationality

The following definition of Bayesian "individual rationality" naturally extends Definition 1.

Definition 8. *A state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is said to be **Bayesian individually rational** if for each agent $k = i, j$ and each $\pi \in \Pi_i \vee \Pi_j$ such that $\pi \subset \Pi_k(\mathbf{t})$, her/his expected payoff conditioning on π is nonnegative, i.e., for each i and each j,*

$$\mathbb{E}_{\tilde{\mathbf{t}}}\left[\nu_{\tilde{\mathbf{t}}(i),\tilde{\mathbf{t}}(\mu(i))}|\pi\right] + \mathbf{p}_{i,\mu(i)} \geq 0 \text{ for all } \pi \text{ such that } \pi \in \Pi_i \vee \Pi_j \text{ and } \pi \subset \Pi_i(\mathbf{t}), \text{ and}$$
$$\mathbb{E}_{\tilde{\mathbf{t}}}\left[\phi_{\tilde{\mathbf{t}}(\mu(j)),\tilde{\mathbf{t}}(j)}|\pi\right] - \mathbf{p}_{\mu(j),j} \geq 0 \text{ for all } \pi \text{ such that } \pi \in \Pi_i \vee \Pi_j \text{ and } \pi \subset \Pi_j(\mathbf{t}).$$

5.2 Bayesian Blocking

Now we consider a potential Bayesian blocking combination $(i, j; p)$ for a state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$. For each $\pi \in \Pi_i \cup \Pi_j \cup (\Pi_i \vee \Pi_j)$, we investigate i and j's perception to participate in the potential Bayesian blocking $(i, j; p)$. Intuitively, we use Y for Yes (willing to participate in the Bayesian blocking) and N for No (unwilling to participate in the Bayesian blocking). We define indicator correspondences ξ_k, whose codomain is $\{\emptyset, \{Y\}, \{N\}, \{Y, N\}\}$, by two steps:

First, for each $\pi \in \Pi_i \vee \Pi_j$, agent $k = i, j$ is willing to participate in the Bayesian blocking if she/he expects a higher payoff from the potential rematching $(i, j; p)$, i.e.,

$$\xi_i(\pi) := \begin{cases} \{Y\} & \text{if } \mathbb{E}_{\tilde{\mathbf{t}}}\left[\nu_{\tilde{\mathbf{t}}(i),\tilde{\mathbf{t}}(j)}|\pi\right] + p > \mathbb{E}_{\tilde{\mathbf{t}}}\left[\nu_{\tilde{\mathbf{t}}(i),\tilde{\mathbf{t}}(\mu(i))}|\pi\right] + \mathbf{p}_{i,\mu(i)}, \\ \{N\} & \text{otherwise;} \end{cases}$$
$$\xi_j(\pi) := \begin{cases} \{Y\} & \text{if } \mathbb{E}_{\tilde{\mathbf{t}}}\left[\phi_{\tilde{\mathbf{t}}(i),\tilde{\mathbf{t}}(j)}|\pi\right] - p > \mathbb{E}_{\tilde{\mathbf{t}}}\left[\phi_{\tilde{\mathbf{t}}(\mu(j)),\tilde{\mathbf{t}}(j)}|\pi\right] - \mathbf{p}_{\mu(j),j}, \\ \{N\} & \text{otherwise.} \end{cases}$$

For notational simplicity, the dependence of ξ_i and ξ_j on $(\mu, \mathbf{p}, \Pi)$ and $(i, j; p)$ is suppressed in the notation.

Second, for each $k = i, j$ and each $\pi \in \Pi_k$,

$$\xi_k(\pi) := \bigcup_{\pi' \in \Pi_i \vee \Pi_j : \pi' \subset \pi} \xi_k(\pi').$$

If for some $\pi \in \Pi_k$ we have $\xi_k(\pi) = \{Y, N\}$, then agent k's perception is not determined: it is possible that conditioning on her/his opponent's willingness, she/he does not need to consider the type assignments that lead to N, which implies that agent k's perception may be Y at the set π.

The refinement of the agents' consideration sets under the Bayesian criterion is the analogous to the one under the worst-case criterion. Formally, we define the initial indicator correspondences $\xi_i^{[0]} = \xi_i$, $\xi_j^{[0]} = \xi_j$, and recursively for $l = 1, 2, \dots$ that

(i) for each $\pi \in \Pi_i \vee \Pi_j$, agent $k = i, j$ will not consider π if her/his opponent expects no higher payoff on π, i.e.,

$$\xi_i^{[l]}(\pi) := \begin{cases} \xi_i^{[l-1]}(\pi) & \text{if } Y \in \xi_j^{[l-1]}(\pi'), \text{ where } \pi' \in \Pi_j \text{ and } \pi \subset \pi', \\ \emptyset & \text{otherwise,} \end{cases}$$
$$\xi_j^{[l]}(\pi) := \begin{cases} \xi_j^{[l-1]}(\pi) & \text{if } Y \in \xi_i^{[l-1]}(\pi'), \text{ where } \pi' \in \Pi_i \text{ and } \pi \subset \pi', \\ \emptyset & \text{otherwise;} \end{cases}$$

(ii) for each $k = i, j$ and each $\pi \in \Pi_k$,

$$\xi_k^{[l]}(\pi) := \bigcup_{\pi' \in \Pi_i \vee \Pi_j : \pi' \subset \pi} \xi_k^{[l]}(\pi').$$

To see that $\xi_i^{[l]}(\pi) = \emptyset$ means "her/his opponent expects no higher payoff on π," we note that $\xi_i^{[l]}(\pi) = \emptyset$ occurs only when $\xi_j^{[l-1]}(\pi') = \{N\}$ or $\xi_j^{[l-1]}(\pi') = \emptyset$, where $\pi' \in \Pi_j$ and $\pi \subset \pi'$. It suffices to discuss the former. Suppose at some type assignment $\mathbf{t}'$, the set $\Pi_i(\mathbf{t}')$ contains two distinct subsets $\pi^1, \pi^2 \in \Pi_i \vee \Pi_j$. If

$$\mathbb{E}_{\tilde{\mathbf{t}}}\left[\nu_{\tilde{\mathbf{t}}(i),\tilde{\mathbf{t}}(j)}|\pi^1\right] + p \leq \mathbb{E}_{\tilde{\mathbf{t}}}\left[\nu_{\tilde{\mathbf{t}}(i),\tilde{\mathbf{t}}(\mu(i))}|\pi^1\right] + \mathbf{p}_{i,\mu(i)} \text{ and}$$

$$\mathbb{E}_{\tilde{\mathbf{t}}}\left[\nu_{\tilde{\mathbf{t}}(i),\tilde{\mathbf{t}}(j)}|\pi^2\right] + p \leq \mathbb{E}_{\tilde{\mathbf{t}}}\left[\nu_{\tilde{\mathbf{t}}(i),\tilde{\mathbf{t}}(\mu(i))}|\pi^2\right] + \mathbf{p}_{i,\mu(i)},$$

then

$$\mathbb{E}_{\tilde{\mathbf{t}}}\left[\nu_{\tilde{\mathbf{t}}(i),\tilde{\mathbf{t}}(j)}|\pi^1 \cup \pi^2\right] + p \leq \mathbb{E}_{\tilde{\mathbf{t}}}\left[\nu_{\tilde{\mathbf{t}}(i),\tilde{\mathbf{t}}(\mu(i))}|\pi^1 \cup \pi^2\right] + \mathbf{p}_{i,\mu(i)}.$$

Therefore, if firm j expects no higher payoff at every $\pi \in \Pi_i \vee \Pi_j$ that is contained in $\Pi_i(\mathbf{t}')$, i.e., $\xi_j^{[l-1]}(\Pi_i(\mathbf{t}')) = \{N\}$, then she expects no higher payoff conditioning any combinations of these π's. Similar argument applies to $\xi_j^{[l]} = \emptyset$.

Let l^* be the smallest integer such that $\xi_k^{[l^*+1]} = \xi_k^{[l^*]}$ for both $k = i, j$. The following definition says that the state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is Bayesian blocked by a combination $(i, j; p)$ if, at the true type assignment $\mathbf{t}$ and conditioning on the willingness of their opponents, both agents expect a higher payoff from participating in the potential Bayesian blocking.

Definition 9. *A state* $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ *is said to be* ***Bayesian blocked*** *by* $(i, j; p)$ *if both* i *and* j *are willing to participate in the blocking* $(i, j; p)$*, conditioning on the willingness of their opponents, i.e.,*

$$\xi_i^{[l^*]}\left(\Pi_i\left(\mathbf{t}\right)\right) = \xi_j^{[l^*]}\left(\Pi_j\left(\mathbf{t}\right)\right) = \{Y\}.$$

Similar to Proposition 10, we show that Bayesian blocking satisfies a Bayesian version of improvement-at-the-truth: if a state is Bayesian blocked, then satisfying a blocking combination would deliver higher expected payoffs to both agents conditioning on the set $[\Pi_i \vee \Pi_j](\mathbf{t})$, which is the most precise information that agents i and j may possess by aggregating their partitions.

