The structure of Bayesian stable matchings*

Gaoji Hu†

Preliminary draft.

July 10, 2024

Abstract

Incomplete information can invalidate standard results on stable matchings, such as the celebrated lone-wolf theorem and lattice theorem. In this paper, we investigate how these properties may generally fail in Bayesian matching markets. Then for each of them, we study the conditions that restore positive results. The key condition which we call *common information* includes complete information (regarded as a condition) as a special case.

JEL Classification: C78, D40, D82, D83

Keywords: matching, incomplete information, the lone-wolf theorem, the lattice theorem.

*I am grateful to Yi-Chun Chen for many helpful discussions. I would also like to thank (in chronological order) Satoru Takahashi, Xiao Luo, Fuhito Kojima, Ziwei Wang, Qianfeng Tang, Weijie Zhong, Jingyi Xue, Yongchao Zhang, Qingmin Liu, and participants at the IAS Workshop (Wuhan U) and SMiW (SUFE) for helpful comments and suggestions. Quan Li and Jiaxin Wang provided excellent research assistance. I acknowledge the financial support provided by the National Natural Science Foundation of China (No. 72394391, No. 72003121, and No. 72033004). All remaining errors are my own.

†School of Economics, Shanghai University of Finance and Economics, and the Key Laboratory of Mathematical Economics (SUFE), Ministry of Education, Shanghai 200433, China. Email address: hu.gaoji@gmail.com

1 Introduction

Since the seminal work of Gale and Shapley (1962), for several decades, stability has been a key notion used in theoretical study as well as practical design of matching markets.[1] Conceptually, stability has been connected to both equity and efficiency, two of the most important notions in economics.[2] Practically, the design of many matching markets aims to achieve stable matchings.[3]

The structure of the set of stable matchings is well understood in the complete-information setting, where the characteristics and preferences of all market participants are publicly known.[4] Particularly, structural results such as the lone-wolf theorem, the rural-hospital theorem, the lattice theorem, the opposition of interests between the two sides of the market, and the existence of extreme stable matchings have been the cornerstone of matching theory and market design practice.

However, incomplete information can invalidate standard results on stable matchings, including all the aforementioned ones. Even more fundamentally, as emphasized recently by Liu et al. (2014) and Liu (2020), incomplete information drastically changes the way we define stability, particularly in what information can be inferred from various possible observations and how information should be updated.[5]

---

[1]See Roth and Sotomayor (1990) and, more recently, Chiappori, Salanie and Che (2024) for comprehensive surveys.

[2]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.

[3]See, e.g., Roth and Peranson (1999) and Abdulkadiroglu and Sönmez (2013) for the designing of practical markets.

[4]For structural results, see, e.g., Roth (1982), Roth (1984), Gale and Sotomayor (1985*b*), Gale and Sotomayor (1985*a*), Klaus and Klijn (2010), Knuth (1976), Blair (1984, 1988), Martínez et al. (2001), Azevedo and Leshno (2016), Hu, Li and Tang (2020).

[5]This arising literature also includes Bikhchandani (2017), Alston (2020), Liu (2023, 2024), Chen and Hu (2020, 2023, 2024), Wang (2023), Pomatto (2022), and Peralta (2024). Another

In this paper, we investigate how two celebrated results, i.e., the lone-wolf theorem and the lattice theorem, may generally fail in Bayesian matching markets.[6] Then for each of them, we study the conditions that restore positive results. Our key condition called "common information" includes the condition of complete information as a special case.

Section 2 sets up the model, where we adopt the stability concepts of Liu (2020) and Chen and Hu (2024) in the one-sided incomplete-information setting, but exclude transfers as in Bikhchandani (2017). Bayesian stability is defined over market *states*, which specify a matching (who matches with whom), a realized type profile (what the payoff-relevant types are) and an information structure (what each agent knows). The existence of Bayesian stable states has been established in earlier works.

Section 3 first provides an example showing that, with incomplete information, the lone-wolf theorem generally fails (Example 1). The key insight is that information updating may encourage a previously unmatched firm to approach a worker whom the firm did not approach because of unresolved uncertainty. Then we prove that if two Bayesian stable states have the common information structure, then the set of unmatched agents is the same across the two states (Proposition 1). Since complete information is a special case of common information, the lone-wolf theorem actually holds more generally. Finally, we provide three examples showing that (1) the common information condition is not necessary, (2) more information may lead to less matched agents, and (3) more information may lead to more matched agents (Examples 2-4). So our knowledge about the lone-wolf property is limited beyond those subsets of Bayesian stable states with common information.

---

stream of literature takes the mechanism design approach. See, e.g., Roth (1989), Yenmez (2013), Chakraborty, Citanna and Ostrovsky (2010).

[6] Fernandez, Rudov and Yariv (2022) also studies structural properties in matching under incomplete information, but for equilibrium outcomes instead of stable outcomes.

Section 4 studies the lattice theorem in a similar manner. Without common information, Example 1 shows that the lattice theorem fails. Now, interestingly, even with common information, the lattice theorem cannot be restored. A simple logical reason is that, unlike the lone-wolf theorem that only compares two given states, the lattice theorem examines two generated states from two given states, where the more complicated situation makes the common information condition inadequate. A concrete example is provided to illustrate the economic reason (Example 5). Then, we borrow intuitive sufficient conditions from Liu (2020) to restore the lattice theorem (Proposition 2). Again, since complete information is a special case of common information, the lattice theorem actually holds more generally, at least in markets that satisfy Liu's conditions.

The complete-information matching literature is so vast that it is impossible to address all related questions with incomplete information in one paper. Section 5 concludes with several future directions.

2 Model

We adopt the setup of Liu (2020) but assume away transfers like Bikhchandani (2017). Let $I = \{1, \dots, n\}$ be a set of workers and $J = \{n+1, \dots, n+m\}$ be a set of firms. Let T_i be a finite set of types for worker i. Worker i's type $t_i \in T_i$ is his private information. Denote by $t = (t_1, \dots, t_n) \in T = \times_{i=1}^{n} T_i$ a type profile for the n workers. There is a common prior $\beta \in \Delta(T)$ on workers' type profiles, and β has a full support. Each firm j's type is commonly known and is summarized by the index j. Similarly, each worker i can also have publicly observable, payoff-relevant attributes that are summarized by i.

Let $a_{ij}(t_i) \in \mathbb{R}$ and $b_{ij}(t_i) \in \mathbb{R}$ be the *matching values* worker i (with type t_i) and firm j receive, respectively, when they match. To ease notation, for a profile of workers' types $t = (t_i, t_{-i}) \in T$, we write $a_{ij}(t) \equiv a_{ij}(t_i)$ and $b_{ij}(t) \equiv b_{ij}(t_i)$. Normalize the unmatched values to be zero, i.e., $a_{ii}(t) = b_{jj}(t) = 0$. To facilitate the comparison with complete-information settings, we assume *strict preferences*, i.e.,

for each $i \in I$ and every $t \in T$, $a_{ij}(t) \neq a_{ij'}(t)$ for any $j \neq j'$, and

for each $j \in J$, $\mathbb{E}_\beta[b_{ij}|D] \neq \mathbb{E}_\beta[b_{i'j}|D']$ for any $(i, D) \neq (i', D')$, where $D, D' \subseteq T$.[7]

The latter inequality reduces to $b_{ij}(t) \neq b_{i'j}(t)$, for any $i \neq i'$, if firms have complete information at t.

