A Theory of Stability in Matching with Incomplete Information*

Yi-Chun Chen† Gaoji Hu‡

November 1, 2021

Abstract

We provide a framework for studying two-sided matching markets with incomplete information. The framework accommodates two-sided incomplete information as well as heterogeneous information among the agents. We propose a notion called stability for a market state, which, based upon agents' information structure, requires (i) individual rationality, (ii) no blocking and (iii) information stability. The novelty of our stability notion lies in how the agents evaluate a blocking prospect, in the presence of general two-sided incomplete information. We show that a stable state exists; moreover, if a state is stable, then coarsening agents' information leads to another stable state.

JEL Classification: C78, D83
Keywords: matching, incomplete information, blocking, stability.

*We are indebted to the editor and three anonymous referees for their insightful comments. We are especially grateful to Fuhito Kojima and Qingmin Liu for encouragement. We also thank David Ahn, In-Koo Cho, Laura Doval, Jiangtao Li, Bart Lipman, Xiao Luo, George Mailath, Yusufcan Masatlioglu, Jordi Massó, Luciano Pomatto, Andy Postlewaite, Mike Ostrovsky, Joel Sobel, Ning Sun, Xiang Sun, Satoru Takahashi, Qianfeng Tang, Xi Weng, 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. Chen acknowledges the partial support by Singapore MOE AcRF Tier 1 Grant. Hu acknowledges the support by the National Natural Science Foundation of China (Grant 72003121). All remaining errors are our own.

†Department of Economics and Risk Management Institute, National University of Singapore, Singapore. Email address: ecsycc@nus.edu.sg.

‡School of Economics, Shanghai University of Finance and Economics, Shanghai, China. Email address: hugaoji@sufe.edu.cn.

Contents

1 Introduction

This paper provides a framework for studying two-sided matching markets with incomplete information. Matching is one of the most important functions of markets (Roth, 2008). Since the seminal work of Gale and Shapley (1962) and Shapley and Shubik (1971), for several decades, stability has been a key notion used in theoretical study as well as practical design of matching markets; see Roth and Sotomayor (1990) for a survey, and Abdulkadiroglu and Sönmez (2013) and Kojima (2015) on recent developments. Conceptually, stability has been connected to both equity and efficiency, two of the most important notions in economics.[1] Practically, the design of many matching markets aims to achieve stable matchings; see Roth and Peranson (1999) and Abdulkadiroglu and Sönmez (2013).

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. The assumption makes the analysis tractable but is at best an idealization. A woman or man may have imprecise information about her or his potential partner in a marriage market. A college may have imprecise information about the students' capability and a student may also have imperfect understanding about her/his suitability to a college. Likewise, a firm may not know the productivity of its job candidates, whereas a worker may not know the corporate culture or financial health of his potential employer.

Incomplete information is also important concerning the theory of market design. Specifically, if incentive-compatible stable mechanisms are always available, then there is no need to concern incomplete information—agents simply reveal their own information truthfully. However, there is no mechanism that yields stable matchings and ensures that agents on both sides of the market have dominant-strategy incentive to truthfully reveal their information (Roth, 1982). For example, the man-proposing deferred acceptance mechanism makes it a dominant strategy

[1] See, e.g., Abdulkadiroglu and Sönmez (2013) for how stability implies the elimination of justified envy, a basic fairness axiom; see, e.g., Shapley and Shubik (1971) for how stability leads to efficiency.

for men to truthfully reveal their information, but not for women; and vice versa. In fact, even with weaker notions of incentive compatibility or stability, stable mechanisms can only exist under stringent assumptions; see, e.g., Ehlers and Massó (2007) and Yenmez (2013). Hence, incomplete information, especially incomplete information on both sides of the market, is of substantive importance.[2]

In this paper, we propose a general analytical framework for matching with incomplete information, along with a novel notion of stability.[3] Equipped with the framework, we establish the existence of a stable matching and study the properties and the structure of stable matchings. Our analytical framework is applicable to a broad range of matching setups with incomplete information.

To motivate our framework, consider a setup where workers are matched with firms (Liu et al., 2014).[4] We first propose a solution concept called (incomplete-information) stability. Specifically, stability under complete information requires individual rationality (i.e., no one strictly prefers staying unmatched to the status quo) and no blocking pair (i.e., no worker and firm would both prefer being matched with each other at some wage to the status quo). When a firm has incomplete information about workers' types, however, she may not know whether hiring another worker is better than keeping her current employee. Moreover, the firm may not even know whether keeping her current employee is better than becoming unmatched. Similar difficulties arise on the worker side. In this situation, the standard notions of individual rationality, blocking, and stability in the complete-information environment become inadequate.

In the incomplete-information environment, each agent is associated with a

[2]A stability concept incorporating incomplete information (on both sides) has fundamental importance for both centralized and decentralized matching markets, and it goes before any market design problem. We focus on formulating such stability concepts without touching upon any incentive problem. This also distinguishes our approach from the market/mechanism design literature.

[3]Most of the existing papers assume that the types of agents on one side (say, the firms) are common knowledge. Two notable exceptions are Bikhchandani (2014) and Liu (2017), which we will discuss in Section 1.1.

[4]It will become clear that our notions and analysis are also applicable to other matching setups such as those without transferable utility and/or with ordinal preferences.

type. The types of the agents determine their payoffs from a match. Moreover, an agent's information is described by a partition over the possible type profiles of all agents. In this setting, a state of the market consists of an allocation (i.e., a matching and wage profile) and an information structure (i.e., a true type profile and a partition profile). 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. We establish the existence of stable states; see Theorem 1.

The novelty of our stability notion lies in how the agents evaluate a blocking prospect "with respect to the information structure" in requirement (ii). When the firms' types are common knowledge as in Liu et al. (2014), a firm's evaluation of a new partnership with a worker is conditional on the worker's willingness of joining it. In particular, the worker's evaluation does not depend on the firm's evaluation, since the worker knows the type of the firm and thereby his potential payoff.

When incomplete information prevails on both sides, each agent in a potential blocking pair needs to consider not only her/his potential partner's "willingness" but also how her/his potential partner considers her/his "willingness", and so on. Unlike the case with one-sided incomplete information, it is not a priori clear how "willingness" should be defined to start with. Our key observation is that a firm can safely rule out a worker's type from consideration if *in no circumstance/type profile* will the worker's type be willing to partner with the firm. Indeed, such a worker type's willingness to block will never be restored, even after the worker type factors in the firm's willingness to narrow down the range of possible firm types in consideration. Once the pair of agents rule out such types from each other, they will consider fewer possible types of each other and there could result in more types who in no (now fewer) circumstances will benefit from the blocking prospect and thereby be ruled out by the other side, and so on. We call the limit set of type profiles a *consideration set* of a blocking agent. Like Liu et al. (2014), we require that each of the two agents benefits from the blocking prospect under *every* type profile in her/his consideration set.

This refinement of consideration set is consistent with and generalizes the belief-free formulation in Liu et al. (2014) to matchings with two-sided incomplete information. More precisely, in Liu et al. (2014), a matching outcome is deemed unstable in their iteration procedure (to achieve requirement (iii)) if and only if it shall be blocked for any possible belief of the firm conditional on the willingness of the worker. Similarly, a type profile is ruled out from a firm's consideration if and only if the worker never prefers the new partnership for *any* belief consistent with the worker's consideration.[5]

Our notions of blocking and stability lend themselves to clear epistemic foundations. Namely, a state is blocked by a worker and a firm if and only if each of them knows according to the consideration set that s/he can benefit from the new partnership. In a similar vein, a state is stable if and only if it is common knowledge that the state is individually rational and not blocked; see Propositions 1-2.

We emphasize two properties of blocking, both of which become vacuous when information is complete. First, improvement at the truth (IT) says that agents in the blocking coalition will obtain higher *ex post* payoffs. Obviously, if IT is satisfied, then blocking implies complete-information blocking. IT bridges the connection between the incomplete-information stability and the complete-information stability. Second, information monotonicity (IM) says that blocking opportunities are easier to arise when agents have more precise information. Our blocking notion satisfies both properties; see Theorem 2.

We investigate the structure of stable states and show that if a state is stable, then coarsening agents' information leads to another stable state; see Proposition 3. The result implies that any stable matching outcome (i.e., the allocation and the type profile in a stable state) can be supported as a stable state by a specific

[5]See Sections 1.1 and 3.3 of Liu et al. (2014) for more discussions and Bergemann and Morris (2007) for belief-free solution concepts in games with incomplete information.

partition profile—a partition profile induced by the trivial partition under which agents have the least information; see Theorem 3.

Finally, we discuss alternative blocking/stability notions. The first notion, called naive blocking, represents the situation where the agents do not refine their consideration set, or equivalently, their consideration set contains all type profiles which they cannot rule out under their information partitions. Naive blocking represents the most conservative blocking notion under incomplete information with no consideration refinement. Between blocking (level-∞) and naive blocking (level-0), we also consider level-l blocking which performs consideration refinement for l rounds. By means of examples, we show that each of the level-l stability notions impose distinct and increasingly more restrictions (with l) on the set of matching outcomes. Moreover, we identify conditions under which naive stability is equivalent to stability and thereby every level-l stability; see Theorem 4.

We also consider another blocking notion called sophisticated blocking which represents the situation where agents are even more aggressive in refining their consideration sets than they are under blocking. Specifically, we argue that there are type profiles which the agents may also rule out under some additional (common knowledge) assumptions on consideration set refinement. Unlike all those level-l blockings ($l = 0$ up to ∞) which satisfy both IT and IM, sophisticated blocking satisfies only IT but *not* IM; see Section 5.3. We also prove additional properties for these blocking/stability notions in the Online Appendix.[6]

The rest of the section reviews the related literature. Section 2 introduces the model. Section 3 defines individual rationality, blocking and stability with incomplete information, and also documents the epistemic foundations of those definitions. Section 4 examines the IT and IM properties. Section 5 studies alternative blocking/stability notions. Section 6 concludes with future research directions. All proofs can be found in the Appendix. More discussions on alternative blocking/stability notions can be found in the Online Appendix.

[6]In particular, we also show that under the conditions identified in Theorem 4, all of these blocking notions (naive or sophisticated) are equivalent.

1.1 Related literature

Our paper follows the recent development on matching with incomplete information. Liu et al. (2014) study stable matchings with one-sided incomplete information. They propose a stability notion and show that under certain monotonicity and supermodularity assumptions on the agents' (premuneration) value functions from a match, every stable matching outcome is efficient. Pomatto (2020) considers a noncooperative matching game and applies forward-induction reasoning to derive the set of stable outcomes in Liu et al. (2014). Chen and Hu (2020) extend the model of Liu et al. (2014) and prove that a random matching process leads to stable outcomes with probability one. Bikhchandani (2017) proposes a notion of stability which is similar to that of Liu et al. (2014) but applies to a Bayesian setting with nontransferable utilities. Bikhchandani (2017) also points out that a Bayesian stable matching does not necessarily exist; see also Alston (2020). Liu (2020) proposes a stability criterion that requires the Bayesian consistency of three beliefs; namely, the exogenously given prior beliefs, the off-path beliefs conditional on counterfactual pairwise blockings, and the on-path beliefs for stable matchings in the absence of such blockings.[7]

All these aforementioned papers assume one-sided incomplete information. When there are two-sided incomplete information, Bikhchandani (2014) and Liu (2017) have proposed different notions of blocking and stability. Specifically, the blocking notion of Bikhchandani (2014) imposes assumptions on both the value functions (e.g., monotonicity) and the agents' knowledge within each side (e.g., workers knowing the types of all workers and firms knowing the types of all firms) so as to define the two agents' most optimistic assumptions about each other's type. In defining blocking, Liu (2017) extends Liu (2020) by proposing a fixed-

[7]Another strand of literature focuses on stable mechanisms (instead of stable matchings) which also involve incomplete information. See, e.g., Roth (1989), Chakraborty, Citanna and Ostrovsky (2010), Yenmez (2013) and Ehlers and Massó (2007, 2015). A stable mechanism usually takes a notion of stable matchings as given and specifies information eliciting and matching allocation rules to achieve stable matchings. From another angle, stable mechanisms are associated with centralized markets, while our paper aims to understand stability of matchings regardless of whether a centralized mechanism is available or not.

point notion of "sets of states of the world" based on which the agents update their beliefs and evaluate the payoffs from joining or rejecting a deviation, i.e., the set of states of the world specified for agent i makes the set of states of the world specified for agent j exactly equal to those in which agent j wants to join the blocking, and vice versa. Unlike these two papers, we allow for arbitrary value functions and information partitions and we describe agents' willingness based on the result of their iterative consideration refinements. Our iterative consideration refinement process also leads to a variety of level-l blocking/stability notions.

Immorlica et al. (2020) study matching markets with endogenous costly information acquisition as opposed to an exogenous information structure. They introduce a notion called regret-free stability that requires optimal student information acquisition in a model of college admissions, where students acquire information about their own preferences over colleges. Since students' priorities at colleges are assumed to be common knowledge, the blocking notion in Immorlica et al. (2020) does not involve the idea of consideration nor its refinement which we study in this paper. Rastegari et al. (2014) study a setting where agents are endowed with partial/incomplete preference orderings and investigate which matchings are stable with respect to some completion of the partial orderings. We may think of Rastegari et al. (2014) as a situation where the agents have a common information partition over the set of profiles of complete preferences. In this perspective, our formulation allows for heterogeneous information about preference incompleteness and reveals novel blocking notions based on iterative consideration refinement.

Our exercise can also be related to the literature on communication and core with incomplete information. Specifically, Wilson (1978) proposes two notions of core which correspond to two polar cases of information-sharing rules within a blocking coalition. In one extreme case, the agents in a coalition pool all the information which they possess, and in the other extreme case, the agents in a coalition share no information with each other. Subsequent papers such as Volij (2000) and Dutta and Vohra (2005) have explored different information-sharing rules between the two polar cases. In particular, Volij (2000) formulates a blocking notion in which agents in a coalition iteratively communicate their willingness to join a blocking until all of them reach a consensus of improvement with respect to their updated information from the communication. In contrast, our notions of blocking and stability rely on an iterative consideration refinement process in which information updating is merely hypothetical and involves no communication. This perspective also differentiates our paper from papers on communication complexity such as Segal (2007), Gonczarowski et al. (2019) and Ashlagi et al. (2020). In particular, Gonczarowski et al. (2019) and Ashlagi et al. (2020) study how, broadly speaking, complete-information stable matchings can be achieved through different communication protocols. In contrast, we formulate and study notions of incomplete-information stability without communication.

2 The Model

We consider the following setup of matching with incomplete information, which extends the ones of Liu et al. (2014) and Chen and Hu (2020).

There is a finite set I of workers to be matched with a finite set J of firms. Denote a generic worker by i and a generic firm by j. Also denote 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* $\nu_{w,f} \in \mathbb{R}$ and the *firm premuneration value* $\phi_{w,f} \in \mathbb{R}$.[8] The sum of $\nu_{w,f}$ and $\phi_{w,f}$ is called the *surplus of the match*. The functions $\nu : W \times F \to \mathbb{R}$ and $\phi : W \times F \to \mathbb{R}$ are common knowledge among the agents. Denote these values by $\nu_{\mathbf{t}(i),\mathbf{t}(i)}$ for unmatched worker i and $\phi_{\mathbf{t}(j),\mathbf{t}(j)}$

[8] See Mailath, Postlewaite and Samuelson (2013, 2017) for discussions on premuneration values.

for unmatched firm j, both of which are set to be zero. Given a match between worker i and firm j, with type assignment $\mathbf{t}$, the worker's payoff and the firm's payoff under wage $p \in \mathbb{R}$ are, respectively, $\nu_{\mathbf{t}(i),\mathbf{t}(j)} + p$ and $\phi_{\mathbf{t}(i),\mathbf{t}(j)} - p$.

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* $\mathbf{p} \in \mathbb{R}^{|I \cup J|}$ associated with a matching μ is a vector that specifies a wage payment $\mathbf{p}_{i,\mu(i)} \in \mathbb{R}$ to each worker i, and a wage payment $\mathbf{p}_{\mu(j),j} \in \mathbb{R}$ from 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 allocation is publicly observable.

Beyond the public information that the agents' type assignment belongs to T, the agents may also have their 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 cell 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., $\Pi := (\{\Pi_i\}_{i \in I}, \{\Pi_j\}_{j \in J})$, 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 $\mathbf{t} \in T$.

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

Finally, we introduce the following standard notions from game theory (see, e.g., Osborne and Rubinstein (1994)). An *event* is a subset of T. Say agent k *knows* an event $E \subset T$ at $\mathbf{t}$ if $\Pi_k(\mathbf{t}) \subset E$. Given a partition profile Π, each agent

k has a *knowledge function* $\mathcal{K}_{k,\Pi}(\cdot)$ defined by

$$\mathcal{K}_{k,\Pi}(E) := \{\mathbf{t} \in T : \Pi_k(\mathbf{t}) \subset E\}.$$

An event $E' \subset T$ is *self-evident* among a set of agents $K \subset I \cup J$ if for all $\mathbf{t} \in E'$ we have $\Pi_k(\mathbf{t}) \subset E'$ for all $k \in K$. An event E is said to be *common knowledge* among a set of agents $K \subset I \cup J$ at $\mathbf{t}$ if there is a self-evident event E' among K for which $\mathbf{t} \in E' \subset E$.