Proposition 11. *If* $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ *is Bayesian blocked by* $(i, j; p)$*, then*

$$\mathbb{E}_{\tilde{\mathbf{t}}}\left[\nu_{\tilde{\mathbf{t}}'(i), \tilde{\mathbf{t}}'(j)} | \left[\Pi_i \vee \Pi_j\right](\mathbf{t})\right] + p > \mathbb{E}_{\tilde{\mathbf{t}}}\left[\nu_{\tilde{\mathbf{t}}'(i), \tilde{\mathbf{t}}'(\mu(i))} | \left[\Pi_i \vee \Pi_j\right](\mathbf{t})\right] + \mathbf{p}_{i,\mu(i)} \text{ and}$$
$$\mathbb{E}_{\tilde{\mathbf{t}}}\left[\phi_{\tilde{\mathbf{t}}'(i), \tilde{\mathbf{t}}'(j)} | \left[\Pi_i \vee \Pi_j\right](\mathbf{t})\right] - p > \mathbb{E}_{\tilde{\mathbf{t}}}\left[\phi_{\tilde{\mathbf{t}}'(\mu(j)), \tilde{\mathbf{t}}'(j)} | \left[\Pi_i \vee \Pi_j\right](\mathbf{t})\right] - \mathbf{p}_{\mu(j),j}.$$

The proof of Proposition 11 is almost identical to that of Proposition 1, and thus omitted. From Propositions 1 and 11, one can construct examples to see that Bayesian blocking does not imply blocking. This exercise is left to the reader. Conversely, the following example says that blocking does not imply Bayesian blocking. Moreover, the example also says that Bayesian blocking does not satisfy monotonicity.

Example 7 (Example 2 Revisited)**.** *Recall that in Example 2, the partitions of worker* β *and firm* a *are given by*

$$\Pi_\beta = \left\{\left\{\mathbf{t}^{23,32}\right\}, \left\{\mathbf{t}^{21,34}, \mathbf{t}^{21,54}, \mathbf{t}^{01,54}\right\}\right\} \text{ and}$$
$$\Pi_a = \left\{\left\{\mathbf{t}^{23,32}, \mathbf{t}^{21,34}\right\}, \left\{\mathbf{t}^{21,54}, \mathbf{t}^{01,54}\right\}\right\},$$

and the state $(\mu, \mathbf{p}, \mathbf{t}^{23,32}, \Pi)$ *is blocked by* $(\beta, a; 0)$*.*

Suppose the prior distribution λ *satisfy that* $\lambda\left(\mathbf{t}^{21,54}\right) < \lambda\left(\mathbf{t}^{01,54}\right)$*. Then the state* $(\mu, \mathbf{p}, \mathbf{t}^{23,32}, \Pi)$ *is not Bayesian blocked by* $(\beta, a; 0)$*. Therefore, blocking does not imply Bayesian blocking.*

Suppose the prior distribution λ *satisfy that* $\lambda\left(\mathbf{t}^{21,54}\right) > \lambda\left(\mathbf{t}^{01,54}\right)$*. Then the state* $(\mu, \mathbf{p}, \mathbf{t}^{23,32}, \Pi)$ *is Bayesian blocked by* $(\beta, a; 0)$*. However, if we consider a finer partition profile in which*

$$\hat{\Pi}_\beta = \left\{\left\{\mathbf{t}^{23,32}\right\}, \left\{\mathbf{t}^{21,34}, \mathbf{t}^{01,54}\right\}, \left\{\mathbf{t}^{21,54}\right\}\right\} \text{ and}$$
$$\hat{\Pi}_a = \left\{\left\{\mathbf{t}^{23,32}, \mathbf{t}^{21,34}\right\}, \left\{\mathbf{t}^{01,54}\right\}, \left\{\mathbf{t}^{21,54}\right\}\right\},$$

then the state $(\mu, \mathbf{p}, \mathbf{t}^{23,32}, \hat{\Pi})$ is not Bayesian blocked by $(\beta, a; 0)$. In other words, Bayesian blocking does not satisfy monotonicity.

5.3 Bayesian Stability

Equipped with the Bayesian blocking notion (Definition 9), we can define $N^B_{\mu,\mathbf{p},\Pi}$ and $H^B_{\mu,\mathbf{p}}(\cdot)$ in a way that is similar to how we define $N_{\mu,\mathbf{p},\Pi}$ and $H_{\mu,\mathbf{p}}(\cdot)$ in Section 3.3. For completeness, we state the definition for Bayesian stability below.

Definition 10. *A state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is said to be* ***Bayesian-stable*** *if it satisfies that*

(i) $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ *is Bayesian individually rational.*

(ii) $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ *is not Bayesian blocked.*

(iii) $H^B_{\mu,\mathbf{p}}(\Pi) = \Pi$.

A straightforward observation is that the existence of a Bayesian stable state is guaranteed by the existence of a complete-information stable state.

6 Concluding Remarks

In this paper, we provide a framework to study two-sided matching markets with incomplete information. This general framework facilitates the study of matching processes and other related issues when incomplete information exists. For example, the Supplementary Appendix of the current paper discusses (i) the convergence of matching processes, and (ii) the structural properties and the comparative statics of incomplete-information stable states.

References

Abdulkadiroğlu, A. and T. Sönmez (2003): "School choice: A mechanism design approach," *American Economic Review*, 729–747.

Abdulkadiroglu, A. and T. Sönmez (2013): "Matching markets: Theory and practice," *Advances in Economics and Econometrics*, 1, 3–47.

Aumann, R. J. (1976): "Agreeing to disagree," *The Annals of Statistics*, 1236–1239.

Balinski, M. and T. Sönmez (1999): "A tale of two mechanisms: student placement," *Journal of Economic Theory*, 84, 73–94.

Bikhchandani, S. (2017): "Stability with One-Sided Incomplete Information," *Journal of Economic Theory*, 168, 372–399.

Chakraborty, A., A. Citanna, and M. Ostrovsky (2010): "Two-sided matching with interdependent values," *Journal of Economic Theory*, 145, 85–105.

Chen, Y.-C. and G. Hu (2018): "Learning by matching," *Working Paper*.

Crawford, V. P. and E. M. Knoer (1981): "Job matching with heterogeneous firms and workers," *Econometrica*, 437–450.

Dutta, B. and R. Vohra (2005): "Incomplete information, credibility and the core," *Mathematical Social Sciences*, 50, 148–165.

Ehlers, L. and J. Massó (2007): "Incomplete information and singleton cores in matching markets," *Journal of Economic Theory*, 136, 587–600.

——— (2015): "Matching markets under (in) complete information," *Journal of Economic Theory*, 157, 295–314.

Gale, D. and L. S. Shapley (1962): "College admissions and the stability of marriage," *The American Mathematical Monthly*, 69, 9–15.

Liu, Q. (2018): "Rational Expectations, Stable Beliefs, and Stable Matching," *Working Paper*.

Liu, Q., G. J. Mailath, A. Postlewaite, and L. Samuelson (2014): "Stable Matching with Incomplete Information," *Econometrica*, 82, 541–587.

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

——— (2017): "Premuneration values and investments in matching markets," *The Economic Journal*.

Osborne, M. J. and A. Rubinstein (1994): *A course in game theory*, MIT press.

Pomatto, L. (2018): "Stable Matching under Forward-Induction Reasoning," *Working Paper*.

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

——— (2008): "Deferred acceptance algorithms: History, theory, practice, and open questions," *International Journal of Game Theory*, 36, 537–569.

Roth, A. E. and M. A. O. Sotomayor (1990): *Two-sided matching: A study in game-theoretic modeling and analysis*, 18, Cambridge University Press.

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

Wilson, R. (1978): "Information, efficiency, and the core of an economy," *Econometrica: Journal of the Econometric Society*, 807–816.

Appendix A Omitted Arguments for Section 4.3

Proposition 12. *Fix a state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ and a potential blocking combination $(i, j; p)$. For each agent $k = i, j$, we have $C_k^{(l)}(\mathbf{t}') = C_k^{(2)}(\mathbf{t}')$ for all $\mathbf{t}' \in T$ and all $l \geq 3$.*

Proof. First of all, since $C_j^{(l)}(\mathbf{t}'') \subset C_j^{(l-1)}(\mathbf{t}'')$ for all $\mathbf{t}'' \in T$, we know that

$$\min_{\tilde{\mathbf{t}} \in C_j^{(l-1)}(\mathbf{t}'')} \left[\phi_{\tilde{\mathbf{t}}(i),\tilde{\mathbf{t}}(j)} - p\right] - \left[\phi_{\tilde{\mathbf{t}}(\mu(j)),\tilde{\mathbf{t}}(j)} - \mathbf{p}_{\mu(j),j}\right] > 0$$

implies that

$$\min_{\tilde{\mathbf{t}} \in C_j^{(l)}(\mathbf{t}'')} \left[\phi_{\tilde{\mathbf{t}}(i),\tilde{\mathbf{t}}(j)} - p\right] - \left[\phi_{\tilde{\mathbf{t}}(\mu(j)),\tilde{\mathbf{t}}(j)} - \mathbf{p}_{\mu(j),j}\right] > 0.$$

Therefore,

$$C_i^{(l+1)}(\mathbf{t}') = \left\{\mathbf{t}'' \in C_i^{(l)}(\mathbf{t}') : C_j^{(l)}(\mathbf{t}'') \neq \emptyset\right\} \tag{28}$$

for all $\mathbf{t}' \in T$. Similarly, for firm j we have

$$C_j^{(l+1)}(\mathbf{t}') = \left\{\mathbf{t}'' \in C_j^{(l)}(\mathbf{t}') : C_i^{(l)}(\mathbf{t}'') \neq \emptyset\right\} \tag{29}$$

for all $\mathbf{t}' \in T$.