We denote a *matching market* by

$$\Gamma = (I, J, t, T, \beta, a, b).$$

The solution of a matching market is referred to as *Bayesian stable states*, a concept adopted from Chen and Hu (2024). Roughly speaking, a market state specifies a matching (who matches with whom), a realized type profile (what the payoff-relevant types are) and an information structure (what each agent knows). Below we formally describe a market state and define Bayesian stability of it.

A *matching* is a one-to-one function $\mu : I \cup J \to I \cup J$ that pairs up workers and firms such that the following holds for each $i \in I$ and each $j \in J$:

(1) $\mu(i) \in J \cup \{i\}$,

(2) $\mu(j) \in I \cup \{j\}$, and

(3) $\mu(i) = j$ if and only if $\mu(j) = i$.

If $\mu(i) = i$ or $\mu(j) = j$, we say that the agent is *unmatched*.

Firms' *information structure* is described by a partition profile

$$\Pi = (\Pi_{n+1}, \ldots, \Pi_{n+m}),$$

[7]No indifference here seems to be restrictive. However, since T is finite, it only has finitely many subsets. Then if we consider the space of matching value functions and prior beliefs, indifference can only occur on a measure zero set.

where for each $j \in J$, Π_j is a partition over T. A *market state*, or simply a *state*, (μ, t, Π) specifies a matching μ, a realized type profile t and an information structure Π. The partitional formulation allows for arbitrarily heterogeneous information among firms and, more importantly, facilitates the aggregation of pieces of information when the market is unstable and evolving.[8] Say partition profile Π' is (weakly) finer than partition profile Π if, for each firm j, we have $\Pi'_j(t) \subset \Pi_j(t)$ for every type profile $t \in T$.[9]

Bayesian stability of a market state has three requirements: individual rationality, no blocking and information stability.

Definition 1. *A state (μ, t, Π) is **individually rational (IR)** if*

$$\begin{aligned} a_{i\mu(i)}(t) &\geq 0 \quad \textit{for all } i \in I \quad \textit{and} \\ \mathbb{E}_\beta\left[b_{\mu(j)j} \middle| \Pi_j(t)\right] &\geq 0 \quad \textit{for all } j \in J. \end{aligned}$$

To define a pairwise deviation (i, j) for the state (μ, t, Π) , we first clarify firm j's belief when evaluating such a deviation. Let $D_{\mu,ij}$ be the set of type profiles such that worker i strictly benefits from the rematching with firm j, i.e.,

$$D_{\mu,ij} \;:=\; \left\{t' \in T : a_{ij}(t') > a_{i\mu(i)}(t')\right\}. \tag{1}$$

Intuitively, for the pairwise deviation (i, j) to be viable, firm j must expect to benefit from rematching with worker i. When calculating her expected payoff, a type profile t' is relevant for firm j only when $t' \in D_{\mu,ij}$; any type profile that violates the inequality

[8] The aggregation of two pieces of information is simply the join of two partitions; see Aumann (1976).

[9] Some papers, e.g., Liu et al. (2014), assume observability within matched pairs, i.e., each matched firm j can observe the type of her employee $\mu(j)$, whereas others may not, e.g., Liu (2020). Let Π^μ denote the partition profile that is generated by a matching μ and the observability assumption, i.e., for every j and every t, $t' \in \Pi^\mu_j(t)$ if and only if $t'_{\mu(j)} = t_{\mu(j)}$. Then the observability assumption is equivalent to saying that in each state (μ, t, Π), Π is weakly finer than Π^μ. It is straightforward to verify that our definitions, examples and results would all go through with this assumption.

in (1) is irrelevant due to the worker's objection. Based on firm j's initial knowledge $\Pi_j(t)$ and the hypothetical knowledge $D_{\mu,ij}$, firm j's belief shall be $\beta(\cdot \big| \Pi_j(t) \cap D_{\mu,ij})$, which is referred to as "off-path" belief in Liu (2020).

Definition 2. *A state* (μ, t, Π) *is* ***blocked*** *by* (i,j) *if*

$$\begin{aligned} a_{ij}(t) \quad &> \quad a_{i\mu(i)}(t) \qquad \textit{and} \\ \mathbb{E}_\beta \left[b_{ij} \big| \Pi_j(t) \cap D_{\mu,ij} \right] \quad &> \quad \mathbb{E}_\beta \left[b_{\mu(j)j} \big| \Pi_j(t) \cap D_{\mu,ij} \right]. \end{aligned}$$

For a market state to be stable, *information stability* captures the intuition that the absence of individual/pairwise deviation provides no further information to firms (in addition to what is already described in the market state, i.e., Π). This requirement is necessary because the absence of deviation may provide additional information to firms, which in turn leads to a deviation; see Chen and Hu (2024, Example 1).

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

$$N_{\mu,\Pi} \quad := \quad \{t' \in T : (\mu, t', \Pi) \text{ is IR and not blocked}\}.$$

Moreover, let K_Π denote the meet (i.e., finest common coarsening) of the partition profile Π. Given a state (μ, t, Π), the set $K_\Pi(t)$ is the cell of the common knowledge partition that contains the realized type profile t. Intuitively, upon observing the absence of individual/pairwise deviation, each firm should refine their partitions within the common knowledge cell $K_\Pi(t)$, according to the newly acquired information that the realized type profile must lie in $N_{\mu,\Pi}$. For notational convenience, we denote by $\mathcal{N}_{\mu,\Pi}$ the binary partition that is induced by $N_{\mu,\Pi}$, i.e., $\mathcal{N}_{\mu,\Pi} := \{N_{\mu,\Pi},\ T \setminus N_{\mu,\Pi}\}$. Define an operator $H_\mu(\cdot)$ (History) to represent the information refinement as follows:

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

If $H_\mu(\Pi) = \Pi$, then the fact of individual rationality and no blocking pair provides no further information to firms (in addition to their knowledge $\Pi_j(t)$).

Definition 3. *A state* (μ, t, Π) *is* ***Bayesian stable*** *if it satisfies the following three requirements:*

(1) (μ, t, Π) *is individually rational.*

(2) (μ, t, Π) *is not blocked.*

(3) $H_\mu(\Pi) = \Pi$.

See Example 1 in the next section for an illustration of Bayesian stable states. For any matching market, the following theorem assures the existence of Bayesian stable states. We attribute the result to earlier papers because in proving existence, whether utility is transfer or not does not play a role.

Theorem (Liu, 2020, Proposition 3; Chen and Hu, 2024, Proposition 1)**.** *If* μ *is a complete-information stable matching when* t *is commonly known, then there exits* Π *such that* (μ, t, Π) *is a Bayesian stable state.*

Particularly, any Π with $K_\Pi(t) = \{t\}$ is desirable. When T is a singleton, information becomes complete. Then, stability of a matching is defined merely by individual rationality and no blocking, as in Gale and Shapley (1962). In this case, the theorem reduces to Gale-Shapley's existence result.

In the complete-information setting, the structure of the set of stable matchings is well understood; see Roth and Sotomayor (1990) for a comprehensive survey. Here, we investigate two celebrated structural results, the lone-wolf theorem and the lattice theorem.[10,11] Some related structural results can be discussed once (and only if) we restore these two results.

[10]The term "lone-wolf" is adopted from Gusfield and Irving (1989, Theorem 4.5.2).

[11]The opposition of interests between the two sides of the market (Roth and Sotomayor, 1990, Theorem 2.13, attributed to Knuth) and the existence of extreme matchings (Gale and Shapley, 1962) can be viewed as corollaries of the lattice theorem.