We add a few remarks on the framework as a model for matching with incomplete information. First, the model apparently covers the classical complete information setup. 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. Second, the model also covers the setup of one-sided incomplete information analyzed in Liu et al. (2014), Chen and Hu (2020) and Pomatto (2020). Formally, these papers analyze the situation where the firms' types are common knowledge, i.e., $\mathbf{f}' = \mathbf{f}$ for all $(\mathbf{w}, \mathbf{f}), (\mathbf{w}', \mathbf{f}') \in T$. Third, unlike Liu et al. (2014) and Chen and Hu (2020), the model accommodates but makes no presumption about the agents' knowledge about their own types or their partners' types. For instance, if we assume that agents know their partners' types, then for each state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ with $\mu(i) = j$ and $\mathbf{t} = (\mathbf{w}, \mathbf{f})$, we have $\mathbf{w}'(i) = \mathbf{w}(i)$ for all $(\mathbf{w}', \mathbf{f}') \in \Pi_j((\mathbf{w}, \mathbf{f}))$ and $\mathbf{f}'(j) = \mathbf{f}(j)$ for all $(\mathbf{w}', \mathbf{f}') \in \Pi_i((\mathbf{w}, \mathbf{f}))$.

None of our results will be affected by any of these aforementioned informational/observability assumptions, namely that our results hold regardless of Π. Finally, our model also lends itself to a transparent epistemic description for the blocking/stability notion which we will demonstrate in the next section.

3 Stability: a leading notion

3.1 Individual rationality

A state is individually rational if for every agent, it is possible, based on the agent's information partition, to have a non-negative payoff.

Definition 1 *A state* $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ *is* ***individually rational*** *if for each* $i \in I$ *and each* $j \in J$*, respectively,*

$$\nu_{\mathbf{t}'(i),\mathbf{t}'(\mu(i))} + \mathbf{p}_{i,\mu(i)} \geq 0 \text{ for some } \mathbf{t}' \in \Pi_i(\mathbf{t}) \text{ and} \tag{1}$$

$$\phi_{\mathbf{t}'(\mu(j)),\mathbf{t}'(j)} - \mathbf{p}_{\mu(j),j} \geq 0 \text{ for some } \mathbf{t}' \in \Pi_j(\mathbf{t}). \tag{2}$$

Individual rationality has a natural epistemic foundation:

Fact 1 *A state* $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ *is individually rational if and only if at* $\mathbf{t}$*, no agent knows the event that s/he has a negative payoff, i.e., for all* $i \in I$ *and all* $j \in J$*,*

$$\mathbf{t} \notin \mathcal{K}_{i,\Pi}\left(\left\{\mathbf{t}' \in T : \nu_{\mathbf{t}'(i),\mathbf{t}'(\mu(i))} + \mathbf{p}_{i,\mu(i)} < 0\right\}\right) \text{ and}$$

$$\mathbf{t} \notin \mathcal{K}_{j,\Pi}\left(\left\{\mathbf{t}' \in T : \phi_{\mathbf{t}'(\mu(j)),\mathbf{t}'(j)} - \mathbf{p}_{\mu(j),j} < 0\right\}\right).$$

In Definition 1, a matched agent is individually rational as long as one of the partner's types gives the agent a non-negative payoff. That is, she is pessimistic about the payoff change from leaving the current partner. As a result, individual rationality here is a permissive notion. To see this feature, consider an example with only one worker and one firm, and no transfer.[9] The firm has no private information. The worker has a good type that produces an output of 2 for the firm and a bad type that causes a loss of 2, which the firm cannot distinguish. The worker's matching payoff is always 1 for himself regardless of his type. Assume the payoff from being unmatched is 0. In this situation, without further assumptions, Definition 1 allows for both the state of "matched" and the state of "unmatched" being individually rational. By Fact 1, this is because in either state, the firm does not know that changing the status quo leads to a payoff improvement.[10]

Liu et al. (2014) do not run into the issue of the example, as they assume mutual observability of the agents' types in a match. With the assumption, their individual rationality notion becomes the ex post notion. Alternatively, Liu (2020) does not require observability of partner's type but instead formulates a

[9]We are grateful to an anonymous referee for suggesting this example to illustrate the issue.

[10]Observe that here the lack of knowledge is independent of the firm's probabilistic belief which is outside our model.

Bayesian/belief-based framework to consistently assign on-path and off-path beliefs of stable matchings. To sidestep the issue in our belief-free setting, we could also follow Liu et al. (2014) by assuming the mutual observability of the agents' types in a match.[11] Here we opt for the permissive notion of Definition 1 to maintain its flexibility as well as consistency with the blocking notion that we are about to introduce in the next subsection.

3.2 Blocking

3.2.1 Motivation of consideration

With complete information, a matching is blocked if there exist a worker and a firm such that both agents can benefit from being rematched with each other at some wage. To motivate our notion of blocking under incomplete information, consider a potential rematching $(i, j; p)$ for the state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$. Observe that in evaluating the rematching with worker i at wage p, firm j will not consider the type assignments in $\Pi_j(\mathbf{t})$ at which worker i *must not* be willing to participate in the rematching with firm j. That is, firm j can rule out from her consideration any type assignment $\mathbf{t}'$ at which worker i knows that he will not benefit from being rematched with firm j with wage p. This is the key idea to define blocking when there is one-sided (worker-side) incomplete information as in Liu et al. (2014) and Chen and Hu (2020).

When the firm-side information is also incomplete, similarly, worker i will not consider type assignments in $\Pi_i(\mathbf{t})$ at which firm j must not be willing to participate in the rematching with worker i. Furthermore, with two-sided incomplete information, once the pair of agents take into account each other's ruling out some type assignments from their consideration, they may be able to rule out some more type assignments from their consideration, as we illustrate in the following example.

Example 1 Consider a market with two workers, i.e., $I = \{\alpha, \beta\}$, and two firms,

[11] As will become clear, all of our results hold regardless of whether we impose this assumption or not.

i.e., $J = \{a, b\}$. The set of possible type assignments is given by $T = \{\mathbf{t}, \mathbf{t}', \mathbf{t}''\}$, i.e.,

$$
\begin{array}{cccc}
 & \mathbf{t} & \mathbf{t}' & \mathbf{t}'' \\
\alpha & 2 & 2 & 2 \\
\beta & 3 & 1 & 1 \\
a & 3 & 3 & 5 \\
b & 2 & 4 & 4.
\end{array}
$$

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.

We first describe a state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$. Suppose that firm a hires worker α and firm b hires worker β. In other words, a matching μ is given by $\mu(\alpha) = a$ and $\mu(\beta) = b$. Assume for simplicity that $\mathbf{p} = \mathbf{0}$. Suppose further that the first type assignment $\mathbf{t}$ is true, and that agents' partitions are given as if agents can observe their own types and their own partner's types, i.e.,

$$
\begin{aligned}
\Pi_\alpha &= \{\{\mathbf{t}, \quad \mathbf{t}'\}, \{\mathbf{t}''\}\}, \\
\Pi_\beta &= \{\{\mathbf{t}\}, \{\mathbf{t}', \quad \mathbf{t}''\}\}, \\
\Pi_a &= \{\{\mathbf{t}, \quad \mathbf{t}'\}, \{\mathbf{t}''\}\}, \text{ and} \\
\Pi_b &= \{\{\mathbf{t}\}, \{\mathbf{t}', \quad \mathbf{t}''\}\}.
\end{aligned}
$$

It is straightforward to verify that if there is complete information at $\mathbf{t}$, then β and a will both benefit from being rematched with each other at wage 0. With incomplete information associated Π, however, firm a would be worried about the possible type assignment $\mathbf{t}'$. To be precise, for the possible type 1 of β, firm a would obtain a lower payoff 3 than her status quo payoff 6.

However, we proceed to argue that the potential rematching $(\beta, a; 0)$ should block the state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$. First of all, worker β is willing to participate in the rematching $(\beta, a; 0)$ because he knows that the true type assignment is $\mathbf{t}$, at which he will get a higher payoff if rematched with firm a. Then it suffices to check firm a's willingness. Firm a only worries about the possible type assignment $\mathbf{t}'$. Nevertheless, we argue that she does not need to consider $\mathbf{t}'$. More precisely, she knows that if the true type assignment is $\mathbf{t}'$, then worker β regards both $\mathbf{t}'$ and $\mathbf{t}''$ as possible. Note that worker β would not consider $\mathbf{t}''$ if the true type assignment were $\mathbf{t}'$, because firm a would definitely object to the rematching $(\beta, a; 0)$ if $\mathbf{t}''$ were the true type profile (i.e., there is no possibility for firm a to obtain a higher payoff). Hence, worker β would consider only $\mathbf{t}'$, at which he would object to the rematching $(\beta, a; 0)$. That is, worker β would definitely object to the rematching $(\beta, a; 0)$ should $\mathbf{t}'$ be the true type profile (i.e., there is no possibility under consideration for worker β to obtain a higher payoff). Therefore, firm a does not need to consider $\mathbf{t}'$. Considering only $\mathbf{t}$, firm a is willing to participate in the rematching $(\beta, a; 0)$. Therefore, $(\beta, a; 0)$ should block the state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$.

3.2.2 Consideration and blocking

To formalize the iterative reasoning in Example 1, we introduce a notion called *consideration correspondence*. To simplify notations, we fix a state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ and a potential rematching $(i, j; p)$, and hereafter we omit the reference of the blocking notion to $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ and $(i, j; p)$. For each $\mathbf{t}' \in T$, we define the indicators $\chi(\mathbf{t}')$ for both agents as follows.

$$\chi_i(\mathbf{t}') := \begin{cases} Y & \text{if } \nu_{\mathbf{t}'(i),\mathbf{t}'(j)} + p > \nu_{\mathbf{t}'(i),\mathbf{t}'(\mu(i))} + \mathbf{p}_{i,\mu(i)}, \\ N & \text{otherwise;} \end{cases} \tag{3}$$

$$\chi_j(\mathbf{t}') := \begin{cases} Y & \text{if } \phi_{\mathbf{t}'(i),\mathbf{t}'(j)} - p > \phi_{\mathbf{t}'(\mu(j)),\mathbf{t}'(j)} - \mathbf{p}_{\mu(j),j}, \\ N & \text{otherwise.} \end{cases} \tag{4}$$

For example, $\chi_i(\mathbf{t}') = Y$ (Yes) means that worker i is willing to participate in the blocking $(i, j; p)$ *if* he knows that the true type profile is $\mathbf{t}'$. However, since worker i may not know whether $\mathbf{t}'$ is true, it should be clear that such worker i's willingness is only hypothetical. This kind of hypothetical willingness/unwillingness constitutes the basis of our consideration refinement. We now define the notion of consideration correspondence as follows.

Definition 2 *Set $C^0 = \Pi$. For every $l \geq 1$ and every $\mathbf{t}' \in T$, define the **round-l consideration correspondences***

$$C_i^l(\mathbf{t}') := \left\{\mathbf{t}'' \in \Pi_i(\mathbf{t}') : \exists \mathbf{t}''' \in C_j^{l-1}(\mathbf{t}'') \text{ s.t. } \chi_j(\mathbf{t}''') = Y\right\} \text{ and} \tag{5}$$

$$C_j^l(\mathbf{t}') := \left\{\mathbf{t}'' \in \Pi_j(\mathbf{t}') : \exists \mathbf{t}''' \in C_i^{l-1}(\mathbf{t}'') \text{ s.t. } \chi_i(\mathbf{t}''') = Y\right\}. \tag{6}$$

*The **(limit) consideration correspondences** for the potential rematching $(i, j; p)$ at $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ are defined as $C_i^\infty(\mathbf{t}') = \cap_{l=1}^\infty C_i^l(\mathbf{t}')$ and $C_j^\infty(\mathbf{t}') = \cap_{l=1}^\infty C_j^l(\mathbf{t}')$ for each $\mathbf{t}' \in T$. We also call $C_k^\infty(\mathbf{t}')$ the **consideration set** of agent k at $\mathbf{t}'$.*

Two remarks on the consideration correspondences are in order. First, (5) and (6) mean that the worker i and firm j in the potential rematching only consider type assignments such that their potential partner "may be" willing to join the rematching, i.e., type assignments at which there is at least some "Yes" in their potential partner's consideration set from the previous round. Equivalently, agent k rules out a type assignment from her consideration in round l if and only if agent $-k$ have confirmed his unwillingness at the type assignment in round $l-1$. The confirmation of unwillingness means that agent $-k$'s consideration set in round $l-1$ either consists of uniformly "No" or has become empty.

Second, the following lemma verifies that the consideration correspondences are monotonically decreasing in l. The monotonicity has two implications. First, it does not matter whether we require $\mathbf{t}'' \in \Pi_k(\mathbf{t}')$ or $\mathbf{t}'' \in C_k^{l-1}(\mathbf{t}')$ in (5) and (6). Second, since T is finite, there is some l^* such that $C_i^l(\mathbf{t}') = C_i^\infty(\mathbf{t}')$ and $C_j^l(\mathbf{t}') = C_j^\infty(\mathbf{t}')$ for all $l \geq l^*$ and all $\mathbf{t}' \in T$. Note that $C_k^\infty(\mathbf{t}')$ might be empty.

Lemma 1 *For $k = i, j$ and each $\mathbf{t}' \in T$, $C_k^l(\mathbf{t}')$ is decreasing in l w.r.t. set inclusion.*

Example 1 (Revisited) Incorporating the information conveyed by indicators, we can rewrite the partitions of worker β and firm a as partitions over lists of indicators, i.e.,

$$\begin{array}{rllll} & & \mathbf{t} & \mathbf{t}' & \mathbf{t}'' \\ \Pi_\beta = & \{ & \{Y\}, & \{N & Y\}\} \\ \Pi_a = & \{ & \{Y & N\}, & \{N\}\}. \end{array}$$

Worker β's and firm a's round-1 consideration correspondences are, respectively,

$$\begin{array}{cccc} & \mathbf{t} & \mathbf{t}' & \mathbf{t}'' \\ C^1_\beta: & \{\mathbf{t}\} & \{\mathbf{t}'\} & \{\mathbf{t}'\} \\ C^1_a: & \{\mathbf{t},\mathbf{t}'\} & \{\mathbf{t},\mathbf{t}'\} & \{\mathbf{t}''\} \end{array}$$

We can represent such correspondences by deleting Y or N from the partitions over the lists of indicators, which facilitates the comparison with partitions:

$$\begin{array}{rcccc} & \mathbf{t} & \mathbf{t}' & \mathbf{t}'' \\ C^1_\beta = & \{\{Y\}, & \{N & \}\} \\ C^1_a = & \{\{Y & N\}, & \{N\}\}. \end{array}$$

In similar ways, worker β's and firm a's round-2 consideration correspondences can be written as follows, respectively:

$$\begin{array}{rcccc} & \mathbf{t} & \mathbf{t}' & \mathbf{t}'' \\ C^2_\beta = & \{\{Y\}, & \{N & \}\} \\ C^2_a = & \{\{Y & \}, & \{\ \}\}. \end{array}$$

We say a state is blocked by a combination $(i, j; p)$ if both agents i and j could have higher rematching payoffs under every type assignment that is considered at the true one.

Definition 3 *A state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is **blocked** by $(i, j; p)$ if $C^\infty_i(\mathbf{t}) \neq \emptyset$, $C^\infty_j(\mathbf{t}) \neq \emptyset$ and*

$$\nu_{\mathbf{t}'(i),\mathbf{t}'(j)} + p > \nu_{\mathbf{t}'(i),\mathbf{t}'(\mu(i))} + \mathbf{p}_{i,\mu(i)} \text{ for all } \mathbf{t}' \in C^\infty_i(\mathbf{t}) \text{ and} \tag{7}$$

$$\phi_{\mathbf{t}'(i),\mathbf{t}'(j)} - p > \phi_{\mathbf{t}'(\mu(j)),\mathbf{t}'(j)} - \mathbf{p}_{\mu(j),j} \text{ for all } \mathbf{t}' \in C^\infty_j(\mathbf{t}). \tag{8}$$

*In this case, we say $(i, j; p)$ is a **blocking combination** for $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$. Say a state is **unblocked** if it is not blocked by any combination (i, j, p).*

As a special case, if $C^\infty_i(\mathbf{t}) = \{\mathbf{t}\}$ and $C^\infty_j(\mathbf{t}) = \{\mathbf{t}\}$, then (7)-(8) reduce to the standard conditions for complete-information blocking. In other words, the limit consideration correspondences pin down when our notion of blocking under incomplete information reduces to complete-information blocking. Definition 3 also extends the definitions of blocking in Liu et al. (2014) and Chen and Hu (2020) to two-sided incomplete information while remaining "belief-free" in two important aspects. First, a type assignment is ruled out from an agent's consideration set if and only if the potential partner entertains *no* possibility of joining the blocking combination, i.e., the potential partner's consideration is either empty or consists exclusively of "No", regardless of his belief. Second, two agents form a blocking combination if and only if they secure strict improvement in all possibilities which they deem consistent with their consideration sets, regardless of their beliefs.

3.2.3 Epistemic interpretation of blocking

Our framework also lends itself to an epistemic characterization of the blocking notion in Definition 3 which we provide here. The epistemic characterization yields a descriptive formulation of the blocking notion in terms of the agents' information and knowledge.

