Matching with Incomplete Information*

Qingmin Liu†
Columbia University

November 3, 2023

Abstract

We review the nascent field of matching with incomplete information and its connections to other areas of research. We also discuss open questions and argue that cooperative solution concepts offer new perspectives for applications.

*This is a draft prepared for "Handbook of the Economics of Matching." Comments are welcome at qingmin@gmail.com.

†I am grateful to Larry Samuelson and Ziyang Shen for their comments and suggestions. I also thank (in chronological order) Yan Chen, Hao Li, Andy Postlewaite, Gaoji Hu, Jacob Leshno, Zhiming Feng, and Ziwei Wang for comments. Yuyang Miao provided excellent research assistance. This research is partly supported by NSF grant SES-1824328.

Contents

1 Introduction **3**

1.1 Reduced-form Solution Concepts 3

1.2 Incomplete Information 4

1.3 Overview 6

2 The Preference Revelation Game Approach **8**

3 Canonical Matching Models with Incomplete Information **11**

4 A Prior-Free Approach to Stable Matching **13**

4.1 An Example 14

4.2 Iterative Stability 16

4.3 The Implications of LMPS Stability 20

4.4 Competitive Equilibrium 21

4.5 Extensions 23

4.6 Epistemic Foundations 26

5 A More Refined Bayesian Approach **27**

5.1 Bayesian Stability 27

5.2 Weak and Strong Consistencies 29

5.3 Implications of Stability and Consistency 33

6 Further Research and Open Questions **34**

6.1 Other Solution Concepts 34

6.2 Information Acquisition 36

6.3 Applications 36

6.4 Implementation 37

6.5 Dynamic Games of Matching 40

References **41**

1 Introduction

The rapid growth of scholarly papers and practical applications in matching theory signifies one of the most notable achievements in microeconomics in recent years. The prevalence and flexibility of matching models are evident, as demonstrated in this handbook. They serve as both descriptive frameworks facilitating empirical analysis and policy evaluations, and as normative tools aiding in the design and improvement of market mechanisms.

1.1 Reduced-form Solution Concepts

The rising popularity of matching models is undoubtedly driven by the abundance of their applications. Many practical economic interactions can be formulated as matching problems and analyzed using tools of matching theory. Nevertheless, we should not overlook the key role played by useful solution concepts, such as competitive equilibrium, pairwise stability, and the core. It is worthwhile to first understand the conceptual and methodological development of the basic ideas of matching theory under complete information, which will help us identify the primary issues as we embark on research involving incomplete information.

Two key factors contribute to the prominence of these solution concepts. First, they have normative appeals. Without dwelling on the obvious, it's worth noting that the desirability of the concept of stability is underscored by a number of studies, including Roth (1991) among others. Secondly, these solution concepts offer a simple reduced-form approach to modeling complex strategic interactions. To illustrate this, consider the alternative non-cooperative equilibrium model for the marriage problem of Gale and Shapley (1962), where individuals can meet, date, marry, divorce, remarry, and so on. Formalizing this model necessitates many explicit assumptions regarding strategic interactions: how people meet, who proposes marriage, whether married individuals exit the market, and if they stay, their subsequent interactions with others, etc. These assumptions are highly context-dependent, and without a deep empirical understanding of the problem at hand, our modeling choice may well be speculative. Additionally, this approach can be too flexible and its predictions too sensitive to modeling assumptions.

Setting aside these modeling challenges, applying a non-cooperative model of marriage is far from straightforward. Its complexity hinders its applicability. Thus, it's not surprising that applications of a non-cooperative model of marriages aren't as widespread as the reduced-form approach featuring the solution concepts of competitive equilibrium or pairwise stability. Nevertheless, it is important to acknowledge the value of these non-cooperative models, especially when their assumptions are plausible for the intended applications. Furthermore, they provide explicit strategic foundations for the reduced-form approach. This is consistent with the influential Nash program. For a fair understanding of reduced-form concepts, it's useful to recognize that the fundamental non-cooperative solution concept of Nash equilibrium is also in reduced form, which does not explain how an equilibrium emerges from explicit strategic interactions, and a large literature is devoted to foundations for Nash equilibrium, often resorting to non-equilibrium evolutionary or adaptive processes.[1] With this in mind, a non-cooperative equilibrium foundation for the reduced-form concepts of stable or competitive equilibrium matching shouldn't be seen as the only benchmark; it's essential to also recognize and appreciate viable non-equilibrium processes.

1.2 Incomplete Information

A fundamental dimension that is often missing from our analysis of matching markets is information: players are uncertain about one another's preferences, and they are asymmetrically informed of payoff-relevant parameters. Information asymmetry potentially has a significant impact on the efficiency and surplus distributions. While there is no reason to believe that incomplete information is inherently non-cooperative, non-cooperative game-theoretical models are traditionally thought to be the best way to analyze imperfect and incomplete information. This is because these models precisely define the game forms and strategies for privately informed players, allowing the analysts to easily formulate players' inferences about uncertainty. Although non-cooperative models post no conceptual challenges, their flexibility and complexity, as mentioned earlier, are amplified in the presence of incomplete information, and, as we know from our experiences with signaling and bargaining models, their predictions are

[1] See, for example, Kandori, Mailath, and Rob (1993) and the references therein.

particularly sensitive to modeling choices. Not surprisingly, while imperfect and incomplete information has transformed economics research, most of our models are limited to simple environments for tractability, such as two-player games with one-sided incomplete information, etc.[2] Expanding these models to encompass multiple players or multiple-sided incomplete information, or both, involves a challenging balance between tractability and making reasonable assumptions.

Given these considerations, reduced-form cooperative models of incomplete information should be more appealing to consider, but limited progress has been made in matching problems and more general games. In his seminal contribution, Wilson (1978) defines and studies the core of an exchange economy. Since then, considerable efforts have been dedicated to theoretical questions, especially in the context of exchange economies, see, e.g., Forges, Minelli, and Vohra (2002), Glycopantis and Yannelis (2006), and Forges and Serrano (2013) for reviews. Referring to the literature of cooperative games with incomplete information, Aumann and Heifetz (2002) comment that "this area is to this day fraught with unresolved conceptual difficulties." Surprisingly, practical applications, sometimes the best guides for theory building, have received little to no attention. Two-sided matching is an ideal platform for studying both theoretical and applied questions in cooperative analysis of incomplete information.

As alluded to earlier, there are two intervening tasks for developing a reduced-form solution concept or about any concept, regardless of information. Let's focus on the central concept of pairwise stability. The first task is on defining the criteria for stable outcomes and the second is on the attainability and practical plausibility of these outcomes. The latter can be demonstrated either by centralized ways of implementing stable outcomes or through decentralized interactions. Gale and Shapley (1962) is an exceptional paper in the sense that it scores high on both: it defines an appealing notion of pairwise stability and provides an algorithm that identifies a stable outcome for a given preference profile. The algorithm, commonly known as the deferred acceptance algorithm, is very useful in market design research and practices. A modified version of the algorithm considered by Crawford and Knoer (1981) bears a resemblance

[2]The seminal papers of adverse selection by Akerlof, Spence, Rothschild and Stiglitz are not game-theoretic. Subsequent non-cooperative game-theoretical formulations make their assumptions and analysis precise; see, e.g., Löfgren, Persson, and Weibull (2002). These problems are special cases of two-sided matching, see Myerson (1995) for some additional discussions.

to real-world labor market adjustment processes, adding to the appeal of the algorithm and the concept of stable matching itself. Although the concept is only partially satisfactory if the stringent implementation criterion of strategy-proofness is considered (Roth, 1982), it is shown that in a preference revelation game induced by the deferred acceptance algorithm, stable outcomes can be obtained as Nash equilibria in undominated strategies (Roth, 1984).[3] With incomplete information, the matter is much more complicated. It would be ideal to replicate the achievement of Gale and Shapley (1962) and the endeavors of subsequent papers. However, it's worthwhile to underscore that defining a reasonable solution concept for an environment with incomplete information and identifying its plausible, practical implementations are not one and the same.

1.3 Overview

We first discuss the preference revelation game approach of incomplete-information matching in Section 2. This approach has its roots in the traditional mechanism design literature. This branch of the literature has progressed in roughly three directions. The first direction explores implementability of complete-information stable matching in environments of incomplete information without proposing new solution concepts. The second direction modifies the preference revelation game and defines stability as non-cooperative solutions of the modified game. The third direction studies equilibria of the non-cooperative games of incomplete information induced by well-known algorithms designed to implement the complete-information stable matching.

There are many plausible and practically useful ways to incorporate incomplete information into a matching model. In Section 3, we introduce canonical one-to-one, two-sided matching models under various informational assumptions. These cover scenarios both with and without transfers. We assume that players maximize their expected utility. In the context of complete information, the defining criteria for stability are individual rationality and immunity to pairwise blocking. Establishing these criteria under incomplete information is a natural and vital step toward developing new solution concepts. Both individual and pairwise deviations are conditional on observed matching outcomes, as seen in applications like marriage or the labor market, regardless of the

[3]This result is strengthened by Ma (1995) to rematching-proof Nash equilibria of all mechanisms that select stable matching outcomes.

presence or lack of informational friction. However, with incomplete information, the incentives for deviations and the presumed stable matching outcomes, where deviation incentives are absent, both convey information. This establishes a feedback loop leading to endogenous information for a stable matching. This is a new conceptual issue to deal with.

Section 4 delves into a prior-free approach to address the endogenous information of stable matching. The key idea is to eliminate matching outcomes that display clear deviations for all plausible beliefs. After these outcomes are excluded, the support of players' beliefs becomes more constrained; they must assign positive probabilities only to states aligning with outcomes that haven't been eliminated. These restricted beliefs permit us to further disqualify additional matching outcomes, culminating in an iterative or a fixed-point definition of stability. The detailed configuration of the definition is crucially dependent on the informational assumptions of the model, such as what players can observe, leading to different variants for various applications.

Under complete information, pairwise stability and competitive equilibrium are outcome equivalent. However, an apparent conceptual difference exists: in a competitive equilibrium, players are price takers and disregard potential partners' incentives, whereas in a stable matching, deviating players must consent to the deviation and can negotiate prices. When it comes to incomplete information, this conceptual difference manifests through the information revealed by prices and incentives. For a more detailed discussion of this issue, see Section 4.4.

Section 5 discusses a more refined Bayesian approach, in which players' beliefs are explicitly formulated. This approach has the advantage of incorporating prior beliefs into the analysis and is more consistent with traditional analyses of incomplete-information games. Central to the approach is a criterion of rational counterfactual reasoning: players must formulate beliefs for each counterfactual deviating scenario and behave rationally in accordance with their beliefs. Restrictions are imposed to formulate the consistency of prior and posterior beliefs. Unlike non-cooperative games with fully specified strategies on a game tree, these restrictions take a reduced form and open up cooperative analysis of matching markets to a wealth of ideas for belief-based refinements.

We conclude in Section 6 with other research directions and discuss their connections

with the approaches reviewed earlier in the article. We also present some open questions, both theoretical and applied ones.

2 The Preference Revelation Game Approach

Following in the footsteps of Roth (1982), Roth (1984), and the mechanism design theory, a natural approach to matching problems with incomplete information is to consider direct preference revelation games. Roth (1989) considers a setting where each agent is uncertain about the preferences of others and reports his own preferences. The mechanism then translates these reported preferences into matching outcomes. He investigates whether state-by-state complete-information stable matching can be achieved in an equilibrium of this Bayesian game. The result is negative: in general, interim implementation of ex post stable matching outcomes is impossible. Roth (1989) also considers the deferred acceptance algorithm for reported preference profiles and shows that it is a dominant strategy for the proposing side to report truthfully, reminiscent of the complete information environment. Ehlers and Massó (2015) adopt this approach, examining the direct preference revelation game induced by stable mechanisms—mechanisms that ensure stable matching for every reported preference profile. They consider many-to-one matching and an ordinal version Bayesian Nash equilibrium.[4] They show that reporting strategies constitute a Bayesian Nash equilibrium if and only if the reported profile for any true profile in the support of the common prior form a Nash equilibrium under complete information. Thus, only the support of the common prior is relevant. If all players on one side of the market always agree on their preferences for all states in the support of the prior, then truth-telling is a Bayesian Nash equilibrium. Ehlers and Massó (2007) examine a similar model and focus on truth-telling equilibria.