3 The lone-wolf theorem

We recap the benchmark theorem in the complete-information setting first.

Theorem (Gale and Sotomayor, 1985*a*; 1985*b*)**.** *The set of unmatched agents is the same in all stable matchings.*

Example 1 (Failure of the lone-wolf theorem with incomplete information)**.**
Consider a matching market Γ, where $I = \{i\}$, $J = \{j, j'\}$, the true type profile is t^*, and the other ingredients of Γ are given as follows:

	β	a_{ij}	b_{ij}	$a_{ij'}$	$b_{ij'}$
t^*	0.5	2	6	4	8
t	0.5	2	2	4	-9

In words, T contains two possible type profiles t^* and t, which are equally probable. Worker i prefers firm j' to firm j under both type profiles. Firms' matchings values with worker i are higher under t^* than under t.

Fix the true type profile t^*. There are two possible kinds of Bayesian stable states in market Γ: either i is matched with j or with j', associated with proper information structure. Any state with i unmatched is blocked by (i, j). Two candidate states are described below.

(1) (μ, t^*, Π), where $\mu(i) = j$, $\Pi_j = \{\{t^*\}, \{t\}\}$, and $\Pi_{j'} = \{\{t^*, t\}\}$.

(2) $(\hat{\mu}, t^*, \hat{\Pi})$, where $\hat{\mu}(i) = j'$, $\hat{\Pi}_j = \{\{t^*\}, \{t\}\}$, and $\hat{\Pi}_{j'} = \{\{t^*\}, \{t\}\}$.

We proceed to verify that the two states are indeed Bayesian stable. First, state (μ, t^*, Π) is individually rational, and it is not blocked by (i, j') since j' would *expect* a negative matching value. Moreover, given μ and Π, both (μ, t^*, Π) and (μ, t, Π) are individually rational and not blocked, which means that $N_{\mu,\Pi} = \{t^*, t\}$. As a result, information stability holds as well. Therefore, state (μ, t^*, Π) is Bayesian stable. Similarly, state $(\hat{\mu}, t^*, \hat{\Pi})$ is individually rational, and it is not blocked by (i, j) since i would have a strictly lower matching value, i.e., $D_{\hat{\mu},ij} = \emptyset$. Moreover, it is straightforward to verify that information stability holds as well. Therefore, state $(\hat{\mu}, t^*, \hat{\Pi})$ is Bayesian stable, where $\hat{\mu}$ is actually the complete-information stable matching at t^*.

Obviously, the sets of unmatched agents differ in the two Bayesian stable states.

Below we provide a sufficient condition to restore the lone-wolf theorem.

Proposition 1. *Let (μ, t, Π) and $(\hat{\mu}, t, \hat{\Pi})$ be two Bayesian stable states.*[12] *If $\hat{\Pi} = \Pi$, then the set of unmatched agents is the same in two states.*

There are four remarks. First, to have the lone-wolf theorem, agents do not need to have complete information; instead, "common" information is sufficient.[13] In this sense, Proposition 1 generalizes Gale and Sotomayor's theorem. Furthermore, in the more general setting with incomplete information, the lone-wolf theorem actually does not fail. Instead, it holds more generally: If we stratify the set of Bayesian stable states according to the information structure, then the lone-wolf theorem holds in *every* stratum. The complete-information lone-wolf theorem is only one of them.

Second, one may doubt whether the common information condition imposed on Π is primitive or not, given that it is part of the solution—Bayesian stable states. Just as the set of possible matchings being primitive, the set of possible information structures is also primitive. The information structure in a Bayesian stable state is merely in a special subset of the possible ones that are immune to updating. That is, the outcome coincides with the primitive. It is important to address the theoretical

[12] Market state is an interim concept where the true type profile has been realized. Hence, it makes little sense to compare Bayesian stable states at two different type profiles.

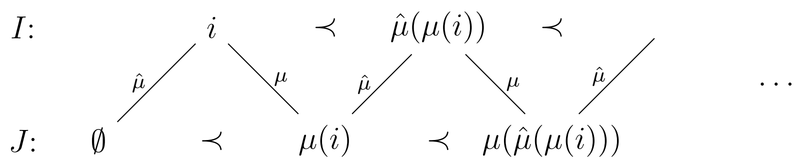
[13] Here, the partition profile is common across two states. We do not mean the strong sense that firms have "common" partitions.

question of the existence of Bayesian stable states, which has been done.

Third, $\hat{\Pi} = \Pi$ is sufficient, but not necessary. In Example 2 (after the proof of Proposition 1 below), the set of matched agents is the same across two Bayesian stable states, but the partition profiles differ.

Fourth, when comparing two Bayesian stable states, more information can lead to either more matched agents or less, as illustrated in Examples 3-4.[14] Nevertheless, regarding different partition profiles, we have at least one implication from Proposition 1: To favor rural hospitals (Roth, 1984), the social planner has to twist the information that agents have so to avoid common (especially complete) information.

Proof of Proposition 1. We prove the result by contradiction. Suppose worker i is matched under μ but unmatched under $\hat{\mu}$. Then by the individual rationality (IR) of (μ, t, Π), worker i prefers $\mu(i)$ to being unmatched, i.e., $a_{i\mu(i)}(t) > 0$. Furthermore, we have $a_{i\mu(i)}(t') > 0$ for all $t' \in K_\Pi(t)$; otherwise, the fact of IR would help firms refine their information within $K_\Pi(t)$, violating information stability of (μ, t, Π).



We claim that firm $\mu(i)$ must be matched in $(\hat{\mu}, t, \hat{\Pi})$. Otherwise, both i and $\mu(i)$ are unmatched under $\hat{\mu}$. Moreover, by individual rationality of (μ, t, Π) and the strictness of preferences, firm $\mu(i)$ enjoys a positive expected payoff in (μ, t, Π), i.e., $\mathbb{E}_\beta \left[b_{i\mu(i)} | \Pi_{\mu(i)}(t)\right] > 0$. Now since firm $\mu(i)$ has the same amount of information in

[14] One may imagine a "maximal-domain" result like the following: Fix I, J, t, T, and β. As long as $\hat{\Pi} \neq \Pi$, there exist a and b such that the sets of unmatched agents in two stable states differ. However, we do not view this as sensible. Particularly, when we compare two Bayesian stable states, $\hat{\Pi}$ and Π are part of the states. Fixing $\hat{\Pi}$ and Π but adjusting the market primitives a and b would undermine the legitimacy of the stable states under comparison.

$(\hat{\mu}, t, \hat{\Pi})$ (particularly about worker i), and

$$D_{\hat{\mu},i\mu(i)} = \{t' \in T : a_{i\mu(i)}(t') > 0\} \supseteq K_{\Pi}(t) = K_{\hat{\Pi}}(t), \tag{2}$$

she would expect a positive matching value with worker i, i.e.,

$$\mathbb{E}_\beta \left[b_{i\mu(i)} | \hat{\Pi}_{\mu(i)}(t) \cap D_{\hat{\mu},i\mu(i)} \right] = \mathbb{E}_\beta \left[b_{i\mu(i)} | \Pi_{\mu(i)}(t) \right] > 0.$$

Therefore, $(i, \mu(i))$ constitutes a blocking pair for $(\hat{\mu}, t, \hat{\Pi})$, contradicting its stability. Hence, firm $\mu(i)$ must be matched in $(\hat{\mu}, t, \hat{\Pi})$.