Fix a state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ and a potential blocking combination $(i, j; p)$. We define B_k as the event in which agent k benefits from being rematched with her or his potential partner, i.e.,

$$B_i := \left\{\mathbf{t}' \in T : \nu_{\mathbf{t}'(i),\mathbf{t}'(j)} + p > \nu_{\mathbf{t}'(i),\mathbf{t}'(\mu(i))} + \mathbf{p}_{i,\mu(i)}\right\} \text{ and}$$
$$B_j := \left\{\mathbf{t}' \in T : \phi_{\mathbf{t}'(i),\mathbf{t}'(j)} - p > \phi_{\mathbf{t}'(\mu(j)),\mathbf{t}'(j)} - \mathbf{p}_{\mu(j),j}\right\}.$$

If firm j knows that B_j does not occur, she knows that she will not benefit from forming the rematching with worker i at wage p and hence will not join the blocking combination. That is, conditional on the blocking combination being formed, worker i could safely ignore any type assignment at which firm j knows that B_j does not occur. This refinement can be formulated with worker i hypothetically updating his partition with the hypothetical information represented by the binary partition $\{\mathcal{K}_{j,\Pi_j}\left(B_j^c\right), T \setminus \mathcal{K}_{j,\Pi_j}\left(B_j^c\right)\}$, namely that

$$\mathcal{U}_i\left(\Pi_i, \Pi_j\right) := \Pi_i \vee \{\mathcal{K}_{j,\Pi_j}\left(B_j^c\right), T \setminus \mathcal{K}_{j,\Pi_j}\left(B_j^c\right)\}.\text{[12]} \tag{9}$$

Likewise, conditional on the blocking combination $(i, j; p)$ being formed, firm j can

[12] Terminologies and notations: for an arbitrary subset B of T, the *complement* of B is denoted by B^c, i.e., $B^c = T \setminus B$. The *join* of two partitions is the coarsest common refinement of them, where the operator is denoted by $\vee$. See Aumann (1976). Analogously, the *meet* of two partitions is the finest common coarsening of them, where the operator is denoted by $\wedge$.

also hypothetically update Π_j to

$$\mathcal{U}_j(\Pi_i, \Pi_j) := \Pi_j \vee \{\mathcal{K}_{i,\Pi_i}(B_i^c), T \setminus \mathcal{K}_{i,\Pi_i}(B_i^c)\}. \tag{10}$$

We write $\mathcal{K}_{j,\Pi_j}$ and $\mathcal{K}_{i,\Pi_i}$ instead of $\mathcal{K}_{j,\Pi}$ and $\mathcal{K}_{i,\Pi}$ because these operators describe hypothetical belief updating of only Π_i and Π_j but not other agents' partitions.

For $k = i, j$, let $\Pi_k^{l+1} \equiv \mathcal{U}_k^{l+1}(\Pi_i, \Pi_j) := \mathcal{U}_k\left(\mathcal{U}_i^l(\Pi_i, \Pi_j), \mathcal{U}_j^l(\Pi_i, \Pi_j)\right)$ for each $l \geq 1$. Since T is finite and (Π_i^l, Π_j^l) becomes finer and finer, there must be an integer l^* such that

$$(\Pi_i^l, \Pi_j^l) = (\Pi_i^\infty, \Pi_j^\infty) \text{ for every } l \geq l^*.$$

We call $(\Pi_i^\infty, \Pi_j^\infty)$ the *hypothetical information structure* for the blocking combination $(i, j; p)$. The information structure is hypothetical as the blocking combination needs not be formed.

Volij (2000) defines a notion of core under incomplete information which incorporates endogenous communication of willingness to trade among the agents. While the definition of hypothetical information structure may appear similar to the definition of core in Volij (2000), it does not require actual communication between worker i and firm j about their willingness to join the blocking combination. More precisely, the "information updating" embodied in (9) and (10) builds only upon the "common knowledge" that each agent will rule out the type assignment according to which the other agent never considers joining the blocking combination. Such "common knowledge" must be formulated via refining the agents' information partitions as per (9) and (10) to reflect the type assignments which the agents rule out. The formulation thus makes the hypothetical information structure similar to the information structure resulted from iterative communication of willingness to join the blocking combination.

The following proposition provides an epistemic characterization of blocking in Definition 3:

Proposition 1 *A state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is blocked by $(i, j; p)$ if and only if under the hypothetical information structure, both i and j know at $\mathbf{t}$ that they will benefit from being rematched, i.e.,* $\mathbf{t} \in \mathcal{K}_{i,\Pi_i^\infty}(B_i) \cap \mathcal{K}_{j,\Pi_j^\infty}(B_j)$.

We may compare Proposition 1 with Fact 1. Indeed, individual rationality can be viewed as absence of the blocking combination $(k, k, 0)$. Since the agent learns nothing about her/himself from her/his willingness of staying unmatched, the hypothetical information structure must be identical to the initial information structure Π.

3.3 Stability

Like the notion of stability with complete information, the notion of stability which we are about to propose also requires individual rationality and the absence of blocking. Unlike the notion of stability with complete information, however, our stability must also embody a notion called information stability. With one-side incomplete information, the notion of information stability is originally formulated in Liu et al. (2014) and reformulated in Chen and Hu (2020) in terms of partitional information structure of the firms. Here we extend information stability to a two-sided incomplete-information environment based upon the novel blocking notion in Definition 3.

Define a set of type assignments as follows:

$$N_{\mu,\mathbf{p},\Pi} := \{\mathbf{t} \in T : (\mu, \mathbf{p}, \mathbf{t}, \Pi) \text{ is individually rational and unblocked}\}.$$

Intuitively, by the public information $(\mu, \mathbf{p}, \Pi)$ and the absence of blocking, agents know that the true type assignment lies in $N_{\mu,\mathbf{p},\Pi}$. Let $\mathcal{CK}_\Pi$ denote the meet (i.e., finest common coarsening) of the partition profile Π. Then, given a state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$, the set $\mathcal{CK}_\Pi(\mathbf{t})$ is the cell of the meet of Π that contains the true type assignment $\mathbf{t}$. Upon observing the absence of blocking, each agent can refine their partitions within $\mathcal{CK}_\Pi(\mathbf{t})$. In contrast to the hypothetical information updating in Section 3.2.3, the information refinement here is based upon the *de facto* absence of blocking.

We now formally define an operator $H_{\mu,\mathbf{p}}(\cdot)$ to represent the information refinement. 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., $\mathcal{N}_{\mu,\mathbf{p},\Pi} := \{N_{\mu,\mathbf{p},\Pi}, T \setminus N_{\mu,\mathbf{p},\Pi}\}$.

$$[H_{\mu,\mathbf{p}}(\Pi)]_j(\mathbf{t}') := \begin{cases} \Pi_j(\mathbf{t}') \cap \mathcal{N}_{\mu,\mathbf{p},\Pi}(\mathbf{t}'), & \text{if } \mathbf{t}' \in \mathcal{CK}_\Pi(\mathbf{t}); \\ \Pi_j(\mathbf{t}'), & \text{otherwise.} \end{cases} \tag{11}$$

A state is said to be stable if it is individually rational and unblocked, and moreover, no further information can be inferred from individual rationality and the absence of blocking.

Definition 4 *A state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is **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) $H_{\mu,\mathbf{p}}(\Pi) = \Pi$.

We provide an epistemic foundation for our notion of stability. In particular, stability is characterized by the common knowledge of individual rationality and no blocking.

Proposition 2 *A state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is stable if and only if at $\mathbf{t}$, it is common knowledge that the state is individually rational and not blocked.*

We close this section by the existence of stable states.

Theorem 1 *The set of stable states is nonempty. In particular, for any $\mathbf{t} \in T$, let $(\mu, \mathbf{p})$ be a complete-information stable allocation and Π be an arbitrary information partition profile. Then $(\mu, \mathbf{p}, \mathbf{t}, H^\infty_{\mu,\mathbf{p}}(\Pi))$ is stable, where $H^l_{\mu,\mathbf{p}}(\cdot)$ is the l-iteration of $H_{\mu,\mathbf{p}}(\cdot)$.*

The result says that every complete-information stable allocation is associated with at least one stable state via the informationally stable partition $H^\infty_{\mu,\mathbf{p}}(\Pi)$. The existence of complete-information stable allocation has been obtained by Shapley and Shubik (1971) and Crawford and Knoer (1981). Moreover, Theorem 1 also extends the existence result in Liu et al. (2014) and Chen and Hu (2020) to a setting with two-sided incomplete information. The proof of Theorem 1 will be presented in Appendix A.2.

4 Properties of blocking and stable states

In this section, we introduce two properties of blocking, i.e., improvement at the truth and information monotonicity. Both of them are specific to the incomplete-information environment and useful in understanding stability with incomplete information.

4.1 Improvement at the truth

Fix a blocking combination $(i, j; p)$ for the state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$. First, we are interested in the essence of blocking—an opportunity to obtain higher payoffs through rematching. Will the two agents i and j get *de facto* higher payoffs? If the answer is yes, then we say blocking satisfies improvement at the truth. This is arguably the most basic desideratum behind any blocking notion in matching theory. We formalize the property as follows:

Improvement at the Truth (IT). If $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is blocked by $(i, j; p)$, then $\chi_i(\mathbf{t}) = Y$ and $\chi_j(\mathbf{t}) = Y$.

We know from the definition of consideration correspondence that at the true type assignment $\mathbf{t}$, only a subset of $\Pi_k(\mathbf{t})$ is considered. If IT holds, then $\mathbf{t}$ is considered at $\mathbf{t}$ by both i and j. More importantly, if IT is satisfied, then blocking with incomplete information implies blocking with complete information. In other words, the set of incomplete-information blocking combinations is a subset of complete-information blocking combinations.

Improvement at the truth holds in the one-sided incomplete-information setting such as Liu et al. (2014) and Chen and Hu (2020). In our more general setup, instead of proving IT directly, we will establish a stronger property in the next subsection which implies IT.

4.2 Information monotonicity

We now introduce the second property of blocking as follows:

Information Monotonicity (IM). If $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is blocked by $(i, j; p)$ and $\hat{\Pi}$ is a finer partition profile than Π, then $(\mu, \mathbf{p}, \mathbf{t}, \hat{\Pi})$ is also blocked by $(i, j; p)$.

Clearly, IM is stronger than IT. More precisely, when $\hat{\Pi}$ is the complete-information partition profile, $(\mu, \mathbf{p}, \mathbf{t}, \hat{\Pi})$ is blocked by $(i, j; p)$ if and only if $\chi_i(\mathbf{t}) = Y$ and $\chi_j(\mathbf{t}) = Y$, which is exactly the definition of IT. Intuitively, IM means that when agents have more precise information, it is easier for them to find blocking opportunities.[13] Indeed, the blocking notion in Definition 3 satisfies IM and, thus, IT.

Theorem 2 *The blocking notion in Definition 3 satisfies IM and IT.*

IM is a generalization of Fact 1 in Chen and Hu (2020).

4.3 Stable states

Equipped with the properties of blocking, we are now ready to investigate properties of stable states. We first derive an analogy to the IM property of blocking. In particular, we show that if $(\mu, \mathbf{p}, \mathbf{t}, \hat{\Pi})$ is stable and $\hat{\Pi}$ is finer than another partition profile Π, then the state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is "essentially" stable in the sense that it will become stable under the refined partition profile $H^{\infty}_{\mu,\mathbf{p}}(\Pi)$. The formal result is as follows.

Proposition 3 *Suppose that $\hat{\Pi}$ is a finer partition profile than Π. If a state $(\mu, \mathbf{p}, \mathbf{t}, \hat{\Pi})$ is stable, then the state $(\mu, \mathbf{p}, \mathbf{t}, H^{\infty}_{\mu,\mathbf{p}}(\Pi))$ is stable.*

The following theorem says that every matching outcome (i.e., an allocation and a true type profile) which arises from a stable state can also be made a stable state via a particular partition profile.

Theorem 3 *Let $\bar{\Pi}$ be the trivial partition profile over T such that agents have the least information, i.e., $\bar{\Pi}_k := \{T\}$ for all $k \in I \cup J$. For an arbitrary matching outcome $(\mu, \mathbf{p}, \mathbf{t})$, the state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is stable for some Π if and only if the state $(\mu, \mathbf{p}, \mathbf{t}, H^{\infty}_{\mu,\mathbf{p}}(\bar{\Pi}))$ is stable.*

[13] Another reason for introducing two separating properties is that IM is not satisfied for some variant blocking notions but IT is satisfied. See Section 5.3.

The sufficiency part is trivial. The necessity part follows from Proposition 3. This result shows that despite the great variety of incomplete-information situations, the entire set of stable allocations across all information structures can be identified via the set of stable allocations under a specific class of information structures obtained from $\bar{\Pi}$.

5 Alternative blocking and stability notions

In this section, we discuss two alternative directions of formulating blocking. In comparison with the blocking notion in Definition 3, the first direction explores situations where the agents are more conservative in forming a blocking pair, whereas the second direction explores situations where the agents are more aggressive in forming a blocking pair. An extreme case in the first direction leads to a so-called naive blocking notion, in which agents' consideration sets conservatively contain all possible type assignments according to their information partitions. Between blocking (level-∞) and naive blocking (level-0), we also consider level-l blocking which performs consideration refinement for l rounds with $1 \leq l < \infty$. In the second direction, we argue that there are type assignments which the agents can also rule out, if we impose additional assumptions on consideration refinement. This leads to what we call a sophisticated blocking notion. Each of these blocking notions, together with the corresponding individual rationality and information stability conditions, corresponds to a different stability notion. We summarize the existence and properties of those stability notions in Section 5.3.

5.1 Level-l blocking/stability

In Section 3.2.2 we define consideration refinement as an iterative process. Reaching the limit of the iteration process may require an arbitrarily large number of rounds and thus a high degree of reasoning complexity. Here we say that a state is *level-l blocked* if there exist a firm, a worker, and a potential wage such that with the rematching of them, both agents would receive a higher payoff under all type assignments which they consider after l rounds of consideration refinement. Formally, recall that C_i^l and C_j^l are round-l consideration correspondences defined in Definition 2.

Definition 5 *A state* $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ *is* ***level-l blocked*** *by* $(i, j; p)$ *if* $C_i^l(\mathbf{t}) \neq \emptyset$, $C_j^l(\mathbf{t}) \neq \emptyset$ *and*

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

Obviously, when $l = \infty$, we have our blocking notion in Definition 3. When $l = 0$, agents evaluate rematching payoffs simply based on $C_i^0 = \Pi_i$ and $C_j^0 = \Pi_j$. We also call level-0 blocking *naive blocking* since agents do not perform consideration refinement at all. This level-l formulation is reminiscent of the level-k model proposed by Stahl and Wilson (1994, 1995) and Nagel (1995), and to our knowledge, Definition 5 is the first blocking notion which incorporates level-l reasoning.

We know by Lemma 1 that if a state is level-l blocked by some worker-firm pair, then it is level-$(l+1)$ blocked by the same pair. However, the converse is not true for any l. To see this, we first notice that in our Example 1 in Section 3.2, the state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is not level-1 blocked by $(\beta, a; 0)$ but it is level-2 blocked by $(\beta, a; 0)$. More generally, for any l, we can construct a similar example in which level-$(l+1)$ blocking strictly refines level-l blocking.

Example 2 Consider a market with only one worker β and one firm a. We assume away transfers throughout the example. There are $l+2$ type assignments with $l \geq 1$ being odd and $\mathbf{t}^1$ being true (the example for an even l is similar):

	$\mathbf{t}^1$	$\mathbf{t}^2$	$\mathbf{t}^3$	$\mathbf{t}^4$	$\mathbf{t}^5$	$\cdots$	$\mathbf{t}^{l-1}$	$\mathbf{t}^l$	$\mathbf{t}^{l+1}$	$\mathbf{t}^{l+2}$
$\mathbf{w}(\beta)$:	$\frac{1}{2}$	$-\frac{1}{4}$	$\frac{1}{4}$	$-\frac{15}{16}$	$\frac{1}{16}$	$\cdots$	$-\frac{2^{l-1}-1}{2^{l-1}}$	$\frac{1}{2^{l-1}}$	$-\frac{2^{l+1}-1}{2^{l+1}}$	$\frac{1}{2^{l+1}}$
$\mathbf{f}(a)$:	$\frac{1}{2}$	$-\frac{1}{2}$	$-\frac{7}{8}$	$\frac{1}{8}$	$-\frac{31}{32}$	$\cdots$	$\frac{1}{2^{l-2}}$	$-\frac{2^l-1}{2^l}$	$\frac{1}{2^l}$	$-\frac{2^{l+2}-1}{2^{l+2}}$.

The premuneration value function for the worker is given by $\nu_{w,f} = wf + w$, and for the firm by $\phi_{w,f} = wf + f$.

Suppose β and a are unmatched. To see agents' willingness/objection to joining a match, we summarize the premuneration values under a match as follows:

	$\mathbf{t}^1$	$\mathbf{t}^2$	$\mathbf{t}^3$	$\mathbf{t}^4$	$\mathbf{t}^5$	$\cdots$	$\mathbf{t}^{l-1}$	$\mathbf{t}^l$	$\mathbf{t}^{l+1}$	$\mathbf{t}^{l+2}$
$\nu_{\mathbf{w}(\beta),\mathbf{f}(a)}$:	$\frac{3}{4}$	$-\frac{1}{8}$	$\frac{1}{32}$	$-\frac{135}{128}$	$\frac{1}{512}$	$\cdots$	$-\frac{\cdots}{2^{2l-3}}$	$\frac{1}{2^{2l-1}}$	$-\frac{\cdots}{2^{2l+1}}$	$\frac{1}{2^{2l+3}}$
$\phi_{\mathbf{w}(\beta),\mathbf{f}(a)}$:	$\frac{3}{4}$	$-\frac{3}{8}$	$-\frac{35}{32}$	$\frac{1}{128}$	$-\frac{527}{512}$	$\cdots$	$\frac{1}{2^{2l-3}}$	$-\frac{\cdots}{2^{2l-1}}$	$\frac{1}{2^{2l+1}}$	$-\frac{\cdots}{2^{2l+3}}$,

where the numerators "$\cdots$" are positive integers whose magnitudes do not matter. We specify agents' partitions over type assignments as if they can only observe the denominator of their types. As in Example 1 (Revisited), we can specify agents' indicators of being matched with each other, which, together with the partition profile, result in the following partitions over lists of indicators:

$$\begin{array}{rcccccccccc}
 & \mathbf{t}^1 & \mathbf{t}^2 & \mathbf{t}^3 & \mathbf{t}^4 & \mathbf{t}^5 & \cdots & \mathbf{t}^{l-1} & \mathbf{t}^l & \mathbf{t}^{l+1} & \mathbf{t}^{l+2} \\
\Pi_\beta = & \{\{Y\}, & \{N & Y\}, & \{N & Y\}, & \cdots, & \{N & Y\}, & \{N & Y\}\} \\
\Pi_a = & \{\{Y & N\}, & \{N & Y\}, & \cdots & \cdots & \cdots, & \{N & Y\}, & \{N\}\}.
\end{array}$$

In this case, (β, a) is a level-$(l+1)$ blocking pair but not a level-l one.