Fernandez, Rudov, and Yariv (2022) examine the preference reporting game induced by the deferred acceptance algorithm and focus on the case where the non-proposing side's preference is identical across states. The incomplete information concerns the non-proposing side's uncertainty about the proposing side's preferences. Their main message

[4] A strategy profile is an ordinal Bayesian Nash equilibrium if it is a Bayesian Nash equilibrium for every utility representation in the preference revelation game induced by the mechanism and a common prior.

is that salient properties of complete-information stable matching are not robust to the introduction of incomplete information.

Chakraborty, Citanna, and Ostrovsky (2010) extend the direct preference revelation approach in several ways. In a school choice context, they consider a setting in which schools have uncertainties about students' qualities, and while students' preferences over schools are known, they are unaware of their own qualities. Each school receives a private signal about a student's quality. Schools' preferences over students are "interdependent" through their signals. Each school submits a message to the mechanism, which outputs a random matching outcome. Each player receives a private message from the mechanism regarding the matching outcome and can then engage in one round of pairwise rematching, with schools making offers. This is a dynamic game to which existing solution concepts apply. They call a matching mechanism stable if the induced game admits a perfect Bayesian equilibrium with truthful reporting of types in the reporting stage and without rematching on the equilibrium path in the rematching stage. They show that such matching mechanisms exist if players only observe their own matches, but they don't necessarily exist if players observe the full matching outcomes. Chakraborty, Citanna, and Ostrovsky (2015) also consider rematching that involves a group of schools and students.

Yenmez (2013) takes a mechanism design approach to matching with transfers and considers ex ante, interim, and ex post notions of blocking, where a pair of agents can rematch, abandoning the mechanism or its outcome. The interim notion was previously studied by Dutta and Vohra (2005) for an exchange economy, known as the credible core, which says that blocking players' incentives must be confirmed by each other in a fixed point fashion; see Forges, Minelli, and Vohra (2002) for a survey of incentive-compatible core in the exchange economy with incomplete information. Dizdar and Moldovanu (2016) study matching with multiple-dimensional types and show that implementing efficient matching in ex post equilibria requires constant surplus sharing rules. However, the paper does not address the question of stability.

Chen and Sönmez (2006) experimentally study a students-schools matching problem where students are uncertain about each other's preferences. They examine how this uncertainty affects truthful reporting and efficiency, conditional on reported preferences, across three mechanisms: the top trading cycles mechanism, the Gale–Shapley mecha-

nism, and the Boston mechanism. This experimental approach is further explored by Pais and Pintér (2008) and Pais, Pintér, and Veszteg (2011).

The preference revelation game approach has the advantage of directly addressing the question of implementing a solution concept. This approach is particularly beneficial for normative market design problems. However, it also presents several challenges. First, the game form may appear ad hoc and inherently non-cooperative. One could argue that various alternative game forms, especially dynamic games involving information exchange, might also be considered.[5] However, evaluating and measuring the performance of existing mechanisms that assume incomplete information away is certainly valuable for the practical design and analysis of markets. Second, the usual treatment of direct revelation games doesn't adequately consider pairwise deviations, particularly during the preference revelation stage. Roth (1989) demonstrates that, in a man-optimal matching mechanism, a coalition of men can benefit from jointly misreporting their types, even when truth-telling is the dominant strategy for each individual man. This problem isn't unique to matching, and some proposed solutions reintroduce non-cooperative games into the direct revelation game (see, for instance, the seminal work of Holmström and Myerson (1983) in the context of collective choices). Third, in the special case of complete information (where the type space is assumed to be a singleton), while the Bayesian incentive compatibility in the revelation game might seem straightforward, the practical implementation might not. This raises questions about practical implementations under incomplete information.

In the remainder of the article, we will discuss various other approaches and review several criteria for stable matching with incomplete information, highlighting recent developments. To substantiate the discussion, we first formally define the model.

[5]The question arises as to whether there is a canonical mechanism; see Sugaya and Wolitzky (2021) for a more general communication revelation principle in multiple-stage games for sequential equilibrium.

3 Canonical Matching Models with Incomplete Information

The set of players is $N = I \cup J$, where $i \in I$ is a worker and $j \in J$ is a firm. Each player $n \in I \cup J$ has a type $\theta_n \in \Theta_n$. If worker i and firm j match together with a transfer p from the firm to the worker, worker i's ex post payoff is $a_{ij}(\theta) + p$ and firm j's ex post payoff is $b_{ij}(\theta) - p$. We shall write $a_{ii}(\theta)$ and $b_{jj}(\theta)$ as i's and j's payoffs from being alone, respectively. In this chapter, we shall assume that a_{ij} and b_{ij} depend on θ only through (θ_i, θ_j), a_{ii} depend on θ through θ_i and b_{jj} depend on θ through θ_j. We denote the matching game with incomplete information by $\Gamma = (\Theta, I, J, a, b)$.

If $\Theta = \{\theta\}$ is a singleton, the matching game has complete information. We denote this complete-information game by $\Gamma^\theta = (\{\theta\}, I, J, a, b)$.

If $\Theta_J = \{\theta_J\}$ is a singleton, the game has one-sided incomplete information, where the firms' types are commonly known. We denote the one-sided incomplete information game as $\Gamma^{\theta_J} = (\Theta_I \times \{\theta_J\}, I, J, a, b)$.

A Bayesian game of incomplete information should also specify prior distributions of types. Let $\beta_n^0 \in \Delta(\Theta)$ be player n's prior, which is assumed to have a full support for notational simplicity. We will write the corresponding Bayesian game as (Γ, β^0). We write the Bayesian game with one-sided incomplete information as $(\Gamma^{\theta_J}, \beta^0)$.

The cooperative game with incomplete information does not specify individual strategies. These are the advantages discussed in the previous sections.

A matching outcome is (μ, τ), where $\mu : I \cup J \to I \cup J$ specifies who each player is matched with so that $\mu_i \in J \cup \{i\}$, $\mu_j \in I \cup \{j\}$, and $\mu_i = j$ iff $\mu_j = i$, and $\tau : I \cup J \to \mathbb{T} \subset \mathbb{R}$ specifies the transfers each player receives such that $\tau_n + \tau_{\mu(n)} = 0$ for $n \in N$. There are two special cases. If the set of transfers is $\mathbb{T} = \{0\}$, the model is a matching without transfers and with cardinal utility; if $\mathbb{T} = \mathbb{R}$, there is no restriction on transfers.

It is also useful to consider stabilized matching outcomes where private types are revealed within each matched pair. For instance, it's more plausible that a married couple would observe each other's type, despite potentially having uncertainty about people outside of their marriage. A stable marriage outcome should account for this

feature. However, this assumption might not be as relevant when studying stable initial allocations in labor markets or other assignment problems. We call a matching game in which a player observes their matched partner's type $\overline{\Gamma}$ and $\overline{\Gamma}^{\theta_J}$. The notation is summarized in Table 1.

I	the set of workers
J	the set of firms
Θ	the set of profiles of types, with a typical element $\theta = ((\theta_i)_{i\in I}, (\theta_j)_{j\in J})$
$a_{ij}(\cdot)$	worker i's payoff from matching with firm j
$b_{ij}(\cdot)$	firm j's payoff from matching with worker i
(μ, τ)	a matching outcome with transfer
Γ	the matching game (Θ, I, J, a, b)
Γ^θ	the matching game with complete information of θ, $(\{\theta\}, I, J, a, b)$
Γ^{θ_J}	the matching game with firms' types θ_J known, $(\Theta_I \times \{\theta_J\}, I, J, a, b)$
$\overline{\Gamma}$	the matching game where players observe each other's type in a matched pair
$\overline{\Gamma}^{\theta_J}$	the matching game $\overline{\Gamma}$ where firms' types θ_J are known
(Γ, β^0)	the Bayesian matching game with priors $\beta_n^0 \in \Delta(\Theta)$
$(\Gamma^{\theta_J}, \beta^0)$	the Bayesian matching game with priors $\beta_n^0 \in \Delta(\Theta_I \times \{\theta_J\})$

Table 1: Notation in this article

The complete-information game is well understood.

Definition 1. *A matching outcome (μ, τ) of Γ^θ is **individually rational** if, for all $i \in I$ and $j \in J$,*

$$a_{i\mu(i)}(\theta) + \tau_i \geq 0,$$
$$b_{\mu(j)j}(\theta) + \tau_j \geq 0.$$

*We say (i, j, p), where $i \in I$, $j \in J$, and $p \in \mathbb{T}$, **blocks** (μ, τ) if*

$$a_{ij}(\theta) + p > a_{i\mu(i)}(\theta) + \tau_i,$$
$$b_{ij}(\theta) - p > b_{\mu(j)j}(\theta) + \tau_j.$$

*We call (i, j, p) a **blocking pair** with transfer if it blocks (μ, τ). A matching outcome of Γ^θ is **stable** if it is individually rational and there does not exist a blocking pair.*

Remark 1. The concept of pairwise stability is agnostic about how the putative matching (μ, τ) arises, how a pair of players (i, j) meet each other, and how they figure out the transfer p together. The concept tests a putative outcome against all possible pairwise deviations. This simplicity is the advantage of this reduced-form concept. Some

of the features here are not unique to stability. For example, the non-cooperative concept of Nash equilibrium doesn't specify how players arrive at an equilibrium outcome (even though players' contemplation of all possible unilateral deviations from a putative equilibrium outcome appear much more straightforward than pairwise deviations). An intuitive justification of considering all possible pairwise blockings is that the market is frictionless, and hence blocking opportunities can be easily identified and carried out. Incomplete information acts as a friction, constraining the ability of players to identify blocking opportunities. □

Remark 2. Individual rationality is synonymous with the absence of unilateral deviations; pairwise blocking is synonymous with bilateral deviations. Basically, a stable outcome is one that withstands unilateral and bilateral deviations. The semantics seem more important than they should be. We will come back to this point later. □

Remark 3. Deviations are conditional on an observed matching outcome. Under incomplete information, a new conceptual issue arises. Since deviation incentives rely on, and hence reveal, players' private information, the absence of such deviations—which defines a stable situation—should also reveal information. A matching outcome is stable with respect to this information, refining the incentives for deviations. This feedback loop shapes stable outcomes. Given this complication, the first obvious question to ask is: what constitutes a stable situation? Just as in the case of complete information, a stable matching should be tantamount to an immunity to unilateral and pairwise deviations. Thus, formulating unilateral and pairwise deviations is key to the development. How a putative matching outcome is formed in the first place is a different question. This contrasts with the preference revelation game approach. □

4 A Prior-Free Approach to Stable Matching

We start with Liu, Mailath, Postlewaite, and Samuelson (2014) (henceforth LMPS). They consider the following setting. First, there is one-sided incomplete information. Second, in a putative match, a matched pair of players observe each other's private information. The paper thus considers the matching game $\overline{\Gamma}^{\theta_J}$. This means we are looking at stable situation in which matching has been formed and stabilized for a

while, such as marriage or some labor markets where players do not want to deviate even after observing their partners' types. (We consider the extension of LMPS to Γ^{θ_J} and Γ in Section 4.5.)

LMPS formulates a theory of stable matching by taking a prior-free approach. Without pinning down the exact belief, they ask: can we determine what cannot be stable?