Next, we claim that firm $\mu(i)$ prefers her partner under $\hat{\mu}$ to worker i, i.e.,

$$\mathbb{E}_\beta \left[b_{\hat{\mu}(\mu(i)),\mu(i)} | \hat{\Pi}_{\mu(i)}(t) \right] > \mathbb{E}_\beta \left[b_{i\mu(i)} | \hat{\Pi}_{\mu(i)}(t) \right] = \mathbb{E}_\beta \left[b_{i\mu(i)} | \Pi_{\mu(i)}(t) \right]. \tag{3}$$

Otherwise, we have

$$\mathbb{E}_\beta \left[b_{i\mu(i)} | \hat{\Pi}_{\mu(i)}(t) \right] > \mathbb{E}_\beta \left[b_{\hat{\mu}(\mu(i)),\mu(i)} | \hat{\Pi}_{\mu(i)}(t) \right]$$

by the strictness of preferences. Since i is unmatched under $\hat{\mu}$ and $D_{\hat{\mu},i\mu(i)}$ in (2) reveals no more information to firm $\mu(i)$ than $\hat{\Pi}_{\mu(i)}(t)$, we know that $(i, \mu(i))$ constitutes a blocking pair for $(\hat{\mu}, t, \hat{\Pi})$, contradicting its stability. Hence, (3) holds. Furthermore, we have

$$\mathbb{E}_\beta \left[b_{\hat{\mu}(\mu(i)),\mu(i)} | \hat{\Pi}_{\mu(i)}(t') \right] > \mathbb{E}_\beta \left[b_{i\mu(i)} | \hat{\Pi}_{\mu(i)}(t') \right] \text{ for all } t' \in K_{\hat{\Pi}}(t);$$

otherwise, the fact of no blocking for $(\hat{\mu}, t, \hat{\Pi})$ would help firms refine their information within $K_{\hat{\Pi}}(t)$, violating information stability of $(\hat{\mu}, t, \hat{\Pi})$.

Similarly, we can argue that worker $\hat{\mu}(\mu(i))$ has to be matched under μ and he prefers $\mu(\hat{\mu}(\mu(i)))$ to $\mu(i) = \hat{\mu}(\hat{\mu}(\mu(i)))$ at all type profiles in $K_{\Pi}(t)$.

Continuing the arguments above would give us an alternative chain of workers and firms, and the particular preference relations where every agent prefers her/his successor to her/his predecessor at all type profiles in the common knowledge event $K_{\Pi}(t)$. Since I and J are both finite, and all agents along the chain are distinct, the construction of the chain stops in finite steps. By construction, the last agent prefers to be unmatched rather than being matched, violating the individual rationality of either (μ, t, Π) or $(\hat{\mu}, t, \hat{\Pi})$. This is the contradiction we intended to have at the beginning of this proof. Hence, there could not be a worker who is matched under μ but unmatched under $\hat{\mu}$. The argument is symmetric if a firm is matched under μ but unmatched under $\hat{\mu}$. $\square$

We close this section by three more examples, showing that without the sufficient condition, the sets of unmatched agents may be the same and may also differ in various possible ways.

Example 2 (The set of matched agents is the *same*, but $\hat{\Pi} \neq \Pi$)**.**

Consider a matching market Γ, where $I = \{i, i'\}$, $J = \{j, j'\}$, the true type profile is t^*, and the other ingredients of Γ are given as follows:

	β	a_{ij}	b_{ij}	$a_{ij'}$	$b_{ij'}$	$a_{i'j}$	$b_{i'j}$	$a_{i'j'}$	$b_{i'j'}$
t^*	0.5	4	4	2	3	2	3	4	4
t	0.5	4	0	2	3	2	3	4	0

In words, T contains two possible type profiles t^* and t, which are equally probable. Matches (i, j) and (i', j') are desirable for all agents at t^*, whereas at t, firms prefer matches (i', j) and (i, j').

Fix the true type profile t^*. We consider two states in market Γ, as follows:

(1) (μ, t^*, Π), where $\mu(i) = j'$, $\mu(i') = j$, $\Pi_j = \{\{t^*, t\}\}$, and $\Pi_{j'} = \{\{t^*, t\}\}$.

(2) $(\hat{\mu}, t^*, \hat{\Pi})$, where $\hat{\mu}(i) = j$, $\hat{\mu}(i') = j'$, $\hat{\Pi}_j = \{\{t^*\}, \{t\}\}$, and $\hat{\Pi}_{j'} = \{\{t^*\}, \{t\}\}$.

It is straightforward to verify that both of them are Bayesian stable. Obviously, the set of matched agents is the same, but $\hat{\Pi}$ is strictly finer than Π.

Example 3 (More information lead to less matched agents)**.**

Consider a matching market Γ, where $I = \{i, i'\}$, $J = \{j, j'\}$, the true type profile is t^*, and the other ingredients of Γ are given as follows:

	β	a_{ij}	b_{ij}	$a_{ij'}$	$b_{ij'}$	$a_{i'j}$	$b_{i'j}$	$a_{i'j'}$	$b_{i'j'}$
t^*	0.5	4	4	2	3	2	3	-1	-1
t	0.5	4	0	2	3	2	3	-1	0

This is a slight modification of the market in Example 2, only in that (i', j') becomes a really bad match for both i' and j'.

Fix the true type profile t^*. We consider two states in market Γ, as follows:

(1) (μ, t^*, Π), where $\mu(i) = j'$, $\mu(i') = j$, $\Pi_j = \{\{t^*, t\}\}$, and $\Pi_{j'} = \{\{t^*, t\}\}$.

(2) $(\hat{\mu}, t^*, \hat{\Pi})$, where $\hat{\mu}(i) = j$, $\hat{\mu}(i') = i'$, $\hat{\Pi}_j = \{\{t^*\}, \{t\}\}$, and $\hat{\Pi}_{j'} = \{\{t^*\}, \{t\}\}$.

It is straightforward to verify that both states are Bayesian stable. Obviously, $\hat{\Pi}$ is strictly finer than Π, and agents i' and j' who are matched in (μ, t^*, Π) are unmatched in $(\hat{\mu}, t^*, \hat{\Pi})$.

Example 4 (More information lead to more matched agents)**.**
Consider a matching market Γ, where $I = \{i\}$, $J = \{j\}$, the true type profile is t^*, and the other ingredients of Γ are given as follows:

	β	a_{ij}	b_{ij}
t^*	0.5	2	3
t	0.5	2	-4

Clearly, firm j prefers a match with worker i if she knows the true type t^*.

Fix the true type profile t^*. We consider two states in market Γ, as follows:

(1) (μ, t^*, Π), where $\mu(i) = i$ and $\Pi_j = \{\{t^*, t\}\}$.

(2) $(\hat{\mu}, t^*, \hat{\Pi})$, where $\hat{\mu}(i) = j$, and $\hat{\Pi}_j = \{\{t^*\}, \{t\}\}$.

It is straightforward to verify that both states are Bayesian stable. Obviously, $\hat{\Pi}$ is strictly finer than Π and agents i and j who are unmatched in (μ, t^*, Π) are matched in $(\hat{\mu}, t^*, \hat{\Pi})$.