Contrary to the relationship among the blocking notions, if a state is level-$(l+1)$ stable, then it is (essentially) level-l stable.[14] Therefore, the set of level-l stable matching outcomes is decreasing in l in the set inclusion sense. Similar to blocking, Example 2 can also differentiate any two stability notions in the series. Hence, each of the level-l stability notions impose (strictly) distinct restrictions on the set of matching outcomes, with higher-level stability imposing sharper restriction.

Example 2 (Continuing) There are only two possible matchings: $\mu(\beta) = a$ (i.e., the two agents are matched) and $\mu(\beta) = \beta$ (i.e., both agents are unmatched). Given no transfer, the true type assignment $\mathbf{t}^1$ and the partition profile $(\Pi_\beta; \Pi_a)$, it is straightforward to check that both of the two possible states are level-l stable. Although the state with a match is level-$(l+1)$ stable, the state with no match is level-$(l+1)$ blocked by (β, a).

[14]The state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is essentially (level-l) stable if $(\mu, \mathbf{p}, \mathbf{t}, H^\infty_{\mu,\mathbf{p}}(\Pi))$ is (level-l) stable. We put "essentially" here because the state may not be exactly level-l stable due to the violation of information stability. But after *information updating from no blocking*, the new state with the same matching outcome must be level-l stable. See Proposition 7 in Online Appendix B.3 and its proof in Online Appendix B.4.

There are actually two layers of reasons why a higher-level stability notion imposes sharper restriction. The first reason is standard—a higher-level blocking notion admits more blocking pairs and thus excludes more states from being stable. The second reason is more subtle and manifests itself through the information-stability requirement. With a higher-level blocking notion, a state is harder to be unblocked. Therefore, the set $N_{\mu,\mathbf{p},\Pi}$ of type assignments under which the state is unblocked, as defined in Section 3.3, is smaller. As a result, the absence of blocking conveys more precise information to agents so that blocking is easier to arise once information is updated. Again, a higher-level blocking/stability notion admits more blocking pairs, but in the updated state, and thus excludes more states from being stable. We revisit Example 1 from this angle.

Example 1 (Continuing) Given μ, $\mathbf{p} = \mathbf{0}$ and Π, it is straightforward to verify that all three states $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$, $(\mu, \mathbf{p}, \mathbf{t}', \Pi)$ and $(\mu, \mathbf{p}, \mathbf{t}'', \Pi)$ are naively stable. Now we adopt level-1 stability. Obviously, the second state $(\mu, \mathbf{p}, \mathbf{t}', \Pi)$ is level-1 blocked by (α, b), which illustrates the first layer of reason. To see the second layer, note that both $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ and $(\mu, \mathbf{p}, \mathbf{t}'', \Pi)$ are level-1 unblocked. Consider the first state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$. Once agents learn $N_{\mu,\mathbf{p},\Pi} = \{\mathbf{t}, \mathbf{t}''\}$ from no blocking, information becomes complete and the updated state is level-1 blocked.

One may wonder whether the sharpest notion in the level-l series could bring us to, say, complete-information stability and have implications like efficiency. When there is only one-sided incomplete information and agents within a matched pair can observe the true type of each other, all our level-l stability notions with $l \geq 1$ reduce to that of Liu et al. (2014) through Chen and Hu (2020). Then, it follows from the result of Liu et al. (2014) that with monotonic and supermodular premuneration value functions, every stable matching outcome is efficient. However, with two-sided incomplete information, even with these assumptions on premuneration value functions, a (level-∞) stable state may not be efficient in general. To see this, we modify Example 1 in Section 3.2.1 slightly and redefine $\mathbf{t}''(\alpha) = 1/2$. Consequently, the market state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ becomes (level-∞) stable, but it is obviously not efficient.

Despite the possible inefficiency of stable matchings, since the set of level-l stable matching outcomes is decreasing in l in the set inclusion sense, for any partition profile and any true type assignment, the worst-case total surplus across all level-l stable outcomes is increasing in l. In this sense, level-l stable matching outcomes are (weakly) more efficient when l is larger.[15] Example 1 (Continuing) and Example 2 (Continuing) also show that the efficiency improvement can be strict: in the former, naive stability admits two inefficient states $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ and $(\mu, \mathbf{p}, \mathbf{t}', \Pi)$, which are excluded by level-1 stability; and in the latter, level-l stability admits an inefficient state, i.e., the no-match one, which is excluded by level-$(l+1)$ stability.

Example 2 also says that in general we need to push consideration refinement to the limit. However, we may impose some conditions on the model to reduce the number of round of reasoning. One possible direction is to consider the distance between the underlying information structure and the complete-information one, say, measured by $d_\Pi = \max_{\mathbf{t} \in T} |\mathcal{CK}_\Pi(\mathbf{t})|$. If $d_\Pi = 1$, then we have a complete-information case. Intuitively, the smaller d_Π is, the closer Π is to the complete-information partition profile, and the less round of reasoning is needed in consideration refinement.

5.1.1 Equivalence between different stability notions

Under what conditions is the level-l blocking as simple as naive blocking? Theorem 4 below provides an answer by identifying the following sufficient conditions under which all the level-l blockings are equivalent:

- **A1.** *(One-Dimensional Type)* $W \subset \mathbb{R}$ *and* $F \subset \mathbb{R}$.
- **A2.** *(Increasing Utility) The premuneration value functions* $\nu_{w,f}$ *and* $\phi_{w,f}$ *are strictly increasing in* w *and* f.

[15] The maximal total surplus across all level-l stable outcomes is invariant with l. This follows from the facts that (1) the complete-information stable outcomes, which are efficient due to Shapley and Shubik (1971), are always included in the set of (level-∞) stable outcomes according to Theorem 1 and that (2) (level-∞) stability is the sharpest solution concept among all level-l stabilities.

A3. *(Non-Transferable Utility) No transfer is permitted.*

A4. *(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 of these conditions have appeared in the literature: Conditions A1-A2 are imposed in Liu et al. (2014); Conditions A1-A3 in Bikhchandani (2017); and Condition A4 (implicitly) in Liu et al. (2014) and in Pomatto (2020).

Theorem 4 *Suppose that Conditions A1-A4 hold. Then, for any $l, l' \in \mathbb{N}$, a state is level-l blocked if and only if it is level-l' blocked; and, consequently, a state is level-l stable if and only if it is level-l' stable.*[16]

Conditions A1 and A2 are standard. Example 1 illustrates that even if A1-A3 hold simultaneously, without A4, a state can be level-2 blocked but not naively blocked. Example 4 in Appendix A.3 demonstrates the importance of A3 in establishing the equivalence, in which Conditions A1, A2 and A4 are satisfied. The examples show that in those environments even when A1 and A2 hold, it is in general necessary to distinguish different blocking/stability notions.

Note that the conditions in Theorem 4 are all independent of agents' specific partition profile. We may alternatively impose conditions on the partition profile to obtain equivalence. For instance, when all agents have the same information partition, all kinds of blocking reduce to naive blocking. And this is true independently of other primitives of the model.

5.2 Sophisticated blocking

Contrary to naive blocking, we can also think of blocking notions in which the agents are more aggressive in forming a blocking pair. In particular, we may conceive scenarios other than the one illustrated in Example 1 in which the agents

[16] Under Conditions A1-A4, all level-l blockings are also equivalent to the notion of sophisticated blocking to be introduced in the next subsection; the claim will be formally stated in Online Appendix B.3.

can refine their consideration further. With more aggressive refinement of consideration within the potential blocking pair, it is easier for a state to be blocked.

To formulate the idea, we propose two natural restrictions on general consideration correspondences, denoted by $C_i^{\dagger}(\cdot)$ and $C_j^{\dagger}(\cdot)$, for a fixed potential blocking pair (i, j).[17] The first one says that consideration should be consistent with knowledge; and the second one says that each consideration set, when it is nonempty, can be expressed as a union of cells in the join of partitions Π_i and Π_j:

$$C_k^{\dagger}(\mathbf{t}) \subset \Pi_k(\mathbf{t}) \text{ for } k = i, j \text{ and all } \mathbf{t} \in T. \tag{12}$$

$$\mathbf{t}' \in C_k^{\dagger}(\mathbf{t}) \text{ implies } [\Pi_i \vee \Pi_j](\mathbf{t}') \subset C_k^{\dagger}(\mathbf{t}) \text{ for } k = i, j \text{ and all } \mathbf{t}, \mathbf{t}' \in T. \tag{13}$$

Since the consideration correspondences are defined via the refining of agents' information partitions, the intuition for (12) is obvious. Moreover, restriction (13) says that consideration refinement, as a hypothetical information updating process, can never go finer than pooling the information of Π_i and Π_j. Clearly, (12) restricts the upper bound, whereas (13) restricts the lower bound of consideration sets. These two restrictions constitute the basis of consideration refinement for the sophisticated blocking notion which we will define.

Although simple, (12) is the assumption to justify the consideration refinement process (5)-(6). To see this, we revisit Example 1 in Section 3.2. Recall that the key driver of worker β's consideration refinement is that there exists a cell $\Pi_a(\mathbf{t}'')$ of Π_a in which firm a always objects, i.e., $\chi_a(\tilde{\mathbf{t}}) = N$ for all $\tilde{\mathbf{t}} \in \Pi_a(\mathbf{t}'')$. Therefore, as long as (12) holds, we know either that $C_a^{\dagger}(\mathbf{t}'') = \emptyset$ or that $\chi_a(\tilde{\mathbf{t}}) = N$ for all $\tilde{\mathbf{t}} \in C_a^{\dagger}(\mathbf{t}'')$. Whichever $C_a^{\dagger}(\mathbf{t}'')$ is, firm a is definitely not willing to join the blocking at $\mathbf{t}''$. Hence, worker β can safely ignore $\mathbf{t}''$. This is the *uniform* N case we have seen in Section 3.2.2.

From worker i's perspective, the uniform N case is based upon (12) and only the firm's information partition Π_j; and vice versa. If we dig further, (12) can initiate more consideration refinement when the information of both Π_j and Π_i is

[17] Recall our baseline principle is that agents' willingness to block is based on their payoff evaluation on the consideration sets; see, e.g., Definition 3. That is why we seek for natural restrictions on general consideration correspondences.

used. To see this, consider the scenario in the figure below.

$$\begin{array}{lccc} & \mathbf{t} & \mathbf{t}' & \mathbf{t}'' \\ \text{Agent } i & \{Y & N\} & \cdots \\ \text{Potential partner } j & \cdots & \{Y & Y\} \end{array}$$

Namely, Y and N co-exist in agent i's partition cell $\{\mathbf{t}, \mathbf{t}'\}$. At the type assignment $\mathbf{t}'$ where agent i's indicator is N, the potential partner's corresponding indicators are uniformly Y's. By (12), as long as $C_j^\dagger(\mathbf{t}')$ is nonempty, at that cell $\{\mathbf{t}', \mathbf{t}''\}$, the potential partner j shall always be willing to join the blocking regardless of her specific consideration set, i.e., $\chi_j(\tilde{\mathbf{t}}) = Y$ for all $\tilde{\mathbf{t}} \in C_j^\dagger(\mathbf{t}')$, whichever $C_j^\dagger(\mathbf{t}')$ is.[18] Recall our key idea that an agent rules out a type assignment from consideration only if it is safe to do so as in (5)-(6), i.e., an agent considers all type assignments under which the potential partner "may be" willing to join the blocking. Now with uniform Y, the potential partner j is "definitely" willing to join the blocking at $\mathbf{t}'$. Thus, plausibly, agent i would consider $\mathbf{t}'$ at $\{\mathbf{t}, \mathbf{t}'\}$, whichever his specific consideration set is. Then agent i would definitely say no at $\{\mathbf{t}, \mathbf{t}'\}$ because he cares about the worst-case payoff, i.e., the undetermined case should become a determined objection. Because of i's determined objection, just like in the uniform N case, the potential partner j does not need to consider either $\mathbf{t}$ or $\mathbf{t}'$. We refer to this scenario as the *supported* N case.

Now let us exemplify how the agents may refine their consideration based upon (13) and the information of both Π_j and Π_i. Consider the scenario in the figure below.

$$\begin{array}{lccc} & \mathbf{t} & \mathbf{t}' & \mathbf{t}'' \\ \text{Agent } i & \{Y & N\} & \cdots \\ \text{Potential partner } j & \{Y & N & Y\} \end{array}$$

[18] Here we assume away the case $C_j^\dagger(\mathbf{t}') = \emptyset$ to make the intuition more straightforward. An empty consideration set means that some consideration refinement has already occurred. Yet here we focus on the "starting point" to initiate consideration refinement, with the two restrictions (12)-(13) in mind. The same explanation applies to the scenario in the next paragraph. See Online Appendix B.1 for a more formal and comprehensive discussion.

Agent i's partition cell $\{\mathbf{t}, \mathbf{t}'\}$ is contained in her potential partner j's corresponding partition cell $\{\mathbf{t}, \mathbf{t}', \mathbf{t}''\}$. Then obviously $\Pi_i(\mathbf{t}) = [\Pi_i \vee \Pi_j](\mathbf{t})$. Thus, as long as (13) holds and $C_i^\dagger(\mathbf{t}) \equiv C_i^\dagger(\mathbf{t}')$ is nonempty, we have $C_i^\dagger(\mathbf{t}) \equiv C_i^\dagger(\mathbf{t}') = \{\mathbf{t}, \mathbf{t}'\}$, i.e., agent i can never distinguish Y and N when refining consideration. Since agent i cares about the worst-case payoff, this also renders the undetermined case a determined objection. Because of i's determined objection, just like in the uniform N case, the potential partner j does not need to consider either $\mathbf{t}$ or $\mathbf{t}'$. We refer to this scenario as the *inseparable* N case.

We argue in Online Appendix B.1 that uniform N's, inseparable N's and supported N's are the only three "fundamental" sources of consideration refinements which we can identify using the restrictions (12)-(13).[19] Within them, from agent i's perspective, uniform N's are of order zero in the sense that it is identified using just the information of Π_j (and restriction (12)); whereas inseparable N's and supported N's are of order one in the sense that they are identified using both the information of Π_j and the information of Π_i (and restriction (12) or (13)). Clearly, higher-order reasoning will be taken account in the iterative consideration refinement process.

Incorporating all uniform N's, supported N's and inseparable N's into the iterative definition of consideration sets will lead to, what we call, a *sophisticated blocking* notion. Intuitively, in such a situation agents are more aggressive in refining their consideration sets than they are under blocking. As a result, a blocking opportunity becomes (strictly) easier to arise, i.e., if a state is blocked, then it is sophisticatedly blocked.[20]

5.3 Properties of IT and IM for blocking notions

Based on the proof of Theorem 2, one can similarly show that every level-l blocking satisfies IM and thus IT. Unlike those level-l blockings, the sophisticated blocking satisfies IT but not IM. In the following example, IM is not satisfied because the

[19]See Appendix A.4 for examples illustrating that not all scenarios with both Y and N faced by an agent can initiate consideration refinements.

[20]See Proposition 5 in Online Appendix B.3 for a formal treatment.

consideration refinement due to an inseparable N disappears when the information partitions become finer.[21]

Example 3 Consider a potential blocking pair for a state, where there are four possible type assignments and the agents' hypothetical willingness/unwillingness is summarized in the figure below.[22]

$$
\begin{array}{lcccc}
 & \textit{Truth} & & & \\
 & \mathbf{t}^1 & \mathbf{t}^2 & \mathbf{t}^3 & \mathbf{t}^4 \\
Worker & \{Y\} & \{N & Y & Y\} \\
Firm & \{Y & N\} & \{Y & N\}
\end{array}
$$

Clearly, the firm's right partition cell, which contains $\mathbf{t}^3$ and $\mathbf{t}^4$, is a case of inseparable N. The firm knows that the true type assignment is either $\mathbf{t}^1$ or $\mathbf{t}^2$, and she worries about $\mathbf{t}^2$, under which she would not obtain a higher payoff through rematching. However, she would think that if the true type assignment is $\mathbf{t}^2$, then the worker would know his right partition cell $\{\mathbf{t}^2, \mathbf{t}^3, \mathbf{t}^4\}$, where $\mathbf{t}^3$ and $\mathbf{t}^4$ should not be considered due to the inseparable N of the firm. Therefore, the firm does not need to consider $\mathbf{t}^2$, which means that the potential blocking pair is indeed a blocking pair.

Now let us consider an alternative situation in the figure below, where only the partition profile changes.

$$
\begin{array}{lcccc}
 & \textit{Truth} & & & \\
 & \mathbf{t}^1 & \mathbf{t}^2 & \mathbf{t}^3 & \mathbf{t}^4 \\
Worker & \{Y\} & \{N & Y\} & \{Y\} \\
Firm & \{Y & N\} & \{Y\} & \{N\}
\end{array}
$$

[21]We provide another similar example in Online Appendix B.4, where the consideration refinement due to a supported N disappears when information becomes finer. As a byproduct, the original market states in those examples distinguish sophisticated blocking from blocking.

[22]One can easily come up with a comprehensive description of the market and the status quo state such that the situation is exactly what we present in the figure. But here we omit those details to only focus on crucial items.

Clearly, agents have more precise information as both agents' partitions become strictly finer. However, the worker and the firm no longer constitute a blocking pair because the firm would consider $\mathbf{t}^2$ when evaluating the rematching.

In Online Appendix B, we provide the formal definition of sophisticated blocking, the formal discussions of IT and IM properties for different blocking notions, and a formal comparison between different blocking/stability notions.

5.4 Further consideration refinement

Is it possible for agents to make further inference from the limit consideration sets? To illustrate the issue, consider the situation depicted in the figure below.