To see the claim, we suppose to the contrary that $\mathbf{t}'' \in C_i^{(2)}(\mathbf{t}')$ but $\mathbf{t}'' \notin C_i^{(3)}(\mathbf{t}')$. Then by (28), we know that $C_j^{(2)}(\mathbf{t}'') = \emptyset$. If $C_j^{(1)}(\mathbf{t}'') = \emptyset$, then $\mathbf{t}'' \notin C_i^{(2)}(\mathbf{t}')$. Therefore, $C_j^{(1)}(\mathbf{t}'') \neq \emptyset$. This, together with $C_j^{(2)}(\mathbf{t}'') = \emptyset$, implies that for each $\mathbf{t}''' \in C_j^{(1)}(\mathbf{t}'')$,

$$\min_{\tilde{\mathbf{t}} \in C_i^{(1)}(\mathbf{t}''')} \left[\nu_{\tilde{\mathbf{t}}(i),\tilde{\mathbf{t}}(j)} + p\right] - \left[\nu_{\tilde{\mathbf{t}}(i),\tilde{\mathbf{t}}(\mu(i))} + \mathbf{p}_{i,\mu(i)}\right] < 0.$$

Since $C_i^{(1)}(\mathbf{t}''') \subset \Pi_i(\mathbf{t}''')$, we have that for each $\mathbf{t}''' \in C_j^{(1)}(\mathbf{t}'')$,

$$\min_{\tilde{\mathbf{t}} \in \Pi_i(\mathbf{t}''')} \left[\nu_{\tilde{\mathbf{t}}(i),\tilde{\mathbf{t}}(j)} + p\right] - \left[\nu_{\tilde{\mathbf{t}}(i),\tilde{\mathbf{t}}(\mu(i))} + \mathbf{p}_{i,\mu(i)}\right] < 0. \tag{30}$$

Note that $\mathbf{t}''' \in C_j^{(1)}(\mathbf{t}'')$ implies $\Pi_j(\mathbf{t}''') = \Pi_j(\mathbf{t}''')$, which in turn implies $C_j^{(1)}(\mathbf{t}''') = C_j^{(1)}(\mathbf{t}'')$. Thus, $\mathbf{t}''' \in C_j^{(1)}(\mathbf{t}''')$, a contradiction to (30). □

Appendix B Monotonicity of Naive and Conservative Stabilities

The definition of naive and conservative stabilities are analogous to Definition 3 and thus omitted. We discuss the monotonicity of conservative stability in this appendix. In particular, it is easier for a state to be conservatively stable if agents have less precise information, i.e., coaser partitions. The monotonicity of naive stability follows from similar but much simpler arguments, and thus omitted.

Given a conservatively stable state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$, if we coarsen the information, i.e., replace Π with a coarser partition profile $\hat{\Pi}$, then the resulted state may not be conservatively stable any more. See Example 8 below. However, the resulted state is "essentially" conservatively stable in the sense that $\left(\mu, \mathbf{p}, \mathbf{t}, H^l_{\mu,\mathbf{p}}\left(\hat{\Pi}\right)\right)$ is stable for some $l \in \mathbb{N}$, where $H^l_{\mu,\mathbf{p}}(\cdot) = H_{\mu,\mathbf{p}}\left(H^{l-1}_{\mu,\mathbf{p}}(\cdot)\right)$.

Example 8. *Consider a market with two workers, $I = \{\alpha, \beta\}$, and two firms, $J = \{a, b\}$. The set of type assignments is given by $T = \left\{\mathbf{t}^{23,23}, \mathbf{t}^{23,21}\right\}$, where $\mathbf{t}$'s are listed below.*

	α	β	a	b
$\mathbf{t}^{23,23}$:	2	3	2	3
$\mathbf{t}^{23,21}$:	2	3	2	1

The premuneration values for workers and firms are both given by product of types, i.e., $\nu_{w,f} = \phi_{w,f} = wf$. The realized type assignment is $\mathbf{t}^{23,23}$. Consider a matching μ with $\mu(\alpha) = a$ and $\mu(\beta) = b$. The payments are all zero, i.e., $\mathbf{p} = \mathbf{0}$.

For the complete-information partition profile Π, i.e., $\Pi_k = \left\{\left\{\mathbf{t}^{23,23}\right\}, \left\{\mathbf{t}^{23,21}\right\}\right\}$ for all k, the state $\left(\mu, \mathbf{p}, \mathbf{t}^{23,23}, \Pi\right)$ is conservatively stable. However, if we consider a alternative Π' that is coarser *than* Π *as below:*

$$\begin{aligned}\Pi'_\alpha &= \left\{\left\{\mathbf{t}^{23,23}, \mathbf{t}^{23,21}\right\}\right\}\\ \Pi'_\beta &= \left\{\left\{\mathbf{t}^{23,23}\right\}, \left\{\mathbf{t}^{23,21}\right\}\right\}\\ \Pi'_a &= \left\{\left\{\mathbf{t}^{23,23}\right\}, \left\{\mathbf{t}^{23,21}\right\}\right\}\\ \Pi'_b &= \left\{\left\{\mathbf{t}^{23,21}\right\}, \left\{\mathbf{t}^{23,23}\right\}\right\},\end{aligned}$$

then $\left(\mu, \mathbf{p}, \mathbf{t}^{23,23}, \Pi'\right)$ *is not conservatively stable since* $H_{\mu,\mathbf{p}}(\Pi') = \Pi \neq \Pi'$*. Nevertheless, the state* $\left(\mu, \mathbf{p}, \mathbf{t}^{23,23}, H_{\mu,\mathbf{p}}(\Pi')\right)$ *is conservatively stable.*

Even though conservative blocking has the monotonicity property, individual rationality does not. To see this, note that when partitions are coarser, the true partition cells contain more type assignments. Under these type assignments, some agents may obtain negative payoffs. Therefore, to obtain the monotonicity of conservative stability, it is necessary to impose conditions such that individual rationality is monotonic in information. We introduce the following assumption as an intuitive sufficient condition for the monotonicity of individual rationality.

Assumption 5. *Each agent can observe her/his own type. Moreover, within each matched pair, agents can observe the type of their own partners.*

Proposition 13. *Suppose that Assumption 5 holds and that* $\hat{\Pi}$ *is a finer partition profile than* Π*. If* $\left(\mu, \mathbf{p}, \mathbf{t}, \hat{\Pi}\right)$ *is conservatively stable, then* $\left(\mu, \mathbf{p}, \mathbf{t}, H^l_{\mu,\mathbf{p}}(\Pi)\right)$ *is conservatively stable for some* $l \in \mathbb{N}$*.*

Proof for Proposition 13. Suppose that $\left(\mu, \mathbf{p}, \mathbf{t}, \hat{\Pi}\right)$ is conservatively stable. Then $\left(\mu, \mathbf{p}, \mathbf{t}, \hat{\Pi}\right)$ is individually rational and not conservatively blocked. Moreover, by the definition of $\mathcal{N}_{\mu,\mathbf{p},\hat{\Pi}}(\mathbf{t})$, we know that $\left(\mu, \mathbf{p}, \mathbf{t}', \hat{\Pi}\right)$ is individually rational and not conservatively blocked for all $\mathbf{t}' \in \mathcal{N}_{\mu,\mathbf{p},\hat{\Pi}}(\mathbf{t})$. Under Assumption 5, the individual rationality of $\left(\mu, \mathbf{p}, \mathbf{t}', \hat{\Pi}\right)$ implies the individual rationality of $(\mu, \mathbf{p}, \mathbf{t}', \Pi)$. By Proposition 10, $\left(\mu, \mathbf{p}, \mathbf{t}', \hat{\Pi}\right)$ being not conservatively blocked implies that $(\mu, \mathbf{p}, \mathbf{t}', \Pi)$ being not conservatively blocked. Therefore, we have

$$\mathcal{N}_{\mu,\mathbf{p},\hat{\Pi}}(\mathbf{t}) \subset \mathcal{N}_{\mu,\mathbf{p},\Pi}(\mathbf{t}). \tag{31}$$

Note that $\mathcal{N}_{\mu,\mathbf{p},\hat{\Pi}}(\mathbf{t})$ is a self-evident event under $\hat{\Pi}$. Since $\hat{\Pi}_k(\mathbf{t}') \subset \mathcal{N}_{\mu,\mathbf{p},\hat{\Pi}}(\mathbf{t})$ for every $\mathbf{t}' \in \mathcal{N}_{\mu,\mathbf{p},\hat{\Pi}}(\mathbf{t})$ and every k, it follows from (31) that

$$\hat{\Pi}_k(\mathbf{t}') \subset \mathcal{N}_{\mu,\mathbf{p},\Pi}(\mathbf{t}).$$

Since $\hat{\Pi}$ is finer than Π, it follows that for every $\mathbf{t}'$ and every k, particularly for every $\mathbf{t}' \in \mathcal{N}_{\mu,\mathbf{p},\hat{\Pi}}(\mathbf{t})$ and every k,

$$\hat{\Pi}_k(\mathbf{t}') \subset \Pi_k(\mathbf{t}').$$

Therefore, for every k and every $\mathbf{t}' \in \mathcal{N}_{\mu,\mathbf{p},\hat{\Pi}}(\mathbf{t})$, we have

$$\hat{\Pi}_k(\mathbf{t}') \subset [H_{\mu,\mathbf{p}}(\Pi)]_k(\mathbf{t}') = \Pi_k(\mathbf{t}') \cap \mathcal{N}_{\mu,\mathbf{p},\Pi}(\mathbf{t}). \tag{32}$$

We claim that the state $(\mu, \mathbf{p}, \mathbf{t}', H_{\mu,\mathbf{p}}(\Pi))$ is also not conservatively blocked. Suppose not. Then, by monotonicity, $\left(\mu, \mathbf{p}, \mathbf{t}', H_{\mu,\mathbf{p}}(\Pi) \vee \mathcal{N}_{\mu,\mathbf{p},\hat{\Pi}}\right)$ is conservatively blocked. Moreover, for every $\mathbf{t}' \in \mathcal{N}_{\mu,\mathbf{p},\hat{\Pi}}(\mathbf{t})$ and every k, it follows from (32) and $\hat{\Pi}_k(\mathbf{t}') \subset \mathcal{N}_{\mu,\mathbf{p},\hat{\Pi}}(\mathbf{t})$ that

$$\hat{\Pi}_k(\mathbf{t}') \subset [H_{\mu,\mathbf{p}}(\Pi)]_k(\mathbf{t}') \cap \mathcal{N}_{\mu,\mathbf{p},\hat{\Pi}}(\mathbf{t}). \tag{33}$$

In other words, on the self-evident event $\mathcal{N}_{\mu,\mathbf{p},\hat{\Pi}}(\mathbf{t})$, the partition profile $\hat{\Pi}$ is finer than $H_{\mu,\mathbf{p}}(\Pi) \vee \mathcal{N}_{\mu,\mathbf{p},\hat{\Pi}}$. Again, by monotonicity, the state $\left(\mu, \mathbf{p}, \mathbf{t}', \hat{\Pi}\right)$ is conservatively blocked, a contradiction.