4 The lattice theorem

We recap the benchmark theorem in the complete-information setting first. In the complete-information setting with t being the true type profile, for any two matchings μ and $\hat{\mu}$, we define the following function on the set $I \cup J$, which can also be seen as operations on the set of all matchings. Let $\lambda = \mu \vee_I \hat{\mu}$, be defined by

$$\lambda(i) = \begin{cases} \mu(i) & \text{if } a_{i\mu(i)}(t) > a_{i\hat{\mu}(i)}(t) \\ \hat{\mu}(i) & \text{otherwise} \end{cases} \quad \text{for all } i \in I, \text{ and} \tag{4}$$

$$\lambda(j) = \begin{cases} \hat{\mu}(j) & \text{if } b_{\mu(j)j}(t) > b_{\hat{\mu}(i)j}(t) \\ \mu(j) & \text{otherwise} \end{cases} \quad \text{for all } j \in J. \tag{5}$$

In a precisely similar way we can define the function $\nu = \mu \wedge_I \hat{\mu}$, which gives each worker his less preferred firm and each firm her more preferred worker.

Theorem (Knuth, 1976, attributed to John Conwey)**.** *If μ and $\hat{\mu}$ are stable matchings, then the functions $\lambda = \mu \vee_I \hat{\mu}$ and $\nu = \mu \wedge_I \hat{\mu}$ are both matchings. Furthermore, they are both stable.*

Example 1 illustrates that with incomplete information, the lattice theorem fails. Particularly, the definition of $\nu = \mu \wedge_I \hat{\mu}$ leads to $\nu(j) = i$, $\nu(j') = i$, and $\nu(i) = i$, which is not a matching.

We proceed to provide sufficient conditions to restore the lattice theorem. We would like to keep the "common" information condition. Nevertheless, since new states will be produced (unlike in Proposition 1, where we only compare two given states), the condition $\hat{\Pi} = \Pi$ becomes inadequate to restore the theorem. This is illustrated in Example 5 (after Proposition 2 below). We borrow two intuitive conditions from Liu (2020), each of which is sufficient (together with $\hat{\Pi} = \Pi$) for restoring the lattice theorem.[15]

[15] Liu (2020) has another condition that involves independent "on-path" beliefs. Since the idea

Assumption 1. *$a_{ij}(t_i) = a_{ij}(t_i')$ for any $t_i, t_i' \in T_i, i \in I$, and $j \in J$.*

Assumption 2. *There exist functions $g : I \times T_i \to \mathbb{R}$ and $h : I \times J \to \mathbb{R}$ such that $a_{ij}(t_i) = g(i, t_i) + h(i, j)$ for any $t_i \in T_i$, $i \in I$, and $j \in J$.*

Assumption 2 is weaker than Assumption 1. For interpretations, if workers care only about the observable identities of firms, then Assumption 1 is satisfied. Moreover, in many classic adverse-selection models such as signaling and screening, Assumption 2 is satisfied with $a_{ij}(t_i) = g(i, t_i)$, i.e., a worker does not value which firm he works for, but his own types may affect his costs of effort in working.

The following proposition is our restored lattice theorem. (In defining λ and ν, the firms' matches are determined by their expected matching values under Π.)

Proposition 2. *Let (μ, t, Π) and $(\hat{\mu}, t, \hat{\Pi})$ be two Bayesian stable states. If Assumption 2 holds and $\hat{\Pi} = \Pi$, then both $\lambda = \mu \vee_I \hat{\mu}$ and $\nu = \mu \wedge_I \hat{\mu}$ are matchings; moreover, both (λ, t, Π) and (ν, t, Π) are Bayesian stable states.*

There are four remarks. First, again, to have the lone-wolf theorem, agents do not need to have complete information; instead, "common" information is sufficient in those markets satisfying Assumption 2. In this sense, Proposition 2 generalizes Conway's theorem in those markets. Furthermore, in the more general setting with incomplete information, the lattice theorem actually does not fail. Instead, it holds more generally: If we stratify the set of Bayesian stable states according to the information structure Π, then the lattice theorem holds in *every* stratum. The complete-information lattice theorem is only one of them.

Second, one may suspect that the conditions in Proposition 2 is so strong that it may push us back to the complete-information case. However, the remark after

---

of adopting this third one is pretty the same as adopting the other two, and since we prefer not to introduce "on-path" beliefs in this paper, we leave it to interested readers to restore the lattice theorem under Liu's third condition; see Liu (2020, Proposition 5 on page 2647).

Example 5 says that the proposition is indeed more general: A slight modification of Example 5 shows that we could apply the proposition to a complete-information stable state (μ, t, Π) and a Bayesian stable state $(\hat{\mu}, t, \hat{\Pi})$ which is *not* complete-information stable.

Third, since we have introduced both the join between two matchings as in (4)-(5) and the join between two partitions in defining information stability, one may wonder what the combined effect could be. For example, if two Bayesian stable state have *different* partition profiles, can we obtain a Bayesian stable state by taking the join of matchings and the join (or meet) of partitions? We think the operators on different partitions may not help to restore the lattice theorem. To wit, we consider Example 1 again. The failure of the lattice property arises because of the failure of the lone-wolf property.[16] So a proper matching cannot arise, before we worry about stability. By Proposition 1, we need common information to first guarantee the lone-wolf property before we study the lattice property. Beyond common information, Examples 1, 3 and 4 indicate that our knowledge about the lone-wolf property is limited, which is itself an open question before we study the lattice property.

Fourth, following the discussion in the previous remark, can we focus on the structure of the set of Bayesian stable states with a common matching (and different partition profiles)? In the belief-free setting studied by Liu et al. (2014) and generalized by Chen and Hu (2020), if a state (μ, t, Π) is (incomplete-information) stable and $\hat{\Pi}$ is coarser than Π, then the state $(\mu, t, \hat{\Pi})$ is "essentially stable" in the sense that $(\mu, t, H^l(\hat{\Pi}))$ is never blocked for any $l = 0, 1, 2, \ldots$, and thus (incomplete-information) stable for some l. In the Bayesian setting, this is no longer true. Consider the simple

[16] Generally, the lone-wolf property may be a necessary condition for the lattice property. For a quick proof in a generic case, consider an unbalanced market with more firms than workers. Suppose every firm-worker pair is mutually acceptable. Then all stable matchings must have workers fully matched. If the sets of unmatched firms differ in two stable matchings, then the firm-join of the two matchings is not a matching, before we are concerned with any stability issue.

Example 4 but modify it by (1) specifying t as the true type and (2) replacing -4 with -2. Then $(\mu, t, \hat{\Pi})$ with $\mu(i) = i$ and $\hat{\Pi}_j = \{\{t^*\}, \{t\}\}$ is Bayesian stable. However, the state (μ, t, Π), where $\Pi_j = \{\{t^*, t\}\}$ is coarser than $\hat{\Pi}_j$, is blocked by (i, j).

Two corollaries are immediate:

Corollary 1. *Suppose Assumption 2 holds. Then for every Π, if the set of Bayesian stable states with common information Π, i.e.,*

$$S(t, \Pi) := \left\{(\hat{\mu}, t, \hat{\Pi}) : (\hat{\mu}, t, \hat{\Pi}) \text{ is Bayesian stable and } \hat{\Pi} = \Pi\right\}$$

is nonempty, then it admits a worker-optimal state and a firm-optimal state.

Corollary 2. *Suppose Assumption 2 holds. If (μ, t, Π) and $(\hat{\mu}, t, \Pi)$ are two Bayesian stable states, then all workers like (μ, t, Π) at least as well as $(\hat{\mu}, t, \Pi)$ if and only if all firms like $(\hat{\mu}, t, \Pi)$ at least as well as (μ, t, Π).*

Proof of Proposition 2. We prove the claim for $\lambda = \mu \vee_I \hat{\mu}$; that for ν is symmetric.