	Truth		
	$\mathbf{t}^1$	$\mathbf{t}^2$	$\mathbf{t}^3$
Worker	$\{Y$	$N\}$	$\{Y\}$
Firm	$\{Y\}$	$\{N$	$Y\}$

In this situation, there are three type assignments $\mathbf{t}^1$, $\mathbf{t}^2$ and $\mathbf{t}^3$. Obviously, there is no uniform N, inseparable N, or supported N. As a result, at $\mathbf{t}^1$ or $\mathbf{t}^3$, both agents' consideration set coincides with their partition cell and either the worker or the firm would object the rematching, whether we adopt naive blocking, blocking, or sophisticated blocking.

Now consider the following reasoning to "justify" a blocking combination for the situation. As the worker cannot distinguish $\mathbf{t}^1$ and $\mathbf{t}^2$, he may think that, given the limit consideration sets, the firm will object the rematching if the true type profile is $\mathbf{t}^2$. Therefore, the worker does not need to consider the type profile $\mathbf{t}^2$ when the true type assignment is $\mathbf{t}^1$. It follows that the worker (as well as the firm) would be willing to join the rematching at $\mathbf{t}^1$, and symmetrically, at $\mathbf{t}^3$. However, this "consideration refinement at the limit" is inconsistent. To wit, when the worker rules out $\mathbf{t}^2$ at $\mathbf{t}^1$, he must assume that the firm will object the rematching at $\mathbf{t}^2$. The basis for the latter assumption, however, is that the firm has "confirmed" the N at $\mathbf{t}^2$ and hence the decision of not participating in the rematching at type assignment $\mathbf{t}^2$ (and hence $\mathbf{t}^3$). This assumption is not warranted, if at type assignment $\mathbf{t}^3$ (and hence $\mathbf{t}^2$), the firm is also following symmetrically the reasoning of the worker at $\mathbf{t}^1$ to entertain the possibility of joining the rematching.

This example illustrates the subtlety as to why in defining sophisticated blocking, we opt to search for a "starting point" such as the three kinds of N which make no reference to how we define a consideration set and only to the basic properties (12) and (13). Actually, all blocking notions studied in the current paper are along this line—specifying a "starting point", pinning down the consideration sets through iteration and stating blocking conditions based on consideration sets, which sharply contrast the fixed-point blocking notion of Liu (2017).

6 Concluding remarks

In this paper, we provide a general framework for studying two-sided matching markets with incomplete information. We propose a stability notion as a solution concept for these matching markets and also establish its existence and epistemic foundation. We document properties of the blocking notion and stable states which are specific to the incomplete-information environment. We also examine alternative blocking/stability notions which include the two extreme cases of consideration refinement.

Stability in incomplete-information two-sided matching markets remains a largely unexplored and yet important topic in market design. One reason for this gaping hole in the literature on incomplete-information markets is that it is usually difficult to develop tractable and well-founded models for this setting. Tractable models might end up making overly simplistic assumptions on the agents, while well-founded ones might require complex reasoning which is hardly predictive.[23] Our paper attempts to bridge the two extremes with the notion of consideration refinement which enables a natural level-l reasoning and further study. It remains to be investigated whether, and to what extent, standard results on the structure

[23] We thank a referee for raising this point.

of stable matchings hold in the general incomplete-information environments.[24] There might also be structural restrictions on the set of stable states beyond IT and IM in more specific matching environments.

Appendix A Proofs and examples

A.1 Proofs of Lemma 1 and Propositions 1 and 2

Proof of Lemma 1 We prove by induction. Obviously, we have

$$C_i^1(\mathbf{t}') \subset C_i^0(\mathbf{t}') = \Pi_i(\mathbf{t}') \text{ and } C_j^1(\mathbf{t}') \subset C_j^0(\mathbf{t}') = \Pi_j(\mathbf{t}').$$

Suppose the claim of the lemma holds for step $l \geq 1$, i.e.,

$$C_i^l(\mathbf{t}') \subset C_i^{l-1}(\mathbf{t}') \text{ and } C_j^l(\mathbf{t}') \subset C_j^{l-1}(\mathbf{t}'). \tag{14}$$

We proceed to show that

$$C_i^{l+1}(\mathbf{t}') \subset C_i^l(\mathbf{t}') \text{ and } C_j^{l+1}(\mathbf{t}') \subset C_j^l(\mathbf{t}').$$

Suppose $\mathbf{t}'' \in C_i^{l+1}(\mathbf{t}')$. Then there exists $\mathbf{t}''' \in C_j^l(\mathbf{t}'')$ such that $\chi_j(\mathbf{t}''') = Y$. By (14), $C_j^l(\mathbf{t}') \subset C_j^{l-1}(\mathbf{t}')$. Therefore, $\mathbf{t}''' \in C_j^{l-1}(\mathbf{t}'')$ and $\chi_j(\mathbf{t}''') = Y$. Hence, $\mathbf{t}'' \in C_i^l(\mathbf{t}')$ and consequently $C_i^{l+1}(\mathbf{t}') \subset C_i^l(\mathbf{t}')$. The argument for $C_j^{l+1}(\mathbf{t}') \subset C_i^l(\mathbf{t}')$ is symmetric. ■

Proof of Proposition 1 We proceed to show that in either direction,

$$C_i^l(\mathbf{t}) = \Pi_i^l(\mathbf{t}) \text{ and } C_j^l(\mathbf{t}) = \Pi_j^l(\mathbf{t}) \text{ for every } l = 0, 1, 2 \ldots, \tag{15}$$

which particularly lead to $C_i^\infty(\mathbf{t}) = \Pi_i^\infty(\mathbf{t})$ and $C_j^\infty(\mathbf{t}) = \Pi_j^\infty(\mathbf{t})$. Since Π_i^∞ and Π_j^∞ are partitions, we know that $C_i^\infty(\mathbf{t}) \neq \emptyset$ and $C_j^\infty(\mathbf{t}) \neq \emptyset$ must hold if (15) holds. Obviously, (7) holds if and only if $C_i^\infty(\mathbf{t}) \subset B_i$. Similarly, (8) holds if and only if $C_j^\infty(\mathbf{t}) \subset B_j$. Therefore, $C_i^\infty(\mathbf{t}) = \Pi_i^\infty(\mathbf{t})$ would imply that $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is blocked by $(i, j; p)$ if and only if $\Pi_i^\infty(\mathbf{t}) \subset B_i$ and $\Pi_j^\infty(\mathbf{t}) \subset B_j$, which are exactly the same as $\mathbf{t} \in \mathcal{K}_{i,\Pi_i^\infty}(B_i)$ and $\mathbf{t} \in \mathcal{K}_{j,\Pi_j^\infty}(B_j)$ respectively.

To see that (15) holds, we use induction. First of all, we have

$$C_i^0(\mathbf{t}) = \Pi_i = \Pi_i^0(\mathbf{t}) \text{ and } C_j^0(\mathbf{t}) = \Pi_j = \Pi_j^0(\mathbf{t}).$$

[24]For instance, Example 2 in Chen and Hu (2020) illustrates that the celebrated Rural Hospital Theorem (c.f. Theorems 2.22 and 5.12 of Roth and Sotomayor (1990)) no longer holds with incomplete information.

Suppose (15) holds for step $l \geq 0$. Then

$$\begin{aligned}
C_i^{l+1}(\mathbf{t}) &= \left\{\mathbf{t}' \in \Pi_i(\mathbf{t}) : \exists \mathbf{t}'' \in C_j^l(\mathbf{t}') \text{ s.t. } \chi_j(\mathbf{t}'') = Y\right\} \\
&= \left\{\mathbf{t}' \in C_i^l(\mathbf{t}) : \exists \mathbf{t}'' \in C_j^l(\mathbf{t}') \text{ s.t. } \chi_j(\mathbf{t}'') = Y\right\} \\
&= \left\{\mathbf{t}' \in \Pi_i^l(\mathbf{t}) : \exists \mathbf{t}'' \in \Pi_j^l(\mathbf{t}') \text{ s.t. } \chi_j(\mathbf{t}'') = Y\right\} \\
&= \Pi_i^l(\mathbf{t}) \cap \left[T \setminus \mathcal{K}_{j,\Pi_j^l}(B_j^c)\right],
\end{aligned}$$

where the second equality follows from Lemma 1, the third equality follows from the induction hypothesis and fourth equality follows from the definition of $\mathcal{K}_{j,\Pi_j^l}$. We claim that $\mathbf{t} \in T \setminus \mathcal{K}_{j,\Pi_j^l}(B_j^c)$, which leads us to $\Pi_i^l(\mathbf{t}) \cap [T \setminus \mathcal{K}_{j,\Pi_j^l}(B_j^c)] = \Pi_i^{l+1}(\mathbf{t})$, and thus $C_i^{l+1}(\mathbf{t}) = \Pi_i^{l+1}(\mathbf{t})$. To see this, suppose $\mathbf{t} \in \mathcal{K}_{j,\Pi_j^l}(B_j^c)$. Then either $\mathbf{t} \in \mathcal{K}_{j,\Pi_j^\infty}(B_j)$ is violated in the "if" direction because $\Pi_j^l(\mathbf{t}) \subset B_j^c$ and Π_j^∞ is finer than Π_j^l, or (8) is violated in the "only-if" direction because $\Pi_j^l(\mathbf{t}) \subset B_j^c$ and $C_j^\infty(\mathbf{t}) \subset C_j^l(\mathbf{t}) = \Pi_j^l(\mathbf{t})$. The argument for $C_j^{l+1}(\mathbf{t}) = \Pi_j^{l+1}(\mathbf{t})$ is symmetric. This completes the proof of (15) and thus the proposition. ■

Proof of Proposition 2 Note that $H_{\mu,\mathbf{p}}(\Pi) = \Pi$ is equivalent to $\mathcal{CK}_\Pi(\mathbf{t}) \subset N_{\mu,\mathbf{p},\Pi}$. Then

$$\begin{aligned}
(\mu, \mathbf{p}, \mathbf{t}, \Pi) \text{ is a stable state } &\iff \mathbf{t} \in N_{\mu,\mathbf{p},\Pi} \text{ and } H_{\mu,\mathbf{p}}(\Pi) = \Pi \\
&\iff \mathbf{t} \in N_{\mu,\mathbf{p},\Pi} \text{ and } \mathcal{CK}_\Pi(\mathbf{t}) \subset N_{\mu,\mathbf{p},\Pi} \\
&\iff \mathcal{CK}_\Pi(\mathbf{t}) \subset N_{\mu,\mathbf{p},\Pi}.
\end{aligned}$$

Since $\mathcal{CK}_\Pi$ is the meet of the partition profile Π, every cell of it, particularly $\mathcal{CK}_\Pi(\mathbf{t})$, is a self-evident event. Hence, $\mathcal{CK}_\Pi(\mathbf{t}) \subset N_{\mu,\mathbf{p},\Pi}$ is equivalent to saying that at $\mathbf{t}$, it is common knowledge that the state is individually rational and not blocked. ■

A.2 Proofs of Theorem 2, Proposition 3 and Theorem 1

Proof of Theorem 2 Throughout this proof, we let l^* be the smallest integer such that $C_k^{l^*} = C_k^\infty$, where $k = i, j$. We need two lemmata to prove the theorem. First, we show that consideration correspondences in iteration (5)-(6), particularly the limit ones, must be measurable with respect to $\Pi_i \vee \Pi_j$.[25] Therefore, should $\mathbf{t}'$

[25]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 $\Pi_i \vee \Pi_j$.

be the true type profile, the smallest possible consideration set is $[\Pi_i \vee \Pi_j](\mathbf{t}')$.

Lemma 2 (Measurability.) *For each l, the correspondences C_i^l and C_j^l are measurable w.r.t.* $\Pi_i \vee \Pi_j$.

Proof 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 must 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 $\Pi_j(\mathbf{t}'') = \Pi_j(\mathbf{t}''')$, i.e., $C_j^0(\mathbf{t}'') = C_j^0(\mathbf{t}''')$, and that

$$\exists \mathbf{t}'''' \in C_j^0(\mathbf{t}'') \text{ s.t. } \chi_j(\mathbf{t}'''') = Y \Longleftrightarrow \exists \mathbf{t}'''' \in C_j^0(\mathbf{t}''') \text{ s.t. } \chi_j(\mathbf{t}'''') = Y,$$

which precisely implies 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. ■

Next, we prove a stronger version of IT.

Lemma 3 (Strong Improvement at the Truth (SIT).) *If $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is blocked by $(i, j; p)$, then for every $\mathbf{t}' \in [\Pi_i \vee \Pi_j](\mathbf{t})$, $\chi_i(\mathbf{t}') = Y$ and $\chi_j(\mathbf{t}') = Y$.*

Proof We prove by contradiction. Suppose that there exists $\mathbf{t}' \in [\Pi_i \vee \Pi_j](\mathbf{t})$ such that $\chi_i(\mathbf{t}') = N$. Since $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is blocked by $(i, j; p)$, we know that $\chi_i(\mathbf{t}'') = Y$ for every $\mathbf{t}'' \in C_i^{l^*}(\mathbf{t})$. Therefore, measurability (Lemma 2) implies

$$C_i^{l^*}(\mathbf{t}) \cap [\Pi_i \vee \Pi_j](\mathbf{t}) = \emptyset. \tag{16}$$

By the iteration rule (5) and measurability, (16) is true only if one of the following two cases happens (for agent j at round $l^* - 1$):

(a) $C_j^{l^*-1}(\mathbf{t}) \neq \emptyset$ and for every $\mathbf{t}''' \in C_j^{l^*-1}(\mathbf{t})$, we have $\chi_j(\mathbf{t}''') = N$.

(b) $C_j^{l^*-1}(\mathbf{t}) = \emptyset$, which implies $C_j^{l^*-1}(\mathbf{t}) \cap [\Pi_i \vee \Pi_j](\mathbf{t}) = \emptyset$.

Suppose case (a) holds. Then monotonicity, i.e., $C_j^{l^*}(\mathbf{t}) \subset C_j^{l^*-1}(\mathbf{t})$ (Lemma 1), implies that for every $\mathbf{t}''' \in C_j^{l^*}(\mathbf{t})$, we have $\chi_j(\mathbf{t}''') = N$, which contradicts $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ being blocked by $(i, j; p)$. Suppose case (b) holds. This is true only if either case (a) or case (b) holds for agent i at round $l^* - 2$. Similar argument shows that case (a) leads to a contradiction. This continues until round zero is reached and only case (b) is possible, i.e., we have either (case (b) for agent i at round zero)

$$C_i^0(\mathbf{t}) \cap [\Pi_i \vee \Pi_j](\mathbf{t}) = \emptyset$$

or (case (b) for agent j at round zero)

$$C_j^0(\mathbf{t}) \cap [\Pi_i \vee \Pi_j](\mathbf{t}) = \emptyset.$$

However, both of which bring us to a contradiction since $C^0 = \Pi$. Hence, for all $\mathbf{t}' \in [\Pi_i \vee \Pi_j](\mathbf{t})$, we have $\chi_i(\mathbf{t}') = Y$. Symmetric argument shows that $\chi_j(\mathbf{t}') = Y$ holds as well for all $\mathbf{t}' \in [\Pi_i \vee \Pi_j](\mathbf{t})$. ■

Now we prove Theorem 2. Suppose $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is blocked by $(i, j; p)$. Let $\hat{\Pi}$ be finer than Π. Then we know that for each $\mathbf{t}' \in T$,

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

Note that $C_i^l(\mathbf{t}')$ defined in (5) is increasing in both $\Pi_i(\mathbf{t}')$ and $C_j^{l-1}(\mathbf{t}'')$, and that $C_j^l(\mathbf{t}')$ defined in (6) is increasing in both $\Pi_j(\mathbf{t}')$ and $C_i^{l-1}(\mathbf{t}'')$. It follows from induction that for each $\mathbf{t}' \in T$ and each $l = 1, 2, \ldots,$

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

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

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

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

Now it suffices to show $\hat{C}_i^{l^*}(\mathbf{t}) \neq \emptyset$ and $\hat{C}_j^{l^*}(\mathbf{t}) \neq \emptyset$. Since $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is blocked by $(i, j; p)$, we know by SIT (Lemma 3) 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 $\hat{C}_j^l(\mathbf{t})$ along (5)-(6), 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 3 that $(\mu, \mathbf{p}, \mathbf{t}, \hat{\Pi})$ is blocked. ■

Proof of Proposition 3 Suppose $(\mu, \mathbf{p}, \mathbf{t}, \hat{\Pi})$ is stable. Then $(\mu, \mathbf{p}, \mathbf{t}, \hat{\Pi})$ is individually rational and not blocked. Moreover, by the definition of $\mathcal{N}_{\mu,\mathbf{p},\hat{\Pi}}(\mathbf{t})$, we know that $(\mu, \mathbf{p}, \mathbf{t}', \hat{\Pi})$ is individually rational and not blocked for all $\mathbf{t}' \in \mathcal{N}_{\mu,\mathbf{p},\hat{\Pi}}(\mathbf{t})$. The individual rationality of $(\mu, \mathbf{p}, \mathbf{t}', \hat{\Pi})$ implies the individual rationality of $(\mu, \mathbf{p}, \mathbf{t}', \Pi)$ because $\hat{\Pi}$ is finer than Π. By Theorem 2, $(\mu, \mathbf{p}, \mathbf{t}', \hat{\Pi})$ being not blocked implies that $(\mu, \mathbf{p}, \mathbf{t}', \Pi)$ is not blocked. Therefore, we have

$$\mathcal{N}_{\mu,\mathbf{p},\hat{\Pi}}(\mathbf{t}) \subset \mathcal{N}_{\mu,\mathbf{p},\Pi}(\mathbf{t}). \tag{17}$$

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 (by the definition of $H(\cdot)$ in (11)), it follows from (17) that

$$\hat{\Pi}_k(\mathbf{t}') \subset \mathcal{N}_{\mu,\mathbf{p},\Pi}(\mathbf{t}) \text{ for all } \mathbf{t}' \in \mathcal{N}_{\mu,\mathbf{p},\hat{\Pi}}(\mathbf{t}) \text{ and all } k \in I \cup J.$$

Since $\hat{\Pi}$ is finer than Π, it follows that

$$\hat{\Pi}_k(\mathbf{t}') \subset \Pi_k(\mathbf{t}') \text{ for all } \mathbf{t}' \in \mathcal{N}_{\mu,\mathbf{p},\hat{\Pi}}(\mathbf{t}) \text{ and all } k \in I \cup J.$$

Therefore, for every k and every $\mathbf{t}' \in \mathcal{N}_{\mu,\mathbf{p},\hat{\Pi}}(\mathbf{t})$, the definition of $H(\cdot)$ in (11) implies that

$$\hat{\Pi}_k(\mathbf{t}') \subset \Pi^1_k(\mathbf{t}'). \tag{18}$$

We claim that the state $(\mu, \mathbf{p}, \mathbf{t}', \Pi^1)$ is not blocked for every $\mathbf{t}' \in \mathcal{N}_{\mu,\mathbf{p},\hat{\Pi}}(\mathbf{t})$. Otherwise, the state $(\mu, \mathbf{p}, \mathbf{t}', \hat{\Pi})$ is blocked by Theorem 2 (applied locally on the event $\mathcal{N}_{\mu,\mathbf{p},\hat{\Pi}}(\mathbf{t})$), which is a contradiction. Hence, $(\mu, \mathbf{p}, \mathbf{t}', \Pi^1)$ is not blocked. The individual rationality of $(\mu, \mathbf{p}, \mathbf{t}', \hat{\Pi})$ implies the individual rationality of $(\mu, \mathbf{p}, \mathbf{t}', \Pi^1)$ because $\hat{\Pi}$ is finer than Π^1 on the event $\mathcal{N}_{\mu,\mathbf{p},\hat{\Pi}}(\mathbf{t})$.