4.1 An Example

Assume that there are three workers and three firms. The matching values are given by $a_{ij}(\theta_i, \theta_j) = \frac{1}{2}\theta_i\theta_j$ and $b_{ij}(\theta_i, \theta_j) = \frac{1}{2}\theta_i\theta_j$. Additionally, assume $a_{ii} = b_{jj} = 0$. (The observable characteristics of i and j do not affect payoffs in this example). There is no restriction on transfers. Firms' types $\theta_J = (2, 4, 5)$ are commonly known. A worker's type can be either 1, 2, or 3. For simplicity, assume that workers' types are from a permutation. That is, there is exactly one worker of each type: 1, 2, and 3. Consider the following matching arrangement:

θ_i	1	3	2
θ_j	2	4	5
τ	0	2	-2

Table 2

For example, the firm of type 4 is matched with the worker of type 3 and pays the worker 2. The payoff of each player associated with this matching arrangement is given in Table 3.

$a_{i\mu(i)} + \tau_i$	1	8	3
θ_i	1	3	2
θ_j	2	4	5
τ	0	2	-2
$b_{\mu(j)j} + \tau_j$	1	4	7

Table 3

A complete-information stable matching, with supermodular surplus functions, should be assortative. The matching arrangement cannot be stable under complete information. Indeed, the only blocking pair involves worker type 2 and firm type 4. Together,

they can create a surplus of 8, but their payoffs in the candidate matching sum to only 7. They can walk away together and split the surplus in a way that makes both better off. Can they still form a blocking pair under incomplete information about workers' types? The worker of type 2 observes the type the firm.[6] For the worker of type 2 to be happy with the deviation, the transfer p needs to be such that $4 + p > 3$, i.e., $p > -1$.

However, the firm does not know the worker's type. What if the worker's type is not 2 but 1, a lower type? In this case, the matching arrangement and the associated payoffs, from the firm's eyes, are as follows:

$a_{i\mu(i)} + \tau_i$	2	8	0.5
θ_i	2	3	1
θ_j	2	4	5
τ	0	2	-2
$b_{\mu(j)j} + \tau_j$	2	4	4.5

Table 4

The firm of type 4 cannot distinguish Table 4 and Table 3. We emphasize that these are the only two possible scenarios for the firm because it observes its own worker's type 3. In Table 4, firm type 4 and worker type 1 together can produce a surplus of 4 and their payoff in this matching sums up to 4.5, so they cannot both benefit from a deviation. Indeed, worker type 1 would deviate with firm type 4 if the transfer to the worker were such that $2 + p > 0.5$, i.e., $p > -1.5$.

When the worker is truly type 2, as in Table 3, we conclude that the type 2 worker would prefer to deviate with the firm if $p > -1$. So we have a problem: if the type 2 worker favors this deviation, the type 1 worker favors it even more. So it appears that the blocking pair cannot form if the firm is uncertain about the worker's type.

Being uncertain about Table 3 and Table 4, a Bayesian firm needs to form a belief about the worker's type and make decisions accordingly. But the belief here should be endogenous. It is unclear what should the firm's belief should be and how we determine it. This is the question addressed by Liu (2020). LMPS instead takes an approach that avoid taking a stance on beliefs.

[6]However, the worker does not know the payoff of the firm because he does not observe the type of the worker who is matched with the firm.

To see their idea, let's look at the scenario of Table 4. To assess the plausibility of this scenario, the firm of type 4 should analyze not only its own incentive to deviate but also the incentives of other firms. To this end, consider the type 5 firm and the type 2 worker in Table 4. They will block if there is complete information because they can produce a surplus of 10 together, exceeding what they obtain in this putative matching. Under incomplete information, type 5 firm doesn't have to worry about the worker's type—the worker's type cannot be worse than 2 because the type 5 firm has already matched with the type 1 worker and can observe the worker's type—there is exactly one worker whose type is 1. If the worker's type is 3, then it is even better for the type 5 firm. So in the alternative scenario of Table 4, the type 5 firm can form a blocking pair with the type 2 worker, despite the firm's uncertainty about the worker's type. Therefore, Table 4 cannot describe a stable matching.

The argument presented in the previous paragraph reflects the thinking of a type 4 firm. This firm deduces that the scenario in Table 4 cannot be deemed "stable" due to the deviation involving the type 5 firm and the type 2 worker. Hence, the firm should deduce that it should rule out Table 4. The only viable scenario is Table 3, in which the firm should deviate with the type 2 worker.

Therefore, the fundamental idea boils down to the following. If a firm is thinking of an alternative scenario, the firm should assess whether that scenario is "stable," i.e., whether there is a deviation from it. If there is a deviation regardless of players' uncertainty, this scenario should be ruled out. If there is no clear deviation, this scenario should be tentatively retained. Furthermore, everyone in that alternative scenario should also check whether there are clear deviations in further scenarios that they cannot distinguish from the alternative scenario, and everyone should anticipate that everyone else is making this calculation in every possible scenario they cannot rule out...and so forth.

4.2 Iterative Stability

We will now associate the observable matching outcome (μ, τ) with the type profile θ. With some abuse of terminology, we will continue to refer to (μ, τ, θ) as a (full) outcome of the matching game.

The key to defining stability is to identify when a deviation—either unilateral or bilateral—occurs. With incomplete information, uncertainty in both unilateral and bilateral deviations should be treated systematically on the same grounds. This is not easy to do as information is endogenous. LMPS assumes that firms know their own workers' types in a putative stabilized matching, therefore individual rationality in the incomplete-information game is the same as individual rationality in the complete-information game. This assumption is not only useful for simplifying the formulation but also helps us prove strong results.

Definition 2. *A full matching outcome* (μ, τ, θ) *is* ***individually rational*** *if*

(i) $a_{i\mu(i)}(\theta) + \tau_i \geq a_{ii}(\theta)$,

(ii) $b_{\mu(j)j}(\theta) + \tau_j \geq b_{jj}(\theta)$.

Definition 3. *Fix a set of full matching outcomes* Σ. *A full matching outcome* $(\mu, \tau, \theta) \in \Sigma$ *is* ***clearly blocked*** *with respect to* Σ *if there is a worker-firm pair* (i, j) *together with a transfer* $p \in \mathbb{T}$ *such that:*

(i) $a_{ij}(\theta) + p > a_{i\mu(i)}(\theta) + \tau_i$, *and*

(ii) $\mathbf{E}_\rho [b_{ij}(\cdot)] - p > b_{\mu(j)j}(\theta) + \tau_j$ *for any* $\rho \in \Delta(\Theta')$, *where*

$$\Theta' = \left\{ \theta' \in \Theta : \begin{array}{l} \text{(ii.a) } (\mu, \tau, \theta') \in \Sigma, \\ \text{(ii.b) } \theta'_{\mu(j)} = \theta_{\mu(j)}, \\ \text{(ii.c) } a_{ij}(\theta') + p > a_{i\mu(i)}(\theta') + \tau_i. \end{array} \right\}. \tag{4.1}$$

Let's parse this definition. Condition (i) describes worker i's incentive to deviate with firm j at a price p. Since worker i knows his own type θ_i and firms' types are commonly known, the deviation is the same as in complete information. Condition (ii) pertains to firm j's deviation: firm j does not know worker i's type; the firm uses a belief $\rho \in \Delta(\Theta')$ to evaluate workers' types (not only worker i) and this belief only assigns positive probability to θ' that satisfies the following requirements: (ii.a) says that firm j thinks that the possible set of states (μ, τ, θ') has to stay in Σ, which is the constraint we start with. (ii.b) says that firm j observes the type of his own worker $\mu(j)$, so θ' cannot contradict with what he knows. (ii.c) says that firm j knows that worker i must benefit from the deviation.

Remark 4. For the exposition of condition (ii) in Definition 3, the following three statements are mathematically equivalent:

$$\mathbf{E}_\rho\left[b_{ij}(\cdot)\right] - p > b_{\mu(j)j}(\theta) + \tau_j \text{ for any } \rho \in \Delta\left(\Theta'\right); \tag{4.2}$$

$$b_{ij}(\theta') - p > b_{\mu(j)j}(\theta) + \tau_j \text{ for any } \theta' \in \Theta'; \tag{4.3}$$

$$\min_{\theta' \in \Theta'}\{b_{ij}(\theta') - p\} > b_{\mu(j)j}(\theta) + \tau_j. \tag{4.4}$$

A potential misunderstanding from (4.3) and (4.4) is that firms evaluate the worst-case scenario, instead of maximizing expected utility. One may even attempt to impose exogenous restrictions on ρ in (4.2), which is particularly problematic (see, e.g., Alston (2020) and Bikhchandani (2017) for further elaborations). It is important to realize that the operation in Definition 3 is a building block that will be iterated to define a solution concept of stability. The correct interpretation is that LMPS does not take a stance on the exact belief the firm should have—firms' uncertainty in a stable matching should be endogenous, as we have argued before. Any ad hoc assumption on beliefs at this stage will endanger the consistency of the solution concept. This logic shouldn't be a complete surprise if one is familiar with the formulation of never-best responses in iterated elimination of strictly dominated strategies (see Bernheim (1984) and Pearce (1984)). □

Remark 5. LMPS examines the game $\overline{\Gamma}^{\theta_J}$, so individual rationality is straightforward. If Γ^{θ_J} is considered, then the uncertainty faced by the uninformed firms again should be endogenous, with respect to which individual rationality should be defined. It is problematic to impose exogenous restrictions in the definition of individual rationality. We shall come back to this issue later. □

A full matching outcome $(\mu, \tau, \theta) \in \Sigma$ has a clear unilateral deviation (i.e., a violation of individual rationality) or a clear bilateral deviation (i.e., being clearly blocked) irrespective plausible beliefs should never be considered stable. We summarize this property below.

Definition 4. *A full matching outcome $(\mu, \tau, \theta) \in \Sigma$ is **never stable** with respect to Σ if it is not individually rational or is clearly blocked with respect to Σ.*

We will iterate this definition. The idea is that to start with the largest set of outcomes Σ^0, removing from Σ^0 that are never stable according to the definition above to obtain Σ^1 (e.g., the blocking identified in the scenario of Table 4 in Section 4.1). With a smaller set of outcomes Σ^1, the plausible beliefs for firms are refined through (ii.a) in Definition 3, and hence firms and workers can engage in deviations that are not possible before (e.g., the blocking identified in the scenario of Table 3 in Section 4.1). This process will terminate when no further removal is possible.

Definition 5. *Let Σ^0 be the set of all full matching outcomes. For $k \geq 1$, let*

$$\Sigma^k := \Sigma^{k-1} \setminus \left\{(\mu, \tau, \theta) \in \Sigma^{k-1} : (\mu, \tau, \theta) \text{ is never stable w.r.t. } \Sigma^{k-1}\right\}.$$

*The **stable set** of matching outcomes for incomplete information is $\Sigma^\infty = \cap_{k=0}^{\infty} \Sigma^k$.*

Remark 6. The formulation in Definition 5 is a slight variation of LMPS stable sets. In LMPS, Σ^0 represents the set of all individually rational outcomes, while Σ^∞ is determined by the iterative removal of clearly blocked outcomes. The two formulations are of course equivalent, but Definition 5 has the benefit of avoiding confusion and enabling a more transparent extension to Γ^{θ_J}. □

The iterative definition in LMPS offers an algorithm to compute the incomplete-information stable set: in each round, all clearly blocked matching outcomes are removed. The iterative concept has a fixed-point characterization.

Definition 6. *A set of full matching outcomes E is **self-stabilizing** if no $(\mu, \tau, \theta) \in E$ is never stable with respect to E.*

The fixed-point notion of self-stabilizing sets connects stability concepts under complete information and incomplete information. The LMPS stable set Σ^∞ contains complete-information stable matching. Furthermore, it is a useful technique: to prove a certain outcome (μ, τ, θ) is in Σ^∞, it suffices to construct a self-stabilizing set E that contains it.

Theorem 1. (i) *A singleton set $\{(\mu, \tau, \theta)\}$ is self-stabilizing if and only if (μ, τ, θ) is complete-information stable for Γ^θ.* (ii) *If E is a self-stabilizing set, then $E \subset \Sigma^\infty$.* (iii) *Σ^∞ is the largest self-stabilizing set.*

4.3 The Implications of LMPS Stability

The formulation and characterization developed in previous sections apply to general transfer space, in particular, the no-transfer case $\mathbb{T} = \{0\}$. To derive sharper results, it is useful to allow unrestricted transfers and impose restrictions on payoff functions.