Part 1. λ is a matching. It suffices to show that $\lambda(i) = j$ if and only if $\lambda(j) = i$.

To prove the "only-if" direction, suppose that $\lambda(i) = j$ and, without loss of generality, that $j = \mu(i)$. If $\mu(i) = \hat{\mu}(i)$, $\lambda(i) = j$ automatically implies $\lambda(j) = i$. So we suppose $\mu(i) \neq \hat{\mu}(i)$.

By the definition of λ, we have

$$a_{ij}(t) > a_{i\hat{\mu}(i)}(t), \tag{6}$$

which implies that $j \in J$. Then by $\hat{\Pi} = \Pi$ and Proposition 1, we have $\hat{\mu}(i) \in J$ as well. By Assumption 2, we rewrite (6) as

$$g(i, t_i) + h(i, j) > g(i, t_i) + h(i, \hat{\mu}(i)),$$

which then implies $D_{\hat{\mu}, ij}$ is nonempty (which contains t) and thus $D_{\hat{\mu}, ij} = T$.

We claim that

$$\mathbb{E}_\beta\left[b_{ij} | \hat{\Pi}_j(t)\right] < \mathbb{E}_\beta\left[b_{\hat{\mu}(j)j} | \hat{\Pi}_j(t)\right]. \tag{7}$$

Otherwise, the strictness of preferences imply that

$$\begin{aligned}
\mathbb{E}_\beta\left[b_{ij}|\hat{\Pi}_j(t)\cap D_{\hat{\mu},ij}\right] &= \mathbb{E}_\beta\left[b_{ij}|\hat{\Pi}_j(t)\right] \\
&> \mathbb{E}_\beta\left[b_{\hat{\mu}(j)j}|\hat{\Pi}_j(t)\right] \\
&= \mathbb{E}_\beta\left[b_{\hat{\mu}(j)j}|\hat{\Pi}_j(t)\cap D_{\hat{\mu},ij}\right].
\end{aligned}$$

Then (i,j) constitutes a blocking pair for $(\hat{\mu},t,\hat{\Pi})$, contradicting its stability. Hence, (7) holds and $\lambda(j)=i$ follows.

The "if" direction is adopted from Roth and Sotomayor (1990, page 36), which is included here for completeness.[17] Let I' be the set of matched workers under λ, i.e.,

$$I'=\{i:\lambda(i)\in J\}=\{i:\mu(i)\in J \text{ or } \hat{\mu}(i)\in J\}$$

and J' be the set of matched workers under λ, i.e.,

$$J'=\{j:\lambda(j)\in I\}=\{j:\mu(j)\in I \text{ and } \hat{\mu}(j)\in I\}.$$

Three facts are useful:

(1) $\lambda(I')$ is contained in J', since, by the only-if direction, $\lambda(i)=j$ implies $\lambda(j)=i$.

(2) $\lambda(I')$ is the same size as I', since $\lambda(i)=\lambda(i')=j$ implies $i=i'=\lambda(j)$.

(3) I' contains $\mu(J')$, which is the same size as J', since J' is only a *subset* of matched firms under μ whereas I' contains *all* matched workers under μ.

Therefore, $\lambda(I')$ and J' are the same size and thus $\lambda(I')=J'$. Now suppose $\lambda(j)=i$. Then $j\in J'$. Since $\lambda(I')=J'$, there exists $i'\in I'$ such that $\lambda(i')=j$. By the only-if direction, $\lambda(j)=i'$, which implies $i'=i$ and thus $\lambda(i)=j$.

Part 2. (λ,t,Π) is a Bayesian stable state. Obviously, (λ,t,Π) is individually rational as both (μ,t,Π) and $(\hat{\mu},t,\hat{\Pi})$ are.

[17] The identical argument applies because this step does not involve incomplete information.

We show there is no blocking for (λ, t, Π) by contradiction. Suppose (i', j') is a blocking pair for (λ, t, Π), i.e.,

$$a_{i'j'}(t) \;\; > \;\; a_{i'\lambda(i')}(t) \quad \text{and} \tag{8}$$

$$\mathbb{E}_{\beta}\left[b_{i'j'}|\hat{\Pi}_{j'}(t) \cap D_{\lambda,i'j'}\right] \;\; > \;\; \mathbb{E}_{\beta}\left[b_{\lambda(j')j'}|\hat{\Pi}_{j'}(t) \cap D_{\lambda,i'j'}\right]. \tag{9}$$

If $\lambda(i') = i'$, then i' is unmatched in both (μ, t, Π) and $(\hat{\mu}, t, \hat{\Pi})$, which implies notationally that $\lambda(i') = \mu(i') = \hat{\mu}(i')$. Then conditions (8)-(9) would say that either (μ, t, Π) is blocked, when $\lambda(j') = \mu(j')$, or $(\hat{\mu}, t, \hat{\Pi})$ is blocked, when $\lambda(j') = \hat{\mu}(j')$. This contradicts to the Bayesian stability of (μ, t, Π) or $(\hat{\mu}, t, \hat{\Pi})$, respectively.

If $\lambda(i') \in J$, then by $\hat{\Pi} = \Pi$ and Proposition 1, we have $\mu(i') \in J$ and $\hat{\mu}(i) \in J$. By Assumption 2, we rewrite (8) as

$$g\left(i', t_{i'}\right) + h(i', j') \;\; > \;\; g\left(i', t_{i'}\right) + h(i', \mu(i')), \quad \text{and} \tag{10}$$

$$g\left(i', t_{i'}\right) + h(i', j') \;\; > \;\; g\left(i', t_{i'}\right) + h(i', \hat{\mu}(i')). \tag{11}$$

Therefore, $t \in D_{\mu,i'j'}$ and $t \in D_{\hat{\mu},i'j'}$, which imply, respectively, that $D_{\mu,i'j'} = T$ and $D_{\hat{\mu},i'j'} = T$. Consider four cases:

(1) $\lambda(i') = \mu(i')$ and $\lambda(j') = \mu(j')$.

(2) $\lambda(i') = \hat{\mu}(i')$ and $\lambda(j') = \hat{\mu}(j')$.

(3) $\lambda(i') = \mu(i')$ and $\lambda(j') = \hat{\mu}(j')$.

(4) $\lambda(i') = \hat{\mu}(i')$ and $\lambda(j') = \mu(j')$.

In the first case, conditions (8)-(9) indicate that (μ, t, Π) is blocked by (i', j'), a contradiction. Similarly, in the second case, we can derive a contradiction to the stability of $(\hat{\mu}, t, \hat{\Pi})$. In the third case, conditions (11) and (9) indicate that $(\hat{\mu}, t, \hat{\Pi})$ is blocked, a contradiction. Similarly, in the last case, we can derive a contradiction to the stability of (μ, t, Π) by conditions (10) and (9).