Inductively for every integer $l \geq 1$, we have

$$\mathcal{N}_{\mu,\mathbf{p},\hat{\Pi}}(\mathbf{t}) \subset \mathcal{N}_{\mu,\mathbf{p},\Pi^l}(\mathbf{t}). \tag{19}$$

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

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

which, together with (19) implies that $(\mu, \mathbf{p}, \mathbf{t}', \Pi^l)$ is individually rational and not blocked. Particularly, $(\mu, \mathbf{p}, \mathbf{t}, \Pi^l)$ is individually rational and not blocked. Since Π^l is weakly finer than Π^{l-1}, there exists l^* such that $\Pi^{l^*+1} = \Pi^{l^*}$. Therefore, the limit state $(\mu, \mathbf{p}, \mathbf{t}, \Pi^\infty) = (\mu, \mathbf{p}, \mathbf{t}, \Pi^{l^*})$ is stable. ∎

We need the following lemma to prove Theorem 1. The lemma also serves as a proof for the existence of naive stable states and sophisticated stable states, because it is independent of how blocking is defined. Moreover, for naive stability, the existence statement can be as strong as Theorem 1 since IM is satisfied.

Lemma 4 *For any* $\mathbf{t} \in T$*, let* $(\mu, \mathbf{p})$ *be a complete-information stable allocation* *and* Π *be the complete-information partition profile. Then* $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ *is stable.*

Proof of Lemma 4 For any $\mathbf{t} \in T$, let $(\mu, \mathbf{p})$ be a complete-information stable allocation at $\mathbf{t}$, which exists by, say, Shapley and Shubik (1971). Define a partition profile Π such that $\Pi_k(\mathbf{t}') = \{\mathbf{t}'\}$ for every $\mathbf{t}' \in T$ and every k. We proceed to verify that $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is stable. First, since $(\mu, \mathbf{p})$ is a complete-information stable allocation at $\mathbf{t}$ and Π represents complete information, $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is obviously individually rational.

Second, we claim $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is not blocked. Suppose to the contrary that it is blocked by a combination $(i, j; p)$. Then we have $C_i^{\infty}(\mathbf{t}) \neq \emptyset$ and $C_j^{\infty}(\mathbf{t}) \neq \emptyset$. On the other hand, by Lemma 1, both $C_i^l(\mathbf{t})$ and $C_j^l(\mathbf{t})$ are decreasing in l, which implies that $C_i^{\infty}(\mathbf{t}) \subset \Pi_i(\mathbf{t}) = \{\mathbf{t}\}$ and $C_j^{\infty}(\mathbf{t}) \subset \Pi_j(\mathbf{t}) = \{\mathbf{t}\}$. Hence, we have $C_i^{\infty}(\mathbf{t}) = \{\mathbf{t}\}$ and $C_j^{\infty}(\mathbf{t}) = \{\mathbf{t}\}$. As a result, (7)-(8) say that $(\mu, \mathbf{p})$ is complete-information blocked by $(i, j; p)$, a contradiction.

Finally, since $\Pi_k(\mathbf{t}) = \{\mathbf{t}\}$ for every $\mathbf{t} \in T$ and every k, i.e., Π is already the finest partition profile, we know that Π is a fixed point of $H_{\mu,\mathbf{p}}(\cdot)$ regardless of $(\mu, \mathbf{p})$. Therefore, the state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is stable. ■

Proof of Theorem 1 It follows from Lemma 4 and Proposition 3. ■

A.3 Proof of Theorem 4 and an example on the importance of A3

Proof of Theorem 4 We clarify some notations before proving the theorem. Let $\mathbf{t} = (\mathbf{w}, \mathbf{f})$ and $\mathbf{t}' = (\mathbf{w}', \mathbf{f}')$. Under Condition A4, for every $\mathbf{t}' \in \Pi_i(\mathbf{t})$, we have $\mathbf{w}' = \mathbf{w}$ and for every $\mathbf{t}' \in \Pi_j(\mathbf{t})$, we have $\mathbf{f}' = \mathbf{f}$. In this sense, we need to treat $\mathbf{w}$ and $\mathbf{f}$ separately. Taking Condition A3 (no transfer) into account, we shall now write a typical state as $(\mu, \mathbf{w}, \mathbf{f}, \Pi)$, and a blocking combination now is simply a blocking pair.

Suppose $(\mu, \mathbf{w}, \mathbf{f}, \Pi)$ is level-l blocked by a pair (i, j). We proceed to show that the state is also level-l' blocked by the same pair (i, j), by arguing $C_i^{l'}(\mathbf{w}, \mathbf{f}) = C_i^l(\mathbf{w}, \mathbf{f})$ and $C_j^{l'}(\mathbf{w}, \mathbf{f}) = C_j^l(\mathbf{w}, \mathbf{f})$. It suffices to show that the two agents' consideration at $(\mathbf{w}, \mathbf{f})$ cannot be refined at all, i.e., for every $l'' \in \mathbb{N}$,

$$C_i^{l''}(\mathbf{w}, \mathbf{f}) = \Pi_i(\mathbf{w}, \mathbf{f}) \quad \text{and} \quad C_j^{l''}(\mathbf{w}, \mathbf{f}) = \Pi_j(\mathbf{w}, \mathbf{f}). \tag{20}$$

We prove this by induction.

First of all, we have $C_i^0(\mathbf{w}, \mathbf{f}) = \Pi_i(\mathbf{w}, \mathbf{f})$ and $C_j^0(\mathbf{w}, \mathbf{f}) = \Pi_j(\mathbf{w}, \mathbf{f})$ by definition. Suppose (20) holds up to $l'' = k-1$, where $k \geq 1$. To complete the induction, we need to show $C_i^k(\mathbf{w}, \mathbf{f}) = \Pi_i(\mathbf{w}, \mathbf{f})$ and $C_j^k(\mathbf{w}, \mathbf{f}) = \Pi_j(\mathbf{w}, \mathbf{f})$. For an arbitrary type assignment $(\mathbf{w}', \mathbf{f}') \in C_i^{k-1}(\mathbf{w}, \mathbf{f}) = \Pi_i(\mathbf{w}, \mathbf{f})$, we claim that

(a) $\chi_j(\mathbf{w}', \mathbf{f}') = Y$ and

(b) $(\mathbf{w}', \mathbf{f}') \in C_j^{k-1}(\mathbf{w}', \mathbf{f}')$,

which would imply by (5) that $(\mathbf{w}', \mathbf{f}') \in C_i^k(\mathbf{w}, \mathbf{f})$ and thus $C_i^k(\mathbf{w}, \mathbf{f}) = \Pi_i(\mathbf{w}, \mathbf{f})$. The argument for $C_j^k(\mathbf{w}, \mathbf{f}) = \Pi_j(\mathbf{w}, \mathbf{f})$ is symmetric.

Proof of Claim (a). Since level-l blocking implies (level-∞) blocking by definition and the latter satisfies IT (Theorem 2), we know that level-l blocking satisfies IT as well. This, together with A3, imply that

$$\nu_{\mathbf{w}(i),\mathbf{f}(j)} > \nu_{\mathbf{w}(i),\mathbf{f}(\mu(i))} \quad \text{and} \quad \phi_{\mathbf{w}(i),\mathbf{f}(j)} > \phi_{\mathbf{w}(\mu(j)),\mathbf{f}(j)}. \tag{21}$$

Since agents prefer higher types by Conditions A1-A2, the second part of (21) implies that

$$\mathbf{w}(i) > \mathbf{w}(\mu(j)). \tag{22}$$

Since worker i knows the types of all workers (A4), we have $\mathbf{w}' = \mathbf{w}$ for each $(\mathbf{w}', \mathbf{f}') \in \Pi_i(\mathbf{w}, \mathbf{f})$. Finally, inequality (22) and Conditions A1-A2 imply that

$$\phi_{\mathbf{w}'(i),\mathbf{f}'(j)} > \phi_{\mathbf{w}'(\mu(j)),\mathbf{f}'(j)},$$

which is equivalent to saying that $\chi_j(\mathbf{w}', \mathbf{f}') = Y$.

Proof of Claim (b). If $k = 1$, then obviously we have

$$(\mathbf{w}', \mathbf{f}') \in C_j^0(\mathbf{w}', \mathbf{f}') = \Pi_j(\mathbf{w}', \mathbf{f}').$$

Suppose $k \geq 2$. Since $(\mathbf{w}', \mathbf{f}') \in C_i^{k-1}(\mathbf{w}, \mathbf{f}) \subset \Pi_i(\mathbf{w}, \mathbf{f})$, we know that $(\mathbf{w}', \mathbf{f}')$ and $(\mathbf{w}, \mathbf{f})$ are in the same partition cell of Π_i. Therefore, $\Pi_i(\mathbf{w}', \mathbf{f}') = \Pi_i(\mathbf{w}, \mathbf{f})$

and $C_i^{k-2}(\mathbf{w}', \mathbf{f}') = C_i^{k-2}(\mathbf{w}, \mathbf{f})$. By the induction hypothesis, we have $C_i^{k-2}(\mathbf{w}, \mathbf{f}) = \Pi_i(\mathbf{w}, \mathbf{f})$, which implies that $C_i^{k-2}(\mathbf{w}', \mathbf{f}') = \Pi_i(\mathbf{w}', \mathbf{f}')$. Therefore, $(\mathbf{w}, \mathbf{f}) \in \Pi_i(\mathbf{w}', \mathbf{f}')$ implies $(\mathbf{w}, \mathbf{f}) \in C_i^{k-2}(\mathbf{w}', \mathbf{f}')$. Note that the first part of (21) means $\chi_i(\mathbf{w}, \mathbf{f}) = Y$. Hence, we have $(\mathbf{w}', \mathbf{f}') \in C_j^{k-1}(\mathbf{w}', \mathbf{f}')$ by (6). ■

The following example demonstrates the importance of A3 in establishing the equivalence in Theorem 4, when Conditions A1, A2 and A4 are all satisfied.

Example 4 Consider a market with two workers, α and β, and one firm, a. The set of possible type assignments is given by $T = \{\mathbf{t}, \mathbf{t}'\}$, i.e.,

	$\mathbf{t}$	$\mathbf{t}'$
α	3	3
β	4	1
a	2	2

The premuneration values for workers and firms are given by $\nu_{w,f} = \phi_{w,f} = wf$. Obviously, given any wage, every agent prefers a partner with a higher type to a partner with a lower type.

We first describe a state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$. Suppose $\mu(\alpha) = a$, $\mu(\beta) = \beta$, and $\mathbf{p} = \mathbf{0}$. Suppose $\mathbf{t}$ is the true type assignment, and that agents' partitions are given as if agents have the knowledge of the agent types within their own side, i.e., $\Pi_\alpha = \Pi_\beta = \{\{\mathbf{t}\}, \{\mathbf{t}'\}\}$, and $\Pi_a = \{\mathbf{t}, \mathbf{t}'\}$.

Now consider the potential blocking combination $(\beta, a; -3)$. Using the partitions over lists of indicators as in Example 1, we rewrite agents' partitions in the left panel below. The consideration correspondences of a and β could then be represented as in the right panel.

$$\begin{array}{rcc} & \mathbf{t} & \mathbf{t}' \\ \Pi_\beta = & \{\{Y\}, & \{N\}\} \\ \Pi_a = & \{\{Y & N\}\} \end{array} \qquad\qquad \begin{array}{rcc} & \mathbf{t} & \mathbf{t}' \\ C_\beta^1 = & \{\{Y\}, & \{N\}\} \\ C_a^1 = & \{\{Y & \}\} \end{array}$$

Clearly, $(\beta, a; -3)$ (level-1) blocks the status quo state, but not naively.

A.4 Scenarios that cannot initiate consideration refinement

The following examples illustrate that not all scenarios with both Y and N faced by an agent can initiate consideration refinements:

Agent	$\{Y$	$N\}$	$\cdots$		Agent	$\{Y$	$N\}$	$\cdots$	
Potential partner	$\ldots$	$\{Y$	$N\}$		Potential partner	$\ldots$	$\{N$	$Y\}$	

In these two cases, whether the agent can rule out N cannot be deduced from (12) and (13) only and must depend on how the consideration set is defined/refined for both agents. For instance, in the left panel, if the potential partner can rule out her N by, say, uniform N's of the agent, then the potential partner will consider only Y and the agent will have to consider N. Alternatively, if the potential partner's N is supported, say, due to the agent's uniform Y's, then the potential partner's consideration set will have to contain the N and the agent can safely rule out his N. Either case can happen.

While the left panel concerns whether the agent's N is potentially a supported N, the right panel concerns whether the agent's N is potentially not considered because of the potential partner's (refined) uniform N. Both cases illustrate that whether the agent can rule out N cannot be deduced from (12) and (13) only and must depend on how the consideration set is defined/refined for both agents.

References

Abdulkadiroglu, Atila, and Tayfun Sönmez. 2013. "Matching markets: Theory and practice." *Advances in Economics and Econometrics*, 1: 3–47.

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

Ashlagi, Itai, Mark Braverman, Yash Kanoria, and Peng Shi. 2020. "Clearing matching markets efficiently: informative signals and match recommendations." *Management Science*, 66(5): 2163–2193.

Aumann, Robert J. 1976. "Agreeing to disagree." *The Annals of Statistics*, 1236–1239.

Bergemann, Dirk, and Stephen Morris. 2007. "Belief Free Incomplete Information Games." *Cowles Foundation Discussion Paper No. 1629, Available at SSRN: https://ssrn.com/abstract=1015247.*

Bikhchandani, Sushil. 2014. "Two-sided Matching with Incomplete Information." *Working Paper.*

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

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

Chen, Yi-Chun, and Gaoji Hu. 2020. "Learning by matching." *Theoretical Economics*, 15(1): 29–56.

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

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

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

Ehlers, Lars, and Jordi Massó. 2015. "Matching markets under (in) complete information." *Journal of Economic Theory*, 157: 295–314.

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

Gonczarowski, Yannai A, Noam Nisan, Rafail Ostrovsky, and Will Rosenbaum. 2019. "A stable marriage requires communication." *Games and Economic Behavior*, 118: 626–647.

Immorlica, Nicole, Jacob Leshno, Irene Lo, and Brendan Lucier. 2020. "Information Acquisition in Matching Markets: The Role of Price Discovery." *Available at SSRN: https://ssrn.com/abstract=3705049.*

Kojima, Fuhito. 2015. "Recent developments in matching theory and their practical applications." *Advances in Economics and Econometrics*, 1: 138–175.

Liu, Qingmin. 2017. "Rational Expectations Stability and Competitive Equilibrium in Matching with Incomplete Information." *mimeo.*

Liu, Qingmin. 2020. “Stability and Bayesian Consistency in Two-Sided Markets.” *American Economic Review*, 110(8): 2625–2666.

Liu, Qingmin, George J Mailath, Andrew Postlewaite, and Larry Samuelson. 2014. “Stable Matching with Incomplete Information.” *Econometrica*, 82(2): 541–587.

Mailath, George J, Andrew Postlewaite, and Larry Samuelson. 2013. “Pricing and investments in matching markets.” *Theoretical Economics*, 8: 535–590.

Mailath, George J, Andrew Postlewaite, and Larry Samuelson. 2017. “Premuneration values and investments in matching markets.” *The Economic Journal*, 127(604): 2041–2065.

Nagel, Rosemarie. 1995. “Unraveling in guessing games: An experimental study.” *The American Economic Review*, 85(5): 1313–1326.

Osborne, Martin J, and Ariel Rubinstein. 1994. *A course in game theory.* MIT press.

Pomatto, Luciano. 2020. “Stable Matching under Forward-Induction Reasoning.” *Working Paper.*

Rastegari, Baharak, Anne Condon, Nicole Immorlica, Robert Irving, and Kevin Leyton-Brown. 2014. “Reasoning about optimal stable matchings under partial information.” 431–448.