Assumption 1. $\mathbb{T} = \mathbb{R}$ so transfers are unrestricted.

Assumption 2. (i) $a_{ii} \equiv b_{ii} \equiv 0$, (ii) $a_{ij}(\theta_i, \theta_j) = a(\theta_i, \theta_j)$, $b_{ij}(\theta_i, \theta_j) = b(\theta_i, \theta_j)$ for some functions a and b that are strictly supermodular and strictly increasing on $\Theta_i \times \Theta_j$.

Theorem 2. *Suppose Assumption 1 and Assumption 2 hold. Then any* $(\mu, \tau, \theta) \in \Sigma^\infty$ *is ex post efficient.*

Remark 7. Ex post efficiency under supermodularity implies positive assortativity. However, Theorem 2 does not reduce Σ^∞ to the set of complete information stable matchings. This is because firms continue to face uncertainties about the workers they do not employ, and therefore competition is not intense enough to equate transfers with those in a complete information stable matching.

To eliminate these residual uncertainties, certain strong conditions are needed; for instance, that the workers' types are permutations (i.e., there is a one-to-one mapping between θ_I and θ'_I for any $\theta_I, \theta'_I \in \Theta_I$). Given this assumption, if different firms have different types, it follows from Theorem 2 that the types of workers will be identified from the firms they match, and we obtain complete information stable matchings (of course, unmatched workers, being always of low type, cannot have their types perfectly identified).

More subtly, if different workers have different types (even if different firms don't have different types), then the prices will reveal enough information to distinguish high-type workers from low-type workers. As a result, Σ^∞ is precisely the set of complete-information stable matchings. For more details, please refer to LMPS. □

Remark 8. The proof resembles the iteration process demonstrated in the example. It starts with the lowest possible type in workers' type space, and show that if some worker has this type, this worker cannot be matched with a better firm than higher-type workers. □

Remark 9. LMPS also shows that if payoff functions are submodular and positive, stable matchings with incomplete information are negatively assortative. The payoff assumptions in this result and Theorem 2 can be slightly relaxed. □

4.4 Competitive Equilibrium

LMPS also defines a notion of competitive equilibrium. As a benchmark, let's first introduce the well-understood competitive equilibrium notion under complete information. A price matrix P is defined as a mapping $P : I \times J \to \mathbb{R}$, and we also append $P_{ii} = P_{jj} = 0$ to this definition. A price P_{ij} indicates the payment a firm j has to make to the worker i if they match. Importantly, if i and j are not matched together, i.e., $\mu(i) \neq j$, the price P_{ij} will be an off-path price, but it is still relevant because it will be the only price for i and j decide to deviate and rematch.

Definition 7. *A complete-information price-taking matching outcome (μ, P) is a **competitive equilibrium** for Γ^θ if*

1. $a_{i\mu(i)}(\theta) + P_{i\mu(i)} \geq a_{ij}(\theta) + P_{ij}$ *for any* $i \in I$ *and* $j \in J \cup \{i\}$.
2. $b_{\mu(j)j}(\theta) - P_{\mu(j)j} \geq b_{ij}(\theta) - P_{ij}$ *for any* $i \in I \cup \{j\}$ *and* $j \in J$.

The key postulates of a competitive equilibrium are price-taking behavior and each player's ability to deviate without necessitating the partner' consent. Effectively, each pair (i, j) is traded as a unit of an indivisible commodity.

Under complete information, if (μ, P) is a competitive equilibrium, then (μ, τ), where $\tau_i = P_{i\mu(i)}$ and $\tau_j = -P_{\mu(j)j}$, is a stable matching for Γ^θ; conversely, if (μ, τ) is stable, there exist P such that $P_{i\mu(i)} = \tau_i$ and $P_{\mu(j)j} = -\tau_j$ such that (μ, P) is a competitive equilibrium. LMPS extends this concept to $\overline{\Gamma}^{\theta_J}$.

Definition 8. *A price-taking matching outcome $(\mu, P, \theta) \in \Psi$ is **never competitive** w.r.t. Ψ for $\overline{\Gamma}^{\theta_J}$ if*

(i) $a_{ij}(\theta) + P_{ij} > a_{i\mu(i)}(\theta) + P_{i\mu(i)}$ *for some* $i \in I$ *and* $j \in J \cup \{i\}$ *or*

(ii) $\mathbf{E}_\rho\left[b_{ij}(\cdot)\right]-P_{ij} > b_{\mu(j)j}(\theta)+P_{\mu(j)j}$ *for some* $j \in J$, $i \in I\cup\{j\}$, *and any* $\rho \in \Delta\left(\Theta'\right)$, *where*

$$\Theta' = \left\{ \theta' \in \Theta : \begin{array}{l} \text{there exists a price matrix } P' \text{ such that} \\ \text{(ii.a) } (\mu, P', \theta') \in \Psi, \\ \text{(ii.b) } \theta'_{\mu(j)} = \theta_{\mu(j)}, \\ \text{(ii.c) } P'_{i'j} = P_{i'j} \text{ and } P'_{i'\mu(i')} = P_{i'\mu(i')} \text{ for all } i' \in I. \end{array} \right\}. \tag{4.5}$$

The restrictions on Θ' deserve elaboration. Conditions (ii.a) and (ii.b) in (4.5) have the same interpretations as their counterparts in (4.1); Condition (ii.c) in (4.1) is not needed here because only price-taking unilateral deviation is considered in a competitive equilibrium. Condition (ii.c) in (4.5) is a new addition. It captures firm j's observation about the price matrix: the deviating firm observes the prices that involves itself $P_{\cdot j}$ (both on and off path); the firm also observes all on-path prices $P_{\cdot\mu(\cdot)}$ (which is the same as in the case of stability); the firm doesn't observe off-path prices for parities that it is not involved with.

These features of the definition of competitive equilibrium make it directly comparable with stability.

Definition 9. *Let* Ψ^0 *represent all price-taking matching outcomes* (μ, P, θ). *For* $k \geq 1$, *let*

$$\begin{aligned} \Psi^k &:= \Psi^{k-1} \setminus \left\{ (\mu, P, \theta) \in \Psi^{k-1} : (\mu, P, \theta) \text{ is never competitive w.r.t. } \Psi^{k-1} \right\} \\ \Psi^\infty &:= \textstyle\bigcap_{k=1}^{\infty} \Psi^k \end{aligned}$$

We call Ψ^∞ *the set of* ***competitive*** *outcomes.*

There is a fixed point definition.

Definition 10. *A set of price-taking matching outcomes* Ψ *is* ***competitive*** *if no price-taking matching outcome* $(\mu, P, \theta) \in \Psi$ *is never competitive w.r.t.* Ψ.

One can easily verify that competitiveness, as defined in Definition 9 and Definition 10, generalizes the concept of competitive equilibrium in a complete information matching game (see Definition 7).

Theorem 3. (i) *A singleton set* $\{(\mu, P, \theta)\}$ *is competitive if and only if* (μ, P) *is a competitive equilibrium for* Γ^θ. (ii) *If* E *is competitive, then* $E \subset \Psi^\infty$. (iii) Ψ^∞ *is the largest competitive set.*

Conceptually, the more important result is that stability is a refinement of competitive equilibrium in the incomplete information matching game $\overline{\Gamma}^{\theta_J}$.

Theorem 4. *For any* $(\mu, \tau, \theta) \in \Sigma^\infty$, *there exists a price matrix* P *with* $P_{i\mu(i)} = \tau_i$ *and* $P_{\mu(j)j} = -\tau_j$ *for all* $i \in I$ *and* $j \in J$ *such that* $(\mu, P, \theta) \in \Psi^\infty$.

The result contrasts with commonly understood insights. It is also intuitive. With incomplete information, a pairwise deviation allows two players to negotiate a price; the information revealed by the incentive of accepting this price facilitates the deviation. In contrast, in a price-taking competitive equilibrium, prices are fixed; limited information revelation hinders unilateral deviation, and hence more outcomes can be sustained as an equilibrium. It can be shown that any self-stabilizing set Σ can be entirely extended into a competitive set Ψ. However, the permissive iterative definition indeed play a role in the result. We shall return to this point in Section 6.1.

4.5 Extensions

In this section, we discuss the prior-free approach to stability in Γ^{θ_J} and Γ. By now, it should be clear why LMPS does not take a stance on the exact beliefs players have in a stable matching. Their focus on $\overline{\Gamma}^{\theta_J}$ means that individual rationality is ex post. Let's see what happens if we go from $\overline{\Gamma}^{\theta_J}$ to Γ^{θ_J}. If firms do not know their own workers' types, the notion of individual rationality is no longer given by Definition 2. If iteration is used to refine the possibility of blocking opportunities (bilateral deviations), iteration should also be used to refine the notion of individual rationality (or more precisely, the absence of unilateral deviations). Several papers that follow LMPS's iterative approach have attempted to define individual rationality using the worst-case belief. This will lead to a more permissive concept; we have offered related comments in Remark 4 and Remark 5.

Recall that stability is defined as immunity to both unilateral and bilateral deviations. All we need to do is to incorporate both deviations in the iterative process and treat them on the same ground. The following material is new to the literature.

Definition 11. *A matching outcome $(\mu, \tau, \theta) \in \Sigma$ has a **clear unilateral deviation** with respect to Σ if*

(i) *there exists $i \in I$ such that $a_{i\mu(i)}(\theta) + \tau_i < a_{ii}(\theta)$, or*

(ii) *there exists $j \in J$ such that $\mathbf{E}_\rho \left[b_{\mu(j)j}(\cdot) \right] + \tau_j < b_{jj}(\theta)$ for all $\rho \in \Delta(\Theta')$, where*

$$\Theta' = \left\{ \theta' \in \Theta : \begin{array}{l} \text{(ii.a) } (\mu, \tau, \theta') \in \Sigma, \\ \text{(ii.b) } a_{i\mu(i)}(\theta') + \tau_i \geq a_{ii}(\theta') \text{ for all } i \in I \end{array} \right\}.$$

Let $U(\Sigma)$ denote the subset of Σ obtained by removing all matching outcomes from Σ that have clear unilateral deviations with respect to Σ, i.e.,

$$U(\Sigma) := \{(\mu, \tau, \theta) \in \Sigma : (\mu, \tau, \theta) \text{ has no clear unilateral deviations w.r.t. } \Sigma\}.$$

Does the definition of $U(\Sigma)$ capture the individual rationality of firms? We think not. That's why we avoid using the terminology of individual rationality in the definition. Nevertheless, the existence of a clear unilateral deviation is obviously a *violation* of individual rationality.

Definition 12. *For any set of matching outcomes Σ, a matching outcome $(\mu, \tau, \theta) \in \Sigma$ has a **clear bilateral deviation** with respect to Σ if there is a worker-firm pair (i, j) together with a transfer $p \in \mathbb{R}$ such that:*

(i) *$a_{ij}(\theta) + p > a_{i\mu(i)}(\theta) + \tau_i$, and*

(ii) *$\mathbf{E}_\rho [b_{ij}(\cdot)] - p > \mathbf{E}_\rho \left[b_{\mu(j)j}(\cdot) \right] + \tau_j$ for all $\rho \in \Delta(\Theta')$, where*

$$\Theta' = \left\{ \theta' \in \Theta : \begin{array}{l} \text{(ii.a) } (\mu, \tau, \theta') \in \Sigma, \\ \text{(ii.b) } a_{ij}(\theta') + p > a_{i\mu(i)}(\theta') + \tau_i. \end{array} \right\}.$$

Let $B(\Sigma)$ denote the subset of Σ that is obtained after removing from Σ all matching outcomes that have clear bilateral deviations with respect to Σ, i.e.,

$$B(\Sigma) := \{(\mu, \tau, \theta) \in \Sigma : (\mu, \tau, \theta) \text{ has no clear bilateral deviations w.r.t. } \Sigma\}. \quad (4.6)$$

Remark 10. It is useful to pause and return to the issues clarified in Remark 4. It's clear that

$$\mathbf{E}_\rho\left[b_{ij}(\cdot)\right] - p > \mathbf{E}_\rho\left[b_{\mu(j)j}(\cdot)\right] + \tau_j \text{ for any } \rho \in \Delta\left(\Theta'\right) \tag{4.7}$$

is equivalent to

$$b_{ij}(\theta') - p > b_{\mu(j)j}(\theta') + \tau_j \text{ for any } \theta' \in \Theta', \tag{4.8}$$

but they are different from the following two expressions:

$$\min_{\theta' \in \Theta'}\left\{b_{ij}(\cdot)\right\} - p > \min_{\theta' \in \Theta'}\left\{b_{\mu(j)j}(\theta)\right\} + \tau_j \text{ for any } \rho \in \Delta\left(\Theta'\right); \tag{4.9}$$

$$\min_{\theta' \in \Theta'}\left\{b_{ij}(\cdot)\right\} - p > \max_{\theta' \in \Theta'}\left\{b_{\mu(j)j}(\theta)\right\} + \tau_j \text{ for any } \rho \in \Delta\left(\Theta'\right). \tag{4.10}$$

The last expression might have an intuitive behavioral interpretation. □

We first define sets of matching outcomes that do not have unilateral or bilateral deviations.