Let $\Pi^0 = \Pi$ and $\Pi^l = H_{\mu,\mathbf{p}}(\Pi^{l-1})$ for each $l \geq 1$ and Π^∞ be the limit of the increasingly finer partitions Π^l. Inductively, we have

$$\mathcal{N}_{\mu,\mathbf{p},\hat{\Pi}}(\mathbf{t}) \subset \mathcal{N}_{\mu,\mathbf{p},\Pi^l}(\mathbf{t}). \tag{34}$$

Thus, for every $\mathbf{t}' \in \mathcal{N}_{\mu,\mathbf{p},\hat{\Pi}}(\mathbf{t})$, every k and every l, we have

$$\hat{\Pi}_k(\mathbf{t}') \subset \Pi_k^l(\mathbf{t}'),$$

which, together with (34) implies that $(\mu, \mathbf{p}, \mathbf{t}', \Pi^l)$ is not conservatively blocked. Particularly, $(\mu, \mathbf{p}, \mathbf{t}, \Pi^\infty)$ is not conservatively blocked. Obviously, $(\mu, \mathbf{p}, \mathbf{t}, \Pi^\infty)$ is individually rational and $\Pi^\infty = H_{\mu,\mathbf{p}}(\Pi^\infty)$. That is, $(\mu, \mathbf{p}, \mathbf{t}, \Pi^\infty)$ is conservatively stable. $\square$

Supplementary Appendix for

"A Theory of Stability in Matching with Incomplete Information"

Yi-Chun Chen* Gaoji Hu †

November 23, 2018

Appendix C documents an alternative stability notion, termed as ex ante stability. Ex ante stable states are all stable (Definition 3). However, the ex ante stable allocations are supported by particular partition profiles. Appendix D studies the path-to-stability problem, i.e., whether or not learning-blocking paths introduced in Chen and Hu (2018a) converges. Adding agents into a stable market state may trigger learning-blocking paths. Appendix E investigates the welfare effect of adding agents.

Throughout this appendix, we assume that agents can observe their own type and their own partner's type (Assumption 5 in Chen and Hu (2018b)).

Assumption 5. *Each agent can observe her/his own type. Moreover, within each matched pair, agents can observe the type of their own partners.*

Appendix C Ex ante Stability

C.1 Ex ante Stability: Definition

The stability notion that we introduced in Section 3 is closely related to that of Liu et al. (2014) through Chen and Hu (2018a). The notion of Liu et al. (2014) is *ex ante* in that it is defined independently of the true type assignment and firms' heterogeneous belief. We document an analogous ex ante stability notion with two-sided incomplete information.

*Department of Economics, National University of Singapore, S(117570). ecsycc@nus.edu.sg

†Nanyang Business School, Nanyang Technological University, S(639798). hu.gaoji@gmail.com

One can imagine an outside analyst who knows the model except for $\mathbf{t}$ and Π, and who wants to identify possible stable outcomes for the market. As usual, stable outcomes are individually rational and immune to blocking pairs. Formally, a *matching outcome* $(\mu, \mathbf{p}, \mathbf{t})$ specifies an allocation and a type assignment. The individual rationality of a matching outcome is defined as in Definition 1 of Liu et al. (2014), i.e., each agent has a nonnegative payoff. The ex ante blocking notion is designed to exclude only outcomes that the analyst can be certain are "blocked."

Definition 11 (Liu et al. (2014))**.** *A matching outcome* $(\mu, \mathbf{p}, \mathbf{t})$ *is said to be* ***individually rational*** *if*

$$\nu_{\mathbf{t}(i),\mathbf{t}(\mu(i))} + \mathbf{p}_{i,\mu(i)} \geq 0 \textit{ for all } i \in I \textit{ and}$$
$$\phi_{\mathbf{t}(\mu(j)),\mathbf{t}(j)} - \mathbf{p}_{\mu(j),j} \geq 0 \textit{ for all } j \in J.$$

We proceed to define "blocking." Let Σ be a set of ex post outcomes. Define $T_{\Sigma,\mu,\mathbf{p}}$ as the set of type assignments that are associated with the allocation $(\mu, \mathbf{p})$, i.e.,

$$T_{\Sigma,\mu,\mathbf{p}} := \{\mathbf{t} \in T : (\mu, \mathbf{p}, \mathbf{t}) \in \Sigma\}.$$

For notational convenience, let $\mathcal{T}_{\Sigma,\mu,\mathbf{p}}$ be the binary partition that is induced by $T_{\Sigma,\mu,\mathbf{p}}$, i.e.,

$$\mathcal{T}_{\Sigma,\mu,\mathbf{p}} := \{T_{\Sigma,\mu,\mathbf{p}}, T \setminus T_{\Sigma,\mu,\mathbf{p}}\}.$$

Let Π^μ denote the partition profile that is generated by a matching μ, i.e., $\mathbf{t}' \in \Pi_k^\mu(\mathbf{t})$ if and only if $\mathbf{t}'(k) = \mathbf{t}(k)$ and $\mathbf{t}'(\mu(k)) = \mathbf{t}(\mu(k))$ for all k. Indeed, since each agent can observe her/his own type and the type of her/his current partner (Assumption 5), the partition profile Π^μ captures the basic information of agents. Now we define the partition profile $\Pi_{\Sigma,\mu,\mathbf{p}}$ as follows:

$$\Pi_{\Sigma,\mu,\mathbf{p}} := \Pi^\mu \vee \mathcal{T}_{\Sigma,\mu,\mathbf{p}}. \tag{35}$$

Definition 12. *Fix a nonempty set of individually rational outcomes* Σ*. A matching outcome* $(\mu, \mathbf{p}, \mathbf{t}) \in \Sigma$ *is said to be* Σ***-blocked*** *if* $(\mu, \mathbf{p}, \mathbf{t}, \Pi_{\Sigma,\mu,\mathbf{p}})$ *is blocked. A matching outcome* $(\mu, \mathbf{p}, \mathbf{t}) \in \Sigma$ *is* Σ***-stable*** *if it is not* Σ*-blocked.*

The set of outcomes that are immune to the blocking described above is given by the iteration process below. Let Σ^0 be the set of all individually rational outcomes. For $l \geq 1$, define

$$\Sigma^l := \left\{(\mu, \mathbf{p}, \mathbf{t}) \in \Sigma^{l-1} : (\mu, \mathbf{p}, \mathbf{t}) \textit{ is } \Sigma^{l-1} - \textit{stable}\right\}. \tag{36}$$

The set of *ex ante stable outcomes* is given by

$$\Sigma^\infty := \bigcap_{l=1}^{\infty} \Sigma^l.$$

If $(\mu, \mathbf{p}, \mathbf{t})$ is an ex ante stable outcome, we say that the allocation $(\mu, \mathbf{p})$ is an *ex ante stable allocation* at $\mathbf{t}$.

C.2 Fixed Point Characterization

The iterative procedure of (36) describes an algorithm to obtain the set of all ex ante stable matching outcomes. This set has a fixed point characterization, which is often more convenient for verifying that a given matching outcome is ex ante stable. The analysis of this subsection extends that of Section 3.4 of Liu et al. (2014).

Definition 13. *A nonempty set of individually rational matching outcomes E is **self-stabilizing** if every $(\mu, \mathbf{p}, \mathbf{t}) \in E$ is E-stable. The set E **stabilizes a given matching outcome** $(\mu, \mathbf{p}, \mathbf{t})$ if $(\mu, \mathbf{p}, \mathbf{t}) \in E$ and E is self-stabilizing. A set of type assignments $T^* \subset T$ **stabilizes** an allocation $(\mu, \mathbf{p})$ if $\{(\mu, \mathbf{p}, \mathbf{t}) : \mathbf{t} \in T^*\}$ is a self-stabilizing set.*

The following proposition provides a fixed-point characterization of the ex ante stable outcomes. The proof will be almost the same as that for Proposition 2 in Liu et al. (2014), and thus omitted.

Proposition 14. *The following statements are true.*

(i) *If E is a self-stabilizing set, then $E \subset \Sigma^\infty$.*

(ii) *Σ^∞ is a self-stabilizing set and hence the largest self-stabilizing set.*

C.3 Equivalence between Ex ante Stability and Stability

The following proposition connects the stable states with the stable outcomes in Σ^∞.

Proposition 15. $(\mu, \mathbf{p}, \mathbf{t}) \in \Sigma^\infty$ *if and only if there exists a partition profile Π such that the state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is stable.*

Proposition 15 extends Theorem 1 in Chen and Hu (2018a).[11] The proof for the proposition is also a standard extension of that for Theorem 1 in Chen and Hu (2018a), and thus omitted.

Denote the set of stable allocations by $\mathscr{S}(T)$, i.e.,

$$\mathscr{S}(T) := \{(\mu, \mathbf{p}) : \exists \mathbf{t} \in T \text{ and } \Pi \text{ s.t. } (\mu, \mathbf{p}, \mathbf{t}, \Pi) \text{ is stable}\}.$$

The corollary below follows immediately from Proposition 15.

Corollary 1. $\mathscr{S}(T) = \{(\mu, \mathbf{p}) \in \mathscr{A} : \exists \mathbf{t} \in T \textit{ s.t. } (\mu, \mathbf{p}, \mathbf{t}) \in \Sigma^\infty\}$.