Roth, Alvin E. 1982. “The economics of matching: Stability and incentives.” *Mathematics of Operations Research*, 7(4): 617–628.

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

Roth, Alvin E. 2008. “Deferred acceptance algorithms: History, theory, practice, and open questions.” *International Journal of Game Theory*, 36(3-4): 537–569.

Roth, Alvin E, and Elliott Peranson. 1999. “The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic D.” *American Economic Review*, 89(4): 748–782.

Roth, Alvin E, and Marilda A Oliveira Sotomayor. 1990. *Two-sided matching: A study in game-theoretic modeling and analysis.* Cambridge University Press.

Segal, Ilya. 2007. “The communication requirements of social choice rules and supporting budget sets.” *Journal of Economic Theory*, 136(1): 341–378.

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

Stahl, Dale O, and Paul W Wilson. 1994. "Experimental evidence on players' models of other players." *Journal of Economic Behavior and Organization*, 25(3): 309–327.

Stahl, Dale O, and Paul W Wilson. 1995. "On players' models of other players: Theory and experimental evidence." *Games and Economic Behavior*, 10(1): 218–254.

Volij, Oscar. 2000. "Communication, credible improvements and the core of an economy with asymmetric information." *International Journal of Game Theory*, 29(1): 63–79.

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

Yenmez, M. Bumin. 2013. "Incentive-compatible matching mechanisms: consistency with various stability notions." *American Economic Journal: Microeconomics*, 5(4): 120–141.

Appendix B For online publication: Formal analysis for Section 5

In this online appendix, we formally state and prove the results that are briefly discussed in Section 5. Throughout, we fix a potential blocking combination $(i, j; p)$ and let the partition profile be Π.

First, in Appendix B.1, we formally elaborate why we shall pay specific attention to the *uniform* N case, the *supported* N case and the *inseparable* N case, which we will refer as the fundamental sources of consideration refinements. Then we formally define sophisticated blocking in Appendix B.2. Related results will be stated in Appendix B.3 and proved in Appendix B.4.

B.1 Fundamental sources of consideration refinements

Suppose for now that we do not have a formulation of consideration refinement yet, and we are about to define some consideration correspondences that specifies for each agent at each type assignment a set of considered type assignments. Denote the consideration correspondences to be defined by $C_i^\dagger(\mathbf{t})$ and $C_j^\dagger(\mathbf{t})$, where $\mathbf{t} \in T$.

For convenience, here we restate the upper-bound constraint (12) and the lower-bound constraint (13) introduced in Section 5.2:

$$C_i^\dagger(\mathbf{t}) \subset \Pi_i(\mathbf{t}) \quad \text{and} \quad C_j^\dagger(\mathbf{t}) \subset \Pi_j(\mathbf{t}), \quad \text{for all } \mathbf{t} \in T. \tag{12}$$

$$\mathbf{t}' \in C_k^\dagger(\mathbf{t}) \text{ implies } [\Pi_i \vee \Pi_j](\mathbf{t}') \subset C_k^\dagger(\mathbf{t}), \quad \text{for } k = i, j \text{ and all } \mathbf{t}, \mathbf{t}' \in T. \tag{13}$$

Based on these natural restrictions and upon nothing else, we proceed to figure out in which cases worker i (firm j) can refine his (her) consideration. Indeed, taking into account the possibility that refinements may lead to further refinements, we shall focus on cases which initiate consideration refinements, instead of defining $C_i^\dagger(\mathbf{t})$ and $C_j^\dagger(\mathbf{t})$ directly. And the initiating cases will actually pin down $C_i^\dagger(\mathbf{t})$ and $C_j^\dagger(\mathbf{t})$.

We now identify all fundamental sources of consideration refinements, which could serve as the starting points of an iteration process, like (5)-(6), that will precisely define the consideration correspondences $C_i^\dagger$ and $C_j^\dagger$. Given the symmetry between i and j, we will focus on worker i's point of view: When should worker i exclude $\mathbf{t} \in \Pi_i(\mathbf{t})$ from consideration? As has been discussed, we shall classify the cases only according to properties (12) and (13) and, of course, common knowledge of the model.

Consider the following cases:

1. *For every* $\mathbf{t}' \in \Pi_j(\mathbf{t})$, *we have* $\chi_j(\mathbf{t}') = N$.[26]

   In this case, whatever $C_j^\dagger(\mathbf{t})$ is, $C_j^\dagger(\mathbf{t})$ can only be empty or contain just N's by (12) (abusing terminology). The former subcase, as well as similar situations below with empty $C_j^\dagger(\mathbf{t})$, involves more properties of $C_j^\dagger$ other than (12) and (13), which makes the case not fundamental. In the latter subcase, agent i should exclude $\mathbf{t}$. This is the source of uniform N's.

2. *For every* $\mathbf{t}' \in \Pi_j(\mathbf{t})$, *we have* $\chi_j(\mathbf{t}') = Y$.

   Now, $C_j^\dagger(\mathbf{t})$ is empty or contains only Y's by (12). The former subcase is not fundamental. In the latter subcase, agent i has no hope to exclude $\mathbf{t}$ from consideration, and, in fact, has to consider $\mathbf{t}$.

3. *For some* $\mathbf{t}' \in \Pi_j(\mathbf{t})$, *we have* $\chi_j(\mathbf{t}') = N$*; and for some* $\mathbf{t}'' \in \Pi_j(\mathbf{t})$, *we have* $\chi_j(\mathbf{t}'') = Y$.

   Consider the following mutually exclusive subcases:

   a. *There is* $\mathbf{t}''' \in \Pi_j(\mathbf{t})$ *such that* $\chi_j(\mathbf{t}''') \neq \chi_j(\ \mathbf{t})$ *and* $\mathbf{t}''' \in \Pi_i(\mathbf{t})$.

      Whatever $C_j^\dagger(\mathbf{t})$ is, either it does not contain $\mathbf{t}$ (not a fundamental case because it involves further discussion of $C_j^\dagger$ other than (12) and (13)) or we have $\mathbf{t}, \mathbf{t}''' \in C_j^\dagger(\mathbf{t})$ (this is true by (13) and $\mathbf{t}, \mathbf{t}''' \in [\Pi_i \vee \Pi_j](\mathbf{t})$). In the latter subcase, firm j will definitely object the new partnership, and agent i should ignore $\mathbf{t}$. This is the source of inseparable N's.

   b. *For any* $\mathbf{t}''' \in \Pi_j(\mathbf{t})$ *such that* $\chi_j(\mathbf{t}''') \neq \chi_j(\mathbf{t})$, *we have* $\mathbf{t}''' \notin \Pi_i(\mathbf{t})$.

      For worker i to ignore $\mathbf{t}$, we must have $C_j^\dagger(\mathbf{t})$ containing at least one N. Pick an arbitrary $\mathbf{t}' \in \Pi_j(\mathbf{t})$ such that $\chi_j(\mathbf{t}') = N$. Consider the

[26] Recall that the indicator functions χ_i and χ_j are defined by (3)-(4) in Section 3.2.

following cases which may result in $\mathbf{t}' \in C_j^\dagger(\mathbf{t})$:

i. *All Y's (and maybe some N's) will be ignored by agent j, so that $C_j^\dagger(\mathbf{t})$ could at most contain only N's.*
   This is not a fundamental case.

ii. *For every $\mathbf{t}'' \in \Pi_i(\mathbf{t}')$, we have $\chi_i(\mathbf{t}'') = Y$.*
   In this case, whatever $C_i^\dagger(\mathbf{t}')$ is, either it is empty (not a fundamental case) or it must contain only Y's by (12). Suppose the latter happens. Then, firm j will have to consider $\mathbf{t}'$, i.e., $\mathbf{t}' \in C_j^\dagger(\mathbf{t})$, due to worker i's definite willingness to participate in the new partnership. This is the source of supported N's.

iii. *For every $\mathbf{t}'' \in \Pi_i(\mathbf{t}')$, we have $\chi_i(\mathbf{t}'') = N$.*
   In this case, whatever $C_i^\dagger(\mathbf{t}')$ is, either it is empty (not a fundamental case) or it must contain only N's by (12). Suppose the latter happens. Then, firm j will not consider $\mathbf{t}'$, i.e., $\mathbf{t}' \notin C_j^\dagger(\mathbf{t})$.

iv. *For some $\mathbf{t}'' \in \Pi_i(\mathbf{t}')$, we have $\chi_i(\mathbf{t}'') = Y$; and for some $\mathbf{t}''' \in \Pi_i(\mathbf{t}')$, we have $\chi_i(\mathbf{t}''') = N$.*
   Now we are faced with another question of what $C_i^\dagger(\mathbf{t}')$ shall be, just as we started with. Naturally, instead of starting another round of discussion, we shall stop here and define $C_i^\dagger$ and $C_j^\dagger$ iteratively using the fundamental sources of consideration refinement which we just identified.

To sum up, there are only three fundamental sources of consideration refinements, uniform N's, inseparable N's and supported N's, exactly as we have introduced. Within them, from agent i's perspective, uniform N's are of order zero in the sense that it is identified using just the information of Π_j (and restriction (12)); whereas inseparable N's and supported N's are of order one in the sense that they are identified using both the information of Π_j and the information of Π_i (and restriction (12) or (13)). Clearly, higher order reasoning will be taken account in the iterative consideration refinement process.

B.2 Definition of sophisticated blocking

To define sophisticated blocking, we first describe how agents' sophisticated consideration correspondences are formed.

We first demonstrate that agents' willingness/unwillingness (i.e., the indicator functions defined in (3)-(4)) may be adjusted to reflect inseparable N's or supported N's, which helps us to build the discussion upon our uniform-N benchmark in Section 3.2.2. To wit, consider the inseparable N (right panel) and the supported N (left panel) in the figure below:

$$\begin{array}{lllll} \text{Agent} & \{Y & N\} & \cdots & \\ \text{Potential partner} & \ldots & \{Y & Y\} & \end{array} \qquad \begin{array}{lllll} \text{Agent} & \{Y & N\} & \cdots \\ \text{Potential partner} & \{Y & N & Y\} \end{array}$$

In the left panel, since the agent will either consider neither of Y and N, or definitely consider the supported N, the potential partner can treat the agent's Y as an N because it is tied with a supported N in the agent's consideration. In the right panel, since the agent will either consider neither of Y and N, or definitely consider both together, again, the potential partner can treat the agent's Y as an N because it is tied with an inseparable N in the agent's consideration. Therefore, in terms of consideration refinement, the information conveyed by the inseparable N and the supported N above can be reflected in the following figure, where the hypothetical willingness/unwillingness of the agent is adjusted:

$$\begin{array}{lllll} \text{Agent} & \{N & N\} & \cdots & \\ \text{Potential partner} & \ldots & \{Y & Y\} & \end{array} \qquad \begin{array}{lllll} \text{Agent} & \{N & N\} & \cdots \\ \text{Potential partner} & \{Y & N & Y\} \end{array}$$

These adjustments turn the two kinds of sophisticated N's into the familiar uniform N, with which we could build our analysis on the benchmark in Section 3.2.2.

Naturally, as in Section 3.2.2, an agent only considers type assignments such that the potential partner does not have uniform N (up to adjustment). Set $\chi^0 = \chi$, where χ is defined in (3)-(4), and $C^0 = \Pi$. Agents' consideration correspondences are defined as the limit of C_k^l in the following double-iteration process:

Indicator function adjustment. For all $l \geq 1$ and all $\mathbf{t}' \in T$,

$$\chi_i^l(\mathbf{t}') := \begin{cases} N & \text{if either } \exists \mathbf{t}'' \in C_i^{l-1}(\mathbf{t}') \text{ s.t. } \chi_i^{l-1}(\mathbf{t}'') \neq \chi_i^{l-1}(\mathbf{t}') \text{ and } \mathbf{t}', \mathbf{t}'' \in C_j^{l-1}(\mathbf{t}') \\ & \text{or } \exists \mathbf{t}'' \in C_i^{l-1}(\mathbf{t}') \text{ s.t. } \chi_i^{l-1}(\mathbf{t}'') = N \text{ and } \forall \mathbf{t}''' \in C_j^{l-1}(\mathbf{t}''), \chi_j^{l-1}(\mathbf{t}''') = Y, \\ \chi_i^{l-1}(\mathbf{t}') & \text{otherwise;} \end{cases}$$

$$\chi_j^l(\mathbf{t}') := \begin{cases} N & \text{if either } \exists \mathbf{t}'' \in C_j^{l-1}(\mathbf{t}') \text{ s.t. } \chi_j^{l-1}(\mathbf{t}'') \neq \chi_j^{l-1}(\mathbf{t}') \text{ and } \mathbf{t}', \mathbf{t}'' \in C_i^{l-1}(\mathbf{t}') \\ & \text{or } \exists \mathbf{t}'' \in C_j^{l-1}(\mathbf{t}') \text{ s.t. } \chi_j^{l-1}(\mathbf{t}'') = N \text{ and } \forall \mathbf{t}''' \in C_i^{l-1}(\mathbf{t}''), \chi_i^{l-1}(\mathbf{t}''') = Y, \\ \chi_j^{l-1}(\mathbf{t}') & \text{otherwise.} \end{cases}$$

Sophisticated consideration refinement. For all $l \geq 1$ and all $\mathbf{t}' \in T$,

$$C_i^l(\mathbf{t}') := \left\{ \mathbf{t}'' \in \Pi_i(\mathbf{t}') : \exists \mathbf{t}''' \in C_j^{l-1}(\mathbf{t}'') \text{ s.t. } \chi_j^l(\mathbf{t}''') = Y \right\} \tag{23}$$

$$C_j^l(\mathbf{t}') := \left\{ \mathbf{t}'' \in \Pi_j(\mathbf{t}') : \exists \mathbf{t}''' \in C_i^{l-1}(\mathbf{t}'') \text{ s.t. } \chi_i^l(\mathbf{t}''') = Y \right\}. \tag{24}$$

The following lemma verifies that the consideration correspondence is monotonically decreasing in l.

Lemma 5 *For $k = i, j$ and each $\mathbf{t}' \in T$, $C_k^l(\mathbf{t}')$ is decreasing in l w.r.t. set inclusion.*

Since the set T is finite, there exists some l^* such that $C_i^l(\mathbf{t}') = C_i^\infty(\mathbf{t}')$ and $C_j^l(\mathbf{t}') = C_j^\infty(\mathbf{t}')$ for all $l \geq l^*$ and all $\mathbf{t}' \in T$. We say a state is sophisticatedly blocked by a combination $(i, j; p)$ if both agents in it have higher rematching payoffs under every type assignment that is sophisticatedly considered at the true one.

Definition 6 *A state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is **sophisticatedly blocked** by $(i, j; p)$ if $C_i^\infty(\mathbf{t}) \neq \emptyset$, $C_j^\infty(\mathbf{t}) \neq \emptyset$ and*

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

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

B.3 Additional results for Section 5

In this subsection, we list properties and connections of different blocking and stability notions. To facilitate comparison, we also include the complete-information blocking/stability and the results already discussed in Section 5 sometimes.

The following proposition examines IT and IM for all blocking notions.

Proposition 4 *IT and IM of blocking notions are summarized as follows:*

Blocking Notions	*Properties*	
	IT	*IM*
Complete-information blocking	*satisfied*	*not applicable*
Sophisticated blocking	*satisfied*	*not satisfied*
Level-l blocking ($l \in \mathbb{N}$)	*satisfied*	*satisfied*

The following proposition ranks blocking notions.

Proposition 5 *The following statements are true:*

(i) *For every $l \in \mathbb{N}$, if a state is level-l blocked, then it is level-$(l+1)$ blocked.*

(ii) *If a state is (level-∞) blocked, then it is sophisticatedly blocked.*

(iii) *If a state is sophisticatedly blocked, then it is complete-information blocked.*

Moreover, none of the converse is true.

Denote by $\mathcal{B}$ the set of blocking combinations. Then an immediate corollary of Proposition 5 is the following: Fix an arbitrary state, we have

$$\mathcal{B}^0 \subset \mathcal{B}^1 \subset \cdots \subset \mathcal{B}^\infty \subset \mathcal{B}^{sophisticated} \subset \mathcal{B}^{complete-information}.$$

For each of the blocking notions, we have a corresponding stability notion. More precisely, we take individual rationality as in Definition 1. The way to formulate information stability for all blocking notions is the identical to that of Section 3.3 up to notional replacement. The following proposition says that the set of stable states is nonempty for each of the stability notions.

Proposition 6 *The sets of level-l stable states with $l \in \mathbb{N}$ and sophisticatedly stable states are all nonempty. Particularly, for any $\mathbf{t} \in T$, let $(\mu, \mathbf{p})$ be a complete-information stable allocation. Then*

(i) *for any partition profile Π, $(\mu, \mathbf{p}, \mathbf{t}, H^\infty_{\mu,\mathbf{p}}(\Pi))$ is level-l stable, where the $H_{\mu,\mathbf{p}}(\cdot)$ operator is defined by level-l blocking; and*

(ii) *for the complete-information partition profile Π, $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is sophisticatedly stable.*

Now we are ready to rank the stability notions.

Proposition 7 *The following statements are true:*

(i) *For every $l \in \mathbb{N}$, a state is (essentially) level-l stable if it is level-$(l+1)$ stable.*

(ii) *A state is (essentially level-∞) stable if it is sophisticatedly stable.*

(iii) *A state is (essentially) sophisticatedly stable if it is complete-information stable.*

Moreover, none of the converse is true.

Denote by $\mathcal{S}$ the set of (essentially) stable states. Then an immediate corollary of Proposition 7 is the following set-inclusion relation:

$$\mathcal{S}^0 \supset \mathcal{S}^1 \supset \cdots \supset \mathcal{S}^\infty \supset \mathcal{S}^{sophisticate} \supset \mathcal{S}^{complete-information}.$$

We close this subsection by establishing the equivalence between naive blocking and sophisticated blocking, which will imply the equivalence with blocking by Theorem 4. Equivalence between blocking notions implies equivalence between stability notions. The conditions to guarantee equivalence are exactly the ones A1-A4 we introduced in Section 5.1.1.