Definition 13. *For any set of matching outcomes Σ, let $S(\Sigma)$ be the subset of matching outcomes in Σ that have neither clear unilateral nor clear bilateral deviations, i.e., $S(\Sigma) := U(\Sigma) \cap B(\Sigma)$. We say Σ is **self-stabilizing** if $S(\Sigma) = \Sigma$.*

We leave the reader to prove the following claims:

Theorem 5. (i) *Σ is self-stabilizing if and only if $U(\Sigma) = \Sigma$ and $B(\Sigma) = \Sigma$.*
(ii) *A singleton set $\{(\mu, \tau, \theta)\}$ is self-stabilizing if and only if (μ, τ, θ) is stable for Γ^θ.*
(iii) *if both Σ_1 and Σ_2 are self-stabilizing, then $\Sigma_1 \cup \Sigma_2$ is self-stabilizing.* (iv) *there exists a largest self-stabilizing set.*

The largest fixed point can be identified via the following procedure. Let Σ_0 be the set of all matching outcomes. For all $k \geq 0$, let $\Sigma^{k+1} := S(\Sigma^k)$ and define $\Sigma^\infty := \bigcap_{k=1}^\infty \Sigma^k$. We leave the reader to check that Σ^∞ is the largest self-stabilizing set, which we will refer to as the **stable set** for Γ^{θ_J}.

We obtain the operator S by applying B and U simultaneously and then we iterate S. We can also apply B and U sequentially. We claim without proof that the order

of eliminations does not matter: as long as both U and B appear in the tail of the sequence, we can apply U and B to Σ_0 in an arbitrary order to obtain the largest fixed point. One such sequence is as follows. For all $k \geq 0$, let $\Sigma^{2k+1} := U(\Sigma^{2k})$ and for all $k \geq 1$, let $\Sigma^{2k} = B(\Sigma^{2k-1})$. Define $\Sigma^{\infty} = \bigcap_{k=1}^{\infty} \Sigma^k$.

The implications of this solution concept remain open. We do not know under what conditions assortative matching can be obtained.

If there is two-sided incomplete information, modifications are needed for both clear unilateral and bilateral deviations. Take unilateral deviations as an example. Workers also face uncertainties, and therefore, they need to take into account firms' incentives. This back-and-forth referencing makes the problem a non-trivial fixed-point problem. This complexity can be resolved by considering beliefs more explicitly, see, e.g., Liu (2023). Chen and Hu (2023) attempted an extension of LMPS stability to two-sided incomplete information, where individual rationality is defined with respect to the worst-case scenario.

4.6 Epistemic Foundations

The new solution concepts we've reviewed thus far clearly possess epistemic appeal. Formal epistemic languages would help elucidate the connection between stability, beliefs, and rationality in these concepts. There are two approaches. The first approach is to formalize a non-cooperative game and invoke the epistemic language of individual rationality. The other is to formulate a notion of cooperative rationality.

Pomatto (2022) models pairwise deviations explicitly as dynamic games. Starting with a putative matching outcome, workers make offers to firms, and firms respond by accepting or declining the offers they receive. For the putative matching to be stable, no workers should make offers (or they anticipate their offers to be rejected by firms). The paper adopts an extensive-form rationalizability approach. In this framework, each player interprets their observations under the assumption that they arise from rational behavior as much as possible. This concept is iterated upon to establish a common strong belief in rationality, as detailed in Battigalli and Siniscalchi (2002). Pomatto (2022) shows that this iteration leads to the same LMPS stable set, but interestingly, they differ in their finite-order iterations.

Wang (2022) takes a different epistemic approach. He formulates the epistemic state space and defines a notion of pairwise rationality. He defines rationalizable stability using an iterative process and provides an epistemic justification using a common strong belief of pairwise rationality. The paper introduces a belief refinement similar to Liu (2020) and shows that rationalizable stability is outcome equivalent to a correlated notion of Bayesian stability (without a common prior). The equivalence between pairwise rationality and pairwise stability echos the equivalence between rationalizability and correlated equilibrium in non-cooperative games, suggesting a deeper connection.

5 A More Refined Bayesian Approach

To formulate a Bayesian theory of stable matching, our primary concern is understanding the beliefs players hold. The iterative approach only pins down these beliefs in special cases of full revelation. To develop a coherent theory, we must consider the beliefs corresponding to each scenario of deviation as well as those in scenarios without any deviations. Both kinds of beliefs are endogenously determined. Liu (2020) introduces an indirect approach termed the "Kreps–Wilson program." Our focus in this review covers the general two-sided incomplete information problem (Γ, β^0), which is a special case of Liu (2023).

5.1 Bayesian Stability

Let Z be the set of all matching outcomes (μ, τ). A matching is a function $M : \Theta \to Z$. A pairwise deviation is represented by (μ, τ, i, j, p), indicating that i and j deviate from the matching outcome (μ, τ) with a transfer p between them. A scenario of bilateral deviation is $(\mu, \tau, i, j, p, \theta)$ while the perceived scenario of player $n \in \{i, j\}$ is $(\mu, \tau, i, j, p, \theta_n)$. A scenario of unilateral deviation is (μ, τ, i, θ) or (μ, τ, j, θ). Let Σ_n^1 and Σ_n^2 be the perceived scenarios of unilateral and bilateral deviations, respectively. Define Σ_n as $\Sigma_n^1 \cup \Sigma_n^2$. Each player n must have beliefs $\beta_n^1 : \Sigma_n^1 \to \Delta(\Theta)$ and $\beta_n^2 : \Sigma_n^2 \to \Delta(\Theta)$. We write $\beta_n : \Sigma_n^1 \cup \Sigma_n^2 \to \Delta(\Theta)$ as the piecewise function that agrees on β_n^1 and β_n^2 on their respective domains. The basic requirement is that player i assigns probability 1 to his own type θ_i and $M^{-1}(z)$ after each outcome z that is possible under M. We refer to

β_n^1 as player n's **on-path belief** and β_n^2 as player n's **off-path belief**. Lastly, we refer to a matching-function-belief configuration (M, β) as an **assessment**, following Kreps and Wilson (1982).

Definition 14. *An assessment (M, β) is **individually rational** if, for all scenarios of unilateral deviations (μ, τ, i, θ) and (μ, τ, j, θ), the following conditions hold:*

$$\begin{aligned} \mathbf{E}_{\beta_i^1(\mu,\tau,i,\theta_i)}[a_{i\mu(i)}(\cdot) + \tau_i] &\geq a_{ii}(\theta_i); \\ \mathbf{E}_{\beta_j^1(\mu,\tau,j,\theta_j)}[b_{\mu(j)j}(\cdot) + \tau_j] &\geq b_{jj}(\theta_j). \end{aligned}$$

Definition 15. *A scenario of bilateral deviation $(\mu, \tau, i, j, p, \theta)$ is a **blocking scenario** of (M, β) if the following conditions are satisfied:*

$$\begin{aligned} \mathbf{E}_{\beta_i^2(\mu,\tau,i,j,p,\theta_i)}[a_{ij}(\cdot) + p] &> \mathbf{E}_{\beta_i^2(\mu,\tau,i,j,p,\theta_i)}[a_{i\mu(i)}(\cdot) + \tau_i]; \\ \mathbf{E}_{\beta_j^2(\mu,\tau,i,j,p,\theta_j)}[b_{ij}(\cdot) - p] &> \mathbf{E}_{\beta_j^2(\mu,\tau,i,j,p,\theta_j)}[b_{\mu(j)j}(\cdot) + \tau_j]. \end{aligned}$$

*If $(\mu, \tau, i, j, p, \theta)$ is a blocking scenario, we say the bilateral deviation (μ, τ, i, j, p) is **viable**.*

Definition 16. *An assessment (M, β) is **Bayesian stable** if it is individually rational and devoid of any blocking scenarios.*

A matching game has **private value** if, for every $i \in I$, $j \in J$, and $\theta \in \Theta$, there exist functions $a_i : \Theta_i \to \mathbb{R}$ and $b_j : \Theta_j \to \mathbb{R}$ such that $a_{ij}(\theta) = a_i(\theta_i)$ and $b_{ij}(\theta) = b_j(\theta_j)$. In this particular case, the concept of Bayesian stability coincides with complete-information stability.

Theorem 6. *For a private-value matching game, an assessment (M, β) is Bayesian stable if and only if $M(\theta)$ is a stable matching outcome of Γ^θ for every $\theta \in \Theta$.*

Remark 11. The matching function $M : \Theta \to Z$ specifies deterministic outcomes. Modeling stochastic matching functions is a matter of notation. This idea is an analog of the correlated equilibrium. For each player $n \in N$, let S_n be the set of payoff-irrelevant signals. We write $S = \prod_{n \in N} S_n$ and $\hat{\Theta} = \Theta \times S$. Player n's prior is given by $\hat{\beta}_n^0 \in \Delta(\hat{\Theta})$. A correlated matching is a function $M : \hat{\Theta} \to Z$. Given this, it is straightforward to

define belief systems and Bayesian stability. The introduction of S is useful. Indeed, for a fixed Θ, we can look at stable matchings generated by $M : \Theta \times S \to Z$ over all possible S and all $\hat{\beta}_n^0 \in \Delta(\Theta \times S)$ such that their marginals on Θ coincide with a given prior $\beta_n^0 \in \Delta(\Theta)$. Therefore, S captures the uncertainties observed by the agent but not observed by the analysts who use an enlarged type space and solution sets to model their ignorance. This interpretation is adopted, e.g., in Liu (2009). □

5.2 Weak and Strong Consistencies

The plain-vanilla notion of Bayesian stability presented in Definition 16 does not have strong restrictions, as the beliefs can be rather arbitrary. Furthermore, these beliefs are not constrained by prior beliefs. Additional restrictions can be imposed.

For all bilateral deviations (μ, τ, i, j, p), let us define "deviating sets"

$$\begin{aligned} D_i^\beta(\mu,\tau,i,j,p) &= \left\{\theta_i : \mathbf{E}_{\beta_i^2(\mu,\tau,i,j,p,\theta_i)}[a_{ij}(\cdot)+p] > \mathbf{E}_{\beta_i^2(\mu,\tau,i,j,p,\theta_i)}[a_{i\mu(i)}(\cdot)+\tau_i]\right\}; \\ D_j^\beta(\mu,\tau,i,j,p) &= \left\{\theta_j : \mathbf{E}_{\beta_j^2(\mu,\tau,i,j,p,\theta_j)}[b_{ij}(\cdot)-p] > \mathbf{E}_{\beta_j^2(\mu,\tau,i,j,p,\theta_j)}[b_{\mu(j)j}(\cdot)+\tau_j]\right\}. \end{aligned}$$

Both deviating sets can be empty. They represent, respectively, the types of players i and j that stand to benefit from the bilateral deviation. If both deviating sets are non-empty, then the bilateral deviation (μ, τ, i, j, p) is viable.