The following corollary says that the set of stable allocations is increasing in the ground set of possible type assignments T.

Corollary 2. *If $T \subset T'$, then $\mathscr{S}(T) \subset \mathscr{S}(T')$.*

Proof of Corollary 2. Since $T \subset T'$, we have $\Sigma^0(T) \subset \Sigma^0(T')$. Induction shows that $\Sigma^l(T) \subset \Sigma^l(T')$ for all $l \geq 1$. By Corollary 1, $\mathscr{S}(T) \subset \mathscr{S}(T')$. $\square$

Appendix D Convergence of Learning-Blocking Paths

This appendix extends the convergence results in Chen and Hu (2018a) to matching with two-sided incomplete information. For completeness, we will briefly state the setting and the results, and, however, omit most of the detailed discussions which are similar to those in Chen and Hu (2018a). Throughout this appendix and Appendix E, we assume, in addition to Assumption 5, that the payments in the job market are integers.

Assumption 6. *Payments permitted in the job market are integers.*[12]

Consider a job market in which any worker and any firm can freely choose to be matched to each other, and any agent can freely opt to be unmatched. Suppose also that the agents are myopic, i.e., once an agent or a worker-firm pair finds an opportunity to improve their status quo, they will do

[11] As a special case, Chen and Hu (2018a) discussed the stability notion under the following assumption: for any two type assignments $\mathbf{t}', \mathbf{t}'' \in T$ where $\mathbf{t}' = (\mathbf{w}', \mathbf{f}')$ and $\mathbf{t}'' = (\mathbf{w}'', \mathbf{f}'')$, the firm-type assignments are the same, i.e., $\mathbf{f}' = \mathbf{f}'' = \mathbf{f}$, where $\mathbf{f}$ is commonly known.

[12] In practise, salaries must be rounded to the nearest dollar or penny. See Crawford and Knoer (1981), Kelso and Crawford (1982), Chen et al. (2016), and particularly Chen and Hu (2018a) for similar integral assumptions when *finite* matching processes are studied.

so by either switching to be unmatched or finding a new partner. These individual and/or pairwise rematchings lead to a sequence of market states, which is referred to as a matching process.

We show that with probability one an arbitrary, random learning-blocking path converges to an incomplete-information stable state after finitely many rematchings. The result holds for four stability notions in Chen and Hu (2018b), including (i) naive stability, (ii) conservative stability, (iii) stability, and (iv) Bayesian stability. Throughout this section, we will fix a realized type assignment $\mathbf{t}^*$, which will be omitted for notational simplicity, i.e., we will write $(\mu, \mathbf{p}, \Pi)$ for the state $(\mu, \mathbf{p}, \mathbf{t}^*, \Pi)$.

D.1 Learning-Blocking Paths

With incomplete information, a matching process is necessarily associated with a *learning process.* In the current setup, a learning process corresponds to a sequence of partitional information structures. Each information partition is updated from the previous one according to a new observation. More precisely, given a state $(\mu, \mathbf{p}, \Pi)$, firms may observe one of the following two situations:

(i) there is no rematching; or

(ii) there is a rematching which satisfies a blocking combination $(i, j; p)$.

In case (i), agents update their information according to the partition $\mathcal{N}_{\mu,\mathbf{p},\Pi}$, i.e., the information structure now becomes $H_{\mu,\mathbf{p}}(\Pi)$ as defined in Section 3.3. In case (ii), agent $k = i, j$ will observe her/his new partner's type after they are matched. Moreover, all other agents may update their information to exclude type assignments under which $(i, j; p)$ would not be a blocking combination. To allow for flexible belief updating, we assume in condition (iv) below only that the agents update their information-partition profile from Π to another profile Π' that is (weakly) finer than $\Pi \vee \Pi^{\mu'}$.

We denote by $(\mu, \mathbf{p}, \Pi) \uparrow_{(i,j;p)}$ the state which is derived from state $(\mu, \mathbf{p}, \Pi)$ by satisfying a (blocking) combination $(i, j; p)$ for $(\mu, \mathbf{p}, \Pi)$. Formally, we define the state $(\mu', \mathbf{p}', \Pi') = (\mu, \mathbf{p}, \Pi) \uparrow_{(i,j;p)}$ such that:

(i) worker i and firm j are rematched at salary p, i.e., $\mu'(i) = j$ and $\mathbf{p}'_{i,j} = p$;

(ii) the previous partners of i and j, if any, become unmatched, i.e., $\mu'(\mu(j)) = \emptyset$ if $\mu(j) \neq \varnothing$, and $\mu'(\mu(i)) = \emptyset$ if $\mu(i) \neq \varnothing$;

(iii) other parts of the allocation remain the same as $(\mu, \mathbf{p})$, i.e.,

$$\mu'(i') = \mu(i') \text{ and } \mathbf{p}'_{i',\mu'(i')} = \mathbf{p}_{i',\mu(i')} \text{ for any } i' \in I \setminus \{i, \mu(j)\};$$

(iv) each agent updates her/his information according to her/his observation of the rematching, i.e., Π' is finer than $\Pi \vee \Pi^{\mu'}$.

For notational convenience, we also set $(\mu, \mathbf{p}, \Pi)\uparrow_{(\emptyset,\emptyset;0)} := (\mu, \mathbf{p}, \Pi)$.

Definition 14. *A **learning-blocking path** is a sequence of states* $\{(\mu^l, \mathbf{p}^l, \Pi^l)\}_{l=0}^{L}$ *such that for any* $l \geq 0$*, the following hold:*

(i) *if* $(\mu^l, \mathbf{p}^l, \Pi^l)$ *is not blocked, then* $(\mu^{l+1}, \mathbf{p}^{l+1}) = (\mu^l, \mathbf{p}^l)$ *and* $\Pi^{l+1} = H_{\mu^l, \mathbf{p}^l}(\Pi^l)$*;*

(ii) *moreover, if* $(\mu^l, \mathbf{p}^l, \Pi^l)$ *is blocked, then* $(\mu^{l+1}, \mathbf{p}^{l+1}, \Pi^{l+1}) = (\mu^l, \mathbf{p}^l, \Pi^l)\uparrow_{(i,j;p)}$*, where* $(i, j; p)$ *is a blocking combination for* $(\mu^l, \mathbf{p}^l, \Pi^l)$*.*

D.2 Convergence of Learning-Blocking Paths

There may be cycles along a learning-blocking path, i.e., a learning-blocking path may not converge. However, learning-blocking paths are not completely chaotic. Given an arbitrary initial state, we show that by carefully choosing blocking pairs at each state, we can construct one finite learning-blocking path that ends with a stable state. The following Propositions 16 and 16′ extend the result in Roth and Vande Vate (1990) and Chen and Hu (2018a) to accommodate two-sided incomplete information.

Proposition 16. *Suppose that Assumptions 5 and 6 hold. Then starting from an arbitrary initial state, there exists a finite learning-blocking path that leads to a stable state.*

Now, following RV, we consider a random process which starts with an arbitrary state. The process proceeds to generate a random learning-blocking path, i.e., whenever an intermediate state is blocked by many combinations, the process randomly satisfies one of them. In particular, the blocking combination to be satisfied is drawn from a distribution which has full support on the set of blocking combinations. Moreover, the distribution depends only on the state but not on the history. This random process mimics the practical situations in the labor market: agents meet and negotiate randomly until they expect no more improvement. The proposition below follows immediately from Proposition 16.

Proposition 16′. *Suppose that Assumptions 5 and 6 hold. Then the random learning-blocking path starting from an arbitrary state converges with probability one to a stable state.*

Although the information updating process along a learning-blocking path in the current setting is different from that of Chen and Hu (2018a), the proof of Proposition 16 is very similar to that of

Theorem 2 in Chen and Hu (2018a). Particularly, Propositions 16 and 16′ hold for four stability notions: naive stability, conservative stability, stability and Bayesian stability. The proofs for the first three notions are similar to the one in Section 4.4 of Chen and Hu (2018a). The proof for Bayeisan stability is similar to the one discussed in Appendix B of Chen and Hu (2018a).

Appendix E Comparative Statics

In this section, we investigate the effect of adding agents into a stable market state. Suppose the original market is in a stable state. With new agents, the stable status may be broken and learning-blocking paths may be triggered. We compare the initial stable state and the limit stable states (see Proposition 16′).

E.1 Standard Comparative Statics

We first describe the connections between the two markets to be compared. Let $\Gamma = (I, J, \mathbf{t}^*, T, \nu, \phi)$ and $\hat{\Gamma} = (\hat{I}, \hat{J}, \hat{\mathbf{t}}^*, \hat{T}, \hat{\nu}, \hat{\phi})$ be two matching markets. We say the market $\Gamma = (I, J, \mathbf{t}^*, T, \nu, \phi)$ *is consistent with* $\hat{\Gamma} = (\hat{I}, \hat{J}, \hat{\mathbf{t}}^*, \hat{T}, \hat{\nu}, \hat{\phi})$ if the premuneration value functions are the same on the intersection domain, i.e., $\hat{\nu} = \nu$ and $\hat{\phi} = \phi$ on $T \cap \hat{T}$, and the natural restrictions of T and $\hat{T}$ to the set $(I \cap \hat{I}) \cup (J \cap \hat{J})$ coincide, i.e., if for $\bar{I} := I \cap \hat{I}$ and $\bar{J} := J \cap \hat{J}$ the following two condition holds:

$$\mathbf{t}^*|_{\bar{I} \cup \bar{J}} = \hat{\mathbf{t}}^*|_{\bar{I} \cup \bar{J}}, \text{ and}$$
$$\left\{ \mathbf{t}|_{\bar{I} \cup \bar{J}} \in W^{|\bar{I}|} \times F^{|\bar{J}|} : \mathbf{t} \in T \right\} = \left\{ \mathbf{t}|_{\bar{I} \cup \bar{J}} \in W^{|\bar{I}|} \times F^{|\bar{J}|} : \mathbf{t} \in \hat{T} \right\},$$

where $\mathbf{t}|_{\bar{I} \cup \bar{J}}$ is the restriction of the vector $\mathbf{t}$ to $\bar{I} \cup \bar{J}$. Throughout this section, we take $\hat{\Gamma}$ as a one-agent extension of Γ, i.e., a market that is consistent with Γ and has exactly one more worker or one more firm than Γ.