Proposition 8 *Under A1-A4, the following statements are equivalent:*

(i) $(\mu, \mathbf{w}, \mathbf{f}, \Pi)$ *is level-l blocked for some $l \in \mathbb{N}$.*

(ii) $(\mu, \mathbf{w}, \mathbf{f}, \Pi)$ *is sophisticatedly blocked.*

An immediate corollary of Proposition 8 is that under A1-A4, the stability notions are all equivalent.

B.4 Proofs of Propositions 4-8

Proof of Proposition 4 It follows from the proof of Theorem 2 that every level-l blocking satisfies both properties. Given Theorem 2 and Example 3 in Section 5.3, we only need to prove that sophisticated blocking satisfies IT.

We prove SIT, i.e., $\chi_i(\mathbf{t}') = Y$ and $\chi_j(\mathbf{t}') = Y$ for all $\mathbf{t}' \in [\Pi_i \vee \Pi_j](\mathbf{t})$, by contradiction. The proof will be similar to that of Lemma 3, which uses measurability (Lemma 2). We omit the establishment of measurability in the corrent context and refers directly to Lemma 2, as the extension is straightforward without changing the statement. However, we present the rest of the proof completely here, instead of just discussing the difference, to avoid confusion.

Suppose that there exists $\mathbf{t}' \in [\Pi_i \vee \Pi_j](\mathbf{t})$ such that $\chi_i(\mathbf{t}') = N$. Since $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is blocked by $(i, j; p)$, we know that

$$\chi_i(\mathbf{t}'') = Y \text{ for every } \mathbf{t}'' \in C_i^{l^*}(\mathbf{t}).$$

Therefore, measurability (Lemma 2) implies

$$C_i^{l^*}(\mathbf{t}) \cap [\Pi_i \vee \Pi_j](\mathbf{t}) = \emptyset. \tag{25}$$

By the iteration of consideration (23)-(24), (25) is true only if one of the following two cases happens (for agent j at round $l^* - 1$):

(a) $C_j^{l^*-1}(\mathbf{t}) \neq \emptyset$ and for every $\mathbf{t}''' \in C_j^{l^*-1}(\mathbf{t})$, we have $\chi_j^{l^*}(\mathbf{t}''') = N$.

(b) $C_j^{l^*-1}(\mathbf{t}) = \emptyset$, which implies $C_j^{l^*-1}(\mathbf{t}) \cap [\Pi_i \vee \Pi_j](\mathbf{t}) = \emptyset$.

Suppose case (a) holds. Then $C_j^{l^*}(\mathbf{t}) \subset C_j^{l^*-1}(\mathbf{t})$ (Lemma 5), implies that

$$\chi_j^{l^*}(\mathbf{t}''') = N \text{ for every } \mathbf{t}''' \in C_j^{l^*}(\mathbf{t}).$$

Since $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is blocked by $(i, j; p)$, we have

$$\chi_j(\mathbf{t}''') = Y \text{ for every } \mathbf{t}''' \in C_j^{l^*}(\mathbf{t}).$$

Therefore, these indicators Y's under χ_j are adjusted to N's when we update the indicator functions. Measurability (Lemma 2) implies that these Y's never turn to N's by inseparable N. Hence, these Y's are adjusted to N's by some supported N, i.e., for some l, $C_j^l(\mathbf{t})$ contains some type assignment $\mathbf{t}''''$ with $\chi_j(\mathbf{t}'''') = N$ and it is supported by worker i's uniform Y's. Obviously, such a type assignment $\mathbf{t}''''$ must be considered by firm j as long as $C_j^l(\mathbf{t}) \neq \emptyset$.[27] Then we have $\mathbf{t}'''' \in C_j^{l^*}(\mathbf{t})$. However, this contradicts $\chi_j(\mathbf{t}''') = Y$ for every $\mathbf{t}''' \in C_j^{l^*}(\mathbf{t})$. Therefore, case (a) does not hold.

Suppose case (b) holds. This is true only if either case (a) or case (b) holds for agent i at round $l^* - 2$. Similar argument as in the last paragraph shows that case (a) leads to a contradiction. This continues until round zero is reached and

[27] It is generally true that supported N type assignments must be considered as long as the corresponding consideration set is nonempty.

only case (b) is possible, i.e., we have either (case (b) for agent i at round zero)

$$C_i^0(\mathbf{t}) \cap [\Pi_i \vee \Pi_j](\mathbf{t}) = \emptyset$$

or (case (b) for agent j at round zero)

$$C_j^0(\mathbf{t}) \cap [\Pi_i \vee \Pi_j](\mathbf{t}) = \emptyset.$$

However, both of them bring us to a contradiction since $C^0 = \Pi$. Hence, for all $\mathbf{t}' \in [\Pi_i \vee \Pi_j](\mathbf{t})$, we have $\chi_i(\mathbf{t}') = Y$. Symmetric argument shows that $\chi_j(\mathbf{t}') = Y$ holds as well for all $\mathbf{t}' \in [\Pi_i \vee \Pi_j](\mathbf{t})$. ■

In the following example, IM is not satisfied by sophisticated blocking because the consideration refinement due to a supported N disappears when information becomes finer.

Example 5 Consider a potential blocking pair for a state, where there are three possible type assignments and the agents' hypothetical willingness/unwillingness is summarized in the figure below.

$$\begin{array}{cccc} & \textit{Truth} & & \\ & \mathbf{t}^1 & \mathbf{t}^2 & \mathbf{t}^3 \\ \textit{Worker} & \{Y & N\} & \{Y\} \\ \textit{Firm} & \{Y\} & \{Y & N\} \end{array}$$

Clearly, the firm's right partition cell, which contains $\mathbf{t}^2$ and $\mathbf{t}^3$, is a case of supported N. The worker knows that the true type assignment is either $\mathbf{t}^1$ or $\mathbf{t}^2$, and he worries about $\mathbf{t}^2$, under which he would not obtain a higher payoff through rematching. However, he would think that if the true type assignment is $\mathbf{t}^2$, then the firm would know her right partition cell $\{\mathbf{t}^2, \mathbf{t}^3\}$, where the indicators should be adjusted into uniform N's due to the supported N. Therefore, the worker does not need to consider $\mathbf{t}^2$, which means that the potential blocking pair is indeed a blocking pair.

Now let us consider an alternative situation in the figure below, where only the firm's partition changes.

	Truth		
	$\mathbf{t}^1$	$\mathbf{t}^2$	$\mathbf{t}^3$
Worker	$\{Y$	$N\}$	$\{Y\}$
Firm	$\{Y\}$	$\{Y\}$	$\{N\}$

Clearly, agents have more precise information as the firm's partition becomes strictly finer. However, the worker and the firm no longer constitute a blocking pair because the worker would consider $\mathbf{t}^2$ when evaluating the rematching.

Proof of Proposition 5 Straightforward by definitions and examples in the main text. ■

Proof of Proposition 6 Identical to that of Theorem 1 for level-l stable states; and identical to that of Lemma 4 for sophisticatedly stable states. ■

Proof of Proposition 7 We provide a unified proof for statements (i)-(iii). More precisely, let W(eak)-blocking and S(trong)-blocking be any two blocking notions among those mentioned in Proposition 5 such that if a state is W-blocked, then it is S-blocked. For example, we can have W = level-l and S = level-$(l+1)$. We proceed to show that if a state $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is S-stable, then the state $(\mu, \mathbf{p}, \mathbf{t}, H^{W,\infty}_{\mu,\mathbf{p}}(\Pi))$ is W-stable, where $H^{W}_{\mu,\mathbf{p}}$ is the information refinement operator as in (11) but associated with the W-blocking notion. We use similar superscripts to differentiate notations associated with the two blocking/stability notions.

Now suppose $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is S-stable. Then $(\mu, \mathbf{p}, \mathbf{t}, \Pi)$ is individually rational and not S-blocked. Moreover, by the definition of $\mathcal{N}^S_{\mu,\mathbf{p},\Pi}(\mathbf{t})$, we know that $(\mu, \mathbf{p}, \mathbf{t}', \Pi)$ is individually rational and not S-blocked for all $\mathbf{t}' \in \mathcal{N}^S_{\mu,\mathbf{p},\Pi}(\mathbf{t})$. By our assumption on S- and W-blocking, $(\mu, \mathbf{p}, \mathbf{t}', \Pi)$ being not S-blocked implies that $(\mu, \mathbf{p}, \mathbf{t}', \Pi)$ is not W-blocked. Therefore, we have

$$\mathcal{N}^S_{\mu,\mathbf{p},\Pi}(\mathbf{t}) \subset \mathcal{N}^W_{\mu,\mathbf{p},\Pi}(\mathbf{t}). \tag{26}$$

Since $\Pi_k(\mathbf{t}') \subset \mathcal{N}^S_{\mu,\mathbf{p},\Pi}(\mathbf{t})$ for every $\mathbf{t}' \in \mathcal{N}^S_{\mu,\mathbf{p},\Pi}(\mathbf{t})$ and every k (by the definition of $H^S_{\mu,\mathbf{p}}(\cdot)$ and the fact that $H^S_{\mu,\mathbf{p}}(\Pi) = \Pi$), it follows from (26) that

$$\Pi_k(\mathbf{t}') \subset \mathcal{N}^W_{\mu,\mathbf{p},\Pi}(\mathbf{t}) \text{ for all } \mathbf{t}' \in \mathcal{N}^S_{\mu,\mathbf{p},\Pi}(\mathbf{t}) \text{ and all } k \in I \cup J.$$

Therefore, we have

$$\Pi_k(\mathbf{t}') = [H^W_{\mu,\mathbf{p}}(\Pi)]_k(\mathbf{t}') \text{ for all } \mathbf{t}' \in \mathcal{N}^S_{\mu,\mathbf{p},\Pi}(\mathbf{t}) \text{ and all } k \in I \cup J. \tag{27}$$

We claim that the state $(\mu, \mathbf{p}, \mathbf{t}', H^W_{\mu,\mathbf{p}}(\Pi))$ is not W-blocked for every $\mathbf{t}' \in \mathcal{N}^S_{\mu,\mathbf{p},\Pi}(\mathbf{t})$. Otherwise, the state $(\mu, \mathbf{p}, \mathbf{t}', \Pi)$ is S-blocked by (27) and our assumption on S- and W-blocking (applied locally on the event $\mathcal{N}^S_{\mu,\mathbf{p},\Pi}(\mathbf{t})$), which is a contradiction. Hence, $(\mu, \mathbf{p}, \mathbf{t}', H^W_{\mu,\mathbf{p}}(\Pi))$ is not W-blocked. The individual rationality of $(\mu, \mathbf{p}, \mathbf{t}', \Pi)$ is equivalent to that of $(\mu, \mathbf{p}, \mathbf{t}', H^W_{\mu,\mathbf{p}}(\Pi))$ by (27).

Inductively for every integer $l \geq 1$, assuming

$$\Pi_k(\mathbf{t}') = [H^{W,l}_{\mu,\mathbf{p}}(\Pi)]_k(\mathbf{t}') \text{ for all } \mathbf{t}' \in \mathcal{N}^S_{\mu,\mathbf{p},\Pi}(\mathbf{t}) \text{ and all } k \in I \cup J,$$

we have that the state $(\mu, \mathbf{p}, \mathbf{t}', H^{W,l}_{\mu,\mathbf{p}}(\Pi))$ is individually rational and not W-blocked for every $\mathbf{t}' \in \mathcal{N}^S_{\mu,\mathbf{p},\Pi}(\mathbf{t})$. Therefore,

$$\mathcal{N}^S_{\mu,\mathbf{p},\Pi}(\mathbf{t}) \subset \mathcal{N}^W_{\mu,\mathbf{p},H^{W,l}_{\mu,\mathbf{p}}(\Pi)}(\mathbf{t}). \tag{28}$$

Hence, for every $\mathbf{t}' \in \mathcal{N}^S_{\mu,\mathbf{p},\Pi}(\mathbf{t})$ and every k, we know by the fact $H^S_{\mu,\mathbf{p}}(\Pi) = \Pi$ that

$$\Pi_k(\mathbf{t}') = [H^{W,l+1}_{\mu,\mathbf{p}}(\Pi)]_k(\mathbf{t}'),$$

which implies that $(\mu, \mathbf{p}, \mathbf{t}', H^{W,l+1}_{\mu,\mathbf{p}}(\Pi))$ is individually rational and not W-blocked. Particularly, $(\mu, \mathbf{p}, \mathbf{t}, H^{W,l+1}_{\mu,\mathbf{p}}(\Pi))$ is individually rational and not W-blocked. Since $H^{W,l+1}_{\mu,\mathbf{p}}(\Pi)$ is weakly finer than $H^{W,l}_{\mu,\mathbf{p}}(\Pi)$, there exists l^* such that $H^{W,l^*+1}_{\mu,\mathbf{p}}(\Pi) = H^{W,l^*}_{\mu,\mathbf{p}}(\Pi)$. Therefore, the limit state $(\mu, \mathbf{p}, \mathbf{t}, H^{W,\infty}_{\mu,\mathbf{p}}(\Pi)) = (\mu, \mathbf{p}, \mathbf{t}, H^{W,l^*}_{\mu,\mathbf{p}}(\Pi))$ is W-stable. ■

Proof of Proposition 8 By Theorem 4, it suffices to show that (i, j) is a blocking pair for the state $(\mu, \mathbf{w}, \mathbf{f}, \Pi)$ if and only if it is a sophisticated one. The only-if part follows from Proposition 5. We proceed to show that if (i, j) is a sophisticated blocking pair, then it is a blocking pair.

First of all, we claim that under A4, there does not exist inseparable N. More precisely, by A4, the join partition must has a singleton cell at every type assignment, particularly at the true type profile $(\mathbf{w}, \mathbf{f})$, i.e.,

$$[\Pi_i \vee \Pi_j]\,(\mathbf{w}', \mathbf{f}') = \{(\mathbf{w}', \mathbf{f}')\} \text{ for all } (\mathbf{w}', \mathbf{f}') \in T. \tag{29}$$

Then, an immediate implication is that there does not exist inseparable N when the two agents update their consideration.

Now we are ready to show that if (i, j) is a sophisticated blocking pair, then it is a blocking pair. The two blocking notion differs only in the refinement of agents' consideration. Namely, sophisticated blocking takes inseparable N and supported N into account but blocking does not. Since inseparable N never happens, we only need to show that supported N, if not a uniform N, cannot differentiate the two blocking notions (of course, uniform N's cannot distinguish the two blocking notions).

Suppose $C_i^l(\mathbf{w}, \mathbf{f})$ can be refined due to agent j's supported N. More precisely, suppose that for some l, there exists $(\mathbf{w}', \mathbf{f}') \in C_i^l(\mathbf{w}, \mathbf{f}) \setminus C_i^{l+1}(\mathbf{w}, \mathbf{f})$ and some $(\mathbf{w}'', \mathbf{f}'') \in C_j^l(\mathbf{w}', \mathbf{f}')$ such that $\chi_j^l(\mathbf{w}'', \mathbf{f}'') = N$ and

$$\chi_i^l(\mathbf{w}''', \mathbf{f}''') = Y \text{ for all } (\mathbf{w}''', \mathbf{f}''') \in C_i^l(\mathbf{w}'', \mathbf{f}''). \tag{30}$$

By the assumption that we are in a supported N case which is not a uniform N, $C_j^l(\mathbf{w}', \mathbf{f}')$ contains both Y and N.

We claim that $(\mathbf{w}'', \mathbf{f}'') \in C_i^l(\mathbf{w}'', \mathbf{f}'')$. Otherwise, it was either ruled out by uniform N or ruled out by supported N, the latter of which implies that the indicators in $C_j^l(\mathbf{w}', \mathbf{f}')$ were all adjusted to N. Then both cases contradicts $C_j^l(\mathbf{w}', \mathbf{f}')$ containing both Y and N. Hence, $(\mathbf{w}'', \mathbf{f}'') \in C_i^l(\mathbf{w}'', \mathbf{f}'')$.

Obviously, by (30), $(\mathbf{w}'', \mathbf{f}'') \in C_i^l(\mathbf{w}'', \mathbf{f}'')$ implies $\chi_i^l(\mathbf{w}'', \mathbf{f}'') = Y$, which in turn implies by Conditions A1-A3 that

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

By A4, firm j knows the types of firms. Then, for each $(\mathbf{w}'''', \mathbf{f}'''') \in \Pi_j(\mathbf{w}'', \mathbf{f}'')$, we have $\mathbf{f}'''' = \mathbf{f}''$. Since $(\mathbf{w}'', \mathbf{f}'') \in C_j^l(\mathbf{w}', \mathbf{f}')$, we know that $(\mathbf{w}', \mathbf{f}') \in \Pi_j(\mathbf{w}'', \mathbf{f}'')$. As a result, $\mathbf{f}' = \mathbf{f}''$. Therefore, $\chi_i^l(\mathbf{w}', \mathbf{f}') = Y$. Since $(\mathbf{w}', \mathbf{f}') \in C_i^l(\mathbf{w}, \mathbf{f}) \setminus C_i^{l+1}(\mathbf{w}, \mathbf{f})$ is taken arbitrarily, we know that agent j's supported N only rules out Y in $C_i^l(\mathbf{w}, \mathbf{f})$.

By symmetric arguments, agent i's supported N only rules out Y in $C_j^l(\mathbf{w}, \mathbf{f})$. Since (i, j) is a sophisticated blocking pair for $(\mu, \mathbf{w}, \mathbf{f}, \Pi)$, we know that with more Y's at $(\mathbf{w}, \mathbf{f})$ for both agents, (i, j) is also a blocking pair. ■