Definition 17. *An assessment (M, β) is **weakly consistent** if, for any scenario of bilateral deviation $(\mu, \tau, i, j, p, \theta)$, the following conditions hold:*

$$\begin{aligned} \beta_i^2(\mu,\tau,i,j,p,\theta_i) &= \beta_i^0\left(\cdot \middle| \left(\{\theta_i\} \times D_j^\beta(\mu,\tau,i,j,p) \times \Theta_{-ij}\right) \cap M^{-1}(\mu,\tau)\right); \\ \beta_j^2(\mu,\tau,i,j,p,\theta_j) &= \beta_j^0\left(\cdot \middle| \left(\{\theta_j\} \times D_i^\beta(\mu,\tau,i,j,p) \times \Theta_{-ij}\right) \cap M^{-1}(\mu,\tau)\right). \end{aligned}$$

Notice that Bayes' rule places no restriction if the conditional set is empty. Weak consistency imposes restrictions on beliefs, even when there is no blocking scenario, because it restricts the beliefs under which a blocking scenario is evaluated. If $\theta \in D := (D_i^\beta(\mu,\tau,i,j,p) \times D_j^\beta(\mu,\tau,i,j,p) \times \Theta_{-ij}) \cap M^{-1}(\mu,\tau) \neq \emptyset$, then both i and j gain from the deviation. Weak consistency says that, in state θ, the beliefs of i and j assign probability 1 to D, so both believe that both gain from the deviation. Not limited to

first-order beliefs as it may appear, weak consistency implies that, in θ, both believe that both believe both gain from the deviation, ad infinitum. In epistemic jargon, D is a "self-evident" common knowledge event. Definition 16 does not satisfy this common knowledge property.

The following example, which is adapted from Liu (2023), demonstrates the restriction of weak consistency.

Example 1. Consider a two-player matching game with no transfer, so $\mathbb{T} = \{0\}$. Player 1 is the worker and player 2 is the firm. Each player has two types: $\Theta_n = \{\theta_n^1, \theta_n^2\}$. There is a uniform common prior $\beta_1^0 = \beta_2^0 \in \Delta(\Theta)$. Players' payoffs from not matching with each other are always 0. Players' payoffs from matching together, $(a_{12}(\theta), b_{12}(\theta))$, are dependent on their types $\theta = (\theta_1, \theta_2)$ as follows:

	θ_2^1	θ_2^2
θ_1^1	$1, 1$	$1, -2$
θ_1^2	$-2, -2$	$-2, -2$

This configuration means that, for instance, players 1 and 2 receive payoffs of 1 and -2, respectively, from the partnership if their types are $\theta = (\theta_1^1, \theta_2^2)$. Consider the matching function that specifies an autarky for every state, and a belief system β which assigns equal probabilities to the opponent's two types in all scenarios of bilateral deviations. The only bilateral deviation is for the two players to match, which we denote as δ. We denote a scenario of deviation by (δ, θ). Therefore, player 1 of type θ_1^1 prefers the deviation, but player 1 of type θ_1^2 and player 2 of both types prefer not to deviate (as their expected payoff from the deviation is negative). Therefore, having autarky in every state is Bayesian stable. However, player 1 prefers the deviation if and only if his type is θ_1^1, i.e., $D_1^\beta(\delta) = \{\theta_1^1\}$. This incentive is understood by player 2 of both types and the relevant belief for player 2's decision of joining the bilateral deviation should condition on this fact. That is, player 2 should assign probability 1 to player 1 being θ_1^1, instead of equal probability to θ_1^1 and θ_1^2, in any of his two perceived scenarios of deviation, which is captured by weak consistency. Given this belief, player 2 of θ_2^1 will join the deviation with player 1 of θ_1^1. Therefore, weak consistency implies that $(\delta, \theta_1^1, \theta_2^1)$ is a blocking scenario. It can be shown that the only matching that is part of

a weakly consistent and stable assessment is the following:

	θ_2^1	θ_2^2
θ_1^1	match	autarky
θ_1^2	autarky	autarky

So weak consistency identifies the more intuitive outcomes in this game. □

Example 2. Consider the example above, but the payoffs from matching together are modified as follows:

	θ_2^1	θ_2^2
θ_1^1	$1, 1$	$-2, -2$
θ_1^2	$-2, -2$	$-2, -2$

It is easy to check that having autarky in every state with a uniform belief is Bayesian stable and weakly consistent.

However, it is quite intuitive that in this common-interest game, the two players can form a partnership in (θ_1^1, θ_2^1). For instance, player 1 of type θ_1^1 can make the following statement to player 2: "I am θ_1^1, and if you are θ_2^1, let's form a partnership. You should know that if you are θ_2^1, I benefit from the partnership if and only if my type is θ_1^1, so you should trust that I am θ_1^1. My question is whether you are θ_2^1." Similarly, player 2 of type θ_2^1 can make the following mirroring statement to player 1: "I am θ_2^1, and if you are θ_1^1, let's form a partnership. You should know that if you are θ_1^1, I benefit from the partnership if and only if my type is θ_2^1, so you should trust that I am θ_2^1. My question is whether you are θ_1^1." The two statements are "mutually reassuring" in the sense that conditional on that the opponent is the cooperative type, a player benefits from carrying out the stated plan if and only if he himself is the cooperative type, which reassures the cooperative opponent. The existence of such mutual reassurance means that it is plausible that the two players collectively deviate to form a partnership in (θ_1^1, θ_2^1). □

Definition 18. *We say $D_i \subset \Theta_i$ and $D_j \subset \Theta_j$ are* ***mutually reassuring with cer-***

tainty *for a bilateral deviation* (μ, τ, i, j, p) *if* $(D_i \times D_j \times \Theta_{-ij}) \cap M^{-1}(\mu, \tau) \neq \emptyset$ *and*

$$D_i = \left\{ \theta_i : \begin{array}{ll} & \text{(i) } (\{\theta_i\} \times D_j \times \Theta_{-ij}) \cap M^{-1}(\mu, \tau) \neq \emptyset \\ \text{(ii)} & \mathbf{E}_{\beta_i^0}[a_{ij}(\cdot) | (\{\theta_i\} \times D_j \times \Theta_{-ij}) \cap M^{-1}(\mu, \tau)] + p \\ & > \mathbf{E}_{\beta_i^0}[a_{i\mu(i)}(\cdot) | (\{\theta_i\} \times D_j \times \Theta_{-ij}) \cap M^{-1}(\mu, \tau)] + \tau_i \end{array} \right\},$$

$$D_j = \left\{ \theta_j : \begin{array}{ll} & \text{(i) } (\{\theta_j\} \times D_i \times \Theta_{-ij}) \cap M^{-1}(\mu, \tau) \neq \emptyset \\ \text{(ii)} & \mathbf{E}_{\beta_j^0}[b_{ij}(\cdot) | (\{\theta_j\} \times D_i \times \Theta_{-ij}) \cap M^{-1}(\mu, \tau)] - p \\ & > \mathbf{E}_{\beta_j^0}[b_{\mu(j)j}(\cdot) | (\{\theta_j\} \times D_i \times \Theta_{-ij}) \cap M^{-1}(\mu, \tau)] + \tau_j \end{array} \right\},$$

where $\mathbf{E}_{\beta_i^0}$ *and* $\mathbf{E}_{\beta_j^0}$ *are the expectation operators with respect to the prior beliefs* β_i^0 *and* β_j^0*, respectively.*

Definition 19. *An assessment* (M, β) *is* ***strongly consistent*** *if the following two conditions are satisfied:*

(i) *it is weakly consistent;*

(ii) *if there exist mutually reassuring sets with certainty* D_i *and* D_j *for* (μ, τ, i, j, p)*, then*

$$(D_i^\beta(\mu, \tau, i, j, p) \times D_j^\beta(\mu, \tau, i, j, p) \times \Theta_{-ij}) \cap M^{-1}(\mu, \tau) \neq \emptyset.$$

Liu (2023) also introduces a strong consistency concept that does not require that $(\{\theta_i\} \times D_j \times \Theta_{-ij}) \cap M^{-1}(\mu, \tau) \neq \emptyset$ and $(\{\theta_j\} \times D_i \times \Theta_{-ij}) \cap M^{-1}(\mu, \tau) \neq \emptyset$ in Definition 18, but make restrictions on $\mathbf{E}_{\beta_i^0}$ and $\mathbf{E}_{\beta_j^0}$ when they are conditioned on zero-probability events.

Example 2 shows that weak consistency and strong consistency have very different implications, but this hinges on two-sided incomplete information. In matching games with one-sided incomplete information, they coincide.

Theorem 7. *An assessment* (M, β) *is weakly consistent for the game* $(\Gamma^{\theta_J}, \beta^0)$ *if and only if it is strongly consistent with certainty for the game.*

The key for this result is the observation that if type θ_J is commonly known, then

$$\begin{aligned} D_i^\beta(\mu, \tau, i, j, p) &= \left\{ \theta_i : \mathbf{E}_{\beta_i^2(\mu,\tau,i,j,p,\theta_i)}[a_{ij}(\cdot) + p] > \mathbf{E}_{\beta_i^2(\mu,\tau,i,j,p,\theta_i)}[a_{i\mu(i)}(\cdot) + \tau_i] \right\} \\ &= \left\{ \theta_i : a_{ij}(\theta_i, \theta_j) + p > a_{i\mu(i)}(\theta_i, \theta_{\mu(i)}) + \tau_i \right\}. \end{aligned}$$

5.3 Implications of Stability and Consistency

Liu (2023) defines a condition called comonotonic differences that is intuitive for matching problems.

Definition 20. *A matching game has* ***comonotonic differences*** *if* $\mathbf{E}_{\theta_{j'}}\left[a_{ij}(\theta_i,\theta_j)-a_{ij'}(\theta_i,\theta_{j'})\right]$ *and* $\mathbf{E}_{\theta_{i'}}\left[b_{ij}(\theta_i,\theta_j)-b_{i'j}(\theta_{i'},\theta_j)\right]$ *are comonotonic on* Θ_i *and on* Θ_j *for any two pairs* $(i,j)\in I\times J$, $(i',j')\in(I\cup\{j\})\times(J\cup\{i\})$, *and any expectation operators* $\mathbf{E}_{\theta_{i'}}$ *over* $\Theta_{i'}$ *for* $i'\in I$ *and* $\mathbf{E}_{\theta_{j'}}$ *over* $\Theta_{j'}$ *for* $j'\in J$.

The condition of comonotonic differences is always satisfied in a complete-information matching game Γ^θ. There are two important special classes of matching games with comonotonic differences. A matching game Γ has **one-sided interdependence** if either there is no restriction on b_{ij} but there exist functions $A_i:\Theta_i\to\mathbb{R}$ and constants A'_{ij} such that

$$a_{ij}(\theta_i,\theta_j)=A_i(\theta_i)+A'_{ij} \text{ for all } i\in I,\ j\in J\cup\{i\},\ \text{and } \theta\in\Theta,$$

or symmetrically, there is no restriction on a_{ij} but there exist functions $B_j:\Theta_j\to\mathbb{R}$ and constants B'_{ij} such that

$$b_{ij}(\theta_i,\theta_j)=B_j(\theta_j)+B'_{ij} \text{ for all } j\in J,\ i\in I\cup\{j\},\ \text{and } \theta\in\Theta.$$

A matching game Γ has **separable values** if

$$\begin{aligned} a_{ij}(\theta_i,\theta_j)&=A_i(\theta_i)+A'_{ij}(\theta_j) \text{ for all } i\in I,\ j\in J\cup\{i\},\ \text{and } \theta\in\Theta,\\ b_{ij}(\theta_i,\theta_j)&=B_j(\theta_j)+B'_{ij}(\theta_i) \text{ for all } j\in J,\ i\in I\cup\{j\},\ \text{and } \theta\in\Theta, \end{aligned}$$

where $A_i:\Theta_i\to\mathbb{R}$, $B_j:\Theta_j\to\mathbb{R}$, $A'_{ij}:\Theta_j\to\mathbb{R}$ and $B'_{ij}:\Theta_i\to\mathbb{R}$, and A'_{ii} and B'_{jj} are constant.