In the complete information setting, we have the following property if agents have non-transferable utilities. See Section 2.5 of Roth and Sotomayor (1990) and Blum et al. (1997).

Property CS. *Adding one worker (firm) to a stable market state, the result of any (learning-) blocking path makes all other workers (firms) weakly worse off and all firms (workers) weakly better off.*

The intuition is that expanding one side of the market increases the competition within that side. However, these properties may collapse when information is incomplete. We provide some counterexamples in the next subsection.

Non-transferable utility is assumed because of the following example. It says that even in the complete-information environment, property CS cannot be guaranteed if utility is transferable.

Example 9. *Consider a market* $\Gamma = (I, J, \mathbf{t}^*, T)$ *with two workers* $I = \{\alpha, \beta\}$*, one firm* $J = \{a\}$*, and the true type assignment* $\mathbf{t}^* = (\mathbf{t}^*(\alpha), \mathbf{t}^*(\beta), \mathbf{t}^*(a)) = (1, 2; 3)$*. The set of possible type assignment functions is* $T = \{\mathbf{t}^*\}$*, i.e., information is complete. Premuneration functions are given by* $\nu_{wf} = \phi_{wf} = wf$*. Suppose the status quo is given by* $\mu(\beta) = a$*,* $\mu(\alpha) = \emptyset$*, and* $\mathbf{p}_{\beta,a} = 0$*.*

Now a new firm b *with type* 2 *enters the market such that we are faced with a new market* $\hat{\Gamma} = (\hat{I}, \hat{J}, \hat{\mathbf{t}}^*, \hat{T})$*, where the set of firms expands, i.e.,* $\hat{J} = \{a, b\}$*, the set of workers remains the same, i.e.,* $\hat{I} = I$*, the type assignments extends, i.e.,* $\hat{\mathbf{t}}^* = (\hat{\mathbf{t}}^*(\alpha), \hat{\mathbf{t}}^*(\beta); \hat{\mathbf{t}}^*(a), \hat{\mathbf{t}}^*(b), \hat{\mathbf{t}}^*(c)) = (1, 2; 3, 2)$ *and* $\hat{T} = \{\hat{\mathbf{t}}^*\}$*. Note that information is again complete. It is obvious to check that* $\hat{\Gamma}$ *is consistent with* Γ*.*

We explicitly present a learning-blocking path in the table below, which shows that the payoff of worker β *decreases in the new stable state. Particularly, the initial stable state in the original market* Γ *is* S_1*, and the new stable state in the new market* $\hat{\Gamma}$ *is* S_4*.*

States	Matches	Payments	Payoffs				Blocking
			α	β	a	b	Combinations
S_1	$\mu(\beta) = a$ $\mu(\alpha) = \emptyset$	$\mathbf{p}_{\beta,a} = 0$	0	6	6	0	$(\beta, b; 3)$
S_2	$\mu(\beta) = b$ $\mu(\alpha) = \emptyset$	$\mathbf{p}_{\beta,b} = 3$	0	7	0	1	$(\alpha, b; -1)$
S_3	$\mu(\alpha) = b$ $\mu(\beta) = \emptyset$	$\mathbf{p}_{\alpha,b} = -1$	1	0	0	3	$(\beta, a; -1)$
S_4	$\mu(\alpha) = b$ $\mu(\beta) = a$	$\mathbf{p}_{\alpha,b} = -1$ $\mathbf{p}_{\beta,a} = -1$	1	5	7	3	N.A.

E.2 Counterexamples

In this subsection, we provide some counterexamples such that when information is incomplete, property CS fails due to the information updating. In Examples 10 and 11 below, agents' types are correlated. Thus inferences can be made through correlation. In Example 12, while agents' types are independent, inferences can also be made through the colored hat reasoning.[13]

[13]See Liu et al. (2014) for the colored hat reasoning in the context of matching with incomplete information.

Throughout this subsection, the premuneration functions are given by

$$\nu_{wf} = |wf| - 1 \text{ and } \phi_{wf} = 3 + wf. \tag{37}$$

Moreover, we assume there is no transfer between workers and firms (regarded as a constant 0).[14]

Example 10. *(Adding one worker makes another worker better off.) Consider a market $\Gamma = (I, J, \mathbf{t}^*, T)$ with one worker $I = \{\alpha\}$, one firm $J = \{a\}$, and the true type assignment $\mathbf{t}^* = (\mathbf{t}^*(\alpha), \mathbf{t}^*(a)) = (1; 4)$. The set of possible type assignments is $T = \{\mathbf{t}, \mathbf{t}^*\}$, where $\mathbf{t} = (-1; 4)$. We list all possible states below.*

States	Matches	Partition Profile	Blocking Pairs
S_1	$\mu(\alpha) = a$	$\Pi_\alpha = \Pi_a = \{\{\mathbf{t}\}, \{\mathbf{t}^*\}\}$	
S_2	$\mu(\alpha) = \emptyset$	$\Pi_\alpha = \{\{\mathbf{t}\}, \{\mathbf{t}^*\}\}$	
		$\Pi_a = \{\{\mathbf{t}, \mathbf{t}^*\}\}$	
S_3	$\mu(\alpha) = \emptyset$	$\Pi_\alpha = \Pi_a = \{\{\mathbf{t}\}, \{\mathbf{t}^*\}\}$	(α, a)

It is straightforward to verify that there are two stable states for this market: the one with no match and the one matches worker α to firm a, i.e., state S_1 and state S_2. Suppose the status quo is S_2.

Now a new worker β with type $\frac{1}{2}$ and a counterfactual type 1 enters the market such that we are faced with a new market $\hat{\Gamma} = (\hat{I}, \hat{J}, \hat{\mathbf{t}}^*, \hat{T})$, where the set of firms remains the same, i.e., $\hat{J} = J$, the set of workers expands, i.e., $\hat{I} = \{\alpha, \beta\}$, and the true type assignment becomes a three dimensional vector, i.e., $\hat{\mathbf{t}}^* = (\hat{\mathbf{t}}^*(x), \hat{\mathbf{t}}^*(y); \hat{\mathbf{t}}^*(a)) = (1, \frac{1}{2}, 4)$, and there are still two possible type assignments, i.e., $\hat{T} = \left\{\hat{\mathbf{t}}, \hat{\mathbf{t}}^*\right\}$ where $\hat{\mathbf{t}} = (-1, 1; 4)$. It is obvious that $\hat{\Gamma}$ is consistent with Γ. For the new market $\hat{\Gamma}$, there are only four possible states, listed below.

States	Matches	Partition Profile	Blocking Pairs
$\hat{S}_1$	$\hat{\mu}(\alpha) = a$	$\hat{\Pi}_\alpha = \hat{\Pi}_\beta = \hat{\Pi}_a = \left\{\left\{\hat{\mathbf{t}}\right\}, \left\{\hat{\mathbf{t}}^*\right\}\right\}$	
	$\hat{\mu}(\beta) = \emptyset$		
$\hat{S}_2$	$\hat{\mu}(\alpha) = \emptyset$	$\hat{\Pi}_\alpha = \hat{\Pi}_\beta = \hat{\Pi}_a = \left\{\left\{\hat{\mathbf{t}}\right\}, \left\{\hat{\mathbf{t}}^*\right\}\right\}$	(α, a)
	$\hat{\mu}(\beta) = a$		
$\hat{S}_3$ or $\hat{S}_4$	$\hat{\mu}(\alpha) = \emptyset$	$\hat{\Pi}_a = \left\{\left\{\hat{\mathbf{t}}\right\}, \left\{\hat{\mathbf{t}}^*\right\}\right\}$ or $\hat{\Pi}_a = \left\{\left\{\hat{\mathbf{t}}, \hat{\mathbf{t}}^*\right\}\right\}$	(β, a)
	$\hat{\mu}(\beta) = \emptyset$	$\hat{\Pi}_\alpha = \hat{\Pi}_\beta = \left\{\left\{\hat{\mathbf{t}}\right\}, \left\{\hat{\mathbf{t}}^*\right\}\right\}$	

[14]It is straightforward to verify that even if we allow for transfers, it cannot be used by firms to screen workers in our examples. Particularly, whenever a firm intends to screen out a low-type worker by setting a wage that he would not accept, the high type must also find the wage unacceptable. Therefore, the violation of property CS cannot be eased by introducing transfer.

Among the four possible states, the state with a match between worker α and firm a, i.e., state $\hat{S}_1$, is stable. Note that if matched with worker β, firm a could infer that worker α has a type of 1. Thus, with this new piece of information, firm a will be better off if she is rematched with worker α. From the initial stable state S_1 to the unique new stable state $\hat{S}_1$, worker α gets better off.

In Example 10, notably, the Rural Hospital Theorem (and the Lone Wolf Theorem as a special case, see Roth (1984)) is violated in the market Γ. Particularly, worker α is matched with firm a under the stable state S_1 but he is not matched under the stable state S_2.