Now we verify the information stability requirement for (λ, t, Π). Since (μ, t, Π)

and $(\hat{\mu}, t, \hat{\Pi})$ are Bayesian stable, we have, respectively,

$$N_{\mu,\Pi} \supseteq K_{\Pi}(t) \quad \text{and} \quad N_{\hat{\mu},\hat{\Pi}} \supseteq K_{\hat{\Pi}}(t).^{18} \tag{12}$$

We have seen in the above argument that if (i', j') is a blocking pair for (λ, t, Π), then it has to be a blocking pair for either (μ, t, Π) or $(\hat{\mu}, t, \hat{\Pi})$. By the identical argument, if (i', j') is a blocking pair for (λ, t', Π), then it has to be a blocking pair for either (μ, t', Π) or $(\hat{\mu}, t', \hat{\Pi})$. Thus, we have

$$T \setminus N_{\lambda,\Pi} \subseteq [T \setminus N_{\mu,\Pi}] \cup [T \setminus N_{\hat{\mu},\hat{\Pi}}].$$

Therefore, we know that

$$N_{\lambda,\Pi} \supseteq N_{\mu,\Pi} \cap N_{\hat{\mu},\hat{\Pi}} \supseteq K_{\Pi}(t),$$

where the second inclusion follows from (12) and the condition $\hat{\Pi} = \Pi$. Hence, (λ, t, Π) satisfies information stability, and thus is Bayesian stable. □

The following example shows that the *common information* condition alone is insufficient to restore the lattice theorem.

Example 5 (Failure of the lattice theorem when $\hat{\Pi} = \Pi$)**.**

Consider a matching market Γ, where $I = \{i_1, i_2, i_3, i_4, i_5\}$, $J = \{j_1, j_2, j_3, j_4, j_5\}$, the true type profile is t^*, $T = \{t^*, t\}$, $\beta(t^*) = \beta(t) = \frac{1}{2}$, and the matchings values are given as follows, where all omitted matching values are -1:

[18] It is straightforward to verify that this set-inclusion condition is equivalent to our information stability requirement in Definition 3.

$a_{ij}(t^*)$ $b_{ij}(t^*)$ / $a_{ij}(t)$ $b_{ij}(t)$	j_1		j_2		j_3		j_4		j_5	
i_1	3	6	5	4	1	6	**7**	**10**		
	3	6	5	4	1	6	**4**	**6**		
i_2	5	4	3	2	1	4				
	5	**-2**	3	2	1	4				
i_3	5	2	3	6	1	2				
	5	2	3	6	1	2				
i_4							7	11	8	9
							7	11	8	9
i_5							8	9	7	11
							8	9	7	11

As indicated by the matching values, Γ contains two mostly separated sub-markets, i.e., the $l$arger Γ^l with $I^l = \{i_1, i_2, i_3\}$ and $J^l = \{j_1, j_2, j_3\}$, and the $s$maller Γ^s with $I^s = \{i_4, i_5\}$ and $J^s = \{j_4, j_5\}$.

The larger market Γ^l has a unique complete-information stable matching μ, at the true type profile t^*:

$$\mu(i_1) = j_1, \quad \mu(i_2) = j_3, \quad \mu(i_3) = j_2.$$

An alternative matching $\hat{\mu}$ that interests us is as follows:

$$\hat{\mu}(i_1) = j_2, \quad \hat{\mu}(i_2) = j_3, \quad \hat{\mu}(i_3) = j_1.$$

Given the true type profile being t^*, $\hat{\mu}$ admits a unique blocking pair (i_2, j_1). The smaller market Γ^s has two complete-information stable matchings, which we abuse notations and still denote by μ and $\hat{\mu}$:

$$\mu(i_4) = j_5, \quad \mu(i_5) = j_4; \qquad \text{and}$$
$$\hat{\mu}(i_4) = j_4, \quad \hat{\mu}(i_5) = j_5.$$

The two sub-markets are only connected through the pair i_1 and j_4, i.e., any other match between an agent from Γ^l and an agent from Γ^s is not individually rational.

Fix the true type profile t^*. We consider two states in market Γ, as follows:

(1) (μ, t^*, Π), where $\Pi_j = \{\{t^*, t\}\}$ for all $j \in J$.

(2) $(\hat{\mu}, t^*, \hat{\Pi})$, where $\hat{\Pi}_j = \{\{t^*, t\}\}$ for all $j \in J$.

Obviously, both states are individually rational.

Claim 1. (μ, t^*, Π) *is Bayesian stable.*

Proof. We verify that neither (μ, t^*, Π) nor (μ, t, Π) is blocked, and thus (μ, t^*, Π) is Bayesian stable. The only potential blocking pair to consider is (i_1, j_4). However, since

$$D_{\mu, i_1 j_4} = \{t' \in T : a_{i_1 j_4} > a_{i_1 j_1}\} = T,$$

we know that

$$\begin{aligned} \mathbb{E}_\beta\left[b_{i_1 j_4} | \Pi_{j_4}(t'') \cap D_{\mu, i_1 j_4}\right] &= \frac{1}{2} \cdot 6 + \frac{1}{2} \cdot 10 \\ &< 9 \\ &= \mathbb{E}_\beta\left[b_{i_5 j_4} | \Pi_{j_4}(t'') \cap D_{\mu, i_1 j_4}\right] \end{aligned}$$

for both $t'' \in T$. Therefore, neither (μ, t^*, Π) nor (μ, t, Π) is blocked. Then we have $N_{\mu, \Pi} = T$ and thus (μ, t^*, Π) is Bayesian stable. □

Claim 2. $(\hat{\mu}, t^*, \hat{\Pi})$ *is Bayesian stable.*

Proof. We verify that neither $(\hat{\mu}, t^*, \hat{\Pi})$ nor $(\hat{\mu}, t, \hat{\Pi})$ is blocked, and thus $(\hat{\mu}, t^*, \hat{\Pi})$ is Bayesian stable. There are two potential blocking pairs to check: (i_2, j_1) and (i_1, j_4). For (i_2, j_1), we have

$$D_{\hat{\mu}, i_2 j_1} = \{t' \in T : a_{i_2 j_1} > a_{i_3 j_1}\} = T.$$

Then

$$\begin{aligned}\mathbb{E}_\beta\left[b_{i_2j_1}|\Pi_{j_1}(t'')\cap D_{\hat{\mu},i_2j_1}\right] &= \frac{1}{2}\cdot 4+\frac{1}{2}\cdot(-2)\\ &< 2\\ &= \mathbb{E}_\beta\left[b_{i_3j_1}|\Pi_{j_1}(t'')\cap D_{\hat{\mu},i_2j_1}\right]\end{aligned}$$

for both $t''\in T$. Thus, (i_2,j_1) is neither a blocking pair for $(\hat{\mu},t^*,\hat{\Pi})$ nor for $(\hat{\mu},t,\hat{\Pi})$. For (i_1,j_4), we have

$$D_{\hat{\mu},i_1j_4}=\{t'\in T: a_{i_1j_4}>a_{i_1j_2}\}=\{t^*\}.$$

However, $b_{i_1j_4}(t^*)=10<11=b_{i_4j_4}(t^*)$. Therefore, (i_1,j_4) is neither a blocking pair for $(\hat{\mu},t^*,\hat{\Pi})$ nor for $(\hat{\mu},t,\hat{\Pi})$. Then we have $N_{\hat{\mu},\hat{\Pi}}=T$ and thus $(\hat{\mu},t^*,\hat{\Pi})$ is Bayesian stable. □

However, if we derive $\lambda=\mu\vee_I\hat{\mu}$ as

$$\lambda(i_1)=j_2,\quad \lambda(i_2)=j_3,\quad \lambda(i_3)=j_1,\quad \lambda(i_4)=j_5,\quad \lambda(i_5)=j_4,$$

the state (λ,t^*,Π) is no longer stable. Particularly, since

$$D_{\lambda,i_1j_4}=\{t'\in T: a_{i_1j_4}>a_{i_1j_2}\}=\{t^*\}$$

and

$$b_{i_1j_4}(t^*)=10>9=b_{i_5j_4}(t^*),$$

we know that (λ,t^*,Π) is blocked by (i_1,j_4). This completes the example.