Liu (2023) show that all weakly consistent and stable assessments satisfy the following

$$z\in\operatorname*{argmax}_{z'\in Z}\sum_{n\in I\cup J}\mathbf{E}^0(u_n(z',\cdot)|M^{-1}(z)) \text{ for any } z\in M(\Theta). \tag{5.1}$$

for games with one-sided interdependence. Further, all strongly consistent and stable

assessments satisfy (5.1) for games with comonotonic differences if a belief independence condition holds. The property captured in Equation (5.1) can be understood from an outside observer's perspective. The observer, who starts with a prior β_0, observes the outcome z, and believes in the theory of stability we reviewed here, will update his assessment of type distribution to $\beta^0(\cdot|M^{-1}(z))$. If Equation (5.1) holds, then the observer will not be able to recommend a rematching to improve the total surplus, without changing the information structure. Liu's result does not rely on the observer's knowledge of β^0, M, or the exact payoff functions.

6 Further Research and Open Questions

6.1 Other Solution Concepts

Stability and competitive equilibrium are two different ways of looking at a matching problem. With complete information, the two solution concepts are equivalent, but competitive equilibrium is especially simple and useful for economic analysis (e.g., Becker (1973) and Chiappori (2017)). The equivalence is no longer valid under Informational asymmetry as we discussed in Section 4.4, where we see that an iterative notion of competitive equilibrium is refined by iterative stability. A Bayesian notion of competitive equilibrium is easy to define. We will define it as the counterpart of weak consistency; different belief-based refinements can be used. A price matrix P is a mapping $P : I \times J \to \mathbb{R}$, and we also append $P_{ii} = P_{jj} = 0$ to this notion.

Definition 21. *A matching $M : \theta \mapsto (\mu, P)$ is a (rational expectations)* ***competitive equilibrium*** *if the following holds for all $\theta \in \Theta$ and $(\mu, P) = M(\theta)$:*

(i) $\mathbf{E}_{\beta_i^0}[a_{i\mu(i)}|M^{-1}(\mu, P)] + P_{i\mu(i)} \geq \mathbf{E}_{\beta_i^0}[a_{ij}|M^{-1}(\mu, P)] + P_{ij}$ *for all $i \in I$ and $j \in J \cup \{i\}$;*

(ii) $\mathbf{E}_{\beta_j^0}[b_{\mu(j)j}|M^{-1}(\mu, P)] - P_{\mu(j)j} \geq \mathbf{E}_{\beta_j^0}[b_{ij}|M^{-1}(\mu, P)] - P_{ij}$ *for all $j \in J$ and $i \in I \cup \{j\}$.*

The concept satisfies individual rationality. Notice also that only the "on-path belief" $\beta^0(\cdot|M^{-1}(\mu, P))$ is utilized in the definition, because a unilateral deviation does not require mutual agreement. This definition generalizes the notion of a competitive equilibrium under complete information.

A stable matching outcome does not specify a price for an unmatched pair, while a competitive matching outcome does specify a price for every pair. The observability of the full price matrix may seem to suggest that prices in a competitive equilibrium matching reveal more information than on-path prices in a stable matching. On the other hand, flexible off-path prices allow for more information revelation. Understanding the difference between them requires taking into account the *incentives* and *information* embedded in their defining criteria. Indeed, Liu (2020) demonstrates that there is no clear way of comparing stability and competitive equilibrium under incomplete information. Different belief-based refinements could change the conclusion, and might even restore the equivalence. In addition, different informational environments such as no aggregate uncertainty (either permutation as in Section 4.1 or a continuum economy) might make the comparison easier.

The idea described in Section 5 can be useful for defining other solution concepts in the context of incomplete information. In a two-sided matching market with complete information, the core is equivalent to pairwise stability. However, this equivalence breaks down under incomplete information. A coalition involving more than two players can reveal more information than pairwise deviations, even when rematches occur only in pairs. This holds true even in situations with one-sided incomplete information, as demonstrated by Liu (2020) using four-player examples. In general, the core refines the concept of pairwise stability.

The concept of von Neumann-Morgenstern stable sets is another significant solution concept, characterized by both internal and external consistencies. Gretschko and Wambach (2022) applied this concept to the study of the principal-agent problem, where agents have private information. Here again, the idea is to explicitly model beliefs within the solution concept. This reduced-form approach intuitively captures equilibria in dynamic contracting games with renegotiation.

The long history of cooperative game theory is marked by the development of many insightful solution concepts, almost always for games with complete information, such as Shapley values, bargaining solutions, and the nucleolus. These concepts, however, are different as they do not explicitly model deviations by coalitions. It would be interesting to investigate how to further develop these insights under incomplete information and whether and how endogenous belief formation outlined in Section 5 can be useful. It

is curious that these concepts have not been widely used in the study of matching problems. For further reading, see Myerson (1984) and Gul and Pesendorfer (2020).

6.2 Information Acquisition

Information acquisition is natural in the context of incomplete information. Immorlica, Leshno, Lo, and Lucier (2020) consider information acquisition in a school matching problem, and they define a notion of stability, which captures individual optimization given others' choices. Thus, their concept is similar to a competitive equilibrium. Its conceptual difference with stability has been discussed above. Pairwise or group deviations would open up new possibilities. For example, one might consider a group deviation in which a school subsidizes students' information acquisition. These applied theory questions have not been addressed.

Maxey (2021) and Kloosterman and Troyan (2020) for the study of information acquisition in a market design setting. Relatedly, Lee and Schwarz (2017) and Echenique, Gonzalez, Wilson, and Yariv (2022) investigate the role of interviews as a means of preference discovery.

6.3 Applications

Many economic problems can be modeled as matching. We consider transfers in our canonical models Section 3, but more generally players can sign complicated incentive contracts when they match with each other (see, e.g., Liu (2023) for explorations). The solution concepts in Sections Section 4 and Section 5 capture stabilized market interactions. The following feedback information loop plays a crucial role in shaping stable outcomes: the prevailing outcomes inform about the incentives and disincentives for deviating from them, and this information in turn shapes the outcomes. This information perspective is a difference between our approach and familiar cooperative or non-cooperative equilibrium models, which opens up a wealth of new applications and sheds new light on well-studied ones.

Markets with adverse selection are an example in point. For instance, if a separating contract prevails in an insurance market, policyholders will reveal their types. Competing insurance companies or new entrants will then take advantage of this infor-

mation and offer full coverage contracts tailored to specific types, thus destabilizing the separating outcome. Therefore, the framework of Rothschild and Stiglitz (1976) is not suitable for analyzing stable insurance markets. These familiar applications, previously analyzed in non-cooperative models, warrant further investigation. We foresee that the stable matching with incomplete information will become valuable in a wide range of applications.

It's worth noting that this approach differs from the posterior implementation discussed by Green and Laffont (1987), which does not incorporate endogenous belief formation through the information feedback loop. Similarly, the consideration of stability under incomplete information offers an alternative perspective on collusion-proof mechanism design, in contrast to Laffont and Martimort (1997). This goes well beyond matching problems.

6.4 Implementation

As mentioned in the introduction, we should be interested in both the properties that define stabilized matching outcomes and the processes that lead to them. There are several different approaches. In the definition of a matching function $M : \Theta \to Z$, we do not impose a Bayesian incentive compatibility condition, as contrast to the preference revelation approach discussed in Section 2. The Bayesian incentive compatibility condition provides a one-shot implementation of stable matching and serves as a desirable selection criterion of stable matchings, so the results proved without incentive compatibility is robust. However, this approach is valid for special cases, such as when the preferences of firms being independent of workers' private types; in general, there is a conflict between stability and incentive compatibility, so existence is not guaranteed.

On the other hand, the one-shot direct revelation game associated with incentive compatibility is too restrictive for the purpose of having stability as a way of capturing stabilized decentralized interactions. As a reduced-form concept, (M, β) abstracts the underlying process; thus, the incentive compatibility of M omits valuable information. Defining a more suitable notion of incentive compatibility remains an open question. One idea is to expand the outcome space to incorporate essential information regarding the underlying processes (for example, the time it takes to settle a matching is impor-

tant, as we have experienced from dynamic screening problems). Alternatively, if we were to maintain a one-shot implementation through conventional incentive compatibility on M, the concept of stability would likely require modification, possibly to make it more permissive by refining the blocking conditions.

Other approaches include a Nash program for incomplete information games, i.e., implementation of a reduced-form concept through non-cooperative games. Research in this area is currently quite limited and appears non-existent in the context of matching.[7] Although the Nash program under complete information is well-developed, the proposed strategic foundations often rely on ad hoc extensive-form games and are constrained to highly specific setups. A more promising approach is to explore adjustment processes that describe how a particular matching market might work. For a complete-information problem, Roth and Vate (1990) consider an adjustment process, where, in each round, a single blocking pair of players is selected to rematch. They show that starting from any unstable matching, there is a path of pairwise rematching that terminates at a stable matching outcome. In a random rematching process where every blocking pair is selected with positive probability in each round, convergence to a stable matching will occur with probability 1.

Chen and Hu (2020) consider a random matching process under incomplete information, in which a player's initial information, modeled as a partition, is subsequently refined through their observations: a new partner's type in a rematch, rematch of other players, and the tentative lack of rematch. In their formulation, a rematching occurs if two players can clearly block the prevailing matching outcome as in Section 4.2. They show that the process converges to a prior-free notion of stable matching that incorporates partitional information of the firms, which generalizes LMPS. Lazarova and Dimitrov (2017) consider a model where a pair is eligible for rematch if there is a state in the support of their beliefs in which both benefit from the rematch (i.e., they evaluate pairwise deviations using the most optimistic belief). Once they rematch, they observe each other's type and update their belief about others' types based on this new observation. The optimistic belief incentivizes players to learn the types of others through their

[7]For instance, Okada (2012) explores bargaining foundations of the signaling core for an exchange economy with incomplete information, and Kamishiro, Vohra, and Serrano (2022) investigates the sequential core.

temporary matches, making incomplete information transient. Lazarova and Dimitrov (2017) give conditions the rematching paths to terminate, and show that the resulting matching outcome must be complete-information stable. The use of optimistic beliefs to evaluate matching opportunities is in contrast with Chen and Hu (2020), where rematch opportunities are evaluated using the most pessimistic beliefs. In both papers, the support of beliefs, rather than the exact beliefs, matters. Chen and Hu (2022) incorporate belief updating into the adaptive matching process of Chen and Hu (2020), which provides a foundation for the stability of Liu (2020).

Agranov, Dianat, Samuelson, and Yariv (2021) experimentally studied a decentralized matching process with transfers, in which agents can propose to others. After being accepted, they learn their match payoffs and consequently their partners' types. They observe that stable matching is difficult to achieve, especially under incomplete information and submodular payoffs. Even under supermodularity and complete information, subjects are unable to figure out the transfers associated with stable matching.[8] Agranov, Dianat, Samuelson, and Yariv (2021) use complete-information stability as their theoretical benchmark, because players in their experiment learn their partners' types through tentative matches. Although their results appear negative for the concept of stability, the critical questions remain as to whether repeatedly playing the same matching games, especially experiencing the roles of different types, will help subjects converge to stable matching outcomes.

The matching processes described above are specific. It is an open question of how to model a large class of reasonable matching processes and understand when and where they converge, particularly in cases where incomplete information is not transient. LMPS shows that under supermodularity/submodularity assumptions, incomplete information unravels, starting from the lowest type to the highest. It remains to see how this mechanism established with a reduced-form concept unfolds in dynamic market interactions.

[8]See also He, Wu, Zhang, and Zhu (2023) for the difficulty of achieving stability in an experimental study of complete-information matching.

6.5 Dynamic Games of Matching

While we advocate for the utilization of reduced-form concepts like stability and competitive equilibrium, it is important not to dismiss non-cooperative frameworks when we have a good understanding of the institutions or market mechanisms governing certain markets.