Example 11. *(Adding one worker makes a firm worse off.) Consider a market $\Gamma = (I, J, \mathbf{t}^*, T)$ with one worker $I = \{\alpha\}$, two firms $J = \{a, b\}$, the true type assignment $\mathbf{t}^* = (\mathbf{t}^*(\alpha); \mathbf{t}^*(a), \mathbf{t}^*(b)) = (1; 2, 4)$, and two possible type assignments $T = \{\mathbf{t}, \mathbf{t}^*\}$ where $\mathbf{t} = (-1; 2, 4)$. Suppose the status quo stable state is given by $\mu(\alpha) = a$ and $\Pi_\alpha = \Pi_a = \{\{\mathbf{t}\}, \{\mathbf{t}^*\}\}$. The information partition for firm b is not relevant for discussion, and thus omitted.*

Now a new worker β enters the market such that we are faced with a new market $\hat{\Gamma} = (\hat{I}, \hat{J}, \hat{\mathbf{t}}^*, \hat{T})$, where the set of firms remains the same, i.e., $\hat{J} = J$, the set of workers expands, i.e., $\hat{I} = \{\alpha, \beta\}$, the true type assignment extends, i.e., $\hat{\mathbf{t}}^* = (\hat{\mathbf{t}}^*(\alpha), \hat{\mathbf{t}}^*(\beta); \hat{\mathbf{t}}^*(a), \hat{\mathbf{t}}^*(b)) = (1, \frac{1}{2}; 2, 4)$, and there are still two possible type assignments $\hat{T} = \left\{\hat{\mathbf{t}}, \hat{\mathbf{t}}^*\right\}$ where $\hat{\mathbf{t}} = (-1, 1; 2, 4)$. It is obvious to check that $\hat{\Gamma}$ is consistent with Γ. For the market $\hat{\Gamma}$, there are seven possible matchings, listed in the following table.

Cases	Matches	Blocking Pairs
1	$\hat{\mu}(\alpha) = \emptyset$ and $\hat{\mu}(\beta) = \emptyset$	(b, β) and (a, β)
2	$\hat{\mu}(\alpha) = a$ and $\hat{\mu}(\beta) = \emptyset$	(b, β)
3	$\hat{\mu}(\alpha) = \emptyset$ and $\hat{\mu}(\beta) = a$	(b, β)
4	$\hat{\mu}(\alpha) = b$ and $\hat{\mu}(\beta) = \emptyset$	(a, β)
5	$\hat{\mu}(\alpha) = \emptyset$ and $\hat{\mu}(\beta) = b$	(a, α)
6	$\hat{\mu}(\alpha) = a$ and $\hat{\mu}(\beta) = b$	(b, α)
7	$\hat{\mu}(\alpha) = b$ and $\hat{\mu}(\beta) = a$	

Note that there is only one stable state, with $\hat{\mu}(\alpha) = b$, $\hat{\mu}(\beta) = a$ and $\hat{\Pi}_k = \left\{\left\{\hat{\mathbf{t}}\right\}, \left\{\hat{\mathbf{t}}^*\right\}\right\}$ for all k, which makes firm a worse off. We have two additional observations pertain to the matching welfare. One is that the number of matched firms increases and the other is that the firm-side aggregate payoff, as well as the social welfare, increases.

In Example 11, the Lattice Structure (see Knuth (1976)) is violated in the market Γ. To see this, consider two stable matchings of the market Γ, i.e., $\mu(\alpha) = a$ and $\mu(\alpha) = b$. The join of these two matchings, in which the firms pick their preferred partners, results in $\mu(a) = \alpha = \mu(b)$, which is not a well-defined matching.

Example 12. *(Adding one firm makes a firm better off and a worker worse off.) Consider a market $\Gamma = (I, J, \mathbf{t}^*, T)$ with two workers $I = \{\alpha, \beta\}$, one firm $J = \{a\}$, and the set of possible type assignments $T = \{\mathbf{t}^1, \mathbf{t}^2, \mathbf{t}^3, \mathbf{t}^*\}$ where*

	α	β	a
$\mathbf{t}^1$	-1	-4	4
$\mathbf{t}^2$	1	-4	4
$\mathbf{t}^3$	-1	$\frac{3}{2}$	4
$\mathbf{t}^*$	1	$\frac{3}{2}$	$4.$

Suppose the status quo stable state is given by $\mu(\alpha) = a$, and

$$\Pi_\alpha = \left\{\left\{\mathbf{t}^1, \mathbf{t}^3\right\}, \left\{\mathbf{t}^2, \mathbf{t}^*\right\}\right\},$$
$$\Pi_\beta = \left\{\left\{\mathbf{t}^1, \mathbf{t}^2\right\}, \left\{\mathbf{t}^3, \mathbf{t}^*\right\}\right\} \text{ and}$$
$$\Pi_a = \left\{\left\{\mathbf{t}^1, \mathbf{t}^3\right\}, \left\{\mathbf{t}^2, \mathbf{t}^*\right\}\right\}.$$

Now a new firm b with type $\frac{1}{2}$ enters the market such that we are faced with a new market $\hat{\Gamma} = (\hat{I}, \hat{J}, \hat{\mathbf{t}}^*, \hat{T})$, where the set of firms expands, i.e., $\hat{J} = \{a, b\}$, the set of workers remains the same, i.e., $\hat{I} = I$, and there are still four possible type assignments, i.e., $T = \{\hat{\mathbf{t}}^1, \hat{\mathbf{t}}^2, \hat{\mathbf{t}}^3, \hat{\mathbf{t}}^*\}$ where

	α	β	a	b
$\hat{\mathbf{t}}^1$	-1	-4	4	$\frac{1}{2}$
$\hat{\mathbf{t}}^2$	1	-4	4	$\frac{1}{2}$
$\hat{\mathbf{t}}^3$	-1	$\frac{3}{2}$	4	$\frac{1}{2}$
$\hat{\mathbf{t}}^*$	1	$\frac{3}{2}$	4	$\frac{1}{2}.$

It is obvious that $\hat{\Gamma}$ is consistent with Γ. For the market $\hat{\Gamma}$, there are seven possible matchings, listed below.

Cases	States (Partial Description)	Blocking Pairs/Updating/IR
1	$\hat{\mu}(\alpha) = \hat{\mu}(\beta) = \emptyset$ and $\hat{\Pi}_a = \{\{\hat{\mathbf{t}}^1, \hat{\mathbf{t}}^2, \hat{\mathbf{t}}^3, \hat{\mathbf{t}}^*\}\}$	$\hat{\Pi}_a = \{\{\hat{\mathbf{t}}^1, \hat{\mathbf{t}}^2\}, \{\hat{\mathbf{t}}^3, \hat{\mathbf{t}}^*\}\}$
2	$\hat{\mu}(\alpha) = \hat{\mu}(\beta) = \emptyset$ and $\hat{\Pi}_a = \{\{\hat{\mathbf{t}}^1, \hat{\mathbf{t}}^2\}, \{\hat{\mathbf{t}}^3, \hat{\mathbf{t}}^*\}\}$	(a, β)
3	$\hat{\mu}(\alpha) = a$, $\hat{\mu}(\beta) = \emptyset$ and $\hat{\Pi}_a = \{\{\hat{\mathbf{t}}^1, \hat{\mathbf{t}}^3\}, \{\hat{\mathbf{t}}^2, \hat{\mathbf{t}}^*\}\}$	$\hat{\Pi}_a = \{\{\hat{\mathbf{t}}^1, \hat{\mathbf{t}}^3\}, \{\hat{\mathbf{t}}^2\}, \{\hat{\mathbf{t}}^*\}\}$
4	$\hat{\mu}(\alpha) = a$, $\hat{\mu}(\beta) = \emptyset$ and $\hat{\Pi}_a = \{\{\hat{\mathbf{t}}^1, \hat{\mathbf{t}}^3\}, \{\hat{\mathbf{t}}^2\}, \{\hat{\mathbf{t}}^*\}\}$	(a, β)
5	$\hat{\mu}(\alpha) = \emptyset$ and $\hat{\mu}(\beta) = a$	
5	$\hat{\mu}(\alpha) = b$ and $\hat{\mu}(\beta) = \emptyset$	Not IR
6	$\hat{\mu}(\alpha) = \emptyset$ and $\hat{\mu}(\beta) = b$	Not IR
7	$\hat{\mu}(\alpha) = a$ and $\hat{\mu}(\beta) = b$	Not IR
8	$\hat{\mu}(\alpha) = b$ and $\hat{\mu}(\beta) = a$	Not IR

Note that there is only one stable matching with $\hat{\mu}(\alpha) = \emptyset$ and $\hat{\mu}(\beta) = a$, regardless of the partition profile, which makes firm a better off and at the same time makes worker α worse off.

E.3 Restoring Comparative Statics

A common feature of the examples in the last subsection is that all status quo states are stable, and, however, none of the status quo matching outcomes is complete-information stable. Thus a natural question to ask is whether we can restore property CS for the class of markets where (i) agents have non-transferable utilities, and (ii) the status quo stable matching outcome is complete-information stable. For this class of markets, the following proposition restores property CS under the mild assumption of strict preferences, i.e., $\phi_{\hat{\mathbf{t}}^*(i),\hat{\mathbf{t}}^*(j)} \neq \phi_{\hat{\mathbf{t}}^*(i'),\hat{\mathbf{t}}^*(j)}$ when $i \neq i'$ and $\nu_{\hat{\mathbf{t}}^*(i),\hat{\mathbf{t}}^*(j)} \neq \nu_{\hat{\mathbf{t}}^*(i),\hat{\mathbf{t}}^*(j')}$ for $j \neq j'$. Recall that $\hat{\Gamma}$ is a one-agent extension of Γ.