A key feature is as follows: There is a particular blocking pair (i_1,j_4) to consider. In state (μ,t^*,Π), j_4 wants to be rematched to i_1 only at t^*, but D_{μ,i_1j_4} is *uninformative* in distinguishing t^* from an alternative t. In state $(\hat{\mu},t^*,\hat{\Pi})$, although $D_{\hat{\mu},i_1j_4}$ is informative, j_4 *does not* want to be rematched to i_1. Finally, in the derived state (λ,t^*,Π) where $\lambda=\mu\vee_I\hat{\mu}$, D_{λ,i_1j_4} is informative, and j_4 wants rematching; so (λ,t^*,Π) is blocked. Assumption 2 helped in excluding such a case.

Example 5 also serves as an illustration of Proposition 2, *if we replace all* *matchings values between* i_1 *and* j_4 *by* -1. Particularly, the proposition says strictly more than the complete-information lattice theorem: In the *revised* example, μ is a complete-information stable matching, neither $\hat{\mu}$ nor λ is complete-information stable, but both $(\hat{\mu}, t^*, \hat{\Pi})$ and $(\lambda, t^*, \hat{\Pi})$ are Bayesian stable states.

5 Concluding remarks

Here are some potential future directions, though this is by no means exhaustive:

(1) *Transferable utility.* We have focused on the non-transferable utility setting in order to isolate the phenomena initiated by incomplete information, without confounding them with the role of transfers. This facilitates the comparison between our results and those following Gale and Shapley (1962).

However, the transferable utility setting is worth investigating for at least two reasons: First, the setting itself has a broad range of applications; see Shapley and Shubik (1971) and the subsequent literature. Second, with incomplete information, transfers may serve as a screening tool in coalitional deviations. For example, if the matching values a and b are increasing in types, then a worker being willing to accept a low wage indicates that he is likely of a high type; see, e.g., Liu et al. (2014).

In most of our examples, the workers' matching values are *independent* of their types, which implies that the phenomena uncovered by the examples would carry over to the transferable utility setting. Therefore, these examples may serve as natural starting points to investigate matching markets with transferable utility. Moreover, since Assumption 1 or 2 enables Liu (2020) to apply the duality approach, the generalized lattice theorem probably can carry over, as in Shapley and Shubik (1971).

(2) *Matching with contracts.* Standard structural results can be established in the general framework of matching with contracts (Hatfield and Milgrom, 2005) which incorporates Gale and Shapley (1962), Kelso and Crawford (1982) and etc. For instance, the lone-wolf theorem can be strengthened to the rural-hospital theorem (Roth, 1984). More precisely, in the matching between interns from medical schools and hospitals, any hospital that does not fill its quota at *some* stable matching is assigned precisely the same set of students at *every* stable matching. Hatfield and Milgrom (2005) also proves a very general lattice theorem. A first difficulty in this direction may arise in defining Bayesian stable states in the many-to-one setting, even with responsive firm preferences.

(3) *Extreme matchings.* Many structural properties of stable matchings and comparative statics results are built upon the extreme matchings, i.e., worker-optimal or firm-optimal stable matchings. For example, Roth (1982) shows that the worker-optimal stable matching is weakly Pareto optimal for the workers among *all* individually rational matchings.

Since we have established the lattice theorem with incomplete information (Proposition 2), and since the opposition of interests between the two sides of the market (Roth and Sotomayor, 1990, Theorem 2.13, attributed to Knuth) and the existence of extreme matchings (Gale and Shapley, 1962) can be viewed as corollaries of the lattice theorem, it would be a natural exercise to replicate these results in the incomplete-information setting.

(4) *Belief-free settings.* Our examples and results may also carry over to the belief-free settings studied in Liu et al. (2014) and Chen and Hu (2020, 2023). Furthermore, the set of stable states in the belief-free settings has an additional structural property in terms of information structure: If (μ, t, Π) and $(\mu, t, \hat{\Pi})$ are two stable states, then $(\mu, t, \Pi \wedge \hat{\Pi})$ is essentially stable. This is a corollary of Proposition 3 in Chen and Hu (2023).

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.

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

Azevedo, Eduardo M, and Jacob D Leshno. 2016. "A supply and demand framework for two-sided matching markets." *Journal of Political Economy*, 124(5): 1235–1268.

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

Blair, Charles. 1984. "Every finite distributive lattice is a set of stable matchings." *Journal of Combinatorial Theory, Series A*, 37(3): 353–356.

Blair, Charles. 1988. "The lattice structure of the set of stable matchings with multiple partners." *Mathematics of Operations Research*, 13(4): 619–628.

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.

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

Chen, Yi-Chun, and Gaoji Hu. 2024. "Bayesian stable states." *Games and Economic Behavior*, 145: 102–116.

Chiappori, P.A., B. Salanie, and Y.K. Che. 2024. *Handbook of the Economics of Matching. Handbooks in Economics*, Elsevier Science.

Fernandez, Marcelo Ariel, Kirill Rudov, and Leeat Yariv. 2022. "Centralized matching with incomplete information." *American Economic Review: Insights*, 4(1): 18–33.

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

Gale, David, and Marilda Sotomayor. 1985*a*. "Ms. Machiavelli and the stable matching problem." *The American Mathematical Monthly*, 92(4): 261–268.

Gale, David, and Marilda Sotomayor. 1985*b*. "Some remarks on the stable matching problem." *Discrete Applied Mathematics*, 11(3): 223–232.

Gusfield, Dan, and Robert W Irving. 1989. *The stable marriage problem: structure and algorithms.* Vol. 54, MIT press Cambridge.

Hatfield, John William, and Paul R Milgrom. 2005. "Matching with contracts." *American Economic Review*, 95(4): 913–935.

Hu, Gaoji, Jiangtao Li, and Rui Tang. 2020. "The revealed preference theory of stable matchings with one-sided preferences." *Games and Economic Behavior*, 124: 305–318.

Kelso, Alexander S Jr, and Vincent P Crawford. 1982. "Job matching, coalition formation, and gross substitutes." *Econometrica: Journal of the Econometric Society*, 1483–1504.

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

Knuth, Donald Ervin. 1976. *Mariages stables et leurs relations avec d'autres problèmes combinatoires: introduction à l'analyse mathématique des algorithmes.* Presses de l'Université de Montréal.

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

Liu, Qingmin. 2023. "Cooperative Analysis of Incomplete Information." *Working Paper*.

Liu, Qingmin. 2024. "Matching with Incomplete Information." In *Handbook of the Economics of Matching*.

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

Martínez, Ruth, Jordi Massó, Alejandro Neme, and Jorgee Oviedo. 2001. "On the lattice structure of the set of stable matchings for a many-to-one model." *Optimization*, 50(5-6): 439–457.

Peralta, Esteban. 2024. "Not all is lost: Sorting and self-stabilizing sets." *Games and Economic Behavior*.

Pomatto, Luciano. 2022. "Stable matching under forward-induction reasoning." *Theoretical Economics*, 17(4): 1619–1649.

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

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

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

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.

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

Wang, Ziwei. 2023. "Rationalizable Stability in Matching with Incomplete Information." *Working Paper.*

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