There is a distinctive literature that incorporates search frictions to models of labor markets and macroeconomics. This "search-and-matching" literature has been mostly about steady state of equilibrium interactions under complete information. For models of incomplete information, see, e.g., Gonzalez and Shi (2010), Guerrieri, Shimer, and Wright (2010), Lauermann (2013), and more recently, Ferdowsian (2023). The focus, however, is not on pairwise blocking and stability. The connection between this literature and the cooperative approach, for both the frictionless limit and away from it, is still not well understood.

Several papers have successfully integrated non-cooperative games with reduced-form concepts, effectively leveraging the strengths of both approaches. This line of research holds the potential to be a very promising area of applied theory. Li and Rosen (1998) study unraveling in the presence of incomplete information. Individual workers' productivities are initially unknown and only become public in the second period. Given the aggregate uncertainty about whether there are enough productive workers compared to firms and the discontinuity of competitive equilibrium prices, both workers and firms might contract with each other before the productivities of workers become known, which can be ex post inefficient.

In the work by Damiano and Li (2007), the focus shifts to a revenue-maximizing monopolistic matchmaker who employs pricing strategies to sort agents into different markets. The paper establishes conditions for equilibria to exist, where each market is exclusively composed of a single type of male and a single type of female participants. Damiano and Li (2008) consider matching coordinated by two competing matchmakers who create two markets with entry fees and women and men self-select into these markets. The paper shows that for the matchmakers, there is no pure strategy equilibrium if they set their prices simultaneously. However, if they move sequentially and the distribution of types is sufficiently diffused, a pure strategy equilibrium does exist.

They identify environments in which the first mover serves the lower quality matching market.

References

Agranov, Marina, Ahrash Dianat, Larry Samuelson, and Leeat Yariv (2021). "Paying to match: Decentralized markets with information frictions". In.

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

Aumann, Robert J and Aviad Heifetz (2002). "Incomplete information". In: *Handbook of Game Theory with Economic Applications* 3, pp. 1665–1686.

Battigalli, Pierpaolo and Marciano Siniscalchi (2002). "Strong belief and forward induction reasoning". In: *Journal of Economic Theory* 106.2, pp. 356–391.

Becker, Gary S (1973). "A theory of marriage: Part I". In: *Journal of Political economy* 81.4, pp. 813–846.

Bernheim, B Douglas (1984). "Rationalizable strategic behavior". In: *Econometrica*, pp. 1007–1028.

Bikhchandani, Sushil (2017). "Stability with one-sided incomplete information". In: *Journal of Economic Theory* 168, pp. 372–399.

Chakraborty, Archishman, Alessandro Citanna, and Michael Ostrovsky (2010). "Two-sided matching with interdependent values". In: *Journal of Economic Theory* 145.1, pp. 85–105.

— (2015). "Group stability in matching with interdependent values". In: *Review of Economic Design* 19, pp. 3–24.

Chen, Yi-Chun and Gaoji Hu (2020). "Learning by matching". In: *Theoretical Economics* 15.1, pp. 29–56.

— (2022). "Bayesian stable states". In: *Available at SSRN 3709205*.

— (2023). "A theory of stability in matching with incomplete information". In: *American Economic Journal: Microeconomics* 15.1, pp. 288–322.

Chen, Yan and Tayfun Sönmez (2006). "School choice: an experimental study". In: *Journal of Economic theory* 127.1, pp. 202–231.

Chiappori, Pierre-André (2017). *Matching with transfers: The economics of love and marriage*. Princeton University Press.

Crawford, Vincent P and Elsie Marie Knoer (1981). "Job matching with heterogeneous firms and workers". In: *Econometrica*, pp. 437–450.

Damiano, Ettore and Hao Li (2007). "Price discrimination and efficient matching". In: *Economic Theory* 30.2, pp. 243–263.

— (2008). "Competing matchmaking". In: *Journal of the European Economic Association* 6.4, pp. 789–818.

Dizdar, Deniz and Benny Moldovanu (2016). "On the importance of uniform sharing rules for efficient matching". In: *Journal of Economic Theory* 165, pp. 106–123.

Dutta, Bhaskar and Rajiv Vohra (2005). "Incomplete information, credibility and the core". In: *Mathematical Social Sciences* 50.2, pp. 148–165.

Echenique, Federico, Ruy Gonzalez, Alistair J Wilson, and Leeat Yariv (2022). "Top of the Batch: Interviews and the Match". In: *American Economic Review: Insights* 4.2, pp. 223–238.

Ehlers, Lars and Jordi Massó (2007). "Incomplete information and singleton cores in matching markets". In: *Journal of Economic Theory* 136.1, pp. 587–600.

— (2015). "Matching markets under (in) complete information". In: *Journal of economic theory* 157, pp. 295–314.

Ferdowsian, Andrew (2023). "Learning through Transient Matching in Congested Markets". In: *Working Paper, Princeton University*.

Fernandez, Marcelo Ariel, Kirill Rudov, and Leeat Yariv (2022). "Centralized matching with incomplete information". In: *American Economic Review: Insights* 4.1, pp. 18–33.

Forges, Françoise, Enrico Minelli, and Rajiv Vohra (2002). "Incentives and the core of an exchange economy: a survey". In: *Journal of Mathematical Economics* 38.1-2, pp. 1–41.

Forges, Françoise and Roberto Serrano (2013). "Cooperative games with incomplete information: Some open problems". In: *International Game Theory Review* 15.02, p. 1340009.

Gale, David and Lloyd S Shapley (1962). "College admissions and the stability of marriage". In: *The American Mathematical Monthly* 69.1, pp. 9–15.

Glycopantis, Dionysius and Nicholas C Yannelis (2006). *Differential information economies*. Vol. 19. Springer Science & Business Media.

Gonzalez, Francisco M and Shouyong Shi (2010). "An equilibrium theory of learning, search, and wages". In: *Econometrica* 78.2, pp. 509–537.

Green, Jerry R and Jean-Jacques Laffont (1987). "Posterior implementability in a two-person decision problem". In: *Econometrica*, pp. 69–94.

Gretschko, Vitali and Achim Wambach (2022). "Stable contracts under renegotiation". In: *Working Paper, University of Mannheim*.

Guerrieri, Veronica, Robert Shimer, and Randall Wright (2010). "Adverse selection in competitive search equilibrium". In: *Econometrica* 78.6, pp. 1823–1862.

Gul, Faruk and Wolfgang Pesendorfer (2020). "Lindahl equilibrium as a collective choice rule". In: *Working Paper, Princeton University*.

He, Simin, Jiabin Wu, Hanzhe Zhang, and Xun Zhu (2023). "Decentralized matching with transfers: experimental and noncooperative analyses". In: *Available at SSRN 3596703*.

Holmström, Bengt and Roger B Myerson (1983). "Efficient and durable decision rules with incomplete information". In: *Econometrica*, pp. 1799–1819.

Immorlica, Nicole, Jacob Leshno, Irene Lo, and Brendan Lucier (2020). "Information acquisition in matching markets: The role of price discovery". In: *Available at SSRN 3705049*.

Kamishiro, Yusuke, Rajiv Vohra, and Roberto Serrano (2022). "Signaling, Screening, and Core Stability". In: *Working Paper, Brown University*.

Kandori, Michihiro, George J Mailath, and Rafael Rob (1993). "Learning, mutation, and long run equilibria in games". In: *Econometrica*, pp. 29–56.

Kloosterman, Andrew and Peter Troyan (2020). "School choice with asymmetric information: Priority design and the curse of acceptance". In: *Theoretical Economics* 15.3, pp. 1095–1133.

Kreps, David M and Robert Wilson (1982). "Sequential equilibria". In: *Econometrica*, pp. 863–894.

Laffont, Jean-Jacques and David Martimort (1997). "Collusion under asymmetric information". In: *Econometrica: Journal of the Econometric Society*, pp. 875–911.

Lauermann, Stephan (2013). "Dynamic matching and bargaining games: A general approach". In: *American Economic Review* 103.2, pp. 663–689.

Lazarova, Emiliya and Dinko Dimitrov (2017). "Paths to stability in two-sided matching under uncertainty". In: *International Journal of Game Theory* 46, pp. 29–49.

Lee, Robin S and Michael Schwarz (2017). "Interviewing in two-sided matching markets". In: *The RAND Journal of Economics* 48.3, pp. 835–855.

Li, Hao and Sherwin Rosen (1998). "Unraveling in matching markets". In: *American Economic Review*, pp. 371–387.

Liu, Qingmin (2009). "On redundant types and Bayesian formulation of incomplete information". In: *Journal of Economic Theory* 144.5, pp. 2115–2145.

— (2020). "Stability and Bayesian consistency in two-sided markets". In: *American Economic Review* 110.8, pp. 2625–2666.

— (2023). "Cooperative Analysis of Incomplete Information". In: *Working Paper, Columbia University*.

Liu, Qingmin, George J Mailath, Andrew Postlewaite, and Larry Samuelson (2014). "Stable matching with incomplete information". In: *Econometrica* 82.2, pp. 541–587.

Löfgren, Karl-Gustaf, Torsten Persson, and Jörgen W Weibull (2002). "Markets with asymmetric information: the contributions of George Akerlof, Michael Spence and Joseph Stiglitz". In: *The Scandinavian Journal of Economics*, pp. 195–211.

Ma, Jinpeng (1995). "Stable matchings and rematching-proof equilibria in a two-sided matching market". In: *Journal of Economic Theory* 66.2, pp. 352–369.

Maxey, Tyler (2021). "School Choice with Costly Information Acquisition". In: *Available at SSRN 3971158*.

Myerson, Roger B (1984). "Two-person bargaining problems with incomplete information". In: *Econometrica: Journal of the Econometric Society*, pp. 461–487.

— (1995). "Sustainable matching plans with adverse selection". In: *Games and Economic Behavior* 9.1, pp. 35–65.

Okada, Akira (2012). "Non-cooperative bargaining and the incomplete informational core". In: *Journal of Economic Theory* 147.3, pp. 1165–1190.

Pais, Joana and Agnes Pintér (2008). "School choice and information: An experimental study on matching mechanisms". In: *Games and Economic Behavior* 64.1, pp. 303–328.

Pais, Joana, Agnes Pintér, and Róbert F Veszteg (2011). "College admissions and the role of information: An experimental study". In: *International Economic Review* 52.3, pp. 713–737.

Pearce, David G (1984). "Rationalizable strategic behavior and the problem of perfection". In: *Econometrica*, pp. 1029–1050.

Pomatto, Luciano (2022). "Stable matching under forward-induction reasoning". In: *Theoretical Economics* 17.4, pp. 1619–1649.

Roth, Alvin E (1982). "The economics of matching: Stability and incentives". In: *Mathematics of operations research* 7.4, pp. 617–628.

— (1984). "Misrepresentation and stability in the marriage problem". In: *Journal of Economic theory* 34.2, pp. 383–387.

— (1989). "Two-sided matching with incomplete information about others' preferences". In: *Games and Economic Behavior* 1.2, pp. 191–209.

— (1991). "A natural experiment in the organization of entry-level labor markets: Regional markets for new physicians and surgeons in the United Kingdom". In: *The American economic review*, pp. 415–440.

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

Rothschild, Michael and Joseph Stiglitz (1976). "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information". In: *The Quarterly Journal of Economics* 90.4, pp. 629–649.

Sugaya, Takuo and Alexander Wolitzky (2021). "The revelation principle in multistage games". In: *The Review of Economic Studies* 88.3, pp. 1503–1540.

Wang, Ziwei (2022). *Rationalizable Stability in Matching with One-Sided Incomplete Information*. Tech. rep. Working paper.

Wilson, Robert (1978). "Information, efficiency, and the core of an economy". In: *Econometrica*, pp. 807–816.

Yenmez, M Bumin (2013). "Incentive-compatible matching mechanisms: consistency with various stability notions". In: *American Economic Journal: Microeconomics* 5.4, pp. 120–141.