Proposition 17. *Suppose that preferences are strict, that there is no transfer permitted in the market, and that Assumptions 5 and 6 hold. If $(\mu, \mathbf{t}^*, \Pi)$ is a stable Γ-state such that μ is a complete-information stable matching, then for any stable $\hat{\Gamma}$-state $(\hat{\mu}, \hat{\mathbf{t}}^*, \hat{\Pi})$ that is produced by learning-blocking paths, property CS holds. Precisely, when $\hat{\Gamma}$ is a one-firm (resp. one-worker) extension of Γ, the payoffs of all workers (resp. firms) increase and the payoffs of all existing firms (resp. workers) decrease, compared with the payoffs under $(\mu, \mathbf{t}^*, \Pi)$.*

The welfare comparison is between a complete-information stable Γ-state and an incomplete-information stable $\hat{\Gamma}$-state, the latter may not be complete-information stable as the following example shows.

Example 13. *Consider a market* $\Gamma = (I, J, \mathbf{t}^*, T)$ *with two workers* $I = \{\alpha, \beta\}$*, one firm* $J = \{a\}$*, and the true type assignment* $\mathbf{t}^* = (\mathbf{t}^*(\alpha), \mathbf{t}^*(\beta), \mathbf{t}^*(a)) = (5, 4; 1)$*. The set of possible type assignment functions is* $T = \{\mathbf{t}, \mathbf{t}^*\}$*, where* $\mathbf{t} = (5, -4, 1)$*. Premuneration functions are given by (37).*

Suppose that in the status quo firm a *hires worker* α*, firm* b *hires no one, worker* β *knows the true type assignment, and the other two agents do not know the true type assignment. Formally,* $\mu(\alpha) = a$*,* $\mu(\beta) = \emptyset$*,* $\Pi_\beta = \{\{\mathbf{t}\}, \{\mathbf{t}^*\}\}$ *and* $\Pi_a = \Pi_\alpha = \{\{\mathbf{t}, \mathbf{t}^*\}\}$*.*

Now a new firm b *with type* 2 *enters the market such that we are faced with a new market* $\hat{\Gamma} = (\hat{I}, \hat{J}, \hat{\mathbf{t}}^*, \hat{T})$*, where the set of firms expands, i.e.,* $\hat{J} = \{a, b\}$*, the set of workers remains the same, i.e.,* $\hat{I} = I$*, and the type assignments extend, i.e.,* $\hat{\mathbf{t}}^* = (\hat{\mathbf{t}}^*(\alpha), \hat{\mathbf{t}}^*(\beta); \hat{\mathbf{t}}^*(a), \hat{\mathbf{t}}^*(b)) = (5, 4; 1, 2)$ *and* $\hat{T} = \left\{\hat{\mathbf{t}}, \hat{\mathbf{t}}^*\right\}$*, where* $\hat{\mathbf{t}} = (5, -4; 1, 2)$*. It is obvious to check that* $\hat{\Gamma}$ *is consistent with* Γ*.*

We notice that in the new market $\hat{\Gamma}$*, the pair* (α, b) *is indeed a blocking pair for the stable* Γ*-state that serves as the status quo. Satisfying* (α, b) *results in a stable* $\hat{\Gamma}$*-state that is given by* $\hat{\mu}(\alpha) = b$*,* $\hat{\mu}(\beta) = \emptyset$*,* $\hat{\Pi}_\beta = \{\{\mathbf{t}\}, \{\mathbf{t}^*\}\}$ *and* $\hat{\Pi}_a = \hat{\Pi}_b = \hat{\Pi}_\alpha = \{\{\mathbf{t}, \mathbf{t}^*\}\}$*. Obviously, the allocation* $\hat{\mu}$ *is not complete-information stable in the new market* $\hat{\Gamma}$*.*

Proof of Proposition 17. We show the case where $\hat{\Gamma}$ is a one-firm extension of Γ, i.e., $J \subsetneq \hat{J}$. The other case is symmetric. To simplify the notation, we denote the payoff for an agent $k \in \hat{I} \cup \hat{J}$ under $(\mu, \mathbf{t}^*, \Pi)$ as $u(k)$ while the payoff under $(\hat{\mu}, \hat{\mathbf{t}}^*, \hat{\Pi})$ as $\hat{u}(k)$.

Suppose to the contrary that there exist $i \in I$ such that $\hat{u}(i) < u(i)$. Consider the sequence constructed by alternatively applying μ and $\hat{\mu}$ to i until some agent is reached for a second time (shown in Figure 1), i.e., the sequence is either $i, \mu(i), \hat{\mu}(\mu(i)), \ldots, (\hat{\mu} \circ \mu)^k(i)$ or $i, \mu(i), \hat{\mu}(\mu(i)), \ldots, \mu \circ (\hat{\mu} \circ \mu)^k(i)$ for some integer k, which stops when encountering an existing agent. Obviously, the sequence has finitely many agents because $\hat{I} \cup \hat{J}$ is finite, which implies that k is finite.

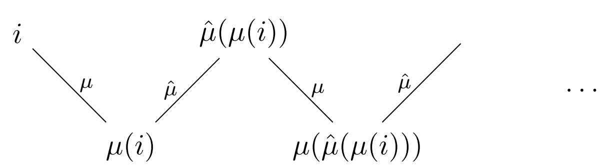

Figure 1: An sequence by alternatively applying μ and $\hat{\mu}$ to i.

We claim that $\hat{u}(\mu(i)) > u(\mu(i))$. Otherwise, $(i, \mu(i))$ is a blocking pair for $(\hat{\mu}, \hat{\mathbf{t}}^*, \hat{\Pi})$ because they have complete information about each other's true type. Next we claim that $\hat{u}(\hat{\mu}(\mu(i))) < u(\hat{\mu}(\mu(i)))$. Otherwise, $(\hat{\mu}(\mu(i)), \mu(i))$ is a blocking pair for μ, i.e., μ is not complete-information stable. We continue this argument until the end of the sequence. Particularly, either $u(\mu((\hat{\mu} \circ \mu)^{k-1}(i))) < \hat{u}(\mu((\hat{\mu} \circ \mu)^{k-1}(i)))$

or $u((\hat{\mu}\circ\mu)^k(i)) > \hat{u}((\hat{\mu}\circ\mu)^k(i))$. Consider two cases.

Case 1. The sequence stops with $(\hat{\mu}\circ\mu)^k(i)$, i.e., firm $\mu((\hat{\mu}\circ\mu)^{k-1}(i))$ is matched under $\hat{\mu}$ with some worker $(\hat{\mu}\circ\mu)^k(i)$ who was unmatched under μ. Since $u(\mu((\hat{\mu}\circ\mu)^{k-1}(i))) < \hat{u}(\mu((\hat{\mu}\circ\mu)^{k-1}(i)))$, we know that $((\hat{\mu}\circ\mu)^k(i), \mu((\hat{\mu}\circ\mu)^{k-1}(i)))$ is a blocking pair for μ, i.e., μ is not complete-information stable, a contradiction.

Case 2. The sequence stops with $\mu\circ(\hat{\mu}\circ\mu)^k(i)$, i.e., firm $\mu\circ(\hat{\mu}\circ\mu)^k(i)$ is unmatched under $\hat{\mu}$. Note that worker $(\hat{\mu}\circ\mu)^k(i)$ and firm $\mu(\hat{\mu}\circ\mu)^{k-1}(i)$ have complete information about each other's true type. Since $u((\hat{\mu}\circ\mu)^k(i)) > \hat{u}((\hat{\mu}\circ\mu)^k(i))$, we know that $((\hat{\mu}\circ\mu)^k(i), \mu((\hat{\mu}\circ\mu)^k(i)))$ is a blocking pair for $(\hat{\mu}, \hat{\mathbf{t}}^*, \hat{\Pi})$, i.e., $(\hat{\mu}, \hat{\mathbf{t}}^*, \hat{\Pi})$ is not a stable $\hat{\Gamma}$-state, a contradition.

Suppose to the contrary that there exists $j\in J$ such that $\hat{u}(j) > u(j)$. We have argued that $\hat{u}(i)\geq u(i)$ for all $i\in I$ and particularly $i=\mu(j)$. Then $(\mu(j), j)$ is a blocking pair for state $(\hat{\mu}, \hat{\mathbf{t}}^*, \hat{\Pi})$, since they have complete information about each other's type. □

References

Blum, Y., A. E. Roth, and U. G. Rothblum (1997): "Vacancy chains and equilibration in senior-level labor markets," *Journal of Economic theory*, 76, 362–411.

Chen, B., S. Fujishige, and Z. Yang (2010): "Decentralized market processes to stable job matchings with competitive salaries," *KIER Discussion Paper*, 749.

——— (2016): "Random decentralized market processes for stable job matchings with competitive salaries," *Journal of Economic Theory*, 165, 25–36.

Chen, Y.-C. and G. Hu (2018a): "Learning by matching," *Working Paper.*

——— (2018b): "A Theory of Stability in Matching with Incomplete Information," *Working Paper.*

Crawford, V. P. and E. M. Knoer (1981): "Job matching with heterogeneous firms and workers," *Econometrica*, 437–450.

Fudenberg, D., F. Drew, D. K. Levine, and D. K. Levine (1998): *The theory of learning in games*, vol. 2, MIT press.

Kelso, A. S. J. and V. P. Crawford (1982): "Job matching, coalition formation, and gross substitutes," *Econometrica: Journal of the Econometric Society*, 1483–1504.

Knuth, D. E. (1976): *Mariages stables et leurs relations avec d&autres problèmes combinatoires*, Presses de l'Université de Montréal.

Liu, Q., G. J. Mailath, A. Postlewaite, and L. Samuelson (2014): "Stable Matching with Incomplete Information," *Econometrica*, 82, 541–587.

Roth, A. E. (1984): "The evolution of the labor market for medical interns and residents: a case study in game theory," *The Journal of Political Economy*, 991–1016.

Roth, A. E. and M. A. O. Sotomayor (1990): *Two-sided matching: A study in game-theoretic modeling and analysis*, 18, Cambridge University Press.

Roth, A. E. and J. H. Vande Vate (1990): "Random paths to stability in two-sided matching," *Econometrica*, 1475–